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

Intelligent Crowd Diversion System (ICDS) – elevating operations and passenger experiences at world-class-event stations

Man Yan Cho, Sherry Yip, Cheuk Hin Harry Wong and Wai Pan Tam
Hong Kong Transport Services Business Unit, MTR Corporation Limited, Hong Kong, People's Republic of China

ABSTRACT

An Intelligent Crowd Diversion System (ICDS) has been developed, utilising Artificial Intelligence (AI) and Machine Learning (ML) technologies to assist in crowd management at Kai Tak (KAT) and Sung Wong Toi (SUW) stations during mega events. The ICDS integrates real-time CCTV data and tailored AI-based video analytics to provide accurate people counting and crowd control metrics. The ICDS predicts real-time platform waiting times for dispersal based on statistics, including previous waiting times, crowd densities, and a unique formula for reliable measurement. Real-time updates are displayed on LCD panels for passenger guidance. A dashboard offers an overview of station busyness status, empowering station managers to make real-time adjustments and provide information for upcoming events during non-event days. The incorporation of AI technologies enables predictive measures for managing sudden increases in crowd volumes during large-scale events. Data exchange with MTR Data Studio allows enhancing the comprehensiveness of databases and increasing the accuracy of predictions. Estimated waiting times benefit by keeping passengers informed with timely information, ensuring a smooth flow of passengers and improving the mobility efficiency. With the implementation of innovative crowd diversion, MTR can replace traditional manual crowd control management with the ICDS and become a modern railway system in the world.

KEYWORDS artificial intelligence; computer vision; crowd control; railway; machine learning; safety; surveillance
CONTACT Man Yan Cho amymycho1@mtr.com.hk
Received 30 January 2026

1. Introduction

MTR is a prominent public transportation network in Hong Kong, known for its efficiency, cleanliness, and extensive coverage across the region. It serves as a crucial component in the daily commute of millions of residents and visitors in the city. By embracing "Go Smart" solutions, MTR aims to provide passengers with seamless and hassle-free railway services through the utilisation of diverse technologies and systems (MTR, 2025).

With the launch of Kai Tak Sports Park (KTSP) in 2025, two nearby stations, Kai Tak (KAT) and Sung Wong Toi (SUW), became critical public transportation hubs. These stations are expected to cater to audiences with up to 50,000 attendees for each world-class event held at KTSP. The surge in passenger traffic during peak hours before and after these mega events causes operational and crowd control challenges for KAT and SUW, including the risk of overcrowding at these stations.

The escalating crowd density in station areas has underscored the importance of effective crowd management. Crowd management helps in controlling overcrowded areas during events and allowing authorities to monitor, manage, and reduce incidents. It essentially comprises a set of cooperative practices and a series of coordinated efforts involving various stakeholders to ensure proper preparation, organisation, and control of events (Struniawski, 2024). Furthermore, leveraging AI and ML technologies can significantly enhance crowd management effectiveness and facilitate real-time problem-solving capabilities (Macriga et al., 2024).

In this paper, an Intelligent Crowd Diversion System (ICDS) is proposed to assist in the crowd management at KAT and SUW during mega events by utilising real-time CCTV data and AI-based video analytics (VA). The ICDS integrates 58 CCTV cameras with the VA system to provide real-time people counting and crowd density during event execution. Highly customised metrics for a crowd control and monitoring model is developed through machine learning using historical data from test events and stress tests, and is poised for future analysis.

During mega events in KTSP, the VA system analyses real-time people counts and densities from CCTV footage using the trained model to predict passenger waiting times for arrivals at platforms. The waiting status is displayed on LCD panels to indicate the platform crowding and encourage patient boarding. The ICDS also offers a dashboard showing an overview of the station busyness status and empowering station managers to manually adjust the station busyness indicators on the dashboard.

Clear instructions for the public and user-friendly interfaces for station staff enhance the crowd management effectiveness at KAT and SUW. By employing AI technologies, predictive measures can be implemented to prepare for a sudden increase in crowdedness during short-term large-scale events. Estimated waiting times for passengers strengthen effective passenger diversion to both stations, ensuring a smooth flow of passengers, allowing a seamless experience, and keeping the city moving.

2. Literature review

Effective crowd management in mass transit systems has long been a critical research area, particularly in high-density urban environments. Traditional approaches relied on manual monitoring and static infrastructure, but recent advances in artificial intelligence (AI), computer vision (CV), and machine learning (ML) have transformed the possibilities for real-time crowd analysis and passenger diversion.

Foundational studies established the basis for understanding pedestrian flows and congestion. Helbing and Molnár introduced the social force model, simulating pedestrian interactions and evacuation dynamics (Helbing & P. Molnár, 1995). Hughes proposed continuum models treating crowds as fluid flows (Hughes, 2002), while Daamen and Hoogendoorn conducted empirical studies on pedestrian walking behaviour in transit stations (Daamen & Hoogendoorn, 2003). These works highlight the importance of modelling crowd density and bottlenecks, which remain central to railway operations.

Dense crowd scenes pose unique challenges for surveillance systems. Early approaches relied on manual counting or coarse estimates from fare collection data. More recent work leverages CNN-based architectures for crowd counting and density estimation. Zhang et al. introduced multi-column CNNs (MCNN) for scale-aware crowd counting (Zhang et al., 2016), while Li et al. developed CSRNet using dilated convolutions for improved density map regression (Li et al., 2018). Sindagi and Patel surveyed CNN-based crowd counting methods, noting their robustness in high-density scenarios (Sindagi & Patel, 2018).

Head detection has emerged as a reliable strategy under heavy occlusion. Datasets such as SCUT-HEAD (Peng et al., 2018) and Brainwash (Vu et al., 2015) enabled specialised detectors that outperform full-body models in crowded environments. This aligns with the ICDS's design choice to focus on head-based detection for accuracy in Hong Kong's packed stations.

The You Only Look Once (YOLO) family of detectors has become the benchmark for real-time detection, balancing speed and accuracy (Jiang et al., 2022; Redmon & Farhadi, 2018; Bochkovskiy et al., 2020). Tracking frameworks such as SORT (Bewley et al., 2016) and DeepSORT (Woike et al., 2017) further reduce duplication errors, enabling reliable passenger counts even when individuals re-enter monitored zones. These advances support the ICDS's reflection filtering and duplication removal strategies.

Beyond detection, routing algorithms are critical for distributing passenger flows. Tolikas, Spyrou, and Kappatos proposed a passenger routing algorithm to minimise overcrowding during COVID-19, demonstrating how dynamic routing can reduce infection risks and congestion (Tolikas et al., 2024). Their work underscores the potential of algorithmic diversion strategies, which the ICDS extends by integrating real-time waiting time predictions and station busyness dashboards.

Other studies explored metro passenger flow monitoring and congestion prediction using multimodal data (Chen et al., 2018; Zheng et al., 2019), highlighting the benefits of combining CCTV analytics with operational levers such as gate controls and signage.

Recent research emphasises multimodal fusion of CCTV, fare collection, and timetable data to improve prediction accuracy (Zhang et al., 2017). Digital twin technologies provide virtual environments for simulating crowd flows and testing interventions before deployment (Negri et al., 2017; Fuller et al., 2020; Garcia et al., 2023). These approaches align with the ICDS's roadmap for integrating with MTR Data Studio and future digital twin simulations.

Studies on evacuation signage and risk communication show that clear, multilingual, and colour-coded displays improve passenger compliance and reduce uncertainty (Wolff et al., 2019; Galesic et al., 2021). The ICDS's LCD panels, which cycle between English, Traditional Chinese, and universal symbols, reflect these best practices by enhancing accessibility and trust.

Finally, ethical considerations are central to deploying AI in public transport. Privacy-preserving analytics emphasise non-identifiable detection methods, such as head-only recognition (European Union, 2018; ICO, 2021; NIST, 2020). Rail safety standards require deterministic failover mechanisms (Cenelec, 2018; IEC, 2019), which the ICDS addresses through pre-recorded fallback messages and operator overrides.

3. System architecture and hardware configuration

This section will introduce the system architecture and hardware configuration of the ICDS. The system comprises three main components: application setup, ICDS architecture, and user interface.

3.1. Application setup

The application setup serves as the bridge between the ICDS and external entities, providing a user interface for station staff and operations managers to monitor the passenger flow in real time and supporting functions such as alarm management, task creation, status queries, and override. It comprises a workstation computer interfaced with the video management (VM) server, a computer display screen, and a set of keyboard and mouse. The resolution for display screens is recommended to be 1920×1080 in this project and interconnected with workstations using video cables.

For public display, 43” LCD display panels were installed in KAT and SUW connecting the media player using a High-Definition Multimedia Interface (HDMI) to output operations managed by the ICDS platform. Customised event schedules and AI algorithm-controlled waiting status can be updated as needed to plan and manage passenger arrivals, dispersals, and activities on both event and non-event days.

3.2. ICDS Architecture

The ICDS is a multi-tiered architecture designed to optimise passenger flow management at KAT and SUW. The system operates across three interconnected networking levels: the Concourse, Telecom Equipment Room (TER), and Station Control Room (SCR). Figure 1 shows the outline of the network flow for the ICDS.

At the Concourse level, 58 CCTV cameras and 12 digital displays form the frontline of data collection and passenger interaction. These 720*p* cameras, powered by Power over Ethernet (PoE) extenders, capture real-time crowd metrics like density and movement patterns, while HDMI optical extenders transmit high-definition video feeds over fibre cables to 43” LCD display panels. These displays provide passengers with dynamic guidance, for example, real-time waiting status for arrivals at the platform and the indicator for crowdedness, ensuring a seamless experience during high-capacity events.

The Telecom Equipment Room acts as the aggregation hub, consolidating data from concourse cameras and distributing power and network connectivity. Equipped with Corporate Data Network (CDN) distribution switches, the data can be transmitted within the whole MTR internal network and perform data exchange with MTR Data Studio. Network switches aggregate raw CCTV feeds and relay them to the VA servers to deploy AI algorithms, and calculate crowd metrics such as passenger counts, station capacity rates, and real-time waiting times. It is equipped with 12 A4000 graphics cards (GPU). Each server comes with three graphics cards that can handle 20 cameras for all three functions. VA processing and output are able to perform within 500 *ms*. The stored analytics results are pushed to the Video Management servers for centralisation. The TER manages the routing of processed analytics data to digital displays and ensures uninterrupted video transmission via fibre links. This level also integrates PoE

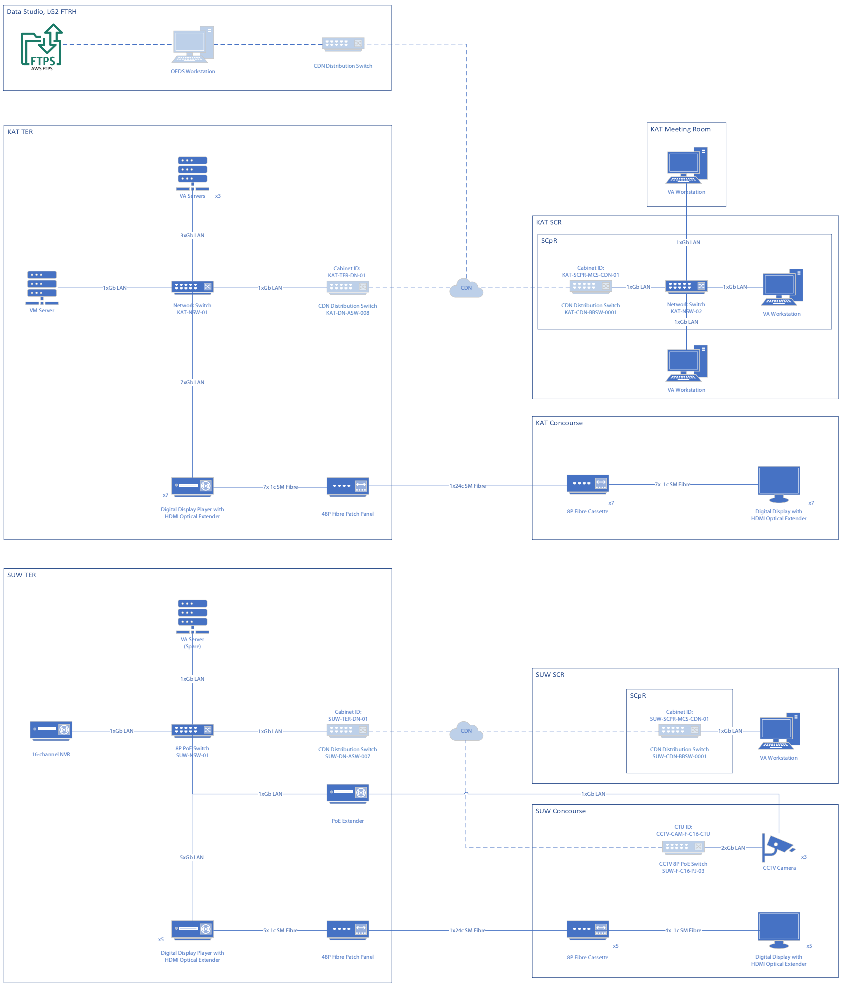

Figure 1. System architecture of the ICDS.

extenders to simplify cabling and power concourse cameras directly from TER.

At the commands operation centre—SCR, station operators can access VA workstations to adjust crowdedness thresholds, for example, red, orange, or green alerts, and trigger crowd control measures like reallocating station staff. Processed insights are pushed back to the TER's CDN switches for display updates, closing the loop between data collection and passenger guidance. A 16-channel Network Video Recorder (NVR) archives footage for compliance and post-event analysis.

Data flows bidirectionally across the three layers. CCTV streams and displays content through fibre cables between the concourse and TER, while raw video data is forwarded to VA servers for AI processing. Processed analytics are then distributed to displays via the TER's CDN. Key design considerations include scalability (modular CDN switches for future display expansions) and low-latency communication (fibre-based HDMI extenders with around a 1 *ms* delay). Security is enforced through LAN segregated and encrypted data links between the SCR and TER.

By integrating CCTV cameras, power-efficient network infrastructure, and AI-driven analytics, the ICDS architecture ensures effective crowd management while enhancing passenger safety and experience. This stratified model not only addresses current operational challenges at KTSP but also provides a scalable framework for future deployments across MTR's network.

3.3. User interface

Two distinct user interfaces are tailored for station staff, operations managers, and passengers. For station staff, the primary interface is a centralised workstation dashboard (shown in Figure 2) that provides real-time oversight of crowd dynamics. This dashboard displays an interactive, colour-coded map of the station, with red, amber, or green zones indicating overcrowded, moderate, and free-flowing areas. An AI-generated headcount overlays the layout to highlight density trends at critical points like exits, ticketing gates, and platforms.

Live metrics such as waiting times and counting data are shown numerically, for example, "C20 Total: 9" referring Camera–C20 detects a total of nine people in its detection zone. With the counting data and waiting time estimation, station staff can have a comprehensive station overview of the passenger arrivals and dispersals at specific zones, adjusting their staff deployment tactics and train service plans.

Additionally, station staff can manually override pre-set scenarios such as "Event Mode" or "Non-event Mode" using an image upload interface to assign custom messages to specific displays. Sliders allow operators to adjust crowdedness thresholds, determining the station crowdedness status when levels exceed predefined limits. Three event grading levels A to C are used for event scale setting, which sets the maximum capacity for specific locations during different scales of event. Station managers

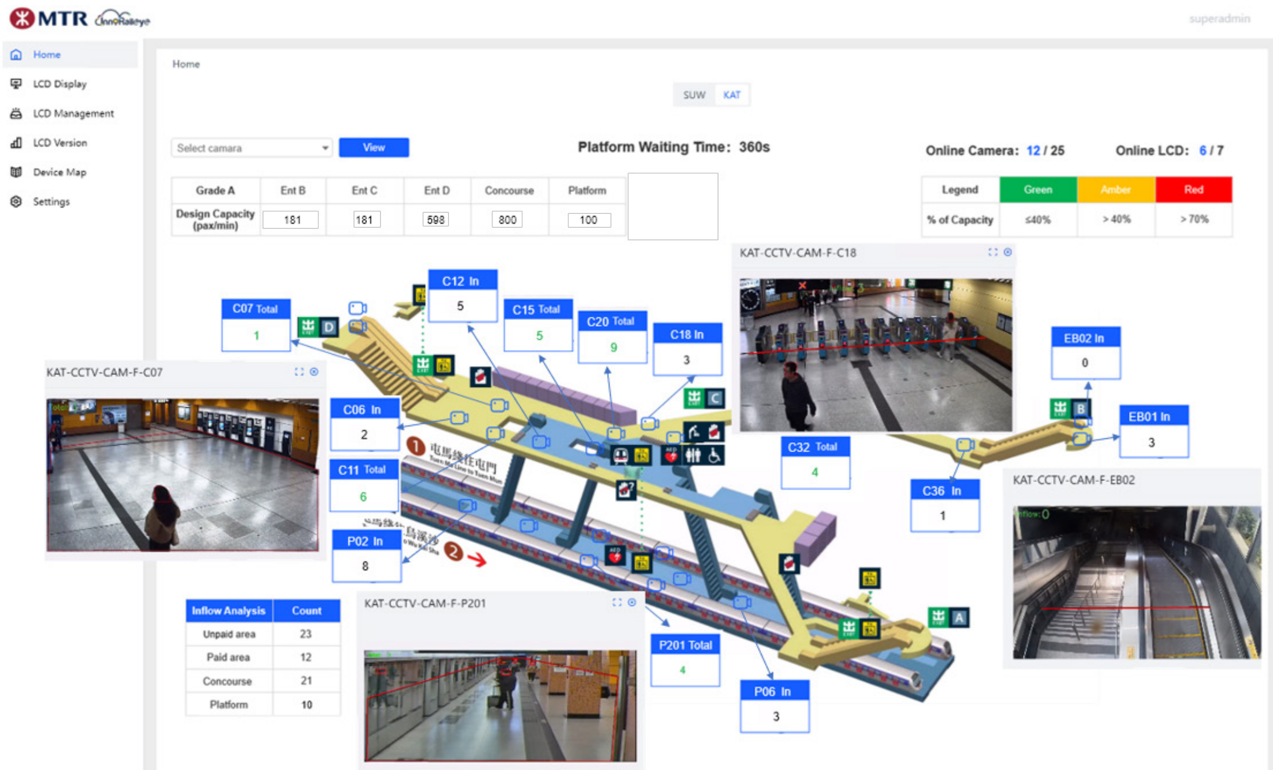

Figure 2. Web dashboard for station staff—station overview.

can choose and modify the grading in the interface shown in Figure 3. The waiting time calculation uses the grading setting.

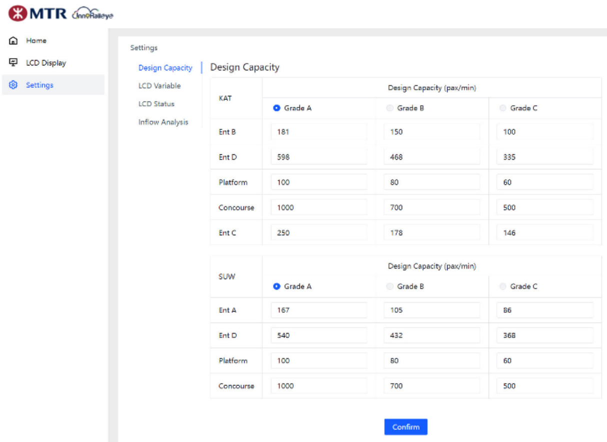

Figure 3. Crowd capacity threshold.

For post-event analysis, the system generates exportable reports detailing crowd patterns, diversion effectiveness, and compliance logs, ensuring accountability and future planning. Figure 4 illustrates the analysis dashboard.

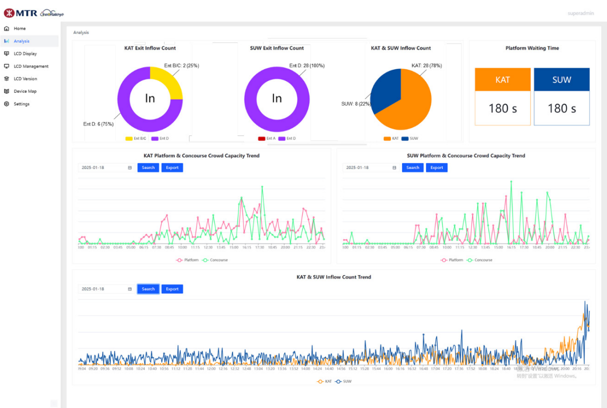

Figure 4. Statistics and graphical representation of passenger flow.

Furthermore, another similar dashboard (shown in Figure 5) was developed for operations managers overseeing the key strategic areas such as entrance and escalator landings to monitor the passenger flow which is updated every 5 minutes as well as spot congestion points which are updated immediately. Real-time crowd counting and live metrics are updated every 5 minutes. By doing so, it allows dynamic adjustments to LCD content scheduling, staffing, and resource allocation, ensuring smoother operations during peak hours, and even on non-event days.

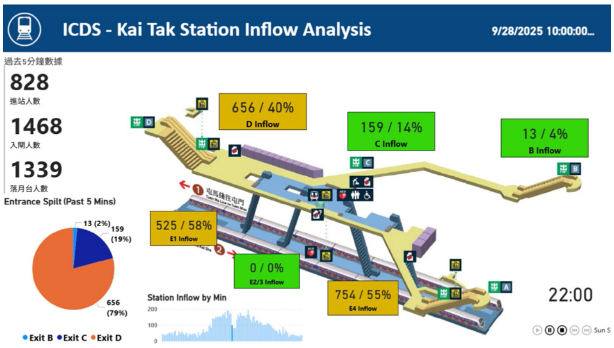

Figure 5. Web dashboard for operations managers.

For passengers, the system employs 43” public displays (shown in Figure 6) positioned across concourses and exits. These screens deliver dynamic, multilingual guidance to optimise flow and safety. Real-time updates include colour-coded icons (Green / Amber / Red) paired with specific platform waiting status, categorised as smooth, normal, or busy. To accommodate diverse users, messages cycle between English, Traditional Chinese, and universal symbols linking to live service updates. The design prioritises clarity, using high-contrast visuals and minimal text to ensure quick comprehension. Content refreshes every second to reflect live AI analytics, maintaining accuracy during rapidly changing conditions with a content display control mechanism. For ease of operation, LCD display management includes “Event Mode” or “Non-event Mode” schedulers to display preloading images and VA results automatically.

Figure 6. LCD content during event days.

During non-event days, displays (demonstrated in Figure 7) show information about the upcoming events for Kai Tak Sports Park (KTSP) and provide additional cautions for crowd management arrangement on event days, such as "Exit D incoming only".

Figure 7. LCD content during non-event days.

The two interfaces operate in tandem: staff use the workstation to modify the usage mode and override the display content, which are then pushed to public displays via the TER's CDN switches. In the event of a system failure, pre-recorded messages are automatically triggered, ensuring uninterrupted passenger guidance. By balancing staff control with automated intelligence, the ICDS ensures adaptive crowd management while maintaining passenger trust and safety.

4. Design of VA System and solution

The main design consideration of the video analytics is reforming station crowd control, embarking on the digital transformation of traditional practice as station operators work to manage crowds in a smarter, safer, and securer way. Crowd management has in the past been provided in the form of visualisation, with real-time updates such as crowd density being counted manually via viewing CCTV cameras or by obtaining people flow information from the automatic fare collection (AFC) system every 15 minutes. A means to provide real-time updates to operators is needed to facilitate timely and effective crowd management. The ICDS based on video analytics automatically detects and counts people in a given area in real time, and transforms the statistics into valuable crowd control insights. Additionally, to prevent overcrowding and congestion, the system also interfaces with newly provisioned overhead LCD panels. Unlike conventional Mobile Display Units (MDUs) and Passenger Information Display Systems (PIDSs) that show public safety messages and advertisements only, these new displays provide crowd control arrangements and platform waiting status, both of which can be updated as the situation changes. With the new system, station staff can utilise the LCD panels as a means to better disseminate information to members of the public.

The ICDS serves as an enabler for station operators to make effective decisions when managing crowd dispersal after mega events held at KTSP to better distribute the crowds leaving between KAT and SUW stations.

4.1. Special requirements of the VA System

Unlike most large cities, a majority of the population in Hong Kong relies on public transport every day. Combined with one of the highest population densities in the world, railway stations regularly become crowded even on normal days. During large-scale events, the sheer crowdedness means that individual passengers cannot be easily seen from the camera views, and the entire field of view might be filled with people. This could possibly confuse or overwhelm off-the-shelf video analytics solutions, as the algorithm might not be able to correctly identify partially obstructed persons, and having too many people in the frame requires high computation power from the system.

Also, most heavy railway stations in Hong Kong are equipped with platform safety doors. When viewing the CCTV footage, the reflections of passengers waiting on the platform can be seen on the platform safety doors, which conventional VA systems could conceivably mistake as a real person. Therefore, an additional requirement for the ICDS is that it needs to distinguish between reflections and actual passengers to prevent overestimating the crowdedness.

To cater for these special requirements, the VA model used in the ICDS is trained to detect heads instead of an entire human body. As heads usually remain visible from the perspective of CCTV footage even when the station is extremely crowded, counting heads instead of bodies is likely to yield much more accurate numbers, especially when most passengers are partially occluded. The system can identify duplicated instances and reflections, which helps to reduce the number of duplicate counts around reflective surfaces. Additionally, the system supports setting up customised regions of interest (ROIs) in any shape. The ROIs could then be drawn to include only relevant key locations such as around station entrances and exits, escalator landings, and on concourse and platform levels. This also helps exclude areas where reflections are found, such as glass panels on railings and platform safety doors to improve accuracy as well.

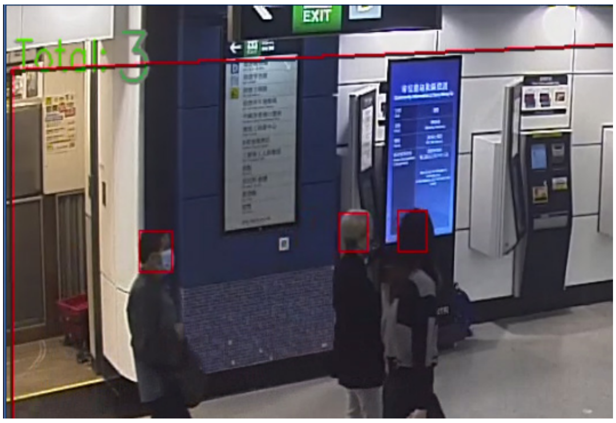

Figure 8. Head detection with crowd density results.

Figure 9. ROI of platform monitoring.

4.2. Technical details

You Only Look Once (YOLO) is amongst the most common algorithms used for object detection in real time (Jiang et al., 2022). The input image is processed through a convolutional neural network (CNN) (Albawi et al., 2017), which attempts to predict bounding boxes for objects on screen, and to categorise the detected objects through regression. The objects that can be detected depend upon what dataset was used. For example, most common YOLO datasets should include humans, animals, cars, and bicycles. In the application of the ICDS, the main focus is on detecting humans. Advanced YOLOv5 and YOLOv8 are used to monitor real-time passenger flow for proactive crowd management in KAT and SUW stations. YOLOv8 provides higher accuracy and is used for people counting and changes in headcounts in monitored areas to estimate waiting times, whereas YOLOv5 is applied to identify and count passengers within a monitored area for real-time detection of passengers with speed and effectiveness.

In brief, YOLOv8's architecture features an input module, backbone, neck, and prediction layer for efficient object detection. The input stage uses data augmentation (random cropping, flipping, colour changes) and adaptive image scaling to boost training efficiency and accuracy. Its backbone leverages convolutional layers with residual connections and attention mechanisms to enhance feature extraction and model focus, addressing gradient issues and improving accuracy. The neck combines a Feature Pyramid Network (FPN) and Path Aggregation Network (PAN) to effectively merge semantic and localisation information, enhancing the recognition of both large and small objects. The prediction module outputs class labels, bounding boxes, and confidence scores, with Non-Maximum Suppression (NMS) removing redundancies.

As YOLOv8 enables real-time detection and tracking of passenger flow by detecting heads with bounding boxes and monitoring head crossings over a virtual line, it greatly brings effectiveness to crowd management during peak hours. Moreover, it is highly optimised for speed, enabling real-time object detection in applications where quick decision making is critical. It also supports multi-task support. Unlike earlier versions which were primarily for object detection, YOLOv8 is a unified framework that natively supports multiple computer vision tasks, allowing the ICDS to perform object detection and AI analysis at the same time. However as mentioned in the previous section, passengers' entire body might not be visible to the camera at all times, especially when crowds begin to form. Therefore, a special dataset and CNN need to be used for the ICDS, which focuses on identifying human heads by extracting features like edges and textures instead of the entire torso. This approach could be challenging; for example, passengers using the MTR network have predominantly black or dark hair, which could be perceived as similar to dark backpacks or other dark objects. A very large number of samples are therefore needed to minimise the probability of misidentification cases.

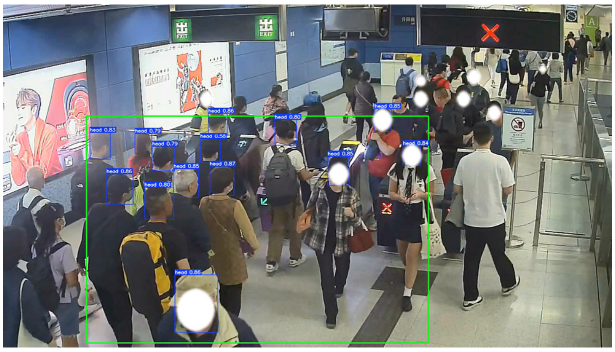

Figure 10. Example of head detection with similarity computation within a defined zone.

YOLOv8 object detection is also used to estimate waiting times by assessing station congestion in specific areas. Its efficient neural network enables fast, accurate detection, generating bounding boxes for passengers in real time across multiple camera frames. For instance, if over 10 heads appear for five consecutive frames, a congestion event is recorded and the timer starts; if the count reaches 20, the waiting time extends to 60 seconds. Its FPN and PAN modules also ensure reliable small-object detection even in complex scenes. A predefined time window of 5 seconds automatically updates waiting times when thresholds are consistently exceeded, supporting real-time crowd monitoring and providing insightful data for management.

On the other hand, YOLOv5 is used for real-time crowd monitoring by detecting and counting passengers within a monitored area. It balances detection accuracy and speed with four core components: input, backbone, neck, and prediction modules, similar to what has been discussed above. The input module applies data augmentation, anchor box calculation, and adaptive scaling to boost robustness and accuracy, especially for far-to-near targets. The backbone integrates the Focus, C3, and SPPF modules to efficiently extract features. The neck module combines FPN and PAN structures to enhance multi-scale feature representation, supporting accurate detection even in crowded scenes. Finally, in the prediction phase, convolutional layers process features, then assign samples, compute loss, and apply NMS to filter the results. Consequently, the system provides precise headcounts for crowd monitoring and triggers congestion alerts (green, amber, and red) when thresholds are exceeded.

4.3. Crowd density and flow calculations

Selected CCTV cameras installed in the station without overlapping coverage are chosen to undergo VA. Crowd densities and passenger flows at key locations are calculated from the video analytics results, with duplicate data omitted from analysis. Based on this approach, a comprehensive crowd management plan can be developed.

To determine the crowd density, threshold parameters are defined for normal, crowded, and overcrowded levels. Each crowd density level is mapped to a corresponding waiting time as well. The VA system continuously monitors and counts the number of passengers within each detection zone in the station, and determines if the detected passenger count in that zone exceeds the parameters. The waiting time (*WT*) estimation then changes according to the crowd density.

Waiting time (in second, s) for each camera

$$(WT_i) = (h_i - d_i - o_i - r_i) \mapsto T_1, \quad (1)$$

Station platform average waiting time (in second, s)

$$(WT_{avg}) = \frac{1}{n}\sum_{i=1}^{n} WT_i, \quad (2)$$

Waiting status displayed on LCD panels

$$(WS) = \frac{1}{C}\sum_{i=1}^{n}(h_i - d_i - o_i - r_i) \mapsto T_2. \quad (3)$$

where

n : Total number of cameras

$h_i = (h_{ix}, h_{iy}, h_{iz})$: Instance of observation from the camera

$d_i = (d_{ix}, d_{iy}, d_{iz})$: Instance of non-human observation from the camera

$o_i = (o_{ix}, o_{iy}, o_{iz})$: Overlapped instance (e.g., image projected on glass panels)

$r_i = (r_{ix}, r_{iy}, r_{iz})$: Repeated instance (e.g., an individual walking in and out from the platform)

T_1 : Time relative to platform crowd density

T_2 : Status of the time range relative to the platform capacity and station busyness status

C : Maximum platform capacity

Figure 11. Example of waiting time for platform camera.

4.4. System performance

The Intelligent Crowd Diversion System (ICDS) has demonstrated strong performance through stress tests and mega events, achieving over 90% accuracy in real-time passenger counting and waiting time predictions. The result is compared to the AFC entry data, ensuring that the number of people count is aligned with the actual passenger inflow.

End-to-end latency from CCTV capture to public display averages just 1–2 seconds, ensuring responsive updates even during peak crowds of more than 65,000 passengers. The system has proved scalable, handling 58 concurrent camera streams and 12 displays without frame loss. Operationally, the ICDS reduces station staff workload and improves safety by lowering overcrowding incidents, and enhances passenger satisfaction with clear, multilingual waiting time information.

4.5 Design of the frontend system

Video management software was developed which allows users to review the analysis results in real time, and to receive any alerts generated by the AI. The video streams used for video analytics are transmitted to the video management software through a Real Time Streaming Protocol (RTSP) (Schulzrinne, 2016), and can be chosen to be displayed on screens in station control rooms like conventional CCTV surveillance systems. The video footage is stored in the local network video recorder for 28 days and the ICDS analysis does not involve facial recognition to protect passenger privacy. This allows station operators to continue to monitor the CCTV feeds through this system.

The inflow and outflow at escalator landings particularly around station entrances, as well as the people count at concourse and platform points of interest are displayed on the homepage of the frontend software. With these information, station operators can observe the crowdedness at different locations in the stations at a glance, along with passenger flow patterns within the station. They can then deploy sufficient manpower to various parts of the station to implement crowd control measures and ensure a better experience, and most importantly, safety, for the waiting passengers. For instance, stanchions, barriers, or queuing lines may be set up to separate people going in different directions, which helps with reducing path conflicts and smoothens the flow of people. Additionally, certain station accesses can be made entry only or exit only, or be temporarily closed as well in coordination with the crowd control measures put in place by the police.

Moreover, the LCD panels are connected to the frontend software as well. While the contents of the LCD panels are designed to change automatically based on VA results, station operators are given the option to access the "LCD Management" page in the software and manually change the contents displayed on the screens. This could be useful in cases where the VA results do not match real-life situations, or if special crowd control measures are required.

5. Conclusion and future work

The Intelligent Crowd Diversion System (ICDS) represents a significant advancement in managing rail transit during large events. It combines AI analysis and user-focused design principles. By using real-time CCTV data, machine learning, and adaptable displays, the ICDS is improving crowd flow at Kai Tak (KAT) and Sung Wong Toi (SUW) stations. It also enhances railway system predictions and reduces congestion risks. The system highlights two key benefits, including reducing staff workload through accurate predictions and providing passengers with useful information. Features like head-detection algorithms, reflection-filtering vision models, and bidirectional links with MTR Data Studio establish a reliable system for real-time decision making. The main breakthrough is the ICDS turning static cameras into a dynamic network for crowd predictions. This puts MTR at the forefront of using AI in the transportation industry.

Future improvements should go beyond minor updates, focusing on the synergy between physical infrastructure, data systems, and passenger behaviours. A key advancement can be a deeper integration with the MTR app through an API-driven routing tool. This tool would combine the ICDS's real-time waiting times with passenger data, like trip history or mobility needs, to send personalised exit advice via alerts. For example, a passenger going to Kowloon Bay might receive the message: "Use SUW Station—15-minute walk, saving 8 minutes vs. KAT's 24-minute route." This feature allows passengers to receive real-time information without entering the station thereby enhancing their travel experience.

Moreover, enhancing the precision of waiting time predictions can also be further developed. The system can integrate with additional data types beyond CCTV, including automatic fare collection transactions and the train timetable managed by Operations Control Center (OCC). A big step would be using digital twins to simulate the people flow during mega events and testing the predicted waiting times. This model offers a virtual environment mirroring the actual station conditions for ICDS training and experimentation.

Numerous opportunities exist for enhancing the comprehensiveness and efficacy of the ICDS. We could also expand the ICDS to more critical stations like Tsim Sha Tsui (TST) and Causeway Bay (CAB) stations after its mature development. By embracing MTR's "Go Smart" initiative, MTR can solidify its position at the forefront of modern railway systems in the world.

Notes on contributors

Miss Man Yan Cho is a Graduate Engineer at MTR and is currently undergoing the HKIE scheme "A" training. She graduated from The Chinese University of Hong Kong with a BEng degree in Electronic Engineering, specialising in electronics product design, software technologies, speech recognition, machine learning, and artificial intelligence (AI) training. Amy actively participates in various innovative projects at MTR that focus on AI detection, algorithms, optimisation, video analytics, and developing diverse applications in the railway industry to enhance the overall railway experience.

Miss Sherry Yip is a Support Engineer at MTR with a focus on artificial intelligence, video analytics, and process optimisation. She graduated from the Stevens Institute of Technology with a BEng. degree in Mechanical Engineering. Despite her background, she participated in various communications and workflow design projects related to extra low voltage systems, varying among CCTV, facial recognition, and LCD technologies before joining MTR. She is experienced in system integration and the use of various software applications for delivering improved system solutions.

Mr Cheuk Hin Harry Wong is a Support Engineer at MTR who recently completed his HKIE scheme "A" training. He received a BEng degree in Transportation Systems Engineering from the Hong Kong Polytechnic University. He currently focuses on works related to video analytics and computer vision, but he also has experience in other applications of artificial intelligence, and has from time to time participated in various innovation projects.

Ir Dr Wai Pan Tam is a professional railway design manager in MTR and an active participant in innovation projects, who 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 was specialised in advanced wireless technologies and communication theory. He is also experienced in the applications of artificial intelligence, computer vision, video analytics, and 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] MTR, "MTR." Accessed: Jan. 21, 2025. [Online]. Available: http://www.mtr.com.hk

[2] J. Struniawski, "Crowd management during mass events," *Zeszyty Naukowe SGSP*, vol. 2, no. 89, pp. 27–38, 2024, doi: 10.5604/01.3001.0054.3100.

[3] G. A. Macriga, S. Bavyatha, S. H. Abhinidhi, and N. Jagadeesh, "Crowd Management using AI & ML," in *Proc. 2024 Int. Conf. Power, Energy, Control Transm. Syst. (ICPECTS)*, Chennai, India, 2024, pp. 1–5, doi: 10.1109/ICPECTS62210.2024.1078050.

[4] D. Helbing and P. Molnár, "Social force model for pedestrian dynamics," *Phys. Rev. E*, vol. 51, no. 5, pp. 4282–4286, May 1995, doi: 10.1103/PhysRevE.51.4282.

[5] R. L. Hughes, "A continuum theory for the flow of pedestrians," *Transp. Res. Part B, Methodol.*, vol. 36, no. 6, pp. 507–535, Jul. 2002, doi: 10.1016/S0191-2615(01)00015-7.

[6] W. Daamen and S. P. Hoogendoorn, "Experimental research of pedestrian walking behavior," *Transp. Res. Rec.*, no. 1828, pp. 20–30, 2003, doi: 10.3141/1828-03.

[7] Y. Zhang, D. Zhou, S. Chen, S. Gao, and Y. Ma, "Single-image crowd counting via multi-column convolutional neural network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 589–597, doi: 10.1109/CVPR.2016.70.

[8] Y. Li, X. Zhang, and D. Chen, "CSRNet: Dilated convolutional neural networks for understanding the highly congested scenes," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Salt Lake City, UT, USA, Jun. 2018, pp. 1091–1100, doi: 10.1109/CVPR.2018.00121.

[9] V. A. Sindagi and V. M. Patel, "A survey of recent advances in CNN-based single image crowd counting and density estimation," *Pattern Recognit. Lett.*, vol. 107, pp. 3–16, May 2018, doi: 10.1016/j.patrec.2017.07.007.

[10] D. Peng et al., "Detecting heads using feature refine net and cascaded multi-scale architecture," arXiv preprint arXiv:1803.09256, 2018.

[11] T. H. Vu, A. Osokin, and I. Laptev, "Context-aware CNNs for person head detection," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCVW)*, Santiago, Chile, Dec. 2015, pp. 2893–2901, doi: 10.1109/ICCVW.2015.88.

[12] P. Jiang, D. Ergu, F. Liu, Y. Cai, and B. Ma, "A review of Yolo algorithm developments," Procedia Comput. Sci., vol. 199, pp. 1066–1073, 2022, doi: 10.1016/j.procs.2022.01.135.

[13] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," arXiv preprint arXiv:1804.02767, 2018.

[14] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," arXiv preprint arXiv:2004.10934, 2020.

[15] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple online and realtime tracking," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Phoenix, AZ, USA, Sep. 2016, pp. 3464–3468, doi: 10.1109/ICIP.2016.7533003.

[16] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Beijing, China, Sep. 2017, pp. 3645–3649, doi: 10.1109/ICIP.2017.8296962.

[17] D. Tolikas, E. D. Spyrou, and V. Kappatos, "Passenger routing algorithm for COVID-19 spread prevention by minimising overcrowding," *Computers*, vol. 13, no. 2, p. 47, Feb. 2024, doi: 10.3390/computers13020047.

[18] X. Chen, H. Li et al., "Real-time metro passenger flow monitoring and forecasting via CCTV," *IEEE Access*, vol. 6, pp. 74638–74650, 2018.

[19] Z. Zheng, Z. Yang et al., "Predicting station congestion using smart card and video data," *Transp. Res. Rec.*, vol. 2673, no. 5, pp. 66–77, 2019.

[20] Y. Zhang, C. Wang et al., "Multimodal transit analytics," *Transp. Res. Part C, Emerg. Technol.*, vol. 77, pp. 275–291, 2017.

[21] E. Negri, L. Fumagalli, and M. Macchi, "A review of the roles of digital twin in CPS-based production systems," *Procedia Manuf.*, vol. 11, pp. 939–948, 2017, doi: 10.1016/j.promfg.2017.07.198.

[22] A. Fuller, Z. Fan, C. Day, and C. Barlow, "Digital twin: Enabling technologies, challenges and open research," *IEEE Access*, vol. 8, pp. 108952–108971, 2020, doi: 10.1109/ACCESS.2020.2998358.

[23] A. Garcia, J. Pérez et al., "Digital twins for metro station crowd management," *Simul. Model. Pract. Theory*, vol. 128, Art. no. 102642, 2023.

[24] M. Wolff, A. Mahdavian et al., "Effectiveness of evacuation signage in public spaces," *Saf. Sci.*, vol. 120, pp. 142–153, Dec. 2019.

[25] M. Galesic, R. Garcia-Retamero et al., "Communicating risk and uncertainty via public displays," *J. Risk Res.*, vol. 24, no. 9, pp. 1133–1153, 2021.

[26] European Union, "General Data Protection Regulation (GDPR) guidance on video analytics in public spaces," 2018.

[27] Information Commissioner's Office (ICO), "Video surveillance code of practice," 2021.

[28] National Institute of Standards and Technology (NIST), "Towards a standard for trustworthy AI," 2020.

[29] CENELEC, "EN 50126 / EN 50128 / EN 50129: Railway Applications – The Specification and Demonstration of Reliability, Availability, Maintainability and Safety (RAMS)," 2017–2018.

[30] International Electrotechnical Commission (IEC), "IEC 62290: Railway applications – Urban guided transport management and command/control systems," 2019.

[31] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in Proc. Int. Conf. Eng. Technol. (ICET), Antalya, Turkey, 2017, pp. 1–6, doi: 10.1109/ICEngTechnol.2017.8308186.

[32] H. Schulzrinne, A. Rao, R. Lanphier, and M. Westerlund, "Real-Time Streaming Protocol Version 2.0," Internet Engineering Task Force, RFC 7826, Dec. 2016, doi: 10.17487/RFC7826.