RailRelax: Enhancing User Experience

Wafiya Shaikh
Student, Third Year Computer Engineering
Vivekanand Education Society's Institute of Technology
Chembur, Mumbai 400074
2022.wafiya.shaikh@ves.ac.in

Anjali Thakrani
Student, Third Year Computer Engineering
Vivekanand Education Society's Institute of Technology Chembur, Mumbai 400074
2022.anjali.thakrani@ves.ac.in

Anisha Shankar
Student, Third Year Computer Engineering
Vivekanand Education Society's Institute of Technology
Chembur, Mumbai 400074
2022.anisha.shankar@ves.ac.in

Himaja Pannati
Student, Third Year Computer Engineering
Vivekanand Education Society's Institute of Technology Chembur, Mumbai 400074
2022.himaja.pannati@ves.ac.in

Prof. Lifna C.S
Assistant Professor, Department of Computer Engineering
Vivekanand Education Society's Institute of Technology
Chembur, Mumbai 400074
lifna.cs@ves.ac.in

***Abstract*—In highly populated urban areas such as Mumbai, the issue of overcrowding in local trains poses a significant challenge and inconvenience for the millions of passengers who travel. This paper introduces a sophisticated solution that employs the YOLO object detection algorithm to accurately count the number of passengers in each train compartment. It utilizes real-time video data from CCTV systems to provide passengers with information about compartment occupancy prior to arrival, enabling them to make informed decisions that optimize boarding and space usage. This also reduces the need for manual video monitoring and increases the efficiency of passenger flow, resulting in a more organized and comfortable journey. The paper signifies a significant advancement in managing public transport systems by using the latest computer vision approach to address the persistent issue of overcrowding in Mumbai's local train network.**

I*ndex Terms*—Crowd Analysis, Overcrowding Management, YOLO Object Detection, Computer Vision, Passenger Flow Optimization, CCTV-Based Monitoring, Automated Passenger Counting, Public Transport Optimization

I. INTRODUCTION

Mumbai, a city with a population exceeding 20 million, heavily relies on its suburban railway system, which transports approximately 7.5 million passengers daily. Despite operating over 2,300 train services across a 390 km network, overcrowding remains a severe issue. During peak hours, compartments designed for 1,750 passengers often accommodate up to 4,500, leading to significant discomfort, delays, and safety concerns (IRCTC, 2018). The system handles 2.64 billion rides annually, making it one of the busiest railway networks in the world. Tragically, overcrowding has led to fatal accidents, with a 2018 estimate reporting 711 deaths due to overloaded trains. The Mumbai Urban Transport Project (MUTP) was initiated to decongest the railway system, but it has failed to meet the rising passenger demand. Projections suggest that by 2030, ridership will increase by another 6–8 million, raising concerns about the sustainability of existing infrastructure. Proposed solutions such as double-decker trains and metro expansions require significant financial investment and long implemen- tation timelines, making them insufficient in addressing the immediate safety and comfort concerns of daily commuters.

Current passenger-flow management systems rely on tra- ditional methods that lack the flexibility to handle dynamic overcrowding scenarios. This paper presents an innovative solution leveraging advanced computer vision technology, specifically the YOLOv3 algorithm, to enable real-time pas- senger counting within train compartments. By analyzing live CCTV footage, the system provides real-time occupancy data, allowing passengers to make informed boarding decisions and optimizing space utilization. Additionally, the technology enhances security by monitoring onboard behavior, deterring theft, and preventing misconduct. Compared to infrastructure- heavy alternatives, this approach is cost-effective, faster to implement, and offers a scalable solution to Mumbai's persis- tent railway congestion. Integrating real-time data analysis into the suburban railway system will not only improve passenger safety but also contribute to a more efficient and organized commuting experience for millions of Mumbaikars.

II. LITERATURE SURVEY

The table provides a comparative overview of various re- search studies on crowd counting and passenger detection in public transport systems, focusing on models like YOLOv3, YOLOv4, YOLOv5, Faster R-CNN, and EspiNet. It highlights the use of diverse datasets such as PAMELA-UANDES, COCO, Hollywood Heads, and Scuthead to evaluate these models. The studies examine parameters like camera setup, environmental adaptability, and object detection accuracy, with performance metrics including precision, recall, F1-scores, and mAP. Key findings show that YOLO-based models, especially YOLOv5, significantly improve detection rates, particularly for small heads in crowded scenes, with detection rates im- proving from 66% to 78%. Overall, the research underscores the potential of deep learning models to enhance real-time crowd counting and passenger flow management, while also acknowledging challenges related to complex environments and varying object sizes.

TABLEI
COMPARISON OF PEOPLE DETECTION AND COUNTING MODELS

Paper Title	Dataset	Models	Performance Metrics	Observations
EspiNet: Passenger Boarding Analysis [1]	PAMELA UANDES (348 videos)	EspiNet, Faster RCNN, YOLOv3	F1 Score: Over 95% (322 videos), MOTA: 80%, MOT Challenge benchmark	Tracking accuracy was affected by platform gaps and door widths (800mm, 1600mm). YOLOv3 performed well, but Faster R-CNN struggled with real-time execution. High passenger density caused occasional miscounts, but overall, the method improved the understanding of boarding patterns.
HOG-SVM Based Passen- ger Counting [2]	PAMELA Train & Bus Dataset	HOG, SVM	Detection Accuracy: 91.2% - 86.24%, Counting Errors: 10% - 13%	The system worked well under different lighting but struggled with closely packed passengers. Automated counting reduced manual effort but lacked the efficiency of deep learning, making real-time deployment difficult.
Deep Learning for Trans- port Crowd Analysis [3]	Pereira Transport System	YOLOv3-tiny, YOLOv4, YOLOv4-tiny, YOLOv8	YOLOv3-tiny:mAP:33.1, FPS 220; YOLOv4: mAP 43.5, FPS 65; YOLOv4-tiny: mAP 22, FPS 443; YOLOv8: mAP :50.2, FPS 500; Precision: 96.8%, Recall: 92.0%, F1- score: 94.4%	YOLO-based models performed well in real-time. YOLOv8 was the fastest and most accurate. Hardware adjustments improved processing. Accuracy was better for dispersed crowds but less consistent in highly dense areas.
Object Detection for Pub- lic Surveillance [4]	COCO, VOC, geNet	Faster R-CNN, SSD, YOLOv8	Model trained on NVIDIA GTX 1650 Ti (4GB RAM)	YOLO efficiently detected and counted people, though small or distant individuals were challenging. Preprocessing and post-processing refinements improved results, but accuracy depended on crowd distribution. Reliable for real-time low-power applications.
Head Detection in Crowds Using YOLO [5]	Casablanca, Hollywood Heads, Scuthead, Merge Dataset	YOLOv5	Casablanca mAP: 83%; Hollywood Heads mAP: 70.9%; Scuthead mAP: 6% (small heads); Overall mAP: 83%; Precision: 95%	YOLOv5 enhanced crowd counting using head detection, particularly in dense settings. The model detected small heads with 78% accuracy (vs. 66% in Fast RCNN).Hollywood Heads dataset accuracy improved from 56% to 75%. mAP increased from 66% to 78% for Scuthead dataset. A robust pre-trained model was developed for various head sizes across datasets.

III. PROPOSEDSYSTEM

To address the challenge of managing passenger movement in congested train compartments, this research proposes a system that leverages real-time video analytics and machine learning to monitor crowd density and optimize passenger distribution.

The system processes video frames using the YOLOv3 object detection model to identify passengers and estimate occupancy levels within different sections of a train compartment. Each detected passenger is marked with a bounding box, enabling an accurate assessment of the crowd density at any given time.

To classify congestion levels, the system categorizes crowd density into three predefined levels—High, Medium, and Low—based on the number of passengers detected per frame. This processed data is then integrated into a mobile application, which provides real-time crowd updates, train route maps, communication features, and alerts to enhance the commuting experience. By offering passengers real-time congestion insights, the system facilitates informed decision-making, thereby promoting a more balanced distribution of passengers across compartments.

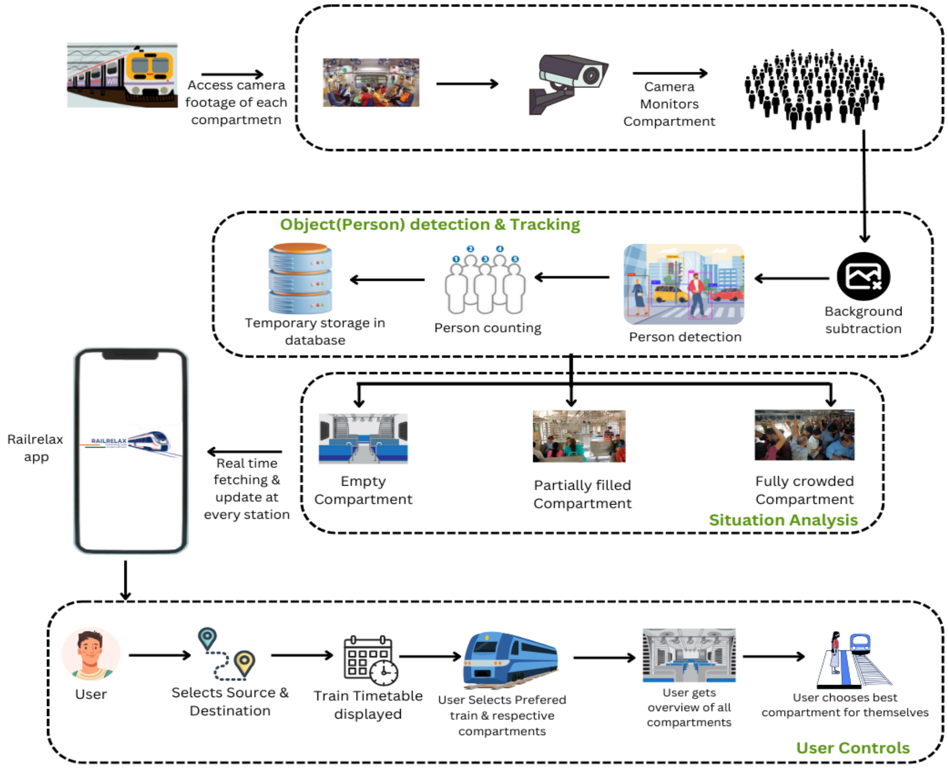

Fig. 1. System Pipeline Structure

A. Collection of Dataset and Preprocessing

To train and evaluate the proposed system, a dataset of annotated images was collected using the Roboflow platform. A total of 1,057 images were sourced from pre-existing datasets, ensuring a diverse representation of real-world train compartment scenarios. These images contained varying levels of crowd density to improve the robustness of the detection model.

Since annotation labels were inconsistent across different datasets (e.g., some labeled individuals as ”person” while others used ”people”), all labels were standardized to a single class, ”person”, to maintain uniformity during model training. This standardization ensured that the object detection model could generalize well across different environments and light- ing conditions.

The data preprocessing phase ensures that raw video data is transformed into a format suitable for efficient and accurate model training. The key steps include:

• Frame Extraction: Video streams are broken down into individual frames for processing.

• Normalization and Resizing: Frames are resized and normalized to optimize computational efficiency without compromising accuracy.

• Background Subtraction: Techniques such as Gaussian Mixture Models (GMM) or deep learning-based methods are applied to distinguish passengers from the background and detect moving objects.

B. Training

To put the plan into action, different versions of YOLO were examined and tested to find out which one works best for identifying how crowded train compartments are in real-time. The results from these tests are shown in Table II. Measurements like accuracy, recall, and mAP@0.50 were used to compare YOLOv4-tiny, YOLOv8, YOLOv11, YOLO-NAS, and YOLOv3.

The models were tested using labeled images from video streams to identify passengers and assess crowd density within the compartments. These tests helped us find the most suitable model for practical use.

While advanced models like YOLOv8 and YOLO-NAS showed greater accuracy, YOLOv3 was chosen for its practical advantages. One main reason for this decision was its recall rate, which was high enough to detect most passengers in crowded compartments, as shown in Figure 2. YOLOv3 is perfect for real-time use on mobile devices and embedded systems because it is lightweight and uses minimal processing resources. The YOLOv3 model achieved a decent compromise between efficiency and detection performance, while the other models required significantly more processing power. As seen in Figure 3, YOLOv3 effectively detects passengers in train cabins. Figure 2 further supports this choice by demonstrating the model’s recall effectiveness.

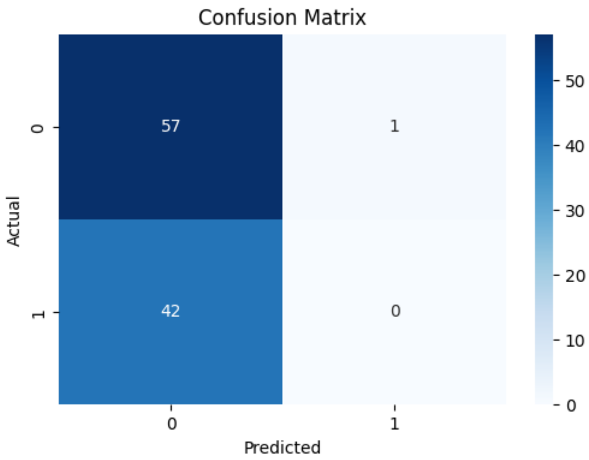

Fig. 2. Yolov3 Model: Confusion Matrix

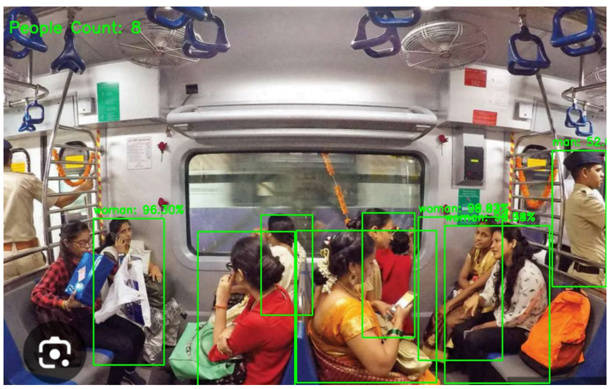

Fig. 3. YOLOv3 Model: People Detection

TABLE II
RESULTS OF YOLO MODELS

Model	Precision (%)	Recall (%)	mAP@0.50 (%)
YOLOv4-tiny	84.0	51.0	62.20
YOLOv8	83.8	68.1	78.2
YOLOv11	81.7	68.3	77.6
YOLO-NAS	26.08	84.44	75.49
YOLOv3	54.0	99.6	35.0

The main performance measurements in Table II are calculated using these equations:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{1}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{2}$$

$$\text{mAP@0.50} = \frac{1}{\text{N}} \sum_{N}^{i=1} AP_i \tag{3}$$

In these formulas, TP represents True Positives, FP represents False Positives, and FN represents False Negatives, based on the model's predictions. APi represents the Average Precision for each category.

C. Crowd Estimation

The RailRelax system estimates crowd density inside Mumbai local train compartments using a YOLOv3-based object detection model. Each compartment is monitored by two cameras that capture images every five minutes, synchronized with station stops. The model detects passengers in these images, and the highest count from the two cameras is recorded in Firebase as the final occupancy for that compartment. By leveraging deep learning, the system effectively identifies passengers while overcoming challenges like occlusion and varying illumination conditions. Crowd density is then classified into three levels-High, Medium, or Low - based on passenger count, facilitating real-time crowd monitoring.

To enhance accuracy, the YOLOv3 model is trained on a dataset that replicates train interiors, ensuring it generalizes well to real-world conditions. Captured images undergo preprocessing techniques such as background subtraction and normalization before detection. The system continuously improves its detection capabilities through periodic model updates and data augmentation. This iterative approach ensures that the RailRelax system provides accurate and reliable occupancy data, aiding in better passenger distribution across compartments.

D. Live Occupancy Monitoring

The RailRelax smartphone application retrieves processed occupancy data from Firebase every five minutes, providing real-time updates to passengers. The app categorizes compartments using a color-coded system - Green for empty, Orange for partially filled, and Red for fully occupied - allowing commuters to make informed boarding decisions. The hierarchical database structure stores both individual camera counts and the final compartment occupancy, ensuring efficient data retrieval. This live monitoring system enhances commuter convenience by eliminating uncertainty and optimizing train capacity.

For seamless connectivity and performance, the system utilizes Firebase Realtime Database for low-latency data exchange. API endpoints facilitate communication between cameras, processing units, and the mobile app, ensuring a steady flow of crowd data. Additionally, role-based access control and Firebase authentication secure the system from unauthorized access, preserving data integrity. By integrating these components, the RailRelax system offers an efficient, real-time crowd monitoring solution tailored to Mumbai's local trains.

IV. CONCLUSION AND FUTUREWORK

RailRelax presents a data-driven approach to enhancing the daily commute experience for Mumbai's local train passengers. By leveraging real-time crowd density analysis and user feedback, the system aims to provide commuters with essential insights for making informed travel decisions. The integration of heatmaps, predictive analytics, and accessibility features ensures a more efficient and comfortable journey.

Future improvements may include expanding the system to cover more railway lines, integrating AI-based predictive modeling, and incorporating additional real-time metrics like weather conditions and train delays. RailRelax has the potential to revolutionize urban commuting by making train travel smarter, more convenient, and user-centric.

REFERENCES

[1] Steffen, J. H. (2008). ”A statistical mechanics model for free-for-all airplane passenger boarding.” Journal of Air Transport Management.

[2] Prasad, D., Kapadni, K., Gadpal, A., Visave, M., & Sultanpure, K. (2021). ”HOG, LBP, and SVM based traffic density estimation at intersection.” arXiv preprint.

[3] Pereira,J.,daSilva,R.,&Santos,F.(2022).”Real-timecrowdcounting using YOLO models for public transport monitoring.” IEEE Transactions on Intelligent Transportation Systems.

[4] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). ”SSD: Single Shot MultiBox Detector.” European Conference on Computer Vision.

[5] Zhang, X., Yu, L., & Wang, J. (2020). ”Head detection-based crowd counting using YOLOv5 for dense urban environments.” Pattern Recognition Letters.

[6] G. Curiel, K. Guerrero, D. Gomez, and D. Charris, ”An Improved Architecture for Automatic People Counting in Public Transport using Deep Learning,” 2023.

[7] C. Wiboonsiriruk, E. Phaisangittisagul, C. Srisurangkul, and I. Kumazawa, ”Efficient Passenger Counting in Public Transport Based on Machine Learning,” 2023.

[8] S. Velastin, R. Fernandez, and J. E. Espinosa Oviedo, ”Detecting, Tracking and Counting People Getting On/Off a Metropolitan Train Using a Standard Video Camera,” 2020.

[9] A. Pankaj, ”A Comprehensive Review on Real-Time Object Detection using Deep Learning Models,” 2022.

[10] H. Khan, H. Yousaf, M. Murtaza, and F. Velastin, ”Video analytics using deep learning for crowd analysis: A review,” *Journal of Visual Communication and Image Representation*, 2019.