Computers, Materials & Continua

Tech Science Press

DOI: 10.32604/cmc.2024.058386

ARTICLE

Optimization of an Artificial Intelligence Database and Camera Installation for Recognition of Risky Passenger Behavior in Railway Vehicles

Min-kyeong Kim[1], Yeong Geol Lee[2], Won-Hee Park[2,*], Su-hwan Yun[2], Tae-Soon Kwon[2] and Duckhee Lee[2]

[1]Railroad Test & Certification Division, Korea Railroad Research Institute (KRRI), Cheoldo Bangmulgwanro, Uiwang-si, 16105, Republic of Korea

[2]Railroad Safety Division, Korea Railroad Research Institute (KRRI), Cheoldo Bangmulgwanro, Uiwang-si, 16105, Republic of Korea

*Corresponding Author: Won-Hee Park. Email: whpark@krri.re.kr

Received: 11 September 2024 Accepted: 06 December 2024 Published: 03 January 2025

ABSTRACT

Urban railways are vital means of public transportation in Korea. More than 30% of metropolitan residents use the railways, and this proportion is expected to increase. To enhance safety, the government has mandated the installation of closed-circuit televisions in all carriages by 2024. However, cameras still monitored humans. To address this limitation, we developed a dataset of risk factors and a smart detection system that enables an immediate response to any abnormal behavior and intensive monitoring thereof. We created an innovative learning dataset that takes into account seven unique risk factors specific to Korean railway passengers. Detailed data collection was conducted across the Shinbundang Line of the Incheon Transportation Corporation, and the Ui-Shinseol Line. We observed several behavioral characteristics and assigned unique annotations to them. We also considered carriage congestion. Recognition performance was evaluated by camera placement and number. Then the camera installation plan was optimized. The dataset will find immediate applications in domestic railway operations. The artificial intelligence algorithms will be verified shortly.

KEYWORDS

AI; railway vehicle; risk factor; smart detection; AI training data

1 Introduction

In 2019, Korean public urban railways were used by an average of 7.97 million people/day, or more than 30% of the 26 million metropolitan population. Usage is expected to increase as urban areas continue to grow. However, various issues on railways, including assaults, fights, theft, property damage, fainting, lingering, drunkenness, sexual harassment, and arson, present ongoing challenges. The 2014 Urban Railway Act mandated the installation of closed-circuit televisions (CCTVs) to prevent crime and identify accidents/incidents quickly. However, currently, monitoring is performed by humans. A smart detection system, trained on a dataset of major risk factors, would enable intensive monitoring and a rapid response to suspicious activities. By 2024, CCTVs will be installed on all

rolling stock of high-speed and urban railways. Since 2015, artificial intelligence (AI) has been used to recognize situations of concern and trigger appropriate action. Intelligent CCTVs have greatly improved.

Several datasets have been built to aid AI-based human behavioral analysis. These serve as essential resources for AI models that recognize and analyze human actions. These datasets are extensively utilized across various research and development sectors. They significantly contribute to advances in AI technology. They also serve as foundational tools enabling the development of sophisticated algorithms that interpret a wide range of human behaviors, thereby enhancing the accuracy and applicability of AI systems in various real-world scenarios. The Kinetics dataset [1], developed by DeepMind, contains an extensive range of human activities captured in YouTube clips. The dataset is a key resource when creating action recognition models. The UCF101 dataset [2] of the University of Central Florida includes 101 distinct action categories and is widely used for action recognition. The Charades dataset [3], a collaborative project led by Carnegie-Mellon University, focuses on daily human activities. These are annotated. The HMDB51 dataset [4] of Brown University covers 51 categories of action and is invaluable when testing recognition algorithms. Finally, the AVA dataset [5] contains detailed annotations of atomic visual actions, facilitating precise action recognition in video clips. All of these datasets have advanced the field of AI-driven human behavioral analysis.

The databases are extensively utilized for action recognition and human behavioral analysis. Their applications range from video-based action recognition that aids model training and evaluation to an understanding of behavioral patterns [6,7] that assist surveillance and healthcare [8,9]. The datasets recognize gestures [10,11] and enhance human–computer interactions. The datasets are used by sports science and fitness experts who design and analyze exercise routines. The datasets are versatile. They advance the capacity of AI to interpret complex human activities.

In this study, the numbers, locations, and angles of cameras installed in railway carriages to monitor risky passenger behaviors were reviewed. Under Korean law, video information-processing equipment can be used when necessary to prevent and investigate fire and crime, and to ensure the safety of facilities. The purpose and location of the installation, the camera range and time of operation, and the name of the manager and his/her contact information must be published. CCTV footage can be used if certain additional legal requirements are met. This study evaluated CCTV applications in the railway system. Integration of AI with CCTV feeds from railway carriages detects situations that may compromise passenger safety. The numbers and locations of CCTVs in existing railway vehicles vary greatly. It was thus essential to evaluate how these differences affected the detection of behaviors of concern. We considered camera numbers, placements, and angles. If a construction robot is to work with people, a camera must recognize the actions of humans close to the robot. Thus, a multi-camera approach was required. The performances of human activity recognition models employing one to four cameras were compared. It was clear that multiple cameras improved recognition performance and allowed accurate identification of various activities [12]. In earlier work, camera numbers and locations were adjusted to optimize the detection of weed removal by humans. A useful algorithm was developed [13]. It has been suggested that camera placement may affect performance. Appropriate placement optimizes object and motion detection. Certain motions are recognized using visual measurements [14]. Another study reported that system performance in terms of measurement, monitoring, and recognition was greatly affected by camera placement. Multiple cameras can be used to detect the location and direction of movement [15]. In one study, Camera Angle-awre Multi-Object Tracking (CAMOT) was used to eliminate occlusion and improve distance estimation; it enhanced data integrity [16].

Overseas, a dataset for an integrated, smart, AI-based passenger risk factor detection system is under construction, but the principal focus is on buildings and external spaces, not domestic railways. This study presents an innovative AI image dataset that enhances the safety of Korean urban railways. Risky behaviors are detected in detail in real time. Preliminary survey data on risky behaviors in urban railway carriages were used to create a dataset. All elements were labeled by reference to high-resolution image data. The constructed dataset can be used to monitor behavior on domestic trains. CCTV numbers, locations, and angles were optimized using case studies that explored risky behaviors of railway passengers. This enhanced surveillance efficiency and passenger safety. The study presents evidence-based recommendations that will aid the deployment of optimal monitoring systems. The work combines theoretical research with practical applications. Railway safety is enhanced using a new AI standard to monitor behavior on public transport.

2 Materials and Methods

2.1 CCTV in Railway Vehicles

This section summarizes the subway cars of the various Korean lines and their CCTV installations (Table 1). All participants gave written informed consent for use and distribution of their personal information captured on camera.

Table 1: Domestic vehicles and CCTV installations

	Vehicle		CCTVs		
	Length (m)	Doors	No./vehicle	Arrangement	Resolution
Ui-Shinseol Light Rail	13.5	4	2	A type	HD
Busan Gimhae Light Rail:	27	4	2	B type	HD
Incheon Line 1 subway	18	8	2	A type	FHD
Incheon Line 2 subway	16	6	2	C type	SD
Daegu subway	17.5	8	2	B type	FHD
Shinbundang subway	19	8	2	B type	SD
Airport Rail	19.5	8	2	A type	FHD
Uijeongbu Light Rail	13	6	2	B type	HD

All vehicles contained two cameras in the different locations shown in Fig. 1. Type A included two cameras in the middle of the vehicle. Type B included a camera pointing directly down the vehicle at either end of the vehicle. Type C included a camera at a diagonal angle at either end of the vehicle.

The Ui-Shinseol Light Rail, Incheon Line 1, and Airport Railway vehicles feature two CCTV cameras in the center of the carriages. These monitor passenger safety. In contrast, the Busan Gimhae Light Rail, Incheon Line 2, and Uijeongbu Light Rail vehicles contain CCTV units at either end, enabling comprehensive surveillance of the entire vehicle. Similarly, the Daegu and Shinbundang Line

vehicles feature CCTV units at each end of the carriages. These continually monitor passenger safety and security. The CCTV locations are determined by the carriage size and the number of doors. The CCTVs are optimally positioned. Surveillance is efficient. All CCTVs are high definition (HD) or full high definition (FHD) models. These deliver high-quality images that enhance safety.

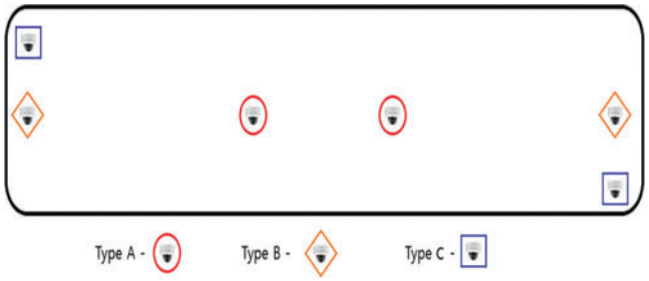


Figure 1: Layouts of CCTVs within railway vehicles

2.2 Database Design

2.2.1 Research Sites and Camera Locations

Trains running on the Shinbundang Line, the Incheon Subway Line 2, and the Ui-Shinseol Light Rail Transit were surveyed (Fig. 2). These urban lines are operated by different organizations. The subway and light rail lines studied included the Shinbundang Line operated by the Incheon Transportation Corporation, and the Ui-Shinseol Line.

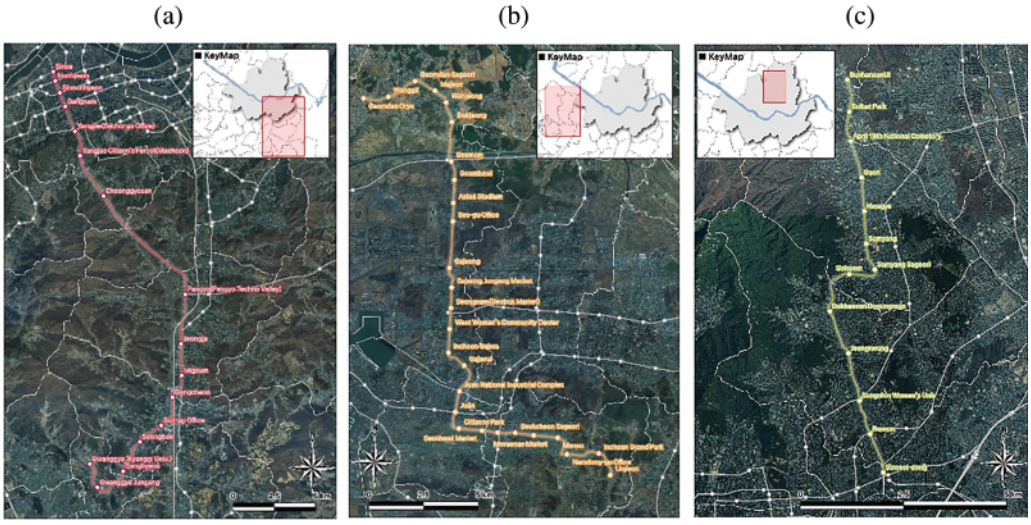


Figure 2: The studied subway lines. (a) The Shinbundang Line (S). (b) The Incheon Subway Line 2(l). (c) The Ui-Shinseol Line (U)

We built a dedicated dataset to train a smart detection system that detected risky behaviors; we used the imaging data of three participating organizations. A carriage of the Shinbundang Line was filmed at the vehicle base from 14 to 22 October 2021. The carriage length was 19.5 m and there were eight entrances. A carriage of the Incheon Subway Line 2 was filmed at the vehicle depot of the Incheon Transportation Corporation from 27 September 2021, to 22 October 2021. The carriage length was

16 m and there were six entrances. A carriage of the Ui-Shinseol light railway was filmed at the Ui-Shinseol vehicle base from 30 October to 04 November 2021; the carriage length was 13.5 m and there were four entrances. Eight Gopro ver. 8 cameras (Cephas, Seoul, Korea) were installed in each carriage to acquire videos from various angles (Fig. 3). The number of cameras, the distances they covered, and the camera angles were varied when assessing the utility of learning materials. The applicability of the findings to real-world lines was prioritized. It is essential to ensure robust assessment of human behavior.

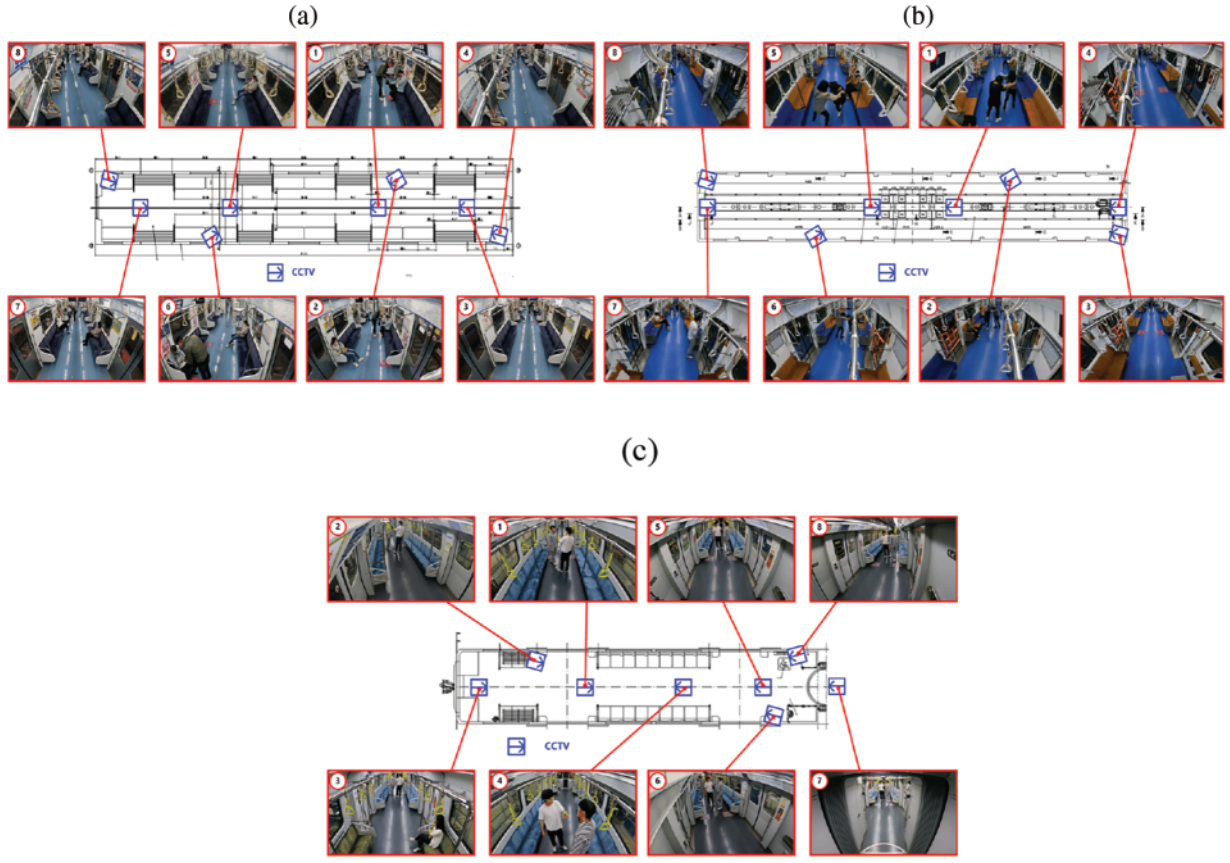


Figure 3: Camera installation prior to dataset construction on (a) the Shinbundang Line. Examples of fights and assaults are shown. (b) Data from a line of the Incheon Transportation Corporation. Examples of loitering are shown. (c) Data from the Ui-Shinseol Line. Examples of fights and assaults are shown

2.2.2 Passenger Risk Factors

In a preliminary study [17], the behaviors of subway passengers that offended the public and compromised safe operation were identified. A questionnaire was used to define risk factors perceived by 1000 railway passengers. These factors were seven in number: loitering, drunkenness, fights and assaults, abnormal door operation, sexual harassment, lingering, and medical issues. In detail, 19 items were selected, including begging, unwanted sales offers, sitting down in the aisle, and lying down (Table 2).

Table 2: Major adverse behaviors

Division	Action	Abbreviation	Details
1	Loitering	Lo	Begging, Selling
2	Drunken behavior	D	Lying down, Sitting down, Spitting, Vomiting
3	Fights and assaults	F	Swinging, Throwing, Kicking, Threatening, Pushing, Stabbing, Pulling, Knocking down, Pointing
4	Abnormal door behavior	A	Door hammering, Forced door opening
5	Reports	R	Reports of sexual harassment by victims or people near the victims
6	Lingering	Li	Remaining in the carriage after train operating hours, Sitting and sleeping after train operating hours.
7	Falling (Medical collapse)	C	Losing consciousness

2.2.3 Scenarios and Dataset

A maximum of eight cameras were installed when constructing the dataset. Images of certain human actions in carriages were collected as follows. Each scenario features one key action that always included at least one detailed sub-action. All scenarios were encountered daily. Third, using the basic scenarios, detailed scenarios tailored to the environment of each train were produced. Scenarios that are rare in everyday life, those in which people were concerned about safety, and scenarios featuring train damage or breakdown were deleted. Ultimately, 217 scenarios were constructed; all 7 major human actions included 19 detailed actions. The ratio of the number of passengers to the maximum vehicle capacity was determined. A typical car can accommodate 160 people, thus 2.7 per m^2. In this study, the AI database evaluated the behaviors of 1.6 people per m^2.

Fig. 4 shows the name of a sample file. This includes the basic metainformation of the captured video data. In terms of initializing, the Incheon Transportation Corporation vehicle depot (the Gyulhyeon Vehicle Depot) is I, the Sinbundang Line vehicle depot (the Gwanggyo Vehicle Depot) is S, and the Ui-Sinseol Line vehicle base (the Ui Vehicle Depot) is U. Each basic scenario was assigned a unique identification number on which the extended (detailed) scenario was built. The seasons (spring/fall, summer, and winter) were added to handle the diversity of clothing worn over the year. The seven major behaviors of interest (at doors, lingering, and falls) are shown in Table 1. The Lo of the human action in Fig. 3 is prowling. M + (a number of actors) and F + (a number of actors) are the numbers of male and female persons of interest respectively. The camera number (01∼08) is included in the file title. The extension is .mp4. The video resolution was 3840 × 2160 pixels (4K ultra-high definition; 4K UHD), the frames per second (FPS) 30, and the video length 3 to 5 min.

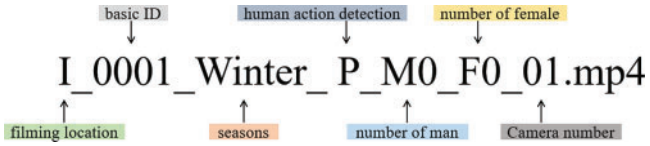


Figure 4: File naming with the basic metainformation

2.2.4 AI Data Labeling: Annotation

Labeling is essential when training an AI using videos from railway carriages. We employed bounding box (bbox), class assignment, and skeleton techniques. A bbox is the smallest box that contains all shapes of a 3D object. Of the positions of the eight points that form the box, the p values of the points with the lowest and highest values on each of the x-, y-, and z-axes were chosen [18,19]. Skeleton techniques confirm the locations of human joints in images and how such joints move in sequential images [20–24]. Bboxs served as the fundamental elements for object detection within railway carriages. bbox annotations were applied to all perceived objects, aiding the training of AI models that detect objects. The class assignments categorize and identify objects and persons. In this study, all objects were classified as "abnormal" or "normal". The skeletal technique addresses the more complicated aspects of labeling, particularly human postures and movements. We used skeletal annotations to detect anomalies that were reflected in human postures, thus during fights and collapses. We used bbox, class assignment, and skeleton techniques for validation and testing. The annotations indicate the timeframes during which abnormal individuals appear in the videos.

3 Results

3.1 Image Data Collection and AI Learning

3.1.1 Collection of Video Data

To enable AI learning, the image data were defined in terms of the detected elements, the detailed behaviors, the season, the ratios of adverse to normal behaviors, and mask-wearing status during the COVID-19 pandemic on all of the Shinbundang Line, the Incheon Subway Line 2, and the Ui-Shinseol light rail transit line (Table 3).

Table 3: Images of each major element and the detailed behaviors over the entire study area

Number	Principal human action	Video length (h/min/s)	Specific actions	Number of scenarios exhibiting specific actions	Number of scenarios recorded
1	Loitering	42:58:40	Begging Attempted sales	24 41	65
2	Reports of anti-social behavior	64:58:56	–	99	99

(Continued)

Table 3 (continued)

Number	Principal human action	Video length (h/min/s)	Specific actions	Number of scenarios exhibiting specific actions	Number of scenarios recorded
3	Fights and assaults	65:10:20	Swinging	18	99
			Throwing	15	
			Kicking	9	
			Threats	51	
			Pushing	9	
			Stinging	3	
			Plucking	18	
			Collapse	3	
			Pointing	24	
4	Abnormal behavior in relation to door	60:51:00	Car door sticking	60	94
			Forced door opening	40	
5	Drunken behavior	71:38:00	Sitting down	33	109
			Lying down	42	
			Spitting/vomiting	28	
			Falling down	6	
6	Lingering	49:39:00	–	75	75
7	Falling (medical)	41:05:00	–	62	62
Total		396:20:56			603

About 126 h of videos were collected on the Incheon Subway Line 2, about 143 h on the Sinbundang Line, and about 126 h on the Ui-Sinseol Line. Notably, filming on the Shinbundang Line was increased because it was difficult to identify lingering passengers on the Ui-Sinseol Line. It is impossible to open and close the doors or operate lights on the Ui-Sinseol Line, so filming focused on other scenarios; some scenarios not evaluated on the Shinbundang Line were filmed on the Ui-Sinseol Line. Table 2 lists the numbers of recorded scenarios, the detailed actions, and the numbers of scenarios featuring multiple actions. One-third of all videos were recorded in each of spring/fall, summer, and winter; mask-wearing was reviewed. The scenarios featured 812 men and 400 women; these were ranked by the scenarios. The dataset reflected all of these factors (Table 4; Fig. 5).

Table 4: Examples of each behavior

Number	Specific action	Details	Scenario
1	Loitering	Begging	A man who walks only with difficulty moves from car to car begging.
2	Sexual harassment	–	A third party reports a man who is sexually harassing a drunk woman.
3	Fights and assaults	Throwing, kicking, pulling, threatening	When a man hits a fallen object, he grabs the owner of the object by the collar and fights the owner.
4	Abnormal door behavior	Door hammering, Forced door opening	A passenger in a wheelchair becomes stuck in a door.
5	Lingering		A woman sleeps lightly while sitting.
6	Drunken behavior	Lying down	A man sleeps in a subway aisle.
7	Medical collapse		A woman stands, stumbles, and falls to the floor.

(a) Loitering and begging.

Figure 5: (Continued)

Figure 5: Representative video data for each action

3.2 AI Data Labeling: Annotation

The 2D and 3D positional coordinates of human joints estimated using the camera images were employed to detect specific behaviors. Any associated objects were identified and their positional coordinates were noted.

An event was defined when at least one person was involved from beginning to end. Bbox was used to perform random sampling and learning. When labeling a bbox, lingering passengers were described as accurately as possible. The visible parts of people were analyzed when multiple people overlapped.

If a person was partially cut off at an edge of the image, that person was nonetheless labeled as a person (Table 5).

Table 5: Annotations of all sensor elements

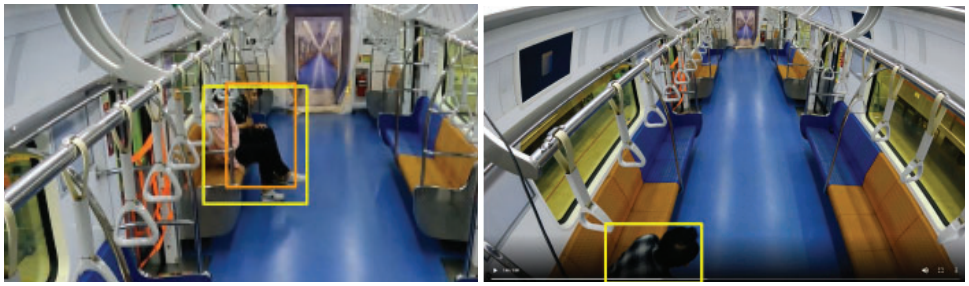
Detected element	Number of videos	FPS	Annotations
Loitering	50	0.77	The training set through to labeling using bbox and the classes of loitering passengers.
Lingering	385	0.2	The training set through to labeling employing bbox and the classes of all passengers.
Falls	240	1	20 s including 5 s before and after each event using bbox and skeletal labeling.
	–	–	The training set through to labeling by the duration of an event.
Assaults	–	–	The training set through to labeling by the duration of an event.
Reports	240	1	20 s including 5 s before and after each event using bbox and skeletal labeling.
	160	0.2	The training set through to labeling employing bbox analysis of the entire video.

Loitering was defined when begging or selling behavior was detected by at least one of the eight cameras, and was labeled in a bbox manner; 50 videos taken at 77 fps were used. Loiterers take about 4 min after boarding to try to sell something and then go; often, they behave like regular passengers, and their behavior does not appear unusual. Loitering times varied greatly; the camera arrangement greatly influenced detection of such behavior (Fig. 6). Loitering included begging and attempted sales. Often such passengers are motionless, as are regular passengers, and it is thus difficult to find anything unusual in their behavior.

Lingering passengers were those detected by at least one of the eight cameras; they were randomly sampled and bbox-labeled in 385 videos obtained at 0.2 fps. Labeling included any bag, any hat, and all accessories. If an object intervened, labeling was confined to all visible components of a person. However, if a person was not well-recognized, the data were excluded. When multiple people lingered, each individual was labeled as accurately as possible based on their visible regions; if the person extended outside of the image, labeling was nonetheless performed if a person was definitely present (Fig. 7).

(a) (b)

Figure 6: Loitering annotations. (a) The difficulty encountered when seeking to distinguish begging from selling. (b) An example of when an object is completely obscured



(a) (b)

Figure 7: Annotations of lingering passengers. (a) If an object is in the way, the label includes only the visible parts of the person. (b) Labeling is performed only when the person is not on the edge of an image and can thus be clearly recognized as a person

A medical fall was scored if it was detected by at least one of the eight cameras. Falls included falls per se, squats, bends, and seizures. Bending and squatting times were recorded from when a person lay completely on the floor; data were sampled from 5 s before the fall to 5 s after the end, and a bbox created. Data were gathered at 1 fps on 240 videos. If an assault was apparent on at least one of the eight cameras, the assault was accepted; we then checked when it occurred, thus whether it was detected within 2 s before and 10 s after the reported time. Assaults included pointing and lightly grabbing, and progressed to punching, kicking, pushing, and pulling. If two people were entangled, an assault was scored; kicking a bag was not an assault. If an assault was prolonged, it was scored over time (Fig. 8).

In terms of sexual harassment and its reporting, an event was scored when it was detected by at least one of the eight cameras. bbox sampling ran from 2 s before harassment commenced to 2 s after the end. A blind spot may be in play when only two cameras report such harassment. The reporting video must include the start and end points of the incident; 5 s bboxs are required for all persons in the video. We collected data at 1 frame per second (FPS) (240 videos).

(a) (b)

Figure 8: Annotations for assault. (a) The point at which a threat commences with the throwing of an object is the start point. (b) The point at which a shoulder is grabbed and pushing commences is a point on the assault timeline

3.3 Detection by Camera Placement

We developed an AI model that reported certain behaviors, including sexual harassment, using a database of risks to railway passengers. FCOS [25] served as the base model for robust human detection by CCTV within a passenger car. This minimized false detections and yielded a model of human detection that was robust against self-occlusion and independent of the camera angles and distances. The CPENML(Classification-based Pose Estimation Network with Multi-task Learning) network structure [26] that engages in robust pose estimation in complex environments with multiple views was used to input all detected human images. Again, this minimized false detections and considered occlusion by objects. This facilitated pose estimation independent of the camera positions in a multi-camera environment. Poses were estimated using cameras placed at various angles. The Re-Identification model [27] was employed to track people and to detect when a change in state value exceeded a threshold value.

The dataset clearly distinguished sexual harassment from the other acts that were recorded. Sexual harassment detection was evaluated by type of camera placement. Again, the AI database was built to recognize risky passenger behaviors in railway vehicles. To verify this, we used AI learning alone to create an algorithm that extracted only incidences of sexual harassment from the database. Sexual harassment is very serious. Well-behaved passengers are greatly inconvenienced and experience anxiety. It is essential to stop such activity. In the future, various algorithms that handle the human movements and poses associated with each database element will be developed. The videos differed by vehicle, season, and camera placement type. The latter was divided into types A, B, and C. A combination of types A and B was considered, as were camera configurations that were not of types A, B, or C. All results were confirmed by video analysis (Fig. 9). Cameras 1 and 5 are those of type A, cameras 3 and 7 are type B, and cameras 4 and 8 are type C. In Fig. 9a, where the camera configuration was of type B, camera 5 detected very few actions. Fig. 9d shows that no images were captured. Table 6 lists the results by situation and season for camera placements of types A, B, and C. Some events were not detected because they occurred too far from the camera, in a blind spot, or at an angle that the camera did not cover. Event detection varied by camera location and angle. The lessons learnt will be reflected in future railway CCTV installation.

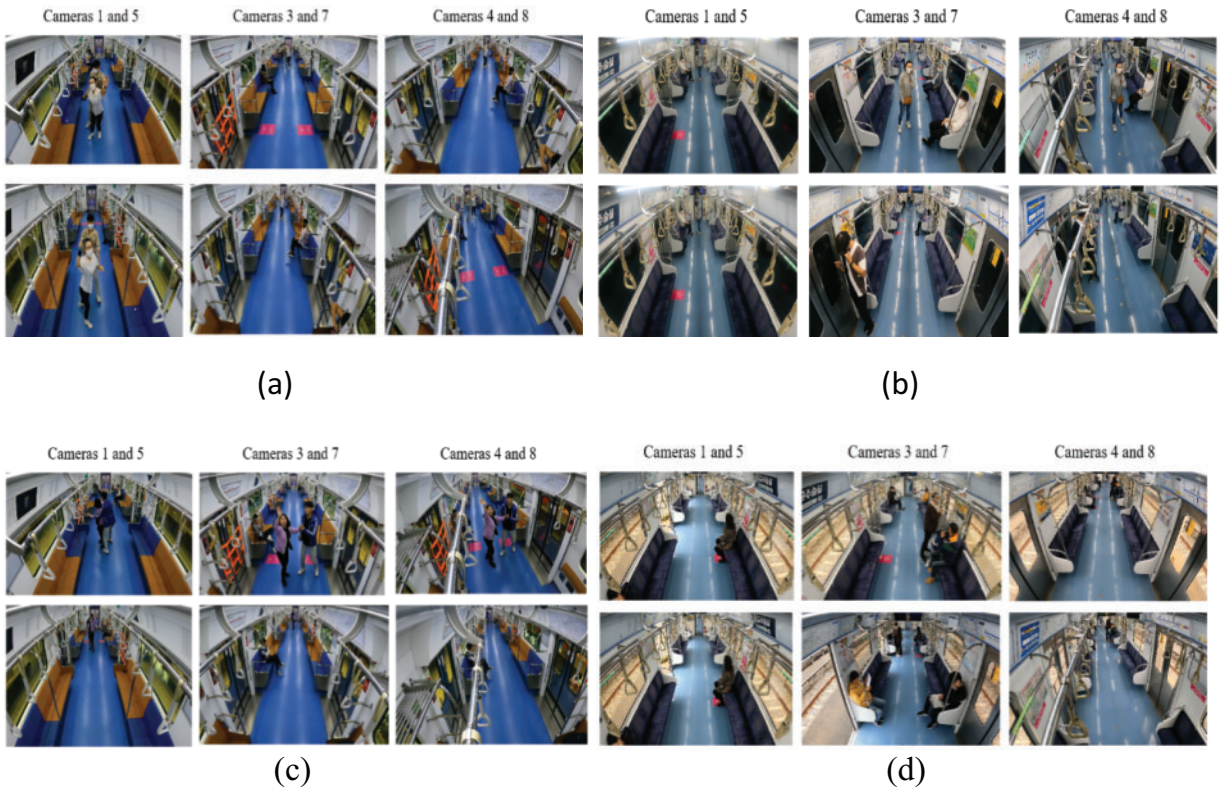


Figure 9: Images by camera angle and placement type. (a) Only type A (D_1026_summer). (b) Only type B (I_1038_spring). (c) Types A and B (D_1026_spring). (d) Type that does not include all (I_1055_winter)

Table 6: Datasets by type of camera deployment

	A type	B type	C type
D_1026_spring	o	o	o
D_1026_summer	o	x	x
D_1026_winter	o	o	o
D_1027_spring	o	x	x
D_1033_spring	x	o	x
D_1033_summer	x	o	x
D_1047_spring	x	o	x
D_1052_spring	o	x	x
D_1053_spring	o	x	x
I_1021_spring	x	o	o
I_1021_summer	x	o	x
I_1021_winter	x	o	o
I_1023_spring	x	o	o
I_1025_spring	x	x	o

(Continued)

Table 6 (continued)

	A type	B type	C type
I_1025_winter	o	x	x
I_1028_summer	o	x	x
I_1038_spring	x	o	x
I_1047_spring	o	o	x
I_1053_spring	x	o	x
I_1055_winter	x	x	x
I_1058_summer	o	x	o
I_1058_winter	o	x	x

4 Conclusions

Open dataset platforms that train AI systems have been developed in Korea and elsewhere, aiding dataset extension and use; a dataset diffusion ecosystem is being developed. In terms of smart detection, certain datasets assist activities of daily living (ADLs) and predict when agricultural products are ripe and should thus be harvested [28,29]. However, most existing datasets have been created in other countries; they do not reflect the daily behaviors of Koreans. In particular, no image dataset yet explores human behavior in a limited space, thus a railway carriage. We built a carriage-specific dataset to facilitate smart integrated detection of passengers acting suspiciously; we used video AI to this end, and CCTVs were installed in all carriages.

In this study, based on previous surveys, seven risk factors: loitering, sexual harassment, fights and assaults, medical conditions, lingering, drunkenness, and abnormal behavior at doors were identified and the detailed behaviors defined. The AI learning database used video cuts from railway carriages in daily operation and labels were assigned to passengers behaving in a risky manner. Our smart detection of such passengers can be immediately applied, with minimal modifications, to aid Korean railway operators. The dataset considers seven risk factors that require detection. Of these, sexual harassment, which is very obvious on a video, was reviewed according to vehicle type and camera placement. We confirmed the positions detected by each camera and applied the findings to the CCTVs of real railway vehicles. In the future, algorithms identifying passengers behaving in a risky manner will be further developed and verified. In particular, we optimized both an AI database and camera installation when detecting risky behaviors of railway passengers. As the algorithm has not yet been applied to railway vehicles, the camera numbers and their placements, angles, and resolution require consideration, as do blind spot concerns. We plan to place both intelligent and fixed cameras in the future. Recognition performance will be improved by minimizing blind spots. We will also consider the effects of passenger density and the time of travel.

This is the first Korean study to employ an AI image dataset to ensure railway safety. The data are immediately applicable. We carefully considered the views of real-world railway operators. Passenger behaviors are affected by cultural factors and the transportation environment. These differ globally. Our system is specific to the Korean railway environment. Future research will explore the systems of other countries. The unique AI image dataset is directed toward the intricacies of how Korean passengers behave on urban transit systems. Such a nuanced understanding sets a new precedent in the field of transportation safety. The detailed identification, classification, and analysis of various risk factors illustrate the comprehensive nature of the dataset, highlighting its capacity to address a wide

variety of safety concerns. By seamlessly integrating theoretical research with practical applications, this study not only contributes to the academic discourse on AI and transportation safety but also pragmatically enhances public transit security. The work will serve as a groundbreaking standard for future research on and the application of AI-driven safety protocols.

Acknowledgement: We would like to thank the three subway operators who provided data collection vehicles and support: The Shinbundang Line, the Incheon Subway Line 2, and the Ui-Shinseol Line.

Funding Statement: This research was supported by a Korean Agency for Infrastructure Technology Advancement (KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (grant no. RS-2023-00239464).

Author Contributions: Study conception and design: Won-Hee Park, draft manuscript preparation: Min-kyeong Kim, data collection: Yeong Geol Lee, review and editing: Su-hwan Yun, Tae-Soon Kwon, Duckhee Lee. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: All original data are included in the article. Further inquiries should be directed to the corresponding author.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

[1] W. Kay *et al.*, "The kinetics human action video dataset," 2017, *arXiv:1705.06950*.
[2] K. Soomro, A. R. Zamir, and M. Shah, "A dataset of 101 human action classes from videos in the wild," *Center Res. Comput. Vis.*, vol. 2, no. 11, pp. 1–7, 2012.
[3] G. A. Sigurdsson, G. Varol, X. Wang, A. Farhadi, I. Laptev and A. Gupta, "Hollywood in homes: Crowdsourcing data collection for activity understanding," 2016, *arXiv:1604.01753*.
[4] H. Kuehne, H. Jhuang, E. Garrote, T. Poggio, and T. Serre, "HMDB: A large video database for human motion recognition," in *Proc. Int. Conf. Comput. Vis. (ICCV)*, 2011.
[5] C. Gu *et al.*, "AVA: A video dataset of spatio-temporally localized atomic visual actions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2018, pp. 6047–6056.
[6] J. A. Labrović, N. Petrović, J. Anđelković, and M. Meršnik, "Patterns of behavioral engagement in an online English language course: Cluster analysis," *J. Comput. High Educ.*, 2023. doi: 10.1007/s12528-023-09382-1.
[7] L. Beinborn and N. Hollenstein, "Behavioral patterns," in *Cognitive Plausibility in Natural Language Processing. Synthesis Lectures on Human Language Technologies*. Cham: Springer, 2024, doi:10.1007/978-3-031-43260-6_4.
[8] A. E. Aiello, A. Renson, and P. N. Zivich, "Social media- and internet-based disease surveillance for public health," *Annu. Rev. Public Health*, vol. 41, no. 1, pp. 101–118, 2020. doi: 10.1146/annurev-publhealth-040119-094402.
[9] N. Antoine-Moussiaux *et al.*, "Valuing health surveillance as an information system: Interdisciplinary insights," *Front. Public Heal.*, vol. 7, 2019, Art. no. 138. doi: 10.3389/fpubh.2019.00138.
[10] J. Cha *et al.*, "Towards single 2D image-level self-supervision for 3D human pose and shape estimation," *Appl. Sci.*, vol. 11, no. 20, 2021, Art.no. 9724. doi: 10.3390/app11209724.
[11] G. V. Kale and V. H. Patil, "A study of vision based human motion recognition and analysis," *Int. J. Ambi. Comput. Intell.*, vol. 7, no. 2, pp. 75–92, 2016. doi: 10.4018/IJACI.2016070104.

[12] Y. Jang, I. Jeong, M. Younesi Heravi, S. Sarkar, H. Shin and Y. Ahn, "Multi-camera-based human activity recognition for human-robot collaboration in construction," *Sensor*, vol. 23, no. 15, 2023, Art. no. 6997. doi: 10.3390/s23156997.

[13] M. N. Khan, M. A. Hasan, and S. Anwar, "Improving the robustness of object detection through a multi-casmera-based fusion algorithm using fuzzy logic," *Front. Artif. Intell.*, vol. 4, 2021, Art. no. 638951. doi: 10.3389/frai.2021.638951.

[14] R. Bodor, A. Drenner, P. Schrater, and N. Papanikolopoulos, "Optimal camera placement for automated surveillance tasks," *J. Intell Robot Syst.*, vol. 50, no. 3, pp. 257–295, 2007. doi: 10.1007/s10846-007-9164-7.

[15] R. Bodor, P. Schrater, and N. Papanikolopoulos, "Multi-camera positioning to optimize task observability," in *IEEE Conf. Adv. Vid. Sig. Based Surveill.*, Como, Italy, 2005, pp. 552–557. doi: 10.1109/AVSS.2005.1577328.

[16] F. Limanta, K. Uto, and K. Shinoda, "CAMOT: Camera angle-aware multi-object tracking," in *IEEE Winter Conf. Appl. Comput. Vis.*, 2024, pp. 6479–6488.

[17] W. -H. Park, S. -J. Park, H. J. Kim, H. S. Kim, and S. C. Oh, "Study on factors for passenger risk in railway vehicle," *J. Soc. Disast. Inform.*, vol. 17, no. 4, pp. 733–746, 2021.

[18] Y. Xing and J. Zhu, "Deep-learning-based action recognition with 3D skeleton: A survey," *CAAI Trans. Intell. Technol.*, vol. 6, pp. 80–92, 2021.

[19] B. Ren, M. Liu, R. Ding, and H. Liu, "A survey on 3D skeleton-based action recognition using learning method," 2020, *arXiv:2002.05907v1*.

[20] H. Ramirez, S. A. Velastin, I. Meza, E. Fabregas, D. Makris and G. Farias, "Fall detection and activity recognition using human skeleton features," *IEEE Access*, vol. 2021, no. 9, pp. 33532–33542, 2021. doi: 10.1109/ACCESS.2021.3061626.

[21] C. J. Taylor, "Reconstruction of articulated objects from point correspondences in a single uncalibrated image," *Comput. Vis. Image Underst.*, vol. 80, pp. 349–363, 2000. doi: 10.1006/cviu.2000.0878.

[22] H. C. Nguyen, T. H. Nguyen, R. Scherer, and V. H. Le, "Deep learning for human activity recognition on 3D human skeleton: Survey and comparative study," *Sensors*, vol. 23, 2023, Art. no. 5121. doi: 10.3390/s23115121.

[23] W. Zhu *et al.*, "Co-occurrence feature learning for skeleton-based action recognition using regularized deep LSTM networks," in *Proc. 30th AAAI Conf. Artif. Intell., AAAI*, Phoenix, AR, USA, Feb. 12–17, 2016, pp. 3697–3703.

[24] J. C. Núñez, R. Cabido, J. J. Pantrigo, A. S. Montemayor, and J. F. Vélez, "Convolutional neural networks and long short-term memory for skeleton-based human activity and hand gesture recognition," *Pattern Recognit.*, vol. 76, pp. 80–94, 2018.

[25] Z. Tian, C. Shen, H. Chen, and T. He, "FCOS: Fully convolutional one-stage object detection," in *2019 IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Seoul, Republic of Korea, 2019, pp. 9626–9635. doi: 10.1109/ICCV.2019.00972.

[26] D. Kang, M. -C. Roh, H. Kim, Y. Kim, and S. -W. Lee, "Classification-based multi-task learning for efficient pose estimation network," in *2022 26th Int. Conf. Pattern Recognit. (ICPR)*, Montreal, QC, Canada, 2022, pp. 3295–3302. doi: 10.1109/ICPR56361.2022.9956539.

[27] N. Aharon, R. Orfaig, and B. Bobrovsky, "BoT-SORT: Robust associations multi-pedestrian tracking," 2022, *arXiv:2206.14651*.

[28] J. An *et al.*, "VFP290K: A large-scale benchmark dataset for vision-based fallen person detection," in *NeurIPS 2021 Track Datasets and Benchmarks Round2*, 2020.

[29] Y. Xiao, J. Chen, Y. Wang, Z. Cao, J. Zhou and X. Bai, "Action recognition for depth video using multi-view dynamic images," *J. Inform. Sci.*, vol. 480, pp. 287–304, 2018.