Research Article *Open Access*

Journal of Scientific Engineering Advances

ISSN: 3071-0723

DOI: doi.org/10.63721/26JSEA0158

Deployment-Oriented Passenger Occupancy Estimation in Double-Decker Buses for Public Transport Monitoring

Binwei Chen[1], Xuanming Chen[1], Leung Yuk Frank Lam[1], Chi Ho Wong[2], and Wai Keung Anthony Loh[2]*

[1]Department of Chemical and Biological Engineering, The Hong Kong University of Science and Technology, Hong Kong, China

[2]Division of Science, Engineering, and Health Studies, School of Professional Education and Executive Development, The Hong Kong Polytechnic University, Hong Kong, China

Citation: *Binwei Chen, Xuanming Chen, Leung Yuk Frank Lam, Chi Ho Wong, Wai Keung Anthony Loh (2026) Deployment-Oriented Passenger Occupancy Estimation in Double-Decker Buses for Public Transport Monitoring. J. of Sci Eng Advances 2(5) 1-8. WMJ/JSEA-158*

Abstract

Efficient urban traffic management and transit HVAC optimisation depend heavily on systematic bus passenger monitoring, a task increasingly delegated to edge-deployed AI. Yet, numerous factors frequently influence the probabilistic confidence output in dynamic transit environments. This study strategically monitors, compares, and adapts the passenger monitoring via several key dimensions: regional supervision, resolution sensitivity scaling, data-efficient learning, inference calibration, and probabilistic confidence outputs in Artifiical intelligence. We investigate the latent bridges within the chosen AI network to understand how these coordinated setups enhance model robustness while systematically identifying and eliminating extraneous configurations. Our multi-vector optimization reveals that integrating uncertainty-aware confidence outputs with adaptive resolution sensitivity unleashes the potential of a deployment-oriented framework, advancing intelligent transportation and providing an adaptive framework for dependable crowd analytics.

***Corresponding author**: Wai Keung Anthony Loh, Division of Science, Engineering, and Health Studies, School of Professional Education and Executive Development, The Hong Kong Polytechnic University, Hong Kong, China.

Submitted: 03.09.2026 **Accepted**: 07.09.2026 **Published**: 24.09.2026

Introduction

Passenger occupancy inside public transportation vehicles changes continuously with route, station, and temporal demand variations, making real-time occupancy estimation challenging. Nevertheless, such information is valuable for operational decision-making, helping operators better understand vehicle utilization and supporting applications such as passenger load monitoring, service planning, crowd management, and occupancy-aware HVAC control [1-3]. Existing occupancy monitoring methods remain limited [1,4-7]. Door counters only record boarding and alighting, while seat sensors increase hardware and maintenance costs. Vision-based regression or density-counting methods can estimate total occupancy but lack spatial localization [5,8-10]. Although vision-based object detection methods can provide explicit passenger localization, traditional two-stage detectors are often too computationally expensive for low-latency edge deployment [11-13]. Unlike open-road or station-level passenger monitoring, in-cabin occupancy estimation in double-decker buses must operate in a confined and highly structured visual environment [14-16]. Seat backs, handrails, narrow aisles, and dense passenger arrangements frequently obscure passengers' heads and upper bodies, while near-field passengers may block those seated farther from the camera. In addition, vehicle vibration, changing illumination, back-lighting, and camera viewpoint jitter introduce motion blur and appearance variation. These conditions make reliable in-cabin occupancy estimation a system-level problem rather than a detector-only task, where data preparation, model configuration, and inference calibration can all affect final counting stability [17].

Despite the growing use of vision-based detection methods for passenger monitoring, existing studies often focus mainly on model accuracy, while giving less attention to the factors that affect counting stability in real in-cabin deployment [1,4,14]. In double-decker bus cabins, where occlusion is severe, camera viewpoints are fixed, and labeled data are limited, annotation strategy, training-data scale, input resolution, and inference-time thresholding or non-maximum suppression (NMS) settings can all directly influence the final occupancy estimate. In addition, many studies report object-detection metrics such as mA, precision, and recall, but provide limited analysis of whether these detections can produce stable passenger counts and occupancy categories under practical operating conditions [18,19]. Therefore, this study develops a deployment-oriented framework for passenger occupancy estimation in double-decker bus cabins. A You Only Look Once (YOLO)-based detector is adopted for passenger localization [11,12,15,20]. However, the focus is placed on system-level design and deployment reliability rather than detector selection alone. The study systematically evaluates how annotation strategy, data scale, input resolution, and post-processing calibration affect detection and counting performance [16]. The final outputs include both frame-level passenger counts and groupwise occupancy estimates, providing practical references for public transport cabin monitoring and other occupancy-aware applications.

Methods

Task Definition: from Detection to Occupancy Sensing

This study formulates passenger monitoring in double-decker bus cabins as an occupancy estimation task. For each in-cabin image I, the system outputs:

$$I \rightarrow (\hat{N}, \hat{G}) \tag{1}$$

where $\hat{N}$ is the estimated passenger count, and $\hat{G}$ is the corresponding occupancy group.

$$\hat{G} = f(\hat{N}), \quad f(\hat{N}) \in \{0-3,\ 4-7,\ 8-11,\ 12-15, \ldots\} \tag{2}$$

This grouped output reflects practical monitoring needs, where stable load categories are often more useful than exact adjacent-count distinctions[1]. Passenger localization is therefore treated as an intermediate step, while the final objective is to produce reliable passenger counts and occupancy groups under occlusion, crowding, fixed viewpoints, and illumination variation.

Supervision Design for Crowded Cabins

In crowded bus cabins, annotation was treated as a supervision-design problem because occlusion and passenger overlap can make the learning target inconsistent [8,21].

Three supervision forms were considered: rectangle-only labels, polygon-to-box labels, and mixed labels. Rectangle-only labels provide a unified detection target.

Polygon labels describe visible regions more closely but must be converted into boxes for detector training. Mixed labels use different annotation geometries within the same dataset. Generative images are used.

Occupancy Estimation Pipeline

The pipeline converts each in-cabin image into two occupancy outputs: total passenger count and groupwise occupancy level.

A YOLO-based detector is used as the instance localization module. Its detections are processed by confidence/ Intersection over Union (IoU) calibration and non-maximum suppression (NMS), then aggregated into the estimated passenger count $\hat{N}$. The count is finally mapped into the groupwise occupancy level $\hat{G}$ using four-passenger intervals.

The detector is therefore an intermediate component, while the final objective is stable occupancy estimation for public transport monitoring.

Deployment-Oriented Configuration

Deployment-related settings were defined around model scale, input resolution, training budget, and inference complexity [11,12,20]. Three detector scales were included: YOLOv8n, YOLOv8s, and YOLOv8m.

All models followed the same training protocol for fair comparison. The default setting used 80 epochs, a batch size of 8, and an input size of 640 × 640 pixels. Additional resolutions and extended training epochs were used to assess deployment sensitivity.

The main training and inference settings are summarized in Table 1.

Table 1: Training and Inference configuration

Category	Configuration
Model variants	YOLOv8n/s/m
Input size	640 default; 600/800/960 for resolution analysis
Epochs	80 default; extended training for budget analysis
Batch size	8
Optimizer	SGD, cosine LR schedule
Augmentation	Mosaic, copy-paste, translation, scaling, shear, perspective
Inference control	Confidence/IoU calibration; NMS variants

Counting Reliability Evaluation

Performance was evaluated at two levels: detection quality and occupancy reliability. Detection quality was measured by mAP50(B), where mAP denotes mean Average Precision and (B) refers to bounding-box evaluation [18].

For occupancy reliability, the estimated count $\hat{N}$ was mapped to the groupwise occupancy level $\hat{G}$ using four-passenger intervals. Groupwise occupancy accuracy measures whether the predicted and true counts fall into the same occupancy group.MAE, RMSE were used for detailed external-error analysis [22].

Inference-Time Reliability Control

Inference-time reliability control was applied to stabilize the final passenger count after detection. Confidence and Intersection over Union (IoU) thresholds were adjusted to balance missed detections and duplicated detections. Non-maximum suppression (NMS) was used to remove highly overlapping boxes around the same passenger [23-25]. In addition, a density-aware confidence schedule was examined to test whether crowded scenes could be handled more robustly without retraining the detector.

The selected settings were evaluated using counting-oriented metrics, especially groupwise occupancy accuracy and external counting error.

Results and Discussion

Evaluation Context and Baseline Performance

Figure. 1 shows the passenger-count distribution of the primary dataset and external test subset, covering low-, medium-, and high-density cabin scenes. This provides the evaluation context for testing occupancy estimation under different passenger loads.

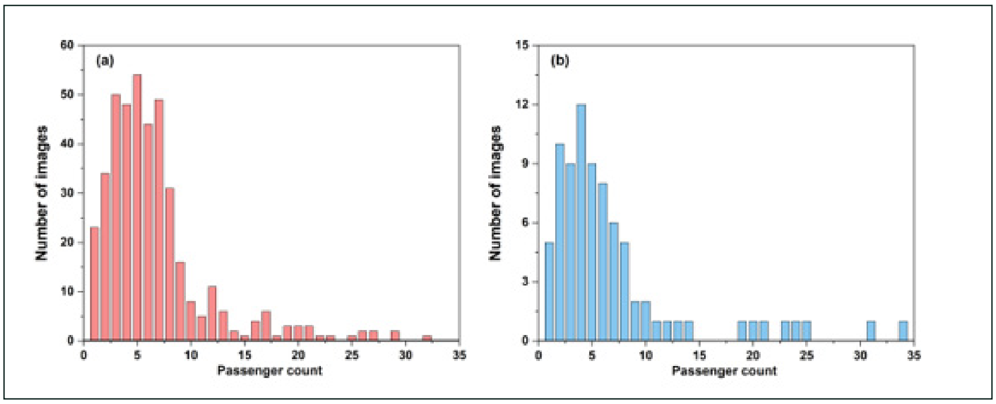

Figure 1: Passenger-Count Distribution in the Primary Dataset and External Test Subset. (a) Primary Dataset Used for Model Development. (b) External Test Subset Used for Independent Evaluation. The Distributions Show the Passenger-Density Range Covered in the Evaluation.

Three detector variants, YOLOv8n, YOLOv8s, and YOLOv8m, were compared under the same training protocol. As shown in Table 2, the medium-scale variant achieved the highest mAP50(B) and recall, while the smaller variants required less training time. Therefore, the medium-scale detector was selected as the main localization baseline, with the small-scale detector retained as a lightweight fallback option.

Table 2: Baseline Performance of Detector Variants

Model	mAP50(B)	Recall	Training time (h)
YOLOv8n	0.9293	0.8616	0.259
YOLOv8s	0.9387	0.8608	0.308
YOLOv8m	0.9423	0.8818	0.642

Supervision Quality and Data Efficiency

Supervision quality was evaluated using a rectangle-only setting with 412 images and a mixed-supervision setting with 319 images. As shown in Figure. 2(a), rectangle-only supervision achieved higher mAP50(B) and groupwise occupancy accuracy, indicating more stable learning signals in crowded cabins.

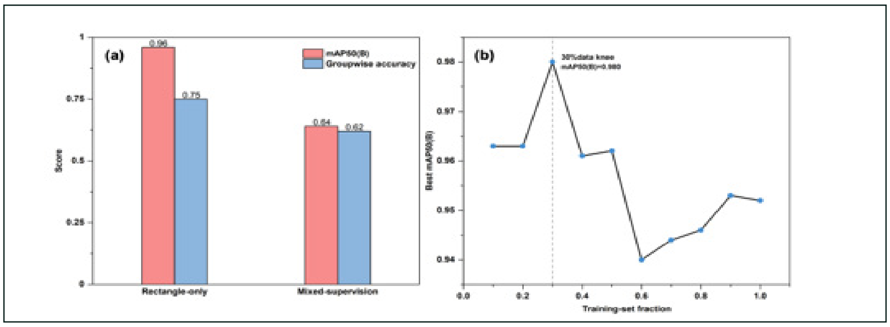

Figure 2: Supervision Quality and Data-Efficiency Analysis.(a) Performance Comparison Between

rectangleOnly and Mixed-Supervision Settings. (b) Best mAP50(B) under Different Training-Set Fractions.

Data efficiency was examined using different fractions of the primary dataset. Figure. 2(b) shows that performance increased rapidly at small data scales and plateaued after approximately 30% of the dataset. This suggests that the current dataset is sufficient for controlled evaluation, while further improvement should focus on targeted dense and low-light samples rather than simply adding more images.

Overall, rectangle-only supervision was adopted for the subsequent analysis.

Deployment Configuration Trade-Offs

Training budget and input resolution were examined as deployment-related settings. Figure. 3 shows that best mAP50(B) changed only slightly after the first 80 epochs, indicating that 80 epochs provide a practical training budget under the fixed protocol.

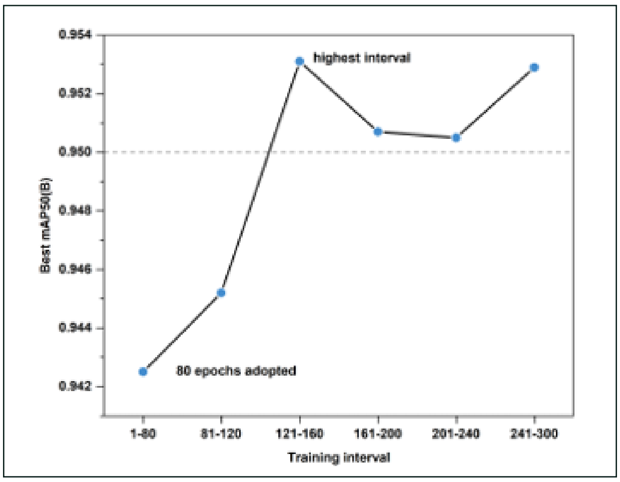

Figure 3: Training-Budget Analysis.Best mAP50(B) Across Training Intervals in the Extended Training Run.

Input resolution was further evaluated using 600, 800, and 960 px settings. As summarized in Table 3, mAP50(B) remained comparable across resolutions, while precision and recall were slightly higher at 960 px. This provides a promising sign that higher-quality image inputs, such as fixed onboard cameras with improved resolution or more stable viewpoints, may further support passenger localization, especially for small or partially occluded passengers.

Table 3: Resolution Sensitivity Under Different Input Sizes

Input size	Precision	Recall	mAP50(B)
600	0.9513	0.9013	0.9587
800	0.9658	0.9122	0.9591
960	0.9704	0.916	0.9571

Overall, excessive training epochs are unnecessary, and input resolution should be selected according to deployment resources and image-quality requirements.

Inference Calibration and External Error Analysis

Inference calibration was evaluated to stabilize the estimated passenger count $\widehat{N}$ and the groupwise occupancy output $\widehat{G}$. Confidence and IoU thresholds were adjusted to balance missed detections and duplicated detections. A stable operating range was found around confidence = 0.12–0.14 and IoU = 0.70–0.75, where groupwise occupancy accuracy remained high.

NMS variants were then compared under the calibrated setting. As summarized in Table 4, standard NMS achieved the best balance, with $\hat{G}$ accuracy of 0.800 and MAE of $\hat{N}$ = 1.050. Soft-NMS and WBF did not improve the final occupancy output, suggesting that more complex suppression strategies were unnecessary for this single-model setting. The density-aware confidence schedule achieved comparable accuracy and may be useful for congestion-prone routes, but fixed thresholding with standard NMS was retained as the default inference setting.

Table 4: Inference Calibration Comparison

Setting	$\hat{G}$ Accuracy	$\hat{N}$ Picture	Interpretation
Standard NMS	0.8	1.05	Default inference setting
Soft-NMS	0.775	1.175	Lower counting stability
WBF	0.7625	1.225	Not beneficial for single-model inference
Density-aware schedule	0.8	1.05	Optional for congestion-prone routes

External testing was used to examine whether the calibrated pipeline remained reliable under practical cabin conditions. On the external subset, the rectangle-only baseline achieved MAE = 1.19, RMSE = 2.19, bias = −1.04, and groupwise occupancy accuracy = 0.75. These results indicate practical $\hat{G}$-level reliability, while larger $\hat{N}$ errors were mainly associated with dense or heavily occluded cabins. Future improvement should therefore focus on overlap-aware inference and targeted dense/low-light data collection.

Promising Deployment Guidelines

The results suggest that deployment reliability is governed less by detector scale alone than by the consistency of the full occupancy-estimation pipeline. Rectangle-only annotation should be retained as the default supervision format, while future data collection should prioritize dense, low-light, and heavily occluded scenes.

For configuration, the 80-epoch setting provides a sufficient training budget under the fixed protocol. Input resolution can be selected according to onboard computing resources and image-quality requirements. At inference, fixed thresholding with standard NMS is recommended as the default setting for estimating $\hat{N}$ and $\hat{G}$, with density-aware scheduling reserved for congestion-prone routes.

Overall, practical deployment should focus on supervision consistency, targeted data coverage, and calibrated inference rather than simply increasing model size.

Conclusions

This study developed a deployment-oriented framework for real-time passenger occupancy estimation in double-decker bus cabins. The framework converts in-cabin images into both frame-level passenger counts and groupwise occupancy estimates, making the output suitable for practical public transport monitoring.

The results show that reliable in-cabin counting depends on the full pipeline rather than detector choice alone. Rectangle-only supervision provided cleaner learning signals, the learning-curve analysis showed that a compact curated dataset can support controlled evaluation, and the configuration and calibration studies confirmed the importance of training settings, input resolution, and confidence/IoU–NMS calibration.Overall, the proposed pipeline provides a reproducible engineering baseline for camera-based passenger occupancy estimation in public transport cabins. Future work should focus on fixed onboard camera validation, targeted dense and low-light data collection, and overlap-aware inference to improve robustness under real deployment conditions.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgments

This work was supported by the Research Grants Council (RGC) of the Hong Kong Special Administrative Region, China under Grant UGC/FDS24/E03/22.

References

1. M Siebert and D Ellenberger (2020) Validation of automatic passenger counting: introducing the t-test-induced equivalence test. Transportation 47: 3031-3045.
2. A Radovan, L Mršić, G Đambić and B Mihaljević (2024) A review of passenger counting in public transport concepts with solution proposal based on image processing and machine learning. Eng 5: 3284-3315.
3. N Zacharof, O Özener, S Broekaert, M Özkan, Z Samaras (2023) The impact of bus passenger occupancy, heating ventilation and air conditioning systems on energy consumption and CO2 emissions Energy 272: 127-155.
4. C Pronello and XR Garzón Ruiz (2023) Evaluating the performance of video-based automated passenger counting systems in real-world conditions: A comparative study. Sensors 23: 77-19.
5. Y Li, X Zhang and D Chen (2023) Csrnet: Dilated convolutional neural networks for understanding the highly congested scenes.1091-1100.
6. C McCarthy, H Ghaderi, F Marti, P Jayaraman and H Dia (2025) Video-based automatic people counting for public transport: On-bus versus off-bus deployment. Computers in Industry 164: 104-195.
7. C Pronello, D Anbarasan, F Spoturno and G Terzolo (2025) A low-cost automatic people-counting system at bus stops using Wi-Fi probe requests and deep learning. Public Transport 17: 71-100.
8. H Bai, J Mao and SHG Chan (2022) A survey on deep learning-based single image crowd counting: Network design, loss function and supervisory signal. Neurocomputing 508: 1-18.
9. H Idrees, M Tayyab, K Athrey, D Zhang, S Al-Maadeed (2022) Composition loss for counting, density map estimation and localization in dense crowds 532-546.
10. Y Zhang, D Zhou, S Chen, S Gao and Y Ma (2016) Single-image crowd counting via multi-column convolutional neural network. 589-597.
11. Y Zhao, W Lv, S Xu, J Wei, G Wang (2024) Detrs beat yolos on real-time object detection. 16965-16974.
12. P Mittal (2024) A comprehensive survey of deep learning-based lightweight object detection models for edge devices. Artificial Intelligence Review 57 : 1.
13. S Ren, K He, R Girshick and J Sun (2015) Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems 28.
14. HY Lin, SF Kao and CC Wang (2024) A passenger detection and action recognition system for public transport vehicles. Journal of Intelligent & Robotic Systems 110: 155.
15. J Zhao, C Li, Z Xu, L Jiao, Z Zhao (2022) Detection of passenger flow on and off buses based on video images and YOLO algorithm. Multimedia Tools and Applications 81: 4669-4692.
16. YW Hsu, YW Chen and JW Perng (2020) Estimation of the number of passengers in a bus using deep learning. Sensors 20: pp. 21-78.
17. Q Wang, J Gao, W Lin and Y Yuan (2020) Learning from synthetic data for crowd counting in the wild. Computer Science 8198-8207.
18. TY Lin, M Maire, S Belongie, J Hays, P Perona (2015) Microsoft coco: Common objects in context. Computer Science 740-755.
19. XY Wei, L Zhang, HY Ma, and XF Zhang (2024) SCU-Counting: A large-scale benchmark dataset for multi-class object counting. Transportation Research Part C: Emerging Technologies 163: 104-608.
20. H Gomes, N Redinha, N Lavado and M Mendes (2022) Counting people and bicycles in real time using YOLO on Jetson Nano. Energies 15: 88-16.
21. Z Ma, X Wei, X Hong and Y Gong (2019) Bayesian loss for crowd count estimation with point supervision. Computer Science 6142-6151.
22. N Cherrier, B Rérolle, M Graive, A Dib and E Schmitt (2024) Context-Aware Automated Passenger Counting Data Denoising. Statistics 1935-1942.
23. R Solovyev, W Wang and T Gabruseva (2021) Weighted boxes fusion: Ensembling boxes from different object detection models. Image and Vision Computing 107: 104-117.
24. N Bodla, B Singh, R Chellappa and LS Davis (2017)

Soft-NMS--improving object detection with one line of code. Computer Science 5561-5569.

25. A Neubeck and L Van Gool (2006) Efficient non-maximum suppression. 850-855.