HKIE THE HONG KONG INSTITUTION OF ENGINEERS 香港工程師學會

Smart Guardian: creating a secure railway through the discovery of suspicious objects

Calvin Cheung, Mickey Yan Po Fong, and Wai Pan Tam
Hong Kong Transport Services Business Unit, MTR Corporation Limited, Hong Kong, People's Republic of China

ABSTRACT

Many railway operations rely on staff vigilance and basic security measures to maintain safety in busy areas like concourses where passengers start their journeys. However, growing passenger numbers and complex railway environments make it harder to spot threats like unattended bags or prohibited items. Safety is paramount in railway operations and critical to public infrastructure. Therefore, an innovative autonomous system called the Smart Guardian, employing artificial intelligence and computer vision, was explored to automatically detect suspicious objects. The system relies on deep learning algorithms trained on railway-specific data to handle crowded and dynamic scenes. It works better than conventional methods because it can overcome challenges like moving foreground objects and busy backgrounds. The system analyses CCTV footage in real time and quickly spots prohibited items and unattended luggage. Once a potential threat is identified, the system immediately alerts station operators, enabling them to take appropriate and timely actions to address the situation, mitigate safety and security risks, and ensure the safety of passengers and staff. With its advanced detection capabilities, the system could become a crucial tool in safeguarding public spaces and transportation while preventing potential security incidents.

KEYWORDS Artificial intelligence; computer vision; video analytic; railway; safety; surveillance
CONTACT Calvin Cheung calvinkscheung@mtr.com.hk
Received 23 February 2026

1. Introduction

Operating a high-capacity railway network requires stringent safety and reliability standards. The train service operates around 19 hours per day, 7 days per week, covering nine main commuter networks and the Airport Express which serve an average of over 5 million passengers daily (MTR, 2024).

Public safety is one of the enabling and critical success factors for being a resilient city, while the adoption of Innovation and Technology under the smart city initiatives is particularly important to create the ability to rapidly adapt to changes and disturbances (ScienceDirect, 2024). Railway stations are not only critical infrastructure, but also busy public areas in Hong Kong with thousands of passengers in a station at a time, and even greater numbers in major interchanging stations. Therefore, it is vital to explore effective measures against serious threats to public safety nowadays. To ensure railway safety, which is also an important element of corporate responsibility and Environmental, Social and Governance (ESG), MTR has various requirements regarding the items that can be carried when travelling on MTR trains; for example, dangerous or over-sized objects are strictly prohibited (MTR, 2024). From a global transportation and city perspective, safety and security are also equally important. According to some statistics, the global expenditure on anti-terrorism reached $1,365 billion in 2017 (Zucchi, 2022). Despite these efforts, terrorism still claimed more than 6,000 lives across the world in 2022 (Global Terrorism Index, 2023). One of the common methods that terrorists use to inflict harm is to plant suspicious unattended objects, such as bombs, flammable chemicals, or poisonous gases in crowded places and detonate them remotely or by using timers, resulting in a high number of casualties. As such, effective and continuous surveillance is essential to ensure public and railway safety.

With the comprehensive planning in station construction and regular reviews on the security provision to uplift the standard, Closed-Circuit Television (CCTV) cameras have been installed in every station to cover major areas. However, there are too many cameras to be monitored, which is a major pain point and challenge for station operators to timely spot suspicious objects, especially when the station is crowded. Therefore, Innovation and Technology was adopted to address the pain point through an automated and intelligent manner. Although there are many Artificial Intelligence (AI) products on the market that have object detection and unattended object detection abilities embedded, they are far from sufficient due to various constraints such as processing capability and detectable object type. According to market research and trials, these products only work well with still backgrounds, detect very limited objects in busy backgrounds, easily lose tracking when occluded by foregrounds, and lack relevant datasets for railway-specific objects (e.g., prohibited items on crowded platforms), differing from general city/road environments.

To address the aforementioned pain point efficiently, an autonomous suspicious object detection system specifically designed for the unique busy railway environment is proposed. Leveraging the power of artificial intelligence, the system employs cutting-edge computer vision techniques and deep learning algorithms (such as Convolutional Neural Networks (CNNs) (Albawi et al., 2017)) to identify potential threats, such as prohibited items, unattended suitcases, handbags, and luggage, via object detection on CCTV footage. It also effectively overcomes the detection difficulties with busy environments, for example dynamic foreground occlusions and complex backgrounds. Datasets are a critical success factor of CNN models. To ensure that sufficient useful and relevant data can be used in the model training, a composite of both real and generated data was adopted. Besides, background modelling was used to filter noisy moving foreground for better tracking of static objects. On the other hand, in order to allow people to react to suspicious objects immediately to prevent serious attacks, fast inference of these CNN models and quick manipulation during pre-processing and post-processing were required. To achieve this, equipment with sufficient processing capability was essential. With the wide availability of compact devices with the demanded capability on the market, the implementation became more practical and effective, especially with the flexibility needed for suiting different station environments.

1.1. Literature review

Prior work on suspicious object detection in surveillance systems has primarily focused on general urban or indoor settings, with limited adaptation to high-density transport environments like railways. Traditional background subtraction techniques, such as the Adaptive Gaussian Mixture Model (AGMM) (Stauffer & Grimson, 1999), enable foreground extraction for motion detection but often fail in dynamic scenes with frequent occlusions, as reviewed by Xu et al. (2016). For instance, Xu et al. (2016) evaluated background modelling methods and found AGMM to be robust to gradual lighting changes but sensitive to sudden crowd-induced perturbations, achieving only ~75% accuracy in crowded videos.

Object detection has advanced with deep learning, particularly You Only Look Once (YOLO) variants. Jocher et al.'s YOLOv5 (Roboflow Universe, 2023) introduced efficient real-time detection via CSPDarknet backbones, outperforming predecessors in regard to speed (up to 140 *fps* on GPUs). While newer models like YOLOv8 (2023) were available during the development of this system, they incorporate complex attention mechanisms and anchor-free architecture designed for better small-object handling. However, deploying continuous pixel-based background modelling (AGMM) in parallel with deep learning inference on live CCTV streams requires the careful management of computational headroom. YOLOv5's streamlined architecture preserves the essential GPU resources required to run AGMM concurrently at 25 *fps* without frame drops, whereas heavier models risk compromising this real-time pipeline. Furthermore, railway surveillance relies on fixed-angle CCTV cameras monitoring objects with predictable aspect ratios, such as suitcases and backpacks. In this context, YOLOv5's anchor-based architecture is highly optimised and efficient. The advanced anchor-free design of YOLOv8, while highly flexible for dynamic camera angles, introduces unnecessary computational overhead with diminishing returns for fixed-view applications. Therefore, YOLOv5 was strategically selected as a baseline. This allowed development efforts to be focused on the novel hybrid foreground filtering pipeline, which yielded the most significant performance gains for this specific environment.

Datasets like COCO or the Person and Luggage Dataset emphasise general luggage but underrepresent railway-specific scenarios (e.g., backpacks near platforms under varying angles/lighting), leading to domain gaps. Hybrid approaches combining background modelling with CNNs improve recall by 15–20% in transport videos, but few studies address duration tracking for unattended items. Commercial AI surveillance tools (e.g., those using off-the-shelf YOLO) perform adequately in static environments but exhibit high false positives (>20%) in busy public spaces due to unfiltered foreground noise (Xu et al., 2016). Gaps persist in railway applications: no prior work integrates AGMM-filtered foregrounds with customised YOLO for occluded, duration-aware detection in real-time CCTV streams at 720p/25 *fps*. This study bridges these gaps by leveraging railway-curated data and hybrid filtering, achieving F1=0.935 in live deployments—surpassing baselines by addressing transport-unique challenges.

The rest of this paper is organised as follows. In Section 2, an overview of the system architecture is provided. Section 3 describes the design of the AI solution. Finally, Section 4 concludes the work and discusses the potential future possibilities.

2. Overview of system architecture

The whole system comprises two parts, namely edge and frontend. Edge refers to the cameras that are installed in different station areas and captures real-time images and transmits them to the CCTV system. The system has a video analytic server, which is a small, compact, but powerful embedded device, that can be flexibly attached to the CCTV system through network connectivity at an interface location that best suits the site environment. The interface location can be at an edge, at a frontend or even in a CCTV central site. This powerful server takes a real-time stream from a CCTV camera via common open protocols such as Real Time Streaming Protocol (RTSP) (Schulzrinne, et al., 1998) and performs suspicious object detection. The server can process up to four concurrent streams (configurable based onsite needs, e.g., number of cameras and functions like ROI filtering), with experiments from preliminary design stages exhibiting an average throughput of 25 fps across streams. It sends out alarms to the frontend whenever a suspicious object is detected. The frontend refers to the monitoring terminal and the respective user interaction in the workflow for monitoring the situation, receiving alarms and producing responses. To enable integration with a third-party video management system at the frontend, the Transmission Control Protocol (TCP) with alarm data in JavaScript Object Notation (JSON) following the REpresentational State Transfer Application Programming Interface (RESTful) protocol (Jiang et al., 2022) was adopted for alarm reporting. Figure 1 shows the video analytic server adopted in the system, Figure 2 illustrates the overall system architecture of the system, and Figure 3 is the frontend user interface.

Figure 1. Video analytic server.

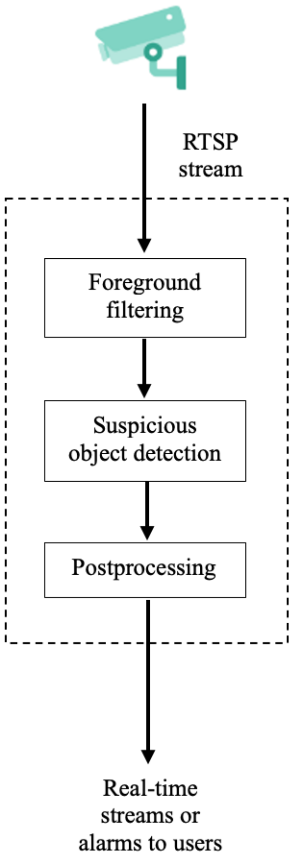

Figure 2. System architecture of the AI solution.

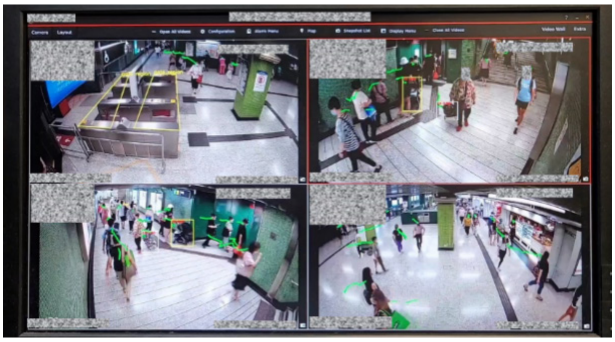

Figure 3. User interface at the frontend.

As a pilot application, this AI solution was implemented in Jordan station. Before the introduction of this system, the monitoring and surveillance workflow mainly relied on human monitoring, for example via the CCTV system or by patrolling. This is not effective during peak hours when the station is crowded and station staff have different tasks to handle. With this solution, the monitoring part can be automated by the AI, and station staff can better arrange the resources and be timely alerted to handle abnormalities detected by the AI, which is part of smart station operations.

3. Design of the AI solution

In this section, the design of the AI solution will be presented. The system comprises the following parts: Foreground Filtering, Enhanced Object Detection Model, Model Training, Performance Evaluation, AI Software architecture, and Design of the frontend system.

3.1. Pain point in the traditional approach

You Only Look Once (YOLO) is a state-of-the-art and real-time object detection system adopted in the traditional approach (Roboflow Universe, 2023). Traditional methods, such as standard YOLO implementations, typically operate in non-crowded environments and struggle with unattended luggage detection, failing to track object duration effectively in busy scenes due to occlusions by moving foreground objects and complex backgrounds—challenges that are unique to railway settings. The traditional object detection dataset is weak in detecting unattended luggage as most luggage is with people (Xu et al., 2016). Figure 4 illustrates the limitations of existing detection methodologies, such as backpacks and suitcases on the floor which cannot be detected, while non unattended items, such as bags carried by people, can be detected. Moreover, even if an object is detected in such busy environment, the traditional approach cannot decide how long the object has remained in place.

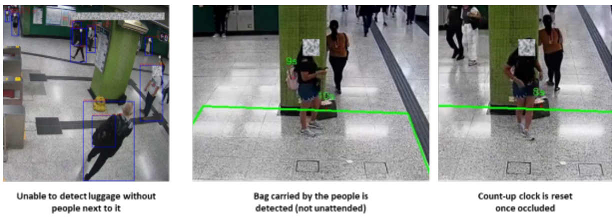

Figure 4. Limitations of detection using existing methodologies.

3.2. Breakthrough by application of this novel AI solution

To develop a solution that addresses the limitations of the traditional approach, a novel integration of foreground filtering is introduced as well as an enhanced model to improve the static object detection and track the unattended duration. The approach introduces novelty through the novel integration of foreground filtering using background modelling (specifically, the Adaptive Gaussian Mixture Model, AGMM (Stauffer & Grimson, 1999)) to filter noisy moving foregrounds in dynamic, crowded railway scenes, combined with an enhanced YOLOv5 model customised using a comprehensive dataset of real and generated railway-specific images. The system overcomes the challenges faced by traditional methods by mapping detected objects across frames even under short occlusions and cumulating the elapsed time for unattended objects.

The technique of background modelling was applied to filter noisy foreground during foreground filtering. In the usual practice, background modelling extracts moving foreground. By subtracting the background, the foreground mask can be obtained. This process is called foreground detection (Stauffer & Grimson, 1999). However, in this case, a novel application that uses foreground detection to filter the moving foreground was used. Although the station environment in the pilot station is indoor, the pixel-based Adaptive Gaussian Mixture Model (AGMM) (Blender, 2025) was adopted to facilitate the changes in lighting conditions as well as future expansion to semi-outdoor or outdoor stations. The "cv2.cuda.createBackgroundSubtractorMOG2" with history = 500 and online updating model in run time was adopted. Figure 5 illustrates the results after the application of foreground detection. With the foreground detected, this means that temporarily moving occlusive objects in a busy environment can be filtered.

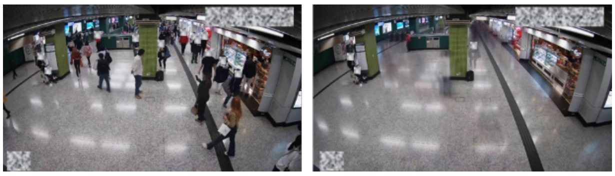

Figure 5. Comparison between before foreground filtering (left) and after (right).

3.3. Enhanced Object Detection Model

The object detection is conducted on the foreground-filtered image. YOLOv5, with the last layer dropped, was adopted but with the enhanced and customised model which was trained with a railway-specified dataset collected, such as the object types that were needed, as well as object image with occlusion and more realistic camera view angles. Obtaining a large quantity of such training data without affecting operations or safety could be difficult or time consuming. In order to have sufficient data for the model training within the project time frame, 3D software was used for generating additional data, which means that both real images and generated images were used. Figure 6 shows an image with backpacks and handbags artificially created by using Blender (Weiss et al., 2016). This helped improve the detection performance of the model for detecting suspicious objects in the scene. Furthermore, detected objects at different frames are mapped to be the same one by noting its object type and position even if the object is occluded for a short time. The composite dataset (real and generated images) enhances detection in real-world CCTV by including occlusions and realistic angles. The domain gap is addressed through data augmentation (e.g., varying lighting/blur), modelling synthetic objects with real-world dimensions (e.g., standard backpack sizes of 40-50 cm in height) to align scales and perspectives with actual CCTV resolutions (1080*p*/720*p*), and fine-tuning of mixed data.

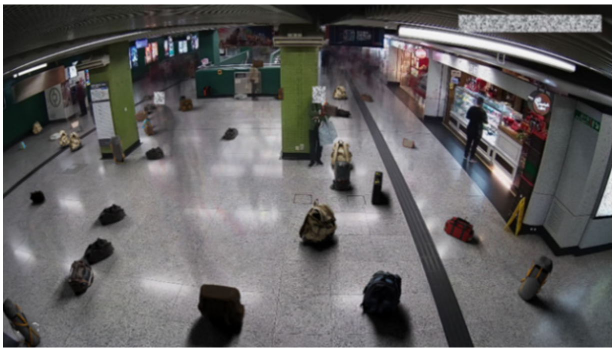

Figure 6. Generated image for model training.

3.4. Model Training

For model training, the process was initiated with pre-trained weights and the model was further trained for 300 epochs using PyTorch. The dataset comprises at least 7,000 images, which were regularly enriched during the project for further tuning of the AI model, with default data augmentation techniques applied. To prevent overfitting, 7% of the data was set aside for validation, and the model with the best Mean Average Precision (mAP) (Henderson & Ferrari, 2017) was selected (refer to Figure 7). If an object remains stationary beyond a predefined threshold, an alarm is triggered. In Figure 7, "@0.5" signifies that an intersection over union (IOU) of 0.5 or greater is considered a correct detection, whereas "@0.5:0.95" indicates the average precision over the range of IOU values from 0.5 to 0.95, in increments of 0.05. The mAP reaches its peak around the 150th epoch; any further training decreases the mAP, indicating that overfitting has occurred, and thus the training should be terminated. The value of mAP is highly dependent on the diversity of the dataset, and a normal value should exceed 0.5 for mAP@0.5. The smallest detectable object size on the screen is approximately 30px by 30px.

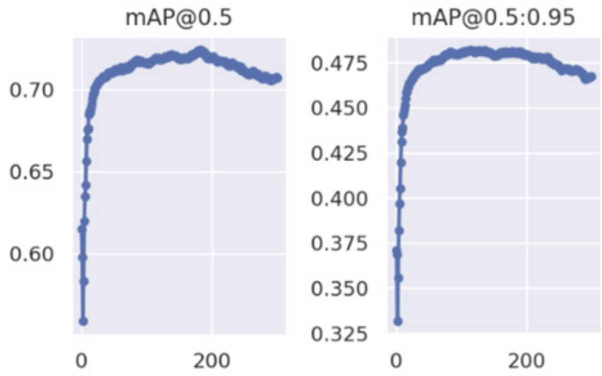

Figure 7. Mean Average Precision (mAP) against epochs at different IOU thresholds.

3.5. Performance evaluation

Upon deployment to the site, the site acceptance test was performed. A total of 100 trials of suspicious detection were conducted in the 100 controlled real-world trials at Jordan station, and the system achieved a good detection rate with no missing detections—every planted unattended object stationary for ≥10 minutes was correctly identified with an alarm triggered. In the subsequent three-month live deployment across operational platforms, the false negative rate was only 2.3% (recall 92%, F1-score 0.935), with the rare misses limited to extreme prolonged total occlusions (>8 minutes) or objects outside the monitoring zones. These cases have since been addressed through zone refinement and enhanced re-identification, confirming that the system's false negative performance comfortably meets MTR's safety requirements. Furthermore, an alarm is activated within 1 second following the acquisition of the relevant video frame. This swift inference is pivotal for immediate action by station operators. The enhanced performance is depicted in Figure 8.

Figure 8. The enhanced model can address the pain point in existing approaches.

3.6. AI Software architecture

The complete AI software architecture is summarised in Figure 9. The video analytic server operates on a Linux (Ubuntu) platform. Python is the primary programming language used in this solution, and relevant Python modules are employed during the development of the AI software module. Based on the environment and project functional requirements, the selected cameras operate at resolutions of at least 720*p* and frame rates of 25/30 *fps*, which are typical specifications for MTR's CCTV systems to balance between bandwidth efficiency and image quality in high-traffic areas. The innovation leverages these specifications by processing real-time RTSP streams at these resolutions and frame rates, ensuring efficient foreground filtering and object detection without excessive computational load. For instance, the AGMM background modelling (Section 3.2) adapts to lighting variations in indoor stations, while the enhanced YOLOv5 model (Section 3.3) handles the smallest detectable objects (≈30*px* by 30*px*) effectively at these resolutions, minimising the bandwidth impact (e.g., via ROI focus) and maintaining inference speeds under 1 second.

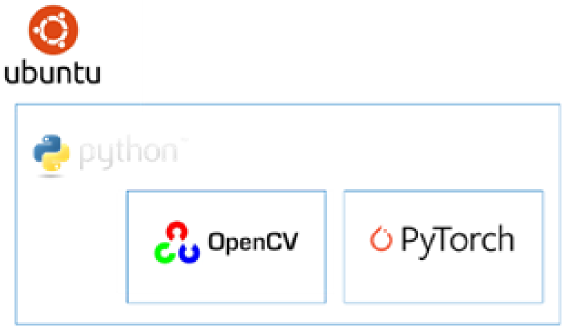

Figure 9. Software architecture for the AI software module.

OpenCV (Villán, 2019) has rich computer vision capability and enables efficient video capturing and image pre-processing which form the foundation of the AI software module. Video input to the AI software module is the main ingredient of the solution, and RTSP is adopted. Once the video stream is captured, the foreground filtering will be conducted. After the object detection is conducted, the overlaid video will be output to the user interface as an RTSP out video stream.

PyTorch (2025) will conduct inference to the pre-processed image and detect any suspicious objects. Pre-defined types of objects (such as dangerous objects) will be detected and alarms will be generated. An unattended object is another scenario of suspicious object. If the object remains in place for more than the pre-defined time, alarms will be generated. With consideration of moving occlusive objects in busy environments, this solution will determine whether an object was already in place previously even though it was occluded, and the elapsed time will be cumulated. To enable efficient, accurate, and robust detection, the Region of Interest (ROI) was defined in each camera view. Only detection results within the ROI will be considered and output to the user.

3.7. Design of the frontend system

To enable station operators to view AI results and receive any generated alarms, a video management system was developed for the frontend system. It was installed in the control room of the pilot station. The system streams RTSP from the video analytic server and displays it in real time on the user interface. Additionally, any alarms produced by the video analytic server are sent to the video management system, which then alerts station operators visually and audibly to ensure timely attention and subsequent action. To facilitate effective operation, human factors have been incorporated into both the workflow and user interface designs. Figure 10 illustrates the final result of the suspicious object detection.

Figure 10. Red bounding box around a suspicious object detected near a column.

3.8. Definition and Validation of Unattended Luggage

An object is considered unattended if it remains stationary without being held by a person beyond a predefined threshold of 10 minutes according to the operational requirements of the station operators and security experts, triggering an alarm. The system tracks the duration by mapping objects across frames based on type and position, cumulating elapsed time even under occlusions from moving foregrounds. As per Sections 3.2 and 3.6, 'unattended baggage' is defined as an object remaining stationary beyond a predefined threshold of 10 minutes, with duration tracked by cumulating the elapsed time across frames, even under occlusions. The type of unattended baggage aligns with user requirements from the station operators and security experts. Validation occurred through live runs at Jordan station (Section

3.5), with 100 trials and three months of deployment, where station staff provided feedback on alert relevance (e.g., reducing false alarms in peaks hours), confirming alignment with real-life experiences like passenger flows near columns (refer to Figure 8 and Figure 10).

4. Conclusion and future work

Undoubtedly, manual operation remains one of the primary challenges for railways, potentially affecting safety and train services. Ensuring passenger safety is the fundamental priority of railway management, and the integration of smart technology is central to advancing modern operational capabilities. Smart Guardian is a cost-effective and robust AI solution for busy railway environments, comprising a backend AI platform that performs object detection and tracking functions using cutting-edge computer vision techniques and deep learning algorithms, and a frontend that displays detected objects and alerts station staff. The system was trained on a composite dataset of real and generated images, which provided sufficient images to train the AI model and thus enhanced the accuracy of suspicious object detection. Background modelling was used to filter out noisy moving foregrounds for better tracking. With these features, the system can function effectively in a unique, challenging, and busy railway environment.

As a next step, recognising the practicality of this technology, it can be further applied and expanded across the railway network to intensify monitoring at stations and enhance railway safety. The potential for this system to operate in other high-traffic areas will also be explored, such as within trains, buses, shopping malls, and properties owned or managed by MTR. Additionally, there is significant potential for its application in other MTR hubs across the Mainland and worldwide, as well as in smart city applications in various industries beyond MTR. Datasets will be expanded for fine-tuning to adapt to these future scenarios. The focus was on supervised deep learning with YOLOv5 trained on a composite dataset, as detailed in Sections 3.3 and 3.4, using real images and generated ones via 3D software like Blender to ensure sufficient railway-specific data within project constraints. While unsupervised/zero-shot models like SAM or Grounding DINO offer promise for general scenes, they were not prioritised due to the need for high precision in safety-critical railway applications, where the labeled dataset (with clear object classes like backpacks, suitcases, and prohibited items) achieved an F1 score of 0.935. Future work could explore hybrid integration for broader adaptability.

Anti-terrorism remains a crucial challenge for the whole world. However, with this revolutionary system, the aim is to ensure that every suspicious object can be detected immediately to minimise the impact brought by them, and uplift railway safety and security to a new standard.

Notes on contributors

Mr Calvin Cheung is a graduate engineer in MTR and currently a trainee in the Scheme "A" Graduate Training of HKIE. He obtained his BEng and MSc degrees in Electronic Information Engineering from City University of Hong Kong with concentrations in networking, artificial intelligence, and big data analysis. He is actively involved in innovation projects at MTR, focusing on the applications of artificial intelligence, computer vision, video analytics, and the Internet of Things in railways, gaining valuable hands-on experience and professional development.

Mr Mickey Yan Po Fong is a graduate engineer at MTR. He holds a Bachelor of Engineering in Computer Engineering from the Hong Kong University of Science and Technology and is currently pursuing a Master of Science in Electrical and Electronics Engineering at the University of Hong Kong. Proficient in programming, he has actively contributed to innovation projects at MTR and demonstrated technical excellence in competitions, including securing the second runner-up and merit award at the 2019 and 2024 HKIE Hong Kong Electronics Project Competition, respectively.

Ir Dr Wai Pan Tam is a professional railway design manager in MTR and an active participant in innovation projects, who has received recognitions in various competitions on innovations both locally and internationally. He received his BEng and Ph.D. degrees in Information Engineering from The Chinese University of Hong Kong and specialised in advanced wireless technologies and communication theory. He is also experienced in the applications of artificial intelligence, computer vision, video analytics, and the Internet of Things in railways. He has been a training tutor of Scheme "A" Graduate Training of HKIE since 2017 and is now an engineering supervisor of the scheme.

References

[1] (2024). MTR. [online]. Available at: <http://www.mtr.com.hk>. [Accessed on 21 January 2025].

[2] (2024). MTR. [online]. Available at: <https://www.sciencedirect.com/topics/social-sciences/urban-resilience>. [Accessed on 21 January 2025].

[3] Zucchi, K. (2022). What countries spend on antiterrorism. [online]. Available at: <https://www.investopedia.com/articles/investing/061215/what-countries-spend-antiterrorism.asp> [Accessed on 21 January 2025].

[4] (2023). Executive Summary. In: Global Terrorism Index 2023. [online]. Available at: <https://www.economicsandpeace.org/wp-content/uploads/2024/02/GTI-2024-web-290224.pdf> [Accessed on 21 January 2025].

[5] Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2017). Understanding of a convolutional neural network. In 2017 international conference on engineering and technology (ICET) (pp. 1-6). Ieee.

[6] Schulzrinne, H., Rao, A., & Lanphier, R. (1998). RFC2326: Real time streaming protocol (RTSP).

[7] Masse, M. (2011). REST API design rulebook: designing consistent RESTful web service interfaces. " O'Reilly Media, Inc.".

[8] Jiang, P., Ergu, D., Liu, F., Cai, Y., & Ma, B. (2022). A Review of Yolo algorithm developments. Procedia Computer Science, 199, 1066-1073.

[9] (2023). Person and Luggage Object Detection Dataset (v2, coco classes - person, backpack, handbag, suitcase). In: Roboflow Universe. 1st ed. [online]. Available at: <https://universe.roboflow.com/playground-v6gip/person-and-luggage/dataset/2>. [Accessed on 21 January 2025].

[10] Xu, Y., Dong, J., Zhang, B., & Xu, D. (2016). Background modeling methods in video analysis: A review and comparative evaluation. CAAI Transactions on Intelligence Technology, 1(1), 43-60.

[11] Stauffer, C., & Grimson, W. E. L. (1999). Adaptive background mixture models for real-time tracking. In Proceedings. 1999 IEEE computer society conference on computer vision and pattern recognition (Cat. No PR00149) (Vol. 2, pp. 246-252). IEEE.

[12] (2025). Blender. [online]. Available at: <https://www.blender.org/>. [Accessed on 21 January 2025].

[13] Weiss, K., Khoshgoftaar, T. M., & Wang, D. (2016). A survey of transfer learning. Journal of Big data, 3(1), 1-40.

[14] Henderson, P., & Ferrari, V. (2017). End-to-end training of object class detectors for mean average precision. In Computer Vision–ACCV 2016: 13th Asian Conference on Computer Vision, Taipei, Taiwan, November 20-24, 2016, Revised Selected Papers, Part V 13 (pp. 198-213). Springer International Publishing.

[15] Yacouby, R., & Axman, D. (2020). Probabilistic extension of precision, recall, and f1 score for more thorough evaluation of classification models. In Proceedings of the first workshop on evaluation and comparison of NLP systems (pp. 79-91).

[16] Villán, A.F. 2019. Mastering OpenCV 4 with Python: A practical guide covering topics from image processing, augmented reality to deep learning with OpenCV 4 and Python 3.7. Packt Publishing.

[17] (2025). PyTorch. [online]. Available at: <https://pytorch.org/>. [Accessed on 21 January 2025].