Sensors and Materials, Vol. 38, No. 1 (2026) 311–328
MYU Tokyo

S & M 4301

AI-based Pedestrian Obstruction Analysis for Safety Assessment of Passenger Terminals

Min-Woo Park,[1] Sung-Sam Hong,[2] and Hwayoung Kim[1*]

[1]Department of Maritime Transportation, Mokpo National Maritime University, Mokpo-si 58628, Korea
[2]Department of Information Security, Hankyung National University, Pyeongtaek-si 17738, Korea

(Received December 25, 2025; accepted January 20, 2026)

Keywords: pedestrian safety, passenger terminal, Mask R-CNN, Swin Transformer, obstruction rate

Passenger terminals are evolving into complex nodes where land and sea transport intersect. However, the safety and convenience of pedestrian environments in access roads are often compromised owing to the development of surrounding commercial areas and inadequate safety facilities. Existing safety assessments for pedestrian environments have primarily relied on qualitative or post-incident analyses centered on static structures, and thus possess fundamental limitations in quantifying real-time risks posed by dynamic obstruction elements. To address these issues, in this study, we propose an AI-based pedestrian obstruction analysis framework. To ensure high-fidelity data acquisition for safety assessment, we utilized optical sensors (action cameras) as wearable sensing units to capture real-time pedestrian dynamics in complex terminal environments. We constructed a dataset of pedestrian obstruction objects based on first-person walking videos recorded with an action camera worn by a pedestrian along the access roads of the Mokpo Port Passenger Terminal. In particular, we adopted the Swin Transformer architecture as the backbone for the mask region-based convolutional neural network (Mask R-CNN) instance segmentation model in order to leverage its previously reported strengths in multiscale object recognition and generalization in complex scenes. Furthermore, we developed the obstruction rate (OBR) measurement algorithm, which utilizes pixel-level mask information of identified objects to calculate their occupancy within designated walking areas. The OBR algorithm was applied to two distinct zones near the terminal, capturing structural differences between sidewalks and mixed pedestrian–vehicle areas. The resulting zone-wise OBR distributions provide a quantitative basis for comparing pedestrian safety conditions and identifying high-risk segments along terminal access routes. In this study, we demonstrate the feasibility of AI-based pedestrian obstruction analysis as a quantitative safety assessment tool and suggest future directions for its integration into real-time monitoring systems and policy decision-making.

*Corresponding author: e-mail: hwayoung@mmu.ac.kr
https://doi.org/10.18494/SAM6147

ISSN 0914-4935
https://myukk.org/

1. Introduction

1.1 Characteristics and problems of pedestrian environments in ports and passenger terminals

Drowning ports are transforming into complex nodal points where human activities are intensifying, expanding beyond traditional logistics functions to include tourism and waterfront spaces. Passenger terminals serve as critical national infrastructure that connect islands with mainland urban centers and are typically constructed in locations with excellent accessibility to city centers.[1,2] However, such high accessibility and the development of surrounding commercial districts pose complex safety risks to the pedestrian environment of terminal access roads. These roads are often characterized by a mix of mobile transportation means such as cars, buses, and motorcycles. Additionally, temporary or fixed obstacles originating from nearby commercial facilities—such as street stalls, signboards, fish tanks, and trash bins—frequently encroach upon pedestrian paths indiscriminately, and in this study, they are collectively labeled under the "obstacle" class in the instance segmentation taxonomy. When the function of walkways is compromised in this manner, the safety and walkability for users are significantly degraded, and the risk of collision accidents increases, which is consistent with previous findings that sidewalk obstructions and narrow effective widths reduce safety and accessibility for pedestrians.[1,3,4]

1.2 Necessity of quantitative assessment of pedestrian safety

Previous research on pedestrian environments at coastal passenger terminals has mainly focused on the static characteristics of facilities, such as the width and condition of walkways, the placement of amenities, and the level of satisfaction reported by users in surveys or complaint records.[1,2] While these approaches are useful for identifying structural deficiencies and perception gaps, they are limited in their ability to capture how dynamic obstruction elements—such as temporarily parked vehicles, informal street stalls, or transient pedestrian clusters—interfere with walking flows over time. In particular, most existing studies either rely on snapshot field surveys or on aggregated indices that do not explicitly represent frame-by-frame changes in the effective width available to pedestrians.[3–7]

In complex transfer environments such as passenger terminals, however, safety risks often arise from instantaneous or short-lived obstructions that may not be visible in static surveys. For example, an illegally parked vehicle or a temporary luggage cart can momentarily block a critical bottleneck segment, creating local crowding and increasing the likelihood of conflicts between pedestrians and vehicles.[3] Capturing such phenomena requires an analysis framework that can link object-level obstruction information from video to quantitative measures of walking-area occupancy. At the same time, the recent development of deep-learning-based instance segmentation models has made it possible to obtain pixel-level masks for pedestrians, vehicles, and fixed facilities directly from video footage, including fixed closed-circuit television

(CCTV) streams and first-person recordings.[8–16] Nevertheless, there is still a methodological gap between these detailed recognition capabilities and the development of interpretable safety indicators that can be used by terminal operators. Most prior works have focused on improving detection accuracy or proposing new network architectures, rather than on defining safety metrics that explicitly quantify the instantaneous loss of effective width owing to obstruction objects.[3,17]

In this context, our aim in this study is to bridge this gap by (1) constructing a passenger-terminal-specific instance segmentation dataset targeting obstruction elements on access walkways[1,18,19] and (2) developing an obstruction rate (OBR) indicator that converts pixel-level mask areas into a quantitative measure of pedestrian-space loss. The proposed framework is applied to first-person walking videos recorded with an action camera along the access roads of the terminal in an offline, frame-by-frame manner to demonstrate how obstruction-induced risk patterns differ across zones with distinct structural and operational characteristics, laying the foundation for future extensions toward real-time monitoring and integrated safety assessment.[2,4–6,13–16] By integrating advanced vision sensors with deep-learning-based spatial analysis, this framework contributes to the development of sensor-based safety monitoring systems within the digital twin environment of passenger terminals.

1.3 Structure and contribution of this paper

Reflecting these requirements, in this paper, we propose an algorithm to calculate the OBR on the basis of the results of an advanced deep-learning-based pedestrian obstruction recognition model. The remainder of this paper is organized as follows: In Sect. 2, we review related work on pedestrian environment analysis and trends in deep-learning-based object recognition. In Sect. 3, we define pedestrian obstruction elements specific to passenger terminals and describe in detail the dataset construction process. In Sect. 4, we present the structure and training environment of the mask region-based convolutional neural network (Mask R-CNN) model applying the Swin Transformer (Swin-T) backbone, focusing on the technical advantages of Swin-T. In Sect. 5, we quantitatively present the concept and calculation logic of the developed OBR algorithm. In Sect. 6, we present the quantitative OBR measurement results applied to actual video data and discuss their significance and limitations in depth. Finally, in Sect. 7, we conclude the paper and outline future research plans.

The main contributions of this study are as follows: First, we constructed a pedestrian obstruction dataset from first-person walking videos recorded with an action camera along the access road to the Mokpo Port Passenger Terminal, explicitly labeling both fixed facilities and dynamic entities that interfere with pedestrian movement. Second, on the basis of this dataset, we adopted a Mask R-CNN instance segmentation model with a Swin-T backbone and verified its applicability to complex pedestrian environments around coastal terminals. Third, we developed the OBR algorithm, which measures obstruction intensity by computing the ratio of obstruction mask areas to predefined walking areas, and applied it to two structurally different zones to demonstrate how OBR can serve as a quantitative safety index for pedestrian environments.

2. Related Work

2.1 Pedestrian environment assessment and obstruction analysis

Research on pedestrian environments has a long history in urban planning, transportation engineering, and public health, addressing topics such as walkability, level of service, and accessibility.[1–6] Traditional approaches often focus on physical infrastructure characteristics (e.g., sidewalk width, surface condition, and presence of crosswalks and signals), land use patterns (e.g., residential density and proximity to destinations), and subjective assessments of safety and comfort gathered through surveys or interviews.[3,4] In the context of coastal passenger terminals, some groups have examined user satisfaction and facility adequacy using methods such as Importance–Performance Analysis (IPA).[1,2] These works have identified issues such as insufficient guidance signage, inadequate waiting spaces, and conflicts between pedestrian and vehicular flows near terminal entrances.

However, these conventional assessments generally treat obstructions as static elements or as qualitative concerns raised in interviews and complaint records. They rarely quantify how specific obstruction objects occupy pedestrian space over time or how such occupancy translates into collision risk or discomfort. In particular, the dynamic interplay among pedestrians, vehicles, and temporary facilities (e.g., street vendors, luggage carts, and informal seating) is challenging to capture using static surveys or periodic field observations alone. As a result, there is a gap between the subjective experience of pedestrians, who often perceive certain segments or time periods as particularly congested or unsafe, and the objective indicators used in facility planning and safety management.

To address these limitations, some researchers have begun to integrate sensor data and computer vision techniques into pedestrian environment analysis. For example, CCTV-based pedestrian counting, density estimation, and trajectory tracking have been used to analyze crowding patterns and evacuation dynamics.[13–16] In urban open spaces and large facilities, such approaches enable the identification of high-density regions and bottleneck points that may be prone to accidents. Nevertheless, most of these studies focus on the pedestrian flow itself rather than on obstruction objects that reduce the effective walking area. While high pedestrian density is certainly a risk factor, situations where fixed or movable objects protrude into walkways, forcing people to detour or squeeze through narrow gaps, also require attention from a safety perspective.

From the viewpoint of passenger terminals, where access roads are often shared among pedestrians, passenger vehicles, and freight operations, the ability to quantitatively evaluate obstruction-induced narrowing of walking space is particularly critical. For example, shuttle buses or taxis may temporarily stop in front of the terminal, and delivery vehicles may occupy parts of the sidewalk during loading and unloading. Meanwhile, commercial facilities and vendors may place signboards, goods, or tables on the sidewalk, further reducing the usable width for pedestrians. When these conditions coincide with peak arrival or departure times, the risk of collision accidents increases, which is consistent with prior findings that sidewalk obstructions and narrow effective widths reduce safety and accessibility for pedestrians.[1,3,4]

2.2 Trends in deep-learning-based instance segmentation

Since the advent of deep learning, AI-based object recognition technology has evolved on the basis of convolutional neural networks (CNNs) and has recently achieved significant advancements in computer vision with the emergence of Transformer architectures.[8] In particular, instance segmentation technology, which classifies objects down to the pixel level, has been actively studied because it can simultaneously provide object category, location, and shape information.[10,17] Mask R-CNN[10] is a representative instance segmentation model capable of precisely identifying object contours by adding a segmentation branch to the bounding-box-based detection framework of Faster R-CNN.[9]

In the field of pedestrian environment analysis, instance segmentation can be used to detect both pedestrians and obstruction objects such as signposts, vehicles, and temporary facilities in CCTV footage.[17] By generating pixel-level masks, the model can distinguish overlapping objects and accurately estimate the area of each element. This is particularly advantageous for quantifying obstruction intensity, as it allows the computation of ratios between obstruction areas and predefined walking areas.

Meanwhile, the introduction of Vision Transformers (ViTs) and related architectures has further improved the performance and flexibility of image recognition models.[8,11,20] Unlike CNNs, which primarily rely on local convolution operations, Transformer-based models capture long-range dependences through self-attention mechanisms. Swin-T, for example, divides images into non-overlapping windows and applies self-attention within each window, periodically shifting the windows to exchange information across regions.[11] This hierarchical design enables Swin-T to achieve high recognition accuracy on various computer vision benchmarks while maintaining computational efficiency.

Recent studies have also explored the integration of Swin-T backbones into object detection and segmentation frameworks. On benchmarks such as MS COCO, models using Swin-T backbones have demonstrated improvements in both box-level and mask-level performance compared with ResNet-based counterparts under similar computational budgets.[11] These results suggest that Swin-T is suitable for complex scenes where objects of various sizes and shapes coexist, such as pedestrian environments around passenger terminals.

Most prior research applying deep learning to pedestrian environments has examined topics such as pedestrian detection, crowd counting, and abnormal behavior recognition.[13–16] While these studies have significantly advanced our ability to monitor and analyze pedestrian flows, they have not fully explored how pixel-level segmentation outputs can be converted into quantitative safety indices. In particular, research that treats “obstruction” as an integrated concept encompassing both fixed and dynamic elements and that defines a metric such as OBR on the basis of the area occupied by these elements within walking areas is still scarce. Therefore, in this study, we focus on the following: (1) establishing a domain-specific obstruction taxonomy for passenger terminal access roads; (2) constructing a corresponding instance segmentation dataset; (3) selecting and applying an advanced backbone (Swin-T) for Mask R-CNN; and (4) developing and validating an OBR algorithm that quantifies obstruction intensity from a pedestrian’s perspective.

3. Definition of Obstruction Elements and Dataset Construction

3.1 Taxonomy of obstruction elements on passenger terminal access roads

From the viewpoint of pedestrian safety and convenience, an obstruction element is defined as any object that occupies part of the walking area, thereby reducing the effective width available for safe movement. In the context of passenger terminals, such obstruction elements include both fixed facilities (e.g., street trees, benches, and signposts) and dynamic entities (e.g., pedestrians, vehicles, luggage, and personal mobility devices). On the basis of field surveys around the Mokpo Port Passenger Terminal and prior studies on pedestrian environment and obstruction analysis,[1,18,19] in this study, we define the obstruction taxonomy as shown in Table 1.

3.2 Construction of instance segmentation-based dataset

On the basis of the 11-class obstruction taxonomy defined in Sect. 3.1,[1,19] first-person walking videos were recorded using an action camera worn by a pedestrian along the main access road to the Mokpo Port Passenger Terminal, covering both the sidewalk section in front of the terminal (Zone A) and the mixed pedestrian–vehicle exit area after disembarkation (Zone B).[2] From these videos, frames were sampled at a fixed interval to capture a wide range of temporal situations, including normal states with minimal obstructions, transient events such as vehicle stopping or boarding, and crowded conditions near peak passenger demand. The exact sampling interval is not critical here, as OBR is computed independently for each sampled frame and the analysis focuses on relative frame-wise patterns rather than absolute time scaling.

Pixel-level instance masks for pedestrians and obstruction elements were annotated using the LabelMe tool.[13–16] All annotations were then converted into the MS COCO dataset format to ensure compatibility with standard instance segmentation frameworks such as Mask R-CNN.[10–12] In this conversion, each object instance was assigned one of the 11 classes, and its segmentation polygons was stored as part of the COCO-style annotation file. This structure enables the direct computation of obstruction areas and supports the training of models that output both bounding boxes and segmentation masks.

During dataset construction, particular attention was paid to correctly labeling pedestrians and vehicles that partially overlap with fixed facilities or with each other. Since OBR is defined

Table 1
Taxonomy of pedestrian obstruction elements around passenger terminal access roads.

Category	Subcategory (Class name)	Description
Mobile objects	person, car, bicycle, motorcycle, mobility_scooter, handcart	- Dynamic objects with collision potential. - Port-specific objects: Handcart (for luggage) and mobility_scooter (for the elderly) defined as separate classes to enhance detection.
Fixed objects	tree, bench, chair, signboard	- Facilities fixed to the walkway that impede traffic flow or reduce effective width.
Others	obstacle	- Unspecified piled objects or barriers not belonging to the above categories but hindering walking.

on the basis of the area occupied by obstruction elements within the walking area, it is essential to accurately distinguish overlapping masks in crowded scenes. When two or more objects overlap in the image, each object's mask retains its own area, and the overlapping regions are later handled at the algorithmic stage of OBR calculation. This approach allows the algorithm to represent multilayered obstruction situations, which are common near terminal entrances and exits.

Finally, the labeled data were divided into training and validation sets in accordance with standard machine learning practice. The training set was used to optimize the parameters of the Mask R-CNN model with a Swin-T backbone, while the validation set was used to monitor performance and prevent overfitting. Frames not used for training were reserved for the qualitative confirmation of segmentation results in the case study.

4. Mask R-CNN with Swin-T Backbone

4.1 Overview of Mask R-CNN

Mask R-CNN is a widely used instance segmentation framework that extends the Faster R-CNN object detection model by adding a branch for predicting pixel-level masks.[10] The model consists of three main components: (1) a backbone network for feature extraction, (2) a region proposal network (RPN) that generates candidate object regions, and (3) classification, regression, and mask branches that operate on region of interest (RoI) features.

In the backbone stage, an input image is processed through a deep neural network (e.g., ResNet and Swin-T) to produce multiscale feature maps. The RPN then slides small convolutional filters over these feature maps to propose candidate bounding boxes (anchors) that may contain objects.[9,10] For each proposed region, RoIAlign is used to extract a fixed-size feature representation that preserves spatial alignment. Finally, the classification branch predicts object categories, the regression branch refines bounding box coordinates, and the mask branch outputs a binary mask for each object–instance pair.

Mask R-CNN has several advantages for pedestrian obstruction analysis. First, its instance-level segmentation capability allows the separation of individual obstruction elements even when they overlap in the image. Second, the combination of bounding box and mask outputs facilitates the calculation of both object counts and areas, which are essential for defining indicators such as OBR. Third, the modular architecture of Mask R-CNN makes it straightforward to swap backbone networks, enabling the use of advanced feature extractors such as the Swin-T without changing the overall detection and segmentation pipeline.[10,11]

4.2 Swin-T backbone

Swin-T is a hierarchical Vision Transformer that applies self-attention within local windows, which are periodically shifted to allow cross-window information exchange.[11] This design addresses the computational inefficiency of global self-attention in standard ViT while preserving the ability to model long-range dependences. In Swin-T, an image is first partitioned into non-overlapping patches that are linearly embedded and passed through a sequence of

Swin-T blocks. Each block consists of a window multihead self-attention (W-MSA) layer and a shifted window multihead self-attention (SW-MSA) layer, enabling the model to capture both local and global contexts.

When used as a backbone for object detection and instance segmentation, Swin-T replaces the convolution-based feature extractor (e.g., ResNet) with a Transformer-based one. Feature maps of different resolutions are generated at multiple stages, which can be fed into feature pyramid networks (FPNs) and subsequent detection heads. Prior studies have revealed that Swin-T backbones achieve higher detection and segmentation accuracy than ResNet-based backbones on benchmarks such as COCO and ADE20K under comparable computational conditions.[11]

In this study, Swin-T is adopted as the backbone of Mask R-CNN to enhance the recognition of pedestrian obstruction elements in the complex environment of the Mokpo Port Passenger Terminal. The access road scenes include objects of diverse scales, shapes, and textures, ranging from small signboards and personal mobility devices to large vehicles and groups of pedestrians. Swin-T's ability to capture multiscale features and long-range dependences is expected to improve the segmentation of such varied objects, especially in scenes where multiple obstruction elements overlap or are partially occluded.

In addition, Swin-T's hierarchical feature maps are well suited for integration with the FPN module in Mask R-CNN, allowing the effective use of both shallow and deep features. Shallow features provide detailed local information useful for delineating object boundaries, whereas deep features encapsulate high-level semantic context that helps distinguish between similar-looking objects (e.g., bicycles vs motorcycles and signboards vs other fixed facilities). This multiscale representation is crucial for accurately identifying and segmenting obstruction elements that directly affect pedestrian walking space, and its advantages are illustrated in Table 2, which summarizes reference instance segmentation results reproduced from Ref. 11 on the COCO dataset and motivated our choice of Swin-T as the backbone in this study.

4.3 Training settings and hyperparameters

The model used for training was Mask R-CNN with a Swin-T backbone. The detailed training hyperparameters are listed in Table 3. To verify the reliability of the OBR metric, we evaluated the model's quantitative performance on the validation set of the Mokpo Port dataset. Unlike previous exploratory stages, we conducted a rigorous performance evaluation using mean average precision (*mAP*) to ensure the accuracy of the segmentation masks.

As shown in Table 4, the trained model achieved a mAP_{segm} (Mask AP) of 0.488 and a mAP_{50} of 0.897. The high mAP_{50} value indicates that the model is exceptionally robust in extracting the

Table 2
Example of COCO instance segmentation results for cascade Mask R-CNN with different backbones (reproduced from Ref. 11).

Backbone	AP_box	AP_mask
ResNet-50	46.3	40.1
Swin-T	50.5	43.7

Table 3
Training model hyperparameters.

Parameter	Value
Backbone	Swin-T
Batch size	2
Learning rate	0.001
Weight decay	0.05
Device	CPU

Table 4
Quantitative performance of Mask R-CNN (Swin-T) on the Mokpo Port dataset.

Metric	Value
mAP_{bbox}	0.485
mAP_{segm}(Mask)	0.488
mAP_{segm}@50	0.897
mAP_{segm}@75	0.545

contours of pedestrians and obstacles within the complex terminal environment. These quantitative results provide a solid foundation for the subsequent OBR-based safety analysis, ensuring that the spatial occupancy measurements are derived from highly accurate instance masks.

5. Development of Quantitative Obstruction Rate Measurement Algorithm

5.1 Concept of pedestrian obstruction and definition of OBR

Pedestrian obstruction in this study is defined as a phenomenon in which an obstruction object recognized by AI—including both structural obstacles (e.g., signboards, fish tanks, and benches) and dynamic instances such as pedestrians and vehicles—occupies a preset walking area, thereby reducing the effective width available for safe and smooth pedestrian movement.[1–3] On the basis of this concept, the OBR is introduced as a quantitative indicator that measures how much of the designated walking area is effectively taken up by obstruction objects at a given moment.

Unlike indicators that simply count the number of objects, OBR is defined in terms of area: it is computed using the pixel-level masks output by the instance segmentation model and represents the ratio between the occupied area and the total area of the walking region.[10–12,17] In this formulation, all obstruction masks that intersect the walking area are considered, and their overlapping regions are summed. Therefore, when multiple obstruction masks overlap in the image (e.g., a pedestrian standing in front of a large signboard), the same pixel can be counted several times. As a result, OBR is not mathematically bounded by 100%; values exceeding 100% are possible and are intentionally interpreted as indicating multilayered obstruction severity, where several obstruction elements are stacked in depth from the camera's viewpoint.

In this study, OBR is thus treated as an obstruction intensity index that directly quantifies the instantaneous loss of effective width, rather than as a simple percentage bounded between 0 and

100. High OBRs, including outliers above 100%, signal situations in which pedestrians are likely to experience a combination of structural blind spots and crowding.[3,13–16] Once embedded in an operational monitoring system, such values can be used by terminal managers to identify high-risk segments or time periods and to trigger environmental improvements or management actions when OBR exceeds predefined thresholds. In this work, however, OBR is primarily used as a second-stage analysis tool that translates detailed instance segmentation outputs into an interpretable safety indicator for specific segments of the terminal access road.

5.2 Frame-based OBR calculation logic

The calculation of OBR begins with the output of the Mask R-CNN model with the Swin-T backbone. For each input frame, the model produces a set of predicted instances, each with a class label, bounding box, and binary segmentation mask. Among these instances, those that belong to the obstruction classes defined in Sect. 3.1 (e.g., person, car, motorcycle, signboard, and obstacle) are selected. Non-obstruction categories, if any, are excluded from the subsequent OBR computation; however, in this study, all defined classes within the modeled scene are treated as obstruction elements since they occupy potential walking space from the pedestrian's perspective.

Next, a walking area mask is defined for each zone of interest (Table 5). In Zone A (the sidewalk in front of the terminal), the walking area is defined as the region that pedestrians are expected to use under normal conditions, excluding roadways for vehicles and areas outside the formal pedestrian path. In Zone B (the mixed pedestrian–vehicle area used after disembarkation), the entire corridor connecting the ship to the terminal is treated as a walking area because pedestrians and vehicles share the same physical space without a clearly segregated sidewalk. This distinction reflects the structural difference between the two zones: Zone A has a relatively clear separation between sidewalk and roadway, whereas Zone B does not.

For each frame and each zone, the OBR (Table 6) is computed by intersecting the obstruction masks with the predefined walking area mask, summing the overlapping pixel counts, and dividing the result by the total number of pixels in the walking area [Eq. (1)].

$$OBR\ (\%) = \frac{\sum_{i=1}^{N} Area\left(Object_i \cap WalkingArea\right)}{Area\left(WalkingArea\right)} \times 100 \tag{1}$$

Here, the intersection areas are summed over all obstruction masks. When masks overlap, pixels in the overlapping region contribute to multiple $Area(Object_i \cap WalkingArea)$ terms, which leads

Table 5
Structure of instance segmentation dataset for case study zones.

Zone	#Video clips	#Annotated frames (total)	#Person instances	#Vehicle instances	#Fixed-object instances (tree/bench/chair/signboard/ obstacle)	Train frames	Val. frames
A	1	166	15	76	332	131	35
B	1	85	341	97	0	69	16

Table 6
Definition of key symbols for AI-based OBR formula.

Symbol	Definition	Description
OBR (%)	Pedestrian obstruction rate	Ratio of the area actually occupied by obstruction objects relative to the designated walking area.
N	Number of detected objects	Total number of object instances classified as obstruction elements among AI detection results, including both structural obstacles and the person class.
$Area(WalkingArea)$	Total walking area	Total pixel area of the pedestrian path set as ROI in advance.
$Object_i$	ith detected object	Instance mask area of the ith detected obstruction object.
$\cap$	Intersection	An operation that extracts only the overlapping part between the object mask and the walking area.

to the possibility that the numerator exceeds the walking-area denominator. Rather than removing this effect, the algorithm preserves it so that situations with multilayered obstruction (e.g., overlapping pedestrians and facilities) are reflected as higher OBRs.[3] This design choice aligns with the study's goal of capturing the intensity of obstruction as experienced by pedestrians walking through a corridor with stacked obstacles.

Once per-frame OBRs are computed for each zone, additional statistics such as mean, median, variance, and histogram distributions can be derived. These derived indicators help to characterize the overall obstruction profile of each zone and to compare zones with different structural features and usage patterns. For example, a zone with a consistently high OBR may be considered structurally constrained, whereas a zone with a highly variable OBR may be subject to intermittent but severe obstruction events.

6. Case Study: OBR-based Safety Assessment of Terminal Access Zones

6.1 Overview of study site and zone definition

The proposed OBR framework was applied to the access road of the Mokpo Port Passenger Terminal, a major coastal passenger terminal that serves both local residents and tourists. The analysis focused on two representative zones: Zone A, the sidewalk in front of the terminal entrance, and Zone B, the mixed pedestrian–vehicle corridor used by disembarking passengers.

Zone A is characterized by a designated sidewalk separated from the roadway, with commercial facilities and street furniture located along the side. Under ideal conditions, pedestrians can use the full width of the sidewalk without entering the road. However, in practice, various obstruction elements such as signboards, benches, street trees, and temporarily parked vehicles frequently encroach upon the walking area, forcing pedestrians to detour or move closer to the roadway. In addition, groups of pedestrians often congregate near entrances to shops or waiting areas, creating local congestion despite the nominal existence of a separate sidewalk.

Zone B, by contrast, is a corridor that connects the ship to the terminal, where pedestrians, vehicles, and luggage flows are not clearly separated. In this area, passengers disembark from

the ship and walk toward the terminal while vehicles such as passenger cars, buses, and freight trucks use the same physical space to enter or leave the port. Temporary guide lines or barriers may be installed, but they do not create a permanent sidewalk-like structure. As a result, the walking area in Zone B is defined as the entire corridor, and any object that occupies this space—including vehicles and groups of passengers—is treated as an obstruction element from the perspective of pedestrians. In the annotated clips used in this study, no fixed obstacles such as piled luggage were actually observed in Zone B, which is reflected in the zero count for fixed-object instances in Table 7. In Zone B, where vehicles and pedestrians coexist, all detected vehicles within the designated walking area were initially categorized as obstruction elements to assess the maximum potential risk. However, this approach may lead to an overestimation of risk by including normal traffic flow. Further refinement to distinguish between moving traffic and hazardous encroachment will be addressed in future studies.

6.2. OBR measurement results and comparative analysis between zones

Using the trained Mask R-CNN with the Swin-T backbone, we performed instance segmentation on first-person walking videos recorded with an action camera covering multiple time periods in Zones A and B. For each frame, obstruction masks were identified, and OBRs were calculated following the logic described in Sect. 5.2. The resulting per-frame OBRs were then aggregated into histograms and summary statistics for each zone.

In Zone B, the OBR distribution was generally shifted toward higher values and exhibited a longer tail than in Zone A. This pattern reflects the structural characteristics of Zone B, where the walking area is shared with vehicles and where disembarking passengers often move in dense groups. In particular, frames with OBRs exceeding 100% were more frequently observed in Zone B, indicating situations where multiple layers of obstruction elements (e.g., overlapping pedestrians and vehicles) are stacked within the walking area. These high-OBR frames correspond to conditions in which pedestrians may experience significant difficulty in finding safe paths through the corridor, with increased potential for close-contact interactions and limited visibility around large vehicles.

When comparing Zones A and B, the mean and median OBRs, as well as the proportion of frames falling into high-OBR ranges, provide a quantitative basis for assessing relative safety

Table 7
Example of frame-level OBR statistics for representative situations in Zones A and B.

Frame ID	Zone	#Persons	#Vehicles	#Fixed obstacles	$\Sigma Area(Object_i \cap WalkingArea)$ (pixels)	$Area(WalkingArea)$ (pixels)	OBR (%)
f_A1 (A_frame_0219)	A	0	1	0	119092	1708641	≈6.97
f_A2 (A_frame_0031)	A	0	0	7	9316707	1708641	≈545.27
f_B1 (B_frame_0126)	B	4	1	0	4080845	8294400	≈49.20
f_B2 (B_frame_0137)	B	6	0	0	6697728	8294400	≈80.75

conditions. In this paper, we mainly illustrate these tendencies using the time-series and histogram plots in Figs. 1–3, and detailed numerical tables for these summary statistics are omitted for brevity. Zone A, despite having a designated sidewalk, occasionally exhibits elevated OBR due to the encroachment of commercial facilities and temporary parking, suggesting that better management of fixed and movable facilities is necessary. Zone B, lacking clear separation between pedestrians and vehicles, consistently shows higher obstruction intensity, implying that structural measures (e.g., dedicated pedestrian lanes AND physical barriers) may be required to ensure safety.

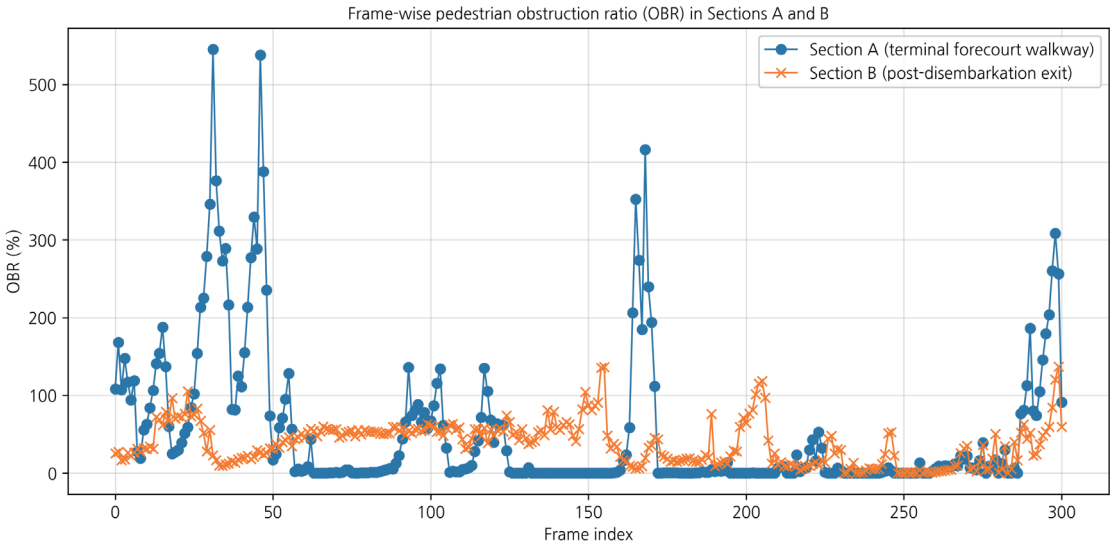

Fig. 1. (Color online) OBR time-series trends (Zone A and Zone B).

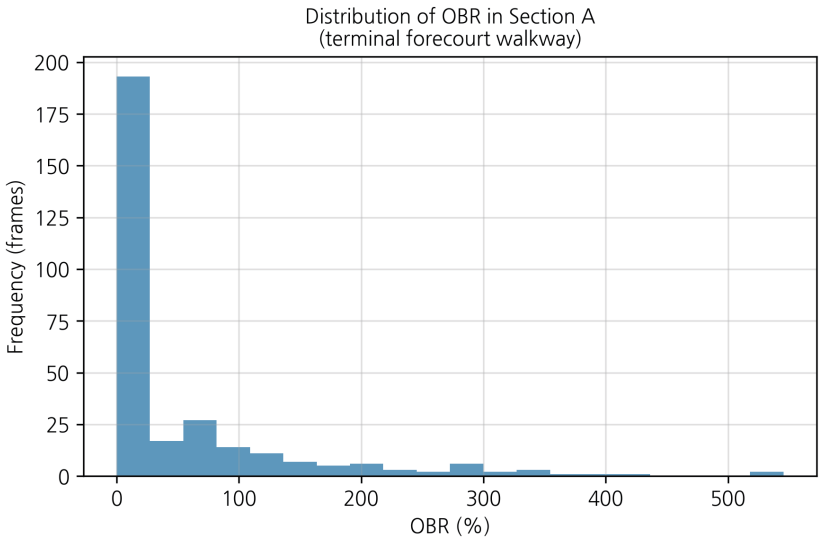

Fig. 2. (Color online) OBR frequency distribution in Zone A (sidewalk in front of terminal).

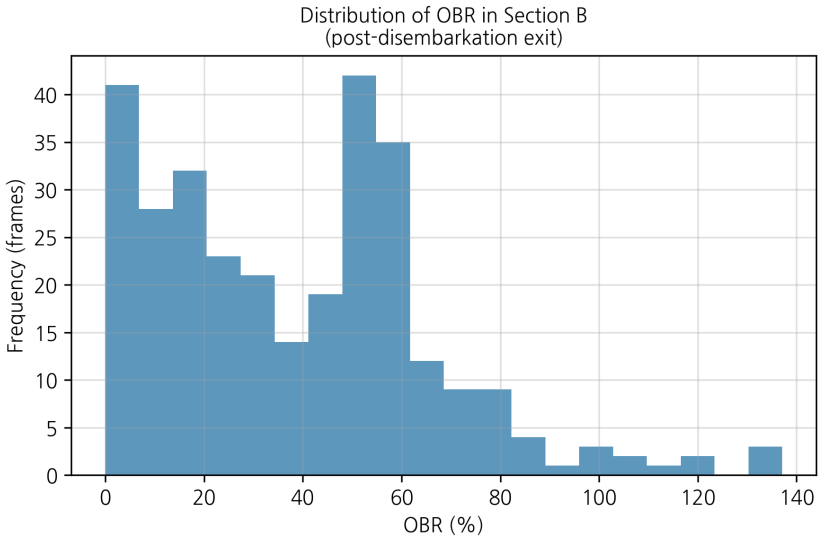

Fig. 3. (Color online) OBR frequency distribution in Zone B (exit after disembarkation).

Representative frames from each zone were also examined to qualitatively interpret OBRs in relation to specific obstruction configurations. For example, frames with low OBRs in Zone A typically showed a relatively clear sidewalk with only a few pedestrians and properly placed street furniture. In contrast, frames with high OBR in Zone B often depicted situations where large vehicles occupied a substantial portion of the corridor while groups of passengers and luggage were clustered around them. These examples confirm that OBR effectively captures the instantaneous obstruction intensity in a manner consistent with intuitive perceptions of safety and comfort.

6.3. OBR frequency distribution and cause analysis by zone

To support the previous time-series analysis, we analyzed in detail how OBRs are distributed in each zone using histograms.

6.3.1 Zone A: Occurrence of outliers owing to structural occlusion

In Zone A, the histogram shows that most frames fall within the 0–10% OBR range, indicating that obstruction elements are minimal under normal conditions. However, a small number of outlier frames exhibit values exceeding 100%. These outliers correspond to time intervals in which large structural elements, such as fish tanks or signboards located in front of shops, completely cover the walking area from the camera's perspective, sometimes with pedestrians standing in front of them. In such cases, overlapping masks cause the same pixels to be counted multiple times in the numerator, leading to OBRs above 100%. These extreme values should therefore be interpreted as flags indicating highly cluttered or structurally occluded

frames, rather than as physically exact ratios of occupied walking width. In this study, we interpret these over-100% values not as errors but as indicators of multilayered obstruction severity that quantitatively reveal the risk of structural blind spots where pedestrians' fields of view are heavily blocked.[2,3] While traditional metrics such as level of service (LOS) focus primarily on pedestrian density, the proposed OBR specifically quantifies the space occupied by physical obstacles. This provides a more direct measure of "walkability" and "safety hazards" from a facility management perspective. Qualitative observations of high-OBR frames confirmed that these values correlate with actual blockages, such as illegally parked vehicles or stacked cargo, which significantly impede pedestrian flow.

6.3.2 Zone B: Continuous congestion in mixed pedestrian environment

On the other hand, the distribution of Zone B shown in Fig. 3 displays a markedly different aspect from Zone A. It forms a relatively broad distribution over the 20–80% range, indicating that OBRs are widely spread rather than concentrated in very low obstruction intervals. This reflects the characteristics of a shared-space environment where disembarking crowds and pick-up vehicles are intermingled, rather than being temporary obstructions. The extremely low frequency of 0% on the graph implies that the area rarely provides a "completely safe time" for pedestrians, statistically indicating that it functions as a high-risk zone requiring constant safety management.

7. Conclusions and Future Work

In this study, we proposed an AI-based pedestrian obstruction analysis framework for the safety assessment of passenger terminal access roads. By constructing a domain-specific instance segmentation dataset for the Mokpo Port Passenger Terminal and adopting a Mask R-CNN model with the Swin-T backbone, we demonstrated the feasibility of accurately recognizing both fixed and dynamic obstruction elements in complex pedestrian environments. On the basis of the segmentation results, we developed the OBR algorithm, which quantifies the area occupied by obstruction elements within predefined walking areas. Applying this algorithm to two structurally distinct zones around the terminal revealed clear differences in obstruction intensity and distribution, highlighting the potential of OBR as a quantitative safety indicator. The proposed framework demonstrates the potential of integrating high-resolution vision sensors with AI to provide a scalable sensing solution for quantifying spatial risks in terminal safety management. However, in this study, OBR is not directly validated against accident records and is therefore used as an exploratory proxy of obstruction intensity, leaving the quantitative linkage to incident risk as future work. However, the current OBR algorithm calculates area on the basis of 2D pixel counts, which may involve perspective distortion where objects closer to the camera appear larger than distant ones. To improve spatial accuracy, incorporating depth-sensing technology or perspective correction algorithms will be considered in future studies.

The main findings of this study can be summarized as follows. First, the integration of Swin-T backbones into Mask R-CNN enabled robust instance segmentation performance in the

complex scenes of passenger terminal access roads, where objects of various scales and types coexist. Second, the proposed OBR algorithm successfully translated pixel-level segmentation outputs into an interpretable index of obstruction intensity from the pedestrian's perspective, capturing both fixed facilities and dynamic entities such as pedestrians and vehicles. Third, the comparative analysis of Zones A and B demonstrated that OBR values reflect structural differences in pedestrian environments, with the mixed pedestrian–vehicle corridor (Zone B) exhibiting higher and more variable obstruction intensity than the designated sidewalk zone (Zone A).

From a practical standpoint, OBR provides terminal managers and policymakers with a quantitative tool to identify high-risk segments and time periods along access routes. By monitoring OBR distributions over time, it becomes possible to evaluate the effectiveness of interventions such as relocating facilities, regulating temporary parking, or installing physical barriers between pedestrians and vehicles. In addition, the OBR framework can be extended to other types of transfer facility, such as subway stations, bus terminals, and airports, where complex interactions between pedestrians and obstruction elements similarly affect safety and comfort.

Future research directions include expanding the dataset—which, in the present study, consists of two representative clips and a small number of annotated frames—to cover a wider variety of weather and lighting conditions and to include additional obstruction classes relevant to different terminal types and operational scenarios, thereby enabling more statistically generalizable analyses and direct validation against safety outcomes. Moreover, while we applied OBR analysis to first-person walking videos recorded with an action camera in an offline manner in this study, in future work the framework can be integrated into real-time monitoring systems based on live CCTV streams, enabling automated alerts when OBR exceeds predefined thresholds. Finally, the combination of OBR with other pedestrian safety metrics, such as crowd density, walking speed, and near-miss detection, may yield a more comprehensive assessment of risks in passenger terminal environments. This study serves as an exploratory analysis focusing on the technical feasibility of the OBR framework. Consequently, temporal factors such as peak-hour congestion or seasonal variations were not fully integrated in this initial stage. Future research will involve long-term data collection across various time slots and diverse terminal environments to enhance the generalizability and predictive reliability of the proposed safety assessment model.

Acknowledgments

This research was supported by the Regional Innovation System & Education (RISE) program through the Jeollanamdo RISE Center, funded by the Ministry of Education (MOE) and the Jeollanamdo, Republic of Korea (2025-RISE-14-002).

References

1 T. H. Song and H. Y. Kim: J. Korea Port Econ. Assoc. **37** (2021) 93. https://doi.org/10.38121/kpea.2021.09.37.3.93
2 H. Kim, J. Choi, Y. Nam, and J.-H. Youn: Sustainability **14** (2022) 2115. https://doi.org/10.3390/su14042115
3 D. Gebremariam and A. Nigussie: Adv. Civ. Eng. **2024** (2024) 5672280. https://doi.org/10.1155/2024/5672280
4 A. M. Mulyadi, A. V. R. Sihombing, H. Hendrawan, A. Vitriana, and A. Nugroho: Transp. Res. Interdiscip. Perspect. **16** (2022) 100695. https://doi.org/10.1016/j.trip.2022.100695
5 W. Gao, M. Yang, and J. Zhang: Sustain. Cities Soc. **85** (2022) 104055. https://doi.org/10.1016/j.scs.2022.104055
6 F. Ulhaq, A. Basfian, and D. Sudibyo: Int. J. Sustain. Constr. Eng. Technol. **15** (2024) 88. https://doi.org/10.30880/ijscet.2024.15.01.024
7 S. Rakshith, S. Mahadeva, and M. R. Santosh: Urban Plan. Transp. Res. **13** (2025) 1. https://doi.org/10.1080/21650020.2025.2557254
8 L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie, and L. Farhan: J. Big Data **8** (2021) 1. https://doi.org/10.1186/s40537-021-00444-8
9 S. Ren, K. He, R. Girshick, and J. Sun: Adv. Neural Inf. Process. Syst. **28** (2015) 91. https://papers.nips.cc/paper/5638-faster-r-cnn-towards-real-time-object-detection-with-region-proposal-networks
10 K. He, G. Gkioxari, P. Dollár, and R. Girshick: Proc. IEEE Int. Conf. Computer Vision (IEEE, 2017) 2961. https://doi.org/10.1109/ICCV.2017.322
11 Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo: Proc. IEEE/CVF Int. Conf. Computer Vision (IEEE, 2021) 10012. https://doi.org/10.1109/ICCV48922.2021.00986
12 T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick: Proc. European Conf. Computer Vision (Springer, 2014) 740. https://doi.org/10.1007/978-3-319-10602-1_48
13 O. Elharrouss, N. Almaadeed, S. Al-Maadeed, and A. Bouridane: J. Visual Commun. Image Represent. **77** (2021) 103116. https://doi.org/10.1016/j.jvcir.2021.103116
14 C. Fan, Y. Li, and M. Feng: IET Image Process. **16** (2022) 1509. https://doi.org/10.1049/ipr2.12446
15 I. V. Pustokhina, D. A. Pustokhin, T. Vaiyapuri, D. Gupta, S. Kumar, and K. Shankar: Saf. Sci. **142** (2021) 105356. https://doi.org/10.1016/j.ssci.2021.105356
16 J. Yoon, J. Choi, and J. Lee: Appl. Sci. **15** (2025) 1823. https://doi.org/10.3390/app15041823
17 M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman: Int. J. Comput. Vis. **111** (2015) 98. https://doi.org/10.1007/s11263-014-0733-5
18 B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman: Int. J. Comput. Vis. **77** (2008) 157. https://doi.org/10.1007/s11263-007-0090-8
19 National Information Society Agency: https://www.aihub.or.kr/ (accessed Nov. 28, 2025).
20 A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby: Proc. Int. Conf. Learning Representations (ICLR, 2021). https://arxiv.org/abs/2010.11929

About the Authors

Min-Woo Park received his B.S. degree from Chongshin University, Korea, in 2019. From 2022 to 2024, he was a software developer in the Research and Development team at Rabahgroow Co., Ltd., an IT service company specializing in AI-based solutions for cyber security and construction safety. Since 2025, he has been a researcher at Mokpo National Maritime University, Korea. His research interests are focused on integrating AI into maritime safety to establish a secure and safe environment. (mwpark@rabahgroow.com)

Sung-Sam Hong received his B.S. degree from Gachon University, Korea, in 2009 and his M.S. and Ph.D. degrees from Gachon University, Korea, in 2011 and 2016, respectively. From 2016 to 2022, he was a research professor at Gachon University, Korea. Since 2025, he has been an instructor at Hankyung University, Korea. His research interests are in AI, data modeling, cyber security, security and safety systems, and AIoT. (sungsamhong@hknu.ac.kr)

Hwayoung Kim received his B.S. and M.S. degrees from Mokpo National Maritime University, Korea, in 1998 and 2002, respectively, and his Ph.D. degree from Kyushu University, Japan, in 2007. From 2014 to 2025, he was an associate professor, and since 2025, he has been a professor at Mokpo National Maritime University. His research interests are in logistics and maritime information systems. (hwayoung@mmu.ac.kr)