Abnormal Activity Detection and Classification of Bus Passengers with In-Vehicle Image Sensing

ROHINI A
(PG Student, Department of CS&E, S.J.M.I.T, Chitradurga, Karnataka, INDIA, and Lecturer, Department of CS&E, *Government Polytechnic, Hiriyur.*
Email: arohini.a1@gmail.com,)
Dr. ARAVINDA T.V
(Professor, Department of CS&E, SJMIT, Chitradurga,Karnataka, INDIA,
Email: traravinda@gmail.com*)*
Dr. KRISHNAREDDY K.R
*(*Professor & HOD, Department of CS&E, S.J.M.I.T, Chitradurga, Karnataka, India,
Email: krkrishnareddy69@gmail.com*)*
Prof. RAMESH B.E
*(*Assistant Professor, Department of CS&E, S.J.M.I.T, Chitradurga, Karnataka, India,
Email: be.ramesh@gmail.com*)*
NAYANA R
*(*Senior Grade Lecturer , Department of CS&E, Government Polytechnic , Chitradurga, Karnataka, India,
Email: arohini.a1@gmail.com*)*

Abstract: Traditional video surveillance systems in public transportation rely heavily on manual monitoring, which is time-consuming, error-prone, and inefficient in detecting anomalous passenger behavior in real time. This project proposes an intelligent, automated system for detecting passenger actions and identifying anomalies within a bus environment using deep learning techniques. The proposed system integrates YOLOv8, a real-time object detection algorithm, to identify and track passengers from live video streams. Once passengers are detected, the system extracts individual regions of interest and classifies their actions using a fine-tuned VGG16 convolutional neural network. The model is trained to distinguish between normal behaviors (e.g., sitting, standing) and anomalous actions (e.g., falling, squanting), enabling timely alerts for safety interventions. This dual-stage approach—combining spatial detection with semantic classification—ensures accurate and robust identification of potentially hazardous situations. The system supports real-time processing and can be enhanced with alert mechanisms such as visual warnings, sound alarms, or event logging. By automating the detection of unsafe or unusual behavior, the system significantly improves passenger safety, reduces the burden on human operators, and contributes to smarter public transportation surveillance systems.

Index Terms: YOLOv8, VGG16, falling, sitting, Video surveillance

I. INTRODUCTION

Public transportation systems, such as buses, are integral to urban mobility, but ensuring the safety and well-being of passengers remains a persistent challenge. Monitoring passenger behavior is essential for identifying incidents like falls, fights, fainting, or other anomalies that could endanger individuals or disrupt service. Traditionally, this monitoring has relied on human surveillance through CCTV systems, which is both labor-intensive and prone to human error due to fatigue or inattentiveness. With the advent of deep learning and computer vision technologies, automated systems can now observe and analyze visual data with high speed and accuracy. In this context, action anomaly detection using artificial intelligence offers a promising solution to enhance passenger safety and operational efficiency in public transport.

This project presents a real-time Bus Passenger Action Anomaly Detection System that combines two powerful deep learning models:

YOLOv8 (You Only Look Once): a fast and accurate object detection algorithm used to detect and localize passengers within each video frame.

VGG16: a Convolutional Neural Network (CNN) employed to classify the actions of detected individuals as either normal (e.g., sitting, standing) or anomalous (e.g., falling, fighting, fainting).

By integrating object detection and action classification, the system continuously monitors passenger behavior inside the bus. It can identify critical events as they occur and trigger automated alerts or logs for further investigation. This not only helps improve response time in emergencies but also reduces the dependency on manual video monitoring. The proposed system can be deployed on live CCTV footage or bus camera feeds, making it suitable for integration into existing public transport infrastructures. Furthermore, its modular design allows it to be extended for different environments or behaviors with minimal modifications. In summary, this system aims to bring smart surveillance to public transportation, enhancing safety, reducing response times, and promoting a safer commuting experience for passengers.

II. Related Work

Mazzeo, P.L., Contino, R., Spagnolo, P., Distante, C., Stella, E., Nitti,M., Renò,V[1],Knowing an accurate passengers attendance estimation on each metro car contributes to the safely coordination and sorting the crowd-passenger in each metro station. In this work we propose a multi-head Convolutional Neural Network (CNN) architecture trained to infer an estimation of passenger attendance in a metro car. The proposed network architecture consists of two main parts: a convolutional backbone, which extracts features over the whole input image, and a multi-head layers able to estimate a density map, needed to predict the number of people within the crowd image. The network performance is first evaluated on publicly available crowd counting datasets, including the ShanghaiTech part_A, ShanghaiTech part_B and UCF_CC_50, and then trained and tested on our dataset acquired in subway cars in Italy. In both cases a comparison is made against the most relevant and latest state of the art crowd counting architectures, showing that our proposed MH-MetroNet architecture outperforms in terms of Mean Absolute Error (MAE) and Mean Square Error (MSE) and passenger-crowd people number prediction.

Karekla, X., Gkiotsalitis, K., Tyler, N[2] Buses are a form of active transportation and can improve people's well-being. However, their high level of acceleration can make them less attractive to users. Even worse, they can be responsible for severe injuries that require hospitalisation or for the development of fear of falling, particularly experienced by older people. Evidence has shown that, bus acceleration up to 1.0 m/s^2 enables passengers to move in a natural way inside the moving vehicle, hence reducing instability and increasing safety. Although operators might be willing to implement such an intervention, they might also be skeptical about its impact on the operation of a service, such as timetabling, travel times, waiting times, etc. The effect of a safety-driven acceleration limit on the operational characteristics of a round trip of a bus service in London is investigated by this study. Data regarding speed, acceleration and journey time were extracted from the engine of a bus and recorded at 2 Hz. Further computations estimated the passenger waiting times and headways between the examined bus and its preceding and following buses. A vehicle movement model was used to test how these operational characteristics would be affected if the acceleration limit of 1.0 m/s^2 were to be implemented. The results suggest that the journey time of the proposed accessible service would be 6 min longer than the current service and passenger waiting time would increase by 2 min. One additional bus would be required to serve the same number of passengers. A discussion of the results is provided.

Zhou, H., Yuan, C., Dong, N.,Wong, S.C., Xu, P[3], Although public buses have been demonstrated as a relatively safe mode of transport, the number of injuries to public bus passengers is far from negligible. Existing studies of public bus safety have focused primarily on injuries caused by collisions. Surprisingly, limited effort has been devoted to identifying factors that increase the severity of passenger injuries in non-collision incidents. Method: Our study therefore investigated the injury risk of public bus passengers involved in collision incidents and non-collision incidents comparatively, based on a police-reported dataset of 17,383 passengers injured on franchised public buses over a 10-year period in Hong Kong. A random parameters logistic model was established to estimate the likelihood of fatal and severe injuries to passengers as a function of various factors. Results: Our results indicated substantial inconsistences in the effects of risk factors between models of non-collision injuries and collision injuries. The severity of passenger injuries tended to increase significantly when non-collision incidents occurred due to excessive speed of bus drivers, on double-decker buses, in less urbanized

areas, in winter, in heavy rains, during daytime, and at night without street lighting. Elderly female passengers were also found more likely to be fatally or severely injured in non-collision incidents if they lost their balance while boarding, alighting from, or standing on a bus. In comparison, the following factors were associated with a greater likelihood of fatal or severe injuries in collision incidents: elderly female passengers, standing passengers who lost balance, buses out of driver control, double-decker buses, collisions with vehicles or objects, and less urbanized areas. Practical Applications: Based on our comparative analysis, more targeted countermeasures, namely "4E" (engineering, enforcement, emergency, and education) and "3A" (awareness, appreciation, and assistance), were recommended to mitigate collision injuries and non-collision injuries to public bus passengers, respectively.
Xiaoyun, Z.,Dongming, Z.,Xiaobo,Y.,Xinyi,H[4], in some typical bus accidents, the causes of passenger injury have presented some common characteristics, which makes it necessary to improve the design of bus. In general, bus traffic accidents can be divided into two typical forms: emergency braking and rollover, both of which could seriously threaten the life security of passengers. In this article, these two kinds of bus accident are studied through numerical simulation and some improving suggestions on bus design and manufacture are given. According to simulation results, passengers may suffer from serious hurt due to emergency braking acceleration and deformation of bus body caused by the impact with highway guardrail. Improving measures include equipping seats with three-point belts, adding board on the handrail and enhancing the material and structure of side uprights of the bus. Simulation results show that these improving measures can provide better protection for the passengers involving bus traffic accidents, which has verified the rationality of these measures
Shirmohammadi, S., Ferrero, A[5], due to continuing and rapid advances of both hardware and software technologies in camera and computing systems, we continue to have access to cheaper, faster, higher quality, and smaller cameras and computing units. As a result, vision based methods consisting of image processing and computational intelligence can be implemented more easily and affordably than ever using a camera and its associated operations units. Among their various applications, such systems are also being used more and more by researchers and practitioners as generic instruments to measure and monitor physical phenomena. In this article, we take a look at this rising trend and how cameras and vision are being used for instrumentation and measurement, and we also cast a glance at the metrological gauntlet thrown down by vision-based instruments.
Lin, H.Y., Tseng, C.H[6], as the self-driving technology is getting mature for public transportation applications, the safety concern of onboard passengers has become an important issue. It is essential to identify inappropriate or hazardous behaviors of passengers for the vehicles without human operators. In this work, we propose a technique to detect and classify the abnormal activities of passengers in a bus environment. Different from the existing human activity classification algorithms, our approach reduces the occlusion and increases the recognition rate by acquiring images from an overhead vision system. To overcome the increased complexity on feature extraction and classification, an action recognition network for top-view images are proposed by incorporating both spatial and temporal information. An image dataset, BUS-HAR, is generated for practical application scenarios with bus passengers. Experiments using real-world scene images have demonstrated the feasibility of our technique compared to existing approaches. The codes and image dataset are made available publicly at https://github.com/richardkuo1999/passenger-action-recognition.
Zhang, J., Shen, D., Tu, L., Zhang, F., Xu, C., Wang, Y., Tian, C., Li, X., Huang, B., Li, Z [7], bus service is the most important function of public transportation. Besides the major goal of carrying passengers around, providing a comfortable travel experience for passengers is also a key business consideration. To provide a comfortable travel experience, effective bus scheduling is essential. Traditional approaches are based on fixed timetables. The wide adoptions of smart card fare collection systems and GPS tracing systems in public transportation provide new opportunities for using the data-driven approaches to fit the demand of passengers. In this paper, we associate these two independent data sets to derive the passengers' origin and destination. As the data are real time, we build a system to forecast the passenger flow in real time. To the best of our knowledge, this is the first paper, which implements a system utilizing smart card data and GPS data to forecast the passenger flow in real time.

III. METHODOLOGY

The proposed system for Bus Passenger Action Anomaly Detection is developed using a two-stage deep learning pipeline that integrates object detection and behavior classification. The key components of the methodology include:

A. Data Collection and Preprocessing

Dataset Composition: The system requires two types of datasets:

Detection Dataset: Used for training the YOLOv8 model to detect passengers.

Classification Dataset: A labeled dataset consisting of cropped images of passengers performing various actions such as *sitting*, *standing*, *fighting*, *falling*, and *fainting*.

B. Preprocessing Steps:

Resize all classification images to 224x224 pixels.

Normalize pixel values (0–1 or VGG16-specific preprocessing).

Augment images using flipping, rotation, zooming to improve model generalization.

C.Passenger Detection using YOLOv8

Model Used: YOLOv8 (You Only Look Once – Version 8)

Objective: Detect and localize passengers in real-time from CCTV or video feed.

Implementation:

Frame-by-frame analysis is performed.

Detected passengers are marked with bounding boxes.

Each bounding box is cropped as a region of interest (ROI) for classification.

D.Action Classification using VGG16

Model Used: Pre-trained VGG16 CNN fine-tuned on action classification dataset.

Workflow:

Cropped passenger images from YOLO are resized and passed to the VGG16 model.

The model outputs class probabilities for actions:

Normal: Sitting, Standing

Anomalous: Fighting, Falling, Fainting

The predicted class is used to determine if the action is normal or an anomaly.

E. Real-Time Detection System

Integration Pipeline:

The live video stream is processed frame by frame.

YOLO detects all passengers.

Each passenger is cropped and classified by the VGG16 model.

If an anomaly is detected, the system:

Draws a red bounding box.

Displays the action label (e.g., “Falling”).

Optionally triggers a siren alert, logs the incident, or records video.

F. Model Training and Evaluation

YOLOv8:

Trained on a custom dataset or COCO-pretrained weights used.

Evaluated using metrics like mAP (mean Average Precision).

YOLOv8 Steps:

Step 1:Input Preprocessing

Input image is resized to a fixed dimension (e.g., 640×640) and normalized.

$$xnorm = 255x$$

Step 2: Feature Extraction

YOLOv8 uses a a custom Ultralytics backbone with C2f modules.

Feature maps of multiple scales are extracted to detect objects of different sizes.

Step 3: Neck: Feature Pyramid Network (FPN)

Combines features from multiple layers using FPN + PAN (Path Aggregation Network).

Helps in preserving semantic and spatial features:

$$Fout = i = 1\sum nUpsample(Fi) + Concat(Fj)$$

Step 3: **Head: Prediction Layer**
YOLOv8 directly predicts:

- Bounding boxes: (x,y,w,h)
- Objectness score
- Class probabilities

🎯 Bounding Box Regression
YOLO predicts the bounding box as:

$$x = \sigma(t_x) + c_x$$
$$y = \sigma(t_y) + c_y$$
$$w = p_w \cdot e^{t_w}$$
$$h = p_h \cdot e^{t_h}$$

Where:

- (t_x,t_y,t_w,t_h) are predicted offsets.
- $(c_x,c_y$ is the top-left coordinate of the grid cell.
- (p_w,p_h) are prior (anchor) box sizes.
- σ is the sigmoid activation.

Step 5:Objectness Score

The objectness score indicates the confidence that a bounding box contains an object:

$$\text{Objectness} = \sigma(t_\circ)$$

Where t_o is the raw objectness logit from the head

Step 6: Class Probabilities

Each bounding box also predicts class probabilities using softmax or sigmoid:

$$P_{\text{class}_i} = \frac{e^{z_i}}{\sum_{j=1}^{C} e^{z_j}} \quad \text{(for softmax)}$$

Or for multi-label classification

$$P_{\text{class}_i} = \sigma(z_i)$$

Step 7: Loss Function

YOLOv8 uses **a** composite loss function

$$\mathcal{L} = \lambda_{\text{box}} \cdot \mathcal{L}_{\text{box}} + \lambda_{\text{obj}} \cdot \mathcal{L}_{\text{obj}} + \lambda_{\text{cls}} \cdot \mathcal{L}_{\text{cls}}$$

Box Loss (CIoU / DIoU / GIoU)
For bounding box regression

$$\mathcal{L}_{\text{box}} = 1 - \text{CIoU}(b, \hat{b})$$

Where CIoU (Complete IoU) includes:

- IoU overlap
- Center distance penalty
- Aspect ratio consistency

Objectness Loss

Binary cross-entropy:

$$\mathcal{L}_{\text{obj}} = -[y \log(p) + (1 - y) \log(1 - p)]$$

Classification Loss

Cross-entropy for multi-class classification:

$$\mathcal{L}_{\text{cls}} = -\sum_{i=1}^{C} y_i \log(p_i)$$

Step 8: Non-Maximum Suppression (NMS)
Removes duplicate detections based on IoU threshold**:**

$$\text{IoU}(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

If IoU between two boxes > threshold (e.g., 0.5), the one with lower confidence is suppressed.

Step 9. Output
For each frame, YOLOv8 outputs:

- Bounding boxes: x,y,w,h
- Confidence score
- Class label

VGG16 Classifier:
Fine-tuned on action dataset.
Evaluated using accuracy, confusion matrix, precision, recall, and F1-score.

G. Alert
Alerts the system if anomaly detects

IV. SYSTEM ARCHITECTURE

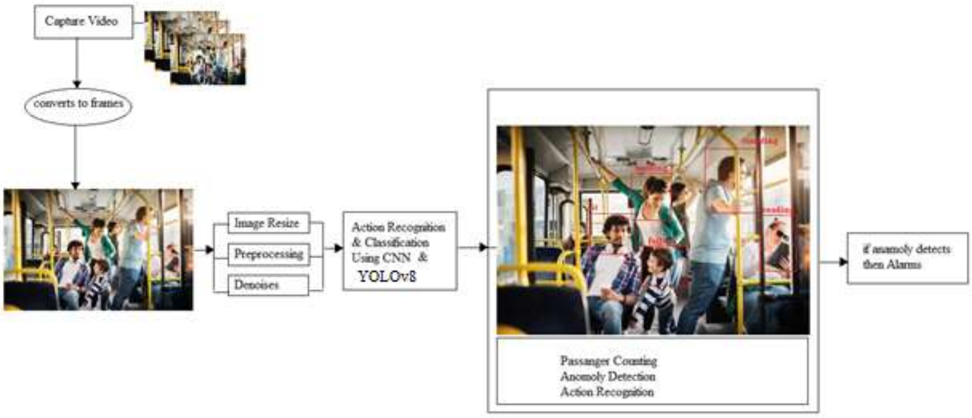


Figure 2: System Architecture

System Architecture working procedure :
Video Input Module: Captures real-time video feed from a CCTV or webcam inside the bus. Used OpenCV, cv2.VideoCapture(). Supports both live stream and recorded video files. Converts frames for downstream processing.
Person Detection Module (YOLOv8): Detects and localizes passengers in each video frame. Ultralytics YOLOv8, torch, onnx, or cv2.dnn. Load pretrained YOLOv8 model. For each frame, detect people and get bounding boxes.
Crop each detected person from the frame for classification.
Action Classification Module (VGG16) : Classifies the cropped passenger image as performing a normal or anomalous action. Classes used Fine-tuned VGG16 CNN with softmax output layer. Action recognises like Sitting, Standing. Anomalous detection like Falling, Fighting
Steps:
Resize image to 224x224
Normalize and feed into the VGG16 model
Get class prediction and confidence
Anomaly Detection & Alert Module: Identifies if an action is abnormal and triggers appropriate response.
Features:
Draw bounding boxes with labels:
Green: Normal
Red: Anomaly
Optional alert mechanisms:
Play siren (via playsound)

V. RESULTS AND DISCUSSION

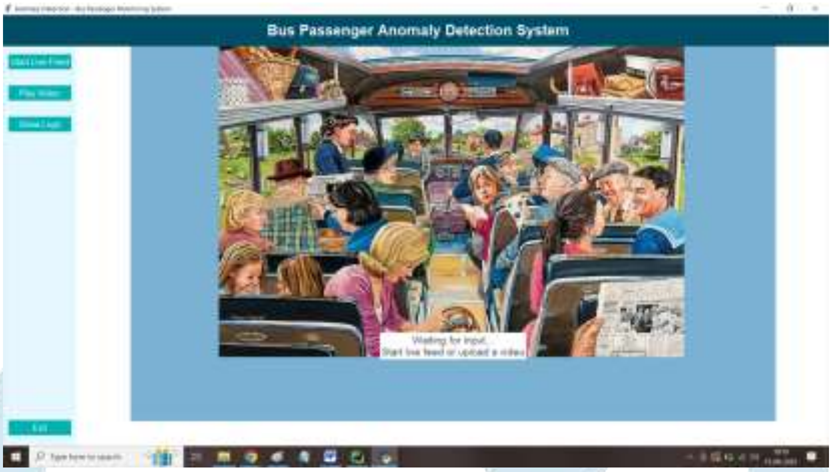
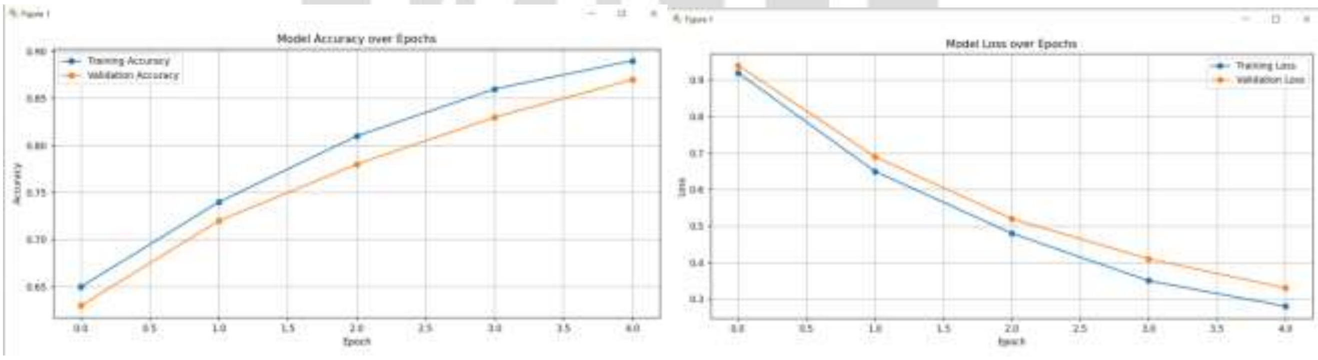


Figure 3: Menu

Figure 4:Action Recognition using

This module: Detects and localizes passengers in each video frame. Classifies the cropped passenger image as performing a **normal** or **anomalous** action using VGG16



Graph 1: Model accuracy and Loss graph while training over epoch

VI. CONCLUSION

The passenger safety issues under unmanned operations are important for the development of autonomous public transport vehicles. In this paper, we propose the algorithms for passenger detection, action recognition using convolutional neural networks. To have a complete coverage of the passenger region, a pair of ceiling-mounted cameras is adopted for top-view image acquisition.

REFERENCES

1. Mazzeo, P.L., Contino, R., Spagnolo, P., Distante, C., Stella, E., Nitti,M., Renò,V.: Mh-metronet—amulti-head cnn for passengercrow attendance estimation. J Imaging 6(7), 62 (2020)
2. Karekla, X., Gkiotsalitis, K., Tyler, N.: The impact of a passengersafety-driven acceleration limit on the operation of a bus service.Accid Anal Prev 148, 105790 (2020)
3. Zhou, H., Yuan, C., Dong, N.,Wong, S.C., Xu, P.: Severity of passenger injuries on public buses:Acomparative analysis of collision injuries and non-collision injuries. J Saf Res 74, 55–69 (2020)

4. Xiaoyun, Z.,Dongming, Z.,Xiaobo,Y.,Xinyi,H.: Improvement on design andmanufacture of bus based on traffic accident prevention. Int J Crashworthiness 24(5), 523–532 (2019)
5. Shirmohammadi, S., Ferrero, A.: Camera as the instrument: the rising trend of vision based measurement. IEEE InstrumMeas Mag 17(3), 41–47 (2014)
6. Lin, H.Y., Tseng, C.H.: Abnormal activity detection and classification of bus passengers with in-vehicle image sensing. IEEE Access (2024)
7. Zhang, J., Shen, D., Tu, L., Zhang, F., Xu, C., Wang, Y., Tian, C., Li, X., Huang, B., Li, Z.: A real-time passenger flow estimation and prediction method for urban bus transit systems. IEEE Trans Intell Transp Syst 18(11), 3168–3178 (2017)
8. Kao, S.F., Lin, H.Y.: Passenger detection, counting, and action recognition for self-driving public transport vehicles. In: 2021 IEEE Intelligent Vehicles Symposium (IV), pp. 572–577 (2021)