ISSN: 2424-8975 (O)
2424-7979 (P)

RESEARCH ARTICLE

Estimating perceived street safety via pairwise comparisons and semantic segmentation with a social-psychological lens

Haoyuan Xiao[*], Yoshinori NATSUME

Architecture and Design, Nagoya Institute of Technology, Nagoya, Japan

*** Corresponding author:** Haoyuan Xiao, haoyuanxiao@yeah.net

ABSTRACT

We integrate pairwise image comparisons with Semantic Segmentation to assess perceived street safety through a social-psychological lens. Drawing on classic findings about natural surveillance, signs of disorder, and risk appraisal, we pre-specified simple directional expectations: brighter and cleaner scenes and those affording visibility should feel safer; visible rubbish and graffiti should depress safety appraisals; moderate human presence should increase perceived safety by signaling guardianship. Using 20 photos from the Shinsakae district (Nagoya, Japan), 69 participants completed 13,110 pairwise choices (all 190 combinations). A Mask2Former model, pretrained on ADE20K and fine-tuned on 263 locally annotated photos, improved mIoU from 34.15% to 66.10% and yielded area ratios for CPTED-relevant elements (lighting, greenery, people, cars, bicycles, rubbish, graffiti). We then estimated a weighted scoring function mapping these visual features to perceived-safety scores. The AI scores broadly tracked human rankings and reproduced expected social-psychological regularities: lighting/cleanliness associated positively with perceived safety, while rubbish/graffiti associated negatively; daylight and a sense of openness mattered across groups; gender, age, and nationality revealed interpretable differences in emphasis (e.g., women prioritized lighting; older adults weighted illumination more strongly; Japanese participants were more sensitive to cleanliness). We discuss how environmental cues shape quick, intuitive judgments of safety and how AI-assisted diagnostics can operationalize CPTED-informed improvements.

Keywords: Urban street; perceived safety; CPTED; semantic segmentation; pairwise-comparison; transfer learning; social perception; environmental cues

1. Introduction

1.1. Research background

We define perceived safety as a low level of anxiety about crime or hazards and the absence of perceived physical or psychological threats. Independent of actual crime rates, visual impressions of streetscapes shape perceived safety and in turn influence people's behavior and spatial range of activity[1]. Independent of actual crime rates, the visual impression of street spaces shapes perceived safety and thereby influences people's behavior and spatial range of activity[2]. In recent years, CPTED (Crime Prevention Through Environmental Design) has received increasing attention[3]; by manipulating environmental

ARTICLE INFO
Received: 20 October 2025 | Accepted: 2 November 2025 | Available online: 13 November 2025

CITATION
Xiao HY, NATSUME Y. Estimating perceived street safety via pairwise comparisons and semantic segmentation with a social-psychological lens. *Environment and Social Psychology* 2025; 10(11): 4236 doi:10.59429/esp.v10i11.4236

conditions, urban design seeks to reduce opportunities for crime[4] and alleviate fear of crime. (We use CPTED thereafter.)Typical CPTED principles include target hardening, access control, natural surveillance, and territorial reinforcement[5], which not only aim to deter crime through environmental effects but also emphasize impacts on psychological safety[6].

A considerable body of work has evaluated perceived safety on urban streets. Conventional approaches have mainly relied on resident questionnaires and field surveys to measure how safe people feel in place. These methods are labor- and time-intensive, lack timeliness, and are vulnerable to interrater variability and bias in subjective assessments. Consequently, there are practical limits to achieving citywide coverage, underscoring the need for more efficient and objective evaluation methods.

Recent advances in deep learning for computer vision have accelerated the automation of environmental perception for driving. In particular, Semantic Segmentation classifies every pixel in images or video frames into predefined classes (e.g., road markings, people, cars, obstacles, traffic signals), enabling automatic analysis of road structure and elements. This technology has progressed primarily within autonomous driving and plays a key role in real-time recognition for safe navigation[7]. Building on this, we apply Semantic Segmentation to the assessment of people' perceived safety in urban streetscapes and seek an objective, efficient analysis of urban safety. Specifically, we integrate subjective safety evaluation using the pairwise-comparison method with objective parsing of street-scene elements by Semantic Segmentation to help establish a new framework for safety-perception assessment. While complementing limitations of conventional subjective surveys, we explore AI-assisted urban safety evaluation with the aim of contributing to more practical urban design and planning.

1.2. Prior research and positioning of this study

In research on perceived safety in urban environments, questionnaire-based methods have been widely adopted[1]. Surveys collect subjective impressions of the safety of specific urban spaces from residents and other participants. Representative techniques include graded safety ratings on Likert scales, the SD (semantic-differential) method using adjective pairs, and photo-based assessments in which respondents rate street photographs[8]. Using these methods, studies have identified key elements that affect perceived safety in street environments—such as lighting, environmental cleanliness, the presence of people and cars, greenery, and types of shops[9].

However, questionnaire-based evaluations face several challenges. First, subjective ratings vary with participants' experiences and values, making it difficult to obtain consistent data and potentially reducing statistical reliability. Second, efficiency in data collection is problematic: evaluating an entire city requires large samples and substantial time and labor. Because urban environments change daily, survey-based assessments also suffer from weak timeliness and may not reflect current conditions. Moreover, because surveys require conscious choices, they struggle to capture relationships with subconscious visual factors. Survey results also tend to be limited in scope, making it difficult to obtain multi-faceted, fine-grained assessments at particular locations. For instance, determining perceived safety on a specific street or intersection ideally requires on-site data, but in practice responses often rely on recall or imagination. Recently, computer-vision approaches have been explored to quantify human perception of urban environments. Dubey et al. (2016)[10] developed a deep learning model to predict urban impressions using more than 110,000 street images from 56 cities and approximately 1.17 million pairwise-comparisons. Compared with questionnaires, AI-based methods can provide real-time, location-specific evaluations of perceived safety[11]. However, existing work often learns end-to-end predictors without opening the black box to estimate interpretable element-level effects aligned with CPTED.

In this paper, we propose a new approach that integrates the pairwise-comparison method with Semantic Segmentation to overcome the limitations of survey-based assessments. In pairwise-comparison, participants are shown two images (here, street photographs from Nagoya) and asked to choose intuitively which appears safer, thereby eliciting relative subjective judgments. Coupling this with Semantic Segmentation enables automatic classification and quantification of visual elements in the images (e.g., lighting, greenery, cleanliness, people, cars, rubbish, graffiti), thereby constructing a more objective and quantitative method for evaluating perceived safety. The approach improves consistency of evaluation and clarifies the relative ordering of street scenes. It also permits decisions even under vague recognition, reduces respondents' cognitive load, and allows checks on evaluators' decision quality, making it an efficient assessment method.

Recent advances in Semantic Segmentation further broaden the algorithmic landscape relevant to urban streets. For road-scene parsing, capsule-network variants (e.g., DDC-Net) target efficient urban-road segmentation[12]. Lightweight Transformer designs such as LKAFormer pursue high accuracy with compact modules grounded in Kolmogorov–Arnold representations[13]. From a systems perspective, federated and prior-knowledge-guided variants of BiSeNet address distributed training and edge extraction[16]. Beyond end-to-end CNNs/Transformers, classical yet competitive pipelines (e.g., threshold segmentation with information fusion for shadow detection) remain useful when computation and interpretability are priorities[15]. In adjacent safety-perception tasks, improved Mask R-CNN pipelines have advanced pedestrian detection under complex scenes[14]; application-driven network designs in intelligent robotics (e.g., DABU-Net with PCA) also echo the push toward efficient, task-aware vision models[17]. These developments motivate our choice to pair an interpretable Semantic Segmentation feature space with pairwise preferences rather than a fully opaque end-to-end predictor.

1.3. Objectives and significance

The objective of this study is to help establish a new framework for evaluating perceived safety on urban streets by combining pairwise-comparison with Semantic Segmentation. To address methodological problems of traditional survey-based approaches, we first enhance the reliability of subjective assessment by asking respondents to intuitively select the safer scene in pairwise-comparisons. We use Semantic Segmentation to extract CPTED-related elements and estimate their weights against pairwise choices, yielding an objective and reproducible scoring function for perceived safety.

Concretely, we use the pairwise-comparison results as the subjective ground truth, extract street-scene elements with Semantic Segmentation, and examine how the resulting features relate to perceived safety. We focus in particular on the effects of lighting, environmental cleanliness, number of people, presence of cars, and presence of graffiti, and we compare AI-derived scores with subjective rankings to verify accuracy.

This study contributes by updating methods for urban safety evaluation and enabling practical applications in urban planning and crime-prevention policy. We examine the potential of an efficient, reproducible method that integrates pairwise-comparison with AI techniques. An additional aim is to test the effectiveness of CPTED-based design and management: we analyze how its core principles—natural surveillance, territorial reinforcement, access control, and target hardening—affect safety-perception evaluations, providing actionable implications for planning practice.

Furthermore, introducing AI-based safety-perception assessment opens the possibility of automated evaluation of street-level safety. Semantic Segmentation allows fine-grained analysis of visual features and quantification of their effects on perceived safety. This can equip planners and public agencies with a practical tool for on-demand assessment of specific areas to inform the design and improvement of crime-prevention measures. In the future, the approach could support real-time monitoring and the optimization of interventions.

1.4. Structure of the paper

This paper integrates subjective safety evaluation via pairwise-comparison with Semantic Segmentation-based image analysis to build an AI-assisted method for assessing safety perception on urban streets. Section 2 analyzes photographs collected in the Shinsakae district of Nagoya using a transfer-learned Semantic Segmentation model; for each image, Semantic Segmentation yields area ratios of elements that influence perceived safety. Section 3 reports a pairwise-comparison experiment on 20 street photographs taken in Shinsakae, deriving relative rankings of safety and analyzing tendencies by participant groups. Section 4 describes the AI evaluation system, which leverages the Section 3 rankings and Semantic Segmentation-derived area ratios to estimate element weights and formulate a safety-perception scoring function. Section 5 compares AI scores with subjective results to validate effectiveness and discusses findings through the lens of CPTED while noting current limitations. Section 6 concludes and outlines future directions toward real-time evaluation, personalization by individual characteristics, and practical deployment.

1.5. Abbreviations and notation

For consistency, we use the following terms throughout:

- CPTED — Crime Prevention Through Environmental Design.
- SS — Semantic Segmentation.
- TL — transfer learning.
- IoU — Intersection over Union; mIoU — mean IoU.
- FOV — field of view.

1.6. Social-psychological framing and hypotheses

Perceived safety is a rapid social judgment informed by visual cues that signal visibility, order, guardianship, and potential threat. Building on social-psychological research on environmental perception, risk appraisal, and cue-based inference, we focus on three families of cues that map naturally onto CPTED:

- Visibility & Natural Surveillance: brighter, more open streetscapes afford monitoring and
- reduce ambiguity, supporting higher perceived safety.
- Order vs. Disorder: visible rubbish or graffiti can be read as cues of weak guardianship
- and social norms, lowering perceived safety.
- Social Presence & Activity: moderate presence of people often signals oversight and
- prosocial control; conversely, occluding vehicles or dense traffic may raise hazard salience.

Accordingly, we advance directional hypotheses (descriptive, not causal):

H1 (Visibility): Higher lighting and a greater sense of openness (e.g., sky/building vistas, unobstructed sightlines) correspond to higher perceived safety.

H2 (Order): Higher rubbish and graffiti area ratios correspond to lower perceived safety.

H3 (Guardianship): A moderate presence of people corresponds to higher perceived safety.

H4 (Heterogeneity): The weighting of these cues varies by gender, age, and nationality, reflecting different risk sensitivities and normative expectations.

Our AI pipeline provides operational measures of these cues via Semantic Segmentation-derived area ratios, and our pairwise design elicits intuitive judgments consistent with fast, cue-based evaluation.

2. Developing a semantic segmentation model for application in Shinsakae, Nagoya

2.1. Objective

This section develops a transfer-learned Semantic Segmentation model tailored to Shinsakae to extract street-scene elements and convert them into features for perceived-safety scoring. Specifically, we extract street-scene elements from photographs of Nagoya, Japan—e.g., road, buildings, greenery, lighting fixtures, people, automobiles, bicycles, graffiti, and rubbish—and classify them at the pixel level to quantify the proportion each element occupies in an image.

(1) Role of Semantic Segmentation

The advantages of Semantic Segmentation are twofold. First, automating street-image analysis reduces dependence on individual subjective cognition and enables objective, efficient classification and measurement of urban features. Second, by quantifying visual elements that influence perceived safety—such as road and sidewalk width, greening ratio, presence/absence of lighting, and people density—Semantic Segmentation can improve the precision with which contributing factors are identified.

In this study, we analyze urban street images with Semantic Segmentation and, through comparative analysis with subjective safety evaluations obtained via the pairwise-comparison method, examine the effectiveness and applicability of an AI-based approach to safety-perception assessment.

(2) Scope of the Work

We construct a high-accuracy, transfer-learned Semantic Segmentation model tailored to the target area of Shinsakae, Nagoya, Japan. In Section 3, this trained Semantic Segmentation model is used to compute area ratios of visual elements in the images employed for the pairwise-comparison experiment. Section 4 then applies the model within a pipeline that parses street images and computes AI-based safety-perception scores.

2.2. Model construction

To extract factors relevant to safety-perception evaluation from urban street images, we build a segmentation model adapted to the streetscape of Shinsakae, Nagoya via transfer learning.

(1) Model Selection

We adopt Mask2Former, a recently proposed model that achieves high accuracy in Semantic Segmentation. Mask2Former is based on a Transformer architecture rather than conventional CNNs and features a unified structure that supports semantic, instance, and panoptic segmentation tasks. It has demonstrated strong performance on diverse datasets containing complex urban and indoor/outdoor scenes, such as ADE20K, with IoU-based scores exceeding 55% (Note 1), making it suitable for extracting street-environment elements in urban spaces. In this study, we classify streetscape elements (e.g., buildings, sidewalk, lighting, greenery) according to the class definitions in ADE20K.

While alternatives such as DDC-Net for urban roads[12] and LKAFormer as a lightweight Transformer[13] are promising, we selected Mask2Former for its unified panoptic/semantic capability and strong performance on ADE20K, then adapted it via transfer learning to Shinsakae. Federated/knowledge-guided BiSeNet variants[16] and classical information-fusion segmentation[15] further contextualize our design choice toward an interpretable, transferable Semantic Segmentation backbone.

(2) Preparation of Training Data

Transfer learning combines the public ADE20K dataset with a database of real-world street photographs collected in Nagoya. ADE20K provides pixel-level annotations for more than 25,000 images across over 150 categories (e.g., buildings, roads, people, vegetation) covering a wide range of indoor and outdoor scenes, and is widely used to train general-purpose Semantic Segmentation models.

We first pretrain Mask2Former on ADE20K to acquire fundamental recognition of generic street scenes. We then fine-tune the model using 263 street images photographed in the Shinsakae district of Nagoya and annotated with pixel-wise class labels. To adapt the model to the local streetscape, the additional data were collected under the conditions specified in **Table 1**. Through this transfer learning, a model trained on ADE20K's globally diverse scenes was adapted to the distinctive characteristics of the target area—Such as signage styles, facade colors, greenery layout, and street structure—Thereby improving recognition accuracy for urban street elements.

Table 1. Shooting conditions for the Nagoya database.

Item	Condition
Shooting environment	Streets in Shinsakae, Nagoya, Japan
Period	Jan 2025 – Mar 2025
Device	Apple iPhone 16 Pro Max
Focal length	1× (≈24 mm equivalent; field of view close to human vision)
Image resolution / aperture	5712 × 4284; f/1.78
Camera height	160 cm (typical adult eye level)
Number of images	263

(3) Pixel-wise Annotation of the Shinsakae Street-Image Database

To quantify elements that influence perceived safety on urban streets, image subjects were categorized into the classes listed in **Table 2**. We annotated 263 street images photographed in Shinsakae at the pixel level using Labelme (**Table 2**; **Figure 1**). This yielded clear delineations of elements pertinent to safety perception—people, cars, lighting facilities, greenery, graffiti, rubbish, parking areas, etc.—And provided training data for transfer learning of the Semantic Segmentation model.

Table 2. Class definitions for the Nagoya database.

Class	Description	Mask color
road	Vehicle travel lane	Orange–brown
Sidewalk	Pedestrian walkway	Deep indigo-violet
Building	Exterior walls of buildings; commercial facilities; houses	Gray
Greenery	Street trees, plantings, grass, flowers	Bright red
Lighting	Streetlights; signboard lighting; vending-machine light sources; building-entrance lights	Olive
People	Pedestrians; people on bicycles	Dark red
Bicycle	City bikes, road bikes, etc.	Dark maroon
Graffiti	—	Deep blue

Class	Description	Mask color
Sky	—	Teal
Car	—	Dark green
Rubbish	Garbage stations; scattered rubbish	Purple
Parking area	Dedicated/household/shop parking areas	Dark olive green
Background	—	Black
Road signs	—	Grayish teal

Table 2. (*Continued*)

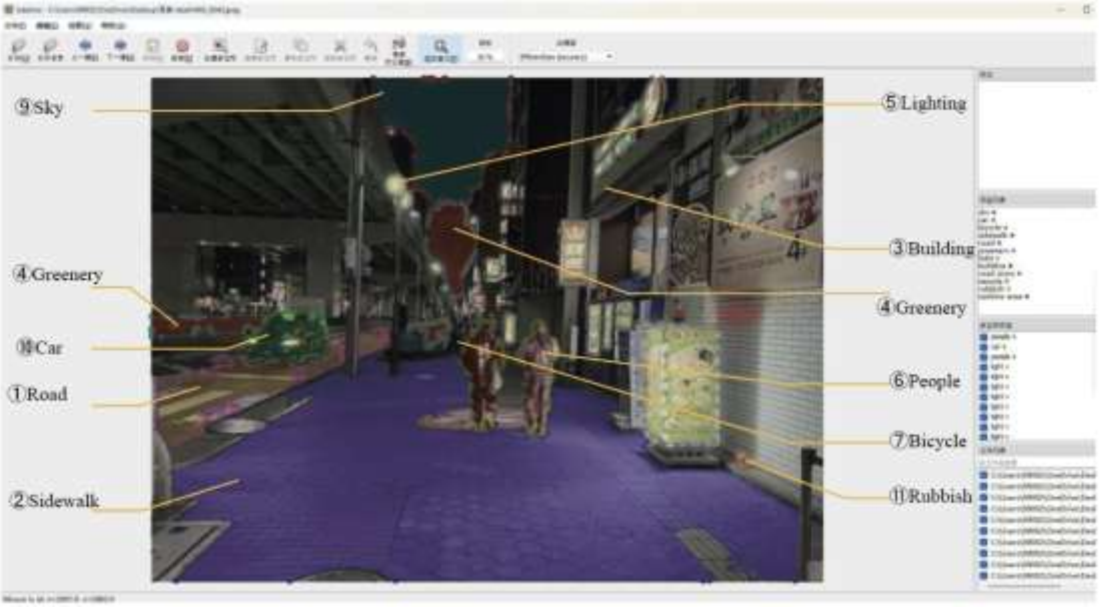


Figure 1. Example of labeling using Labelme.

(4) Transfer Learning for the Shinsakae Target Area

Baseline evaluation on the pretrained model indicated low IoU values (often ≤30) for many classes, revealing inadequate adaptation to Japanese streetscapes. Boundaries among buildings, sidewalk, and carriageways were not reliably recognized, and elements required for safety-perception evaluation could not be robustly segmented (**Table 5**). Because ADE20K pretraining draws heavily on imagery from overseas cities, the model's fit to Japanese streets and architectural styles was insufficient; as a result, elements such as sidewalk vs. road, buildings, rubbish, and graffiti were not accurately distinguished, making it difficult to extract information essential for evaluating perceived safety.

We therefore conducted transfer learning using the 263 annotated, real-scene street images from Shinsakae (**Tables 3** and **4**). Recognition accuracy improved substantially for area-specific elements, enabling high-precision segmentation of multiple factors that affect perceived safety—Lighting, greenery/cleanliness, people, cars, graffiti, advertisements, etc. (**Table 4**). As training progressed (**Figure 2**), overall IoU rose markedly, and the accuracies for people, rubbish, and graffiti stabilized (**Table 5**). The 16th-epoch model achieved the highest mIoU (Note 2) and was consequently adopted for all subsequent analyzes. We selected the 16-epoch checkpoint based on validation mIoU and used it for all downstream analyses. Class-wise IoU gains are summarized in **Table 5**.

Table 3. Hardware and training environment for deep learning.

Item	Details
OS	Windows 11 (64-bit)
Python version	3.8.20 (conda-forge)
CUDA version	11.3 (cuDNN 8.2 compatible)
PyTorch version	1.10.1 + cu113
GPU	NVIDIA RTX 4090 (24 GB)
System memory	64 GB
CPU	Intel Core i9-13900HK

Table 4. Training configuration for deep learning.

Item	Details
Model	Mask2Former
Number of classes	14 (see Table 2)
CrossEntropy loss	weight = 2.0
Dice loss	weight = 5.0
Mask loss	weight = 5.0; sigmoid enabled
Optimizer	AdamW (learning rate = 0.0001; weight decay = 0.05)
Max iterations	160,000
Validation interval	Every 5,000 iterations
Batch size	2
Crop size	512 × 512
LR scheduler	PolyLR (polynomial decay)
Datasets	ADE20K + Nagoya database

Table 5. IoU before and after transfer learning by class (16th epoch).

No.	Class	Before TL IoU	After TL IoU(16th)	Improvement (Δ)
1	Road	22.30	81.25	58.95
2	Sidewalk	45.84	83.60	37.76
3	Building	0.47	46.21	45.74
4	Greenery	26.22	73.76	47.54
5	Lighting	26.09	61.25	35.16
6	People	38.00	49.31	11.31
7	Bicycle	34.03	45.44	11.41
8	Graffiti	15.77	61.45	45.68
9	Sky	55.08	89.64	34.56
10	Car	32.70	84.36	51.66
11	Rubbish	46.48	68.87	22.39
12	Parking area	46.33	66.81	20.48

No.	Class	Before TL IoU	After TL IoU(16th)	Improvement (Δ)
13	Background	58.38	63.64	5.26
14	Road signs	30.42	49.82	19.4
	mIoU(Note2)	34.15	66.10	31.95

Table 5. (*Continued*)

Note: *mIoU is averaged across the 14 classes in* ***Table 2****; values are from the held-out validation set.*

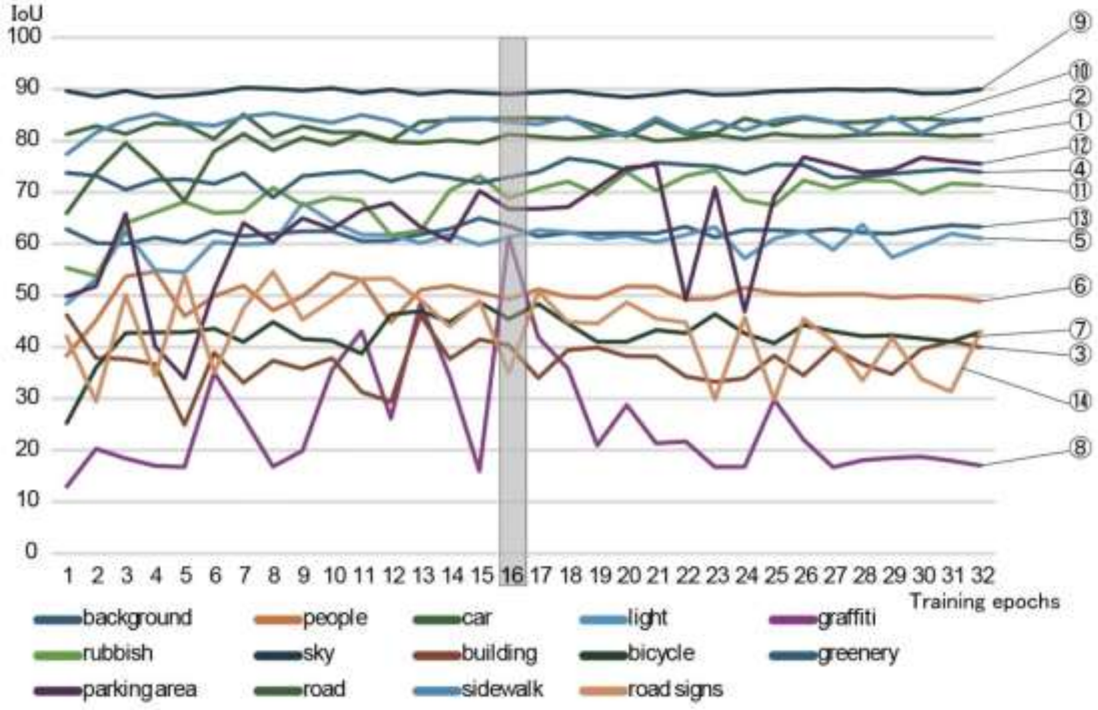


Figure 2. Evolution of segmentation accuracy (IoU) by class over 32 training epochs.

Note 1. *IoU (Intersection over Union) is a standard metric for evaluating segmentation accuracy.*

Note 2. *mIoU denotes mean IoU averaged across classes.*

3. Pairwise-comparison experiment for perceived safety

3.1. Objective

This study conducts a pairwise-comparison experiment to collect safety-perception data: participants are asked to select, on an intuitive basis, which of two street images from Nagoya appears "safer." pairwise-comparison, in which respondents choose the option that better fits the question from two alternatives, helps reduce rating variance and ambiguity and enables clear ranking. Using street photographs from the Shinsakae district of Nagoya, we analyze safety-perception rankings derived from participants' selections (**Table 6**).

Table 6. Overview of the pairwise-comparison experiment on perceived safety.

Item	Details
Objective	To obtain a ranking of photographs based on perceived safety of street scenes.
Task	Participants compare two photos and choose the one that feels safer.
Venue	Building 24, Nagoya Institute of Technology.
Stimuli	Street photographs taken in Shinsakae, Naka Ward, Nagoya City, Aichi Prefecture, Japan (see Table 6,

Item	Details
	original photos).
Method	pairwise-comparison.
Participants	69.
Period	Dec 2024 – Feb 2025.

Table 6. (*Continued*)

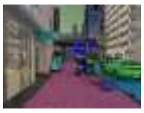
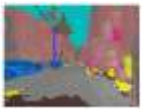
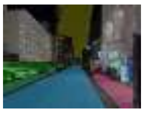
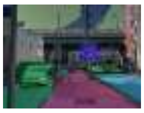
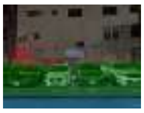
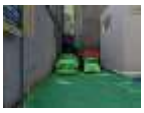
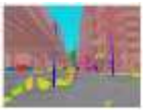
We further examine differences in evaluation tendencies by participant attributes—such as gender, age, and nationality—To consider factors influencing perceived safety from multiple perspectives. In particular, based on CPTED theory, we analyze the effects of elements such as presence/absence of lighting, environmental cleanliness, presence of greenery, numbers of people and cars, and presence of graffiti/posters on perceived safety.

In addition, the experiment is designed to provide foundational data for determining the relative weights of evaluation elements in Semantic Segmentation-based safety-perception scoring. Concretely, we use the subjective results from the pairwise-comparison to estimate weight coefficients from the areas of elements considered influential (e.g., cars, people, greenery, cleanliness), and combine these weights with Semantic Segmentation-extracted area ratios to derive a more reliable safety-perception score.

3.2. Experimental design

We used 20 photographs of streets in Shinsakae, Nagoya (**Table 7**; IDs 1–10 taken in daytime, 11–20 at night). Each participant was shown two randomly selected images and asked to choose which appeared safer. Pair generation and randomization were performed in Excel so as to avoid order bias. Every photograph was presented at least once as a comparison target to obtain relative safety evaluations for all images. Each participant completed 190 comparisons in total (20C2 = 190). This design balanced pair frequencies across images to ensure fairness. At the end of the questionnaire, participants also ranked seven factors previously verified to influence perceived safety[8]—“lighting,” “people,” “cars,” “cleanliness,” “types of shops,” “graffiti,” and “greenery”—From strongest to weakest perceived influence. Responses were collected using Microsoft Forms.

Table 7. Images displayed in the pairwise-comparison experiment (original photos) and semantic segmentation outputs before/after transfer learning.

ID	Original photo	SS output (pre-TL)	SS output (post-TL)	ID	Original photo	SS output (pre-TL)	SS output (post-TL)
1				11			
2				12			
3				13			
4				14			

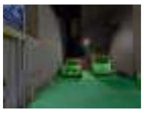
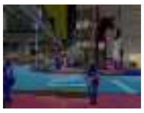
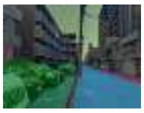
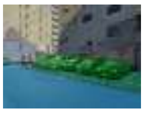
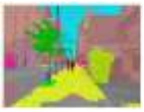
ID	Original photo	SS output (pre-TL)	SS output (post-TL)	ID	Original photo	SS output (pre-TL)	SS output (post-TL)
5				15			
6				16			
7				17			
8				18			
9				19			
10				20			

Table 7. (*Continued*)

Mask colors after transfer learning follow ***Table 2****. SS=Semantic Segmentation.*

3.3. Shooting time coding

In all analyses we explicitly label each image by shooting time: IDs 1–10 are daytime (D), IDs 11–20 are nighttime (N). This indicator is reported in **Table 11** as Shooting Time (Day/Night) and is included as a control variable in the scoring model in Section 4.1. Because local illuminance within each target environment can blur the distinction between daytime and nighttime, we coded the shooting time (day vs. night) and included this qualitative (dummy) variable in the regression/score model to reduce confounding.

(1) Photography

Street photographs for the experiment were taken with a smartphone (iPhone 16 Pro Max). The camera was set to 1× (approximately a 24-mm-equivalent lens), and the shooting height was fixed at typical adult eye level (~160 cm). Photography was conducted on walkable sidewalk in Shinsakae, Nagoya (**Figure 3**). A 24-mm lens offers a field of view close to the human visual angle (Note 3) and produces a natural appearance (Note 4), making it suitable for safety-perception evaluation in this study (**Table 8**; Note 5).

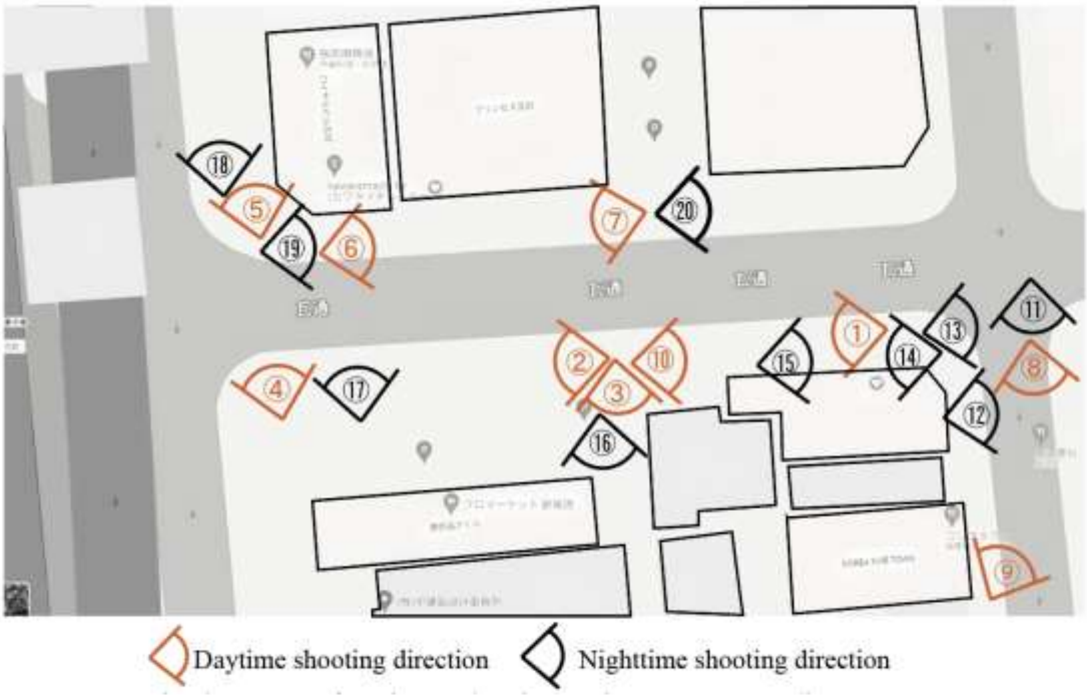


Figure 3. Shooting locations and camera directions for the experimental photos (Map data © Google.).

Table 8. Comparison between camera focal length and human horizontal field of view.

Focal length (mm)	Human horizontal FOV (°)	Interpretation
24	~84°	Approximates a slightly wide human field of view (used in this study).
35	~63°	Close to the angle within which human visual attention typically concentrates.
50	~40°	Similar to the field of view when gaze is focused on a single point.

(2) Procedure

Using a 24-inch display, participants compared two randomly presented images and selected the one they perceived as safer. The experiment laswted about 25 minutes with a 5-minute break midway.

(3) Participants

A total of 69 participants took part. We collected age, gender, and nationality information (**Table 9**) and compared tendencies in perceived safety by gender, nationality, and age groups.

Table 9. Participant demographics.

Attribute	Category		
Gender	Male:37.		Female:32.
Nationality	Japanese:11.		Chinese:58.
Age group	< 20:5.	20s:44.	≥ 30:20.

3.4. Alignment with Social-Psychological Expectations

The aggregate rankings and subgroup patterns are consistent with H1–H4. Images with higher lighting and openness ranked safer, whereas images with salient rubbish/graffiti ranked lower—supporting visibility and order accounts (H1–H2). Moderate human presence tended to co-occur with higher safety rankings (H3).

Group analyses indicated that women gave the highest priority to lighting, older adults showed stronger illumination sensitivity, and Japanese participants emphasized cleanliness more than the Chinese group (H4). These differences plausibly reflect risk sensitivity and norm expectations rather than data artifacts.

3.5. Results and analysis of the pairwise-comparison

This section analyzes the results in detail. We compare evaluation tendencies by participant attributes (gender, nationality, and age group) and examine the ranking of selection frequencies across the 20 target photographs.

(1) Reliability Analysis

To ensure data reliability, we assessed internal consistency using Cronbach's alpha. The alpha coefficient was 0.971, indicating very high internal consistency among items and thus high reliability of the responses.

(2) Analysis of Selection Frequencies

Aggregating responses from all 69 participants, we ranked photographs by the number of times they were selected as "safe." Many low-rated images were night photographs (Figure 4). Because "lighting" in our Semantic Segmentation refers to fixtures rather than illumination intensity, day–night conditions likely confound both human judgments and feature–outcome associations; we therefore interpret lighting-related results with caution. For example, at the same locations and camera angles, daytime/nighttime pairs such as Photos 3 vs. 16, 4 vs. 17, 1 vs. 14, and 8 vs. 11 showed this pattern clearly (**Table 10**). Aggregating responses from all 69 participants, we ranked photographs by the number of times they were selected as "safe." To improve readability, **Figure 4** now uses a 0–1200 y-axis with major ticks every 200 for the Number of selections axis; note that each photo appeared in 19 pairs per participant (69 participants completed all 190 pairs; total choices = 13,110).

Table 10. Descriptive difference between daytime and nighttime images in the pairwise ranking.

Metric	Daytime (IDs 1–10)	Nighttime (IDs 11–20)	Δ (Night − Day)
Mean rank (lower = safer)	6.5	14.5	8.0
	Matched daytime vs nighttime at similar locations (lower rank = safer)		
	Pair	Daytime photo (rank)	Nighttime photo (rank)
	1	1 (3)	14 (13)
	2	4 (5)	17 (9)
	3	8 (8)	11 (14)
	4	3 (18)	16 (19)

n = 69 participants; 13,110 pairwise comparisons.

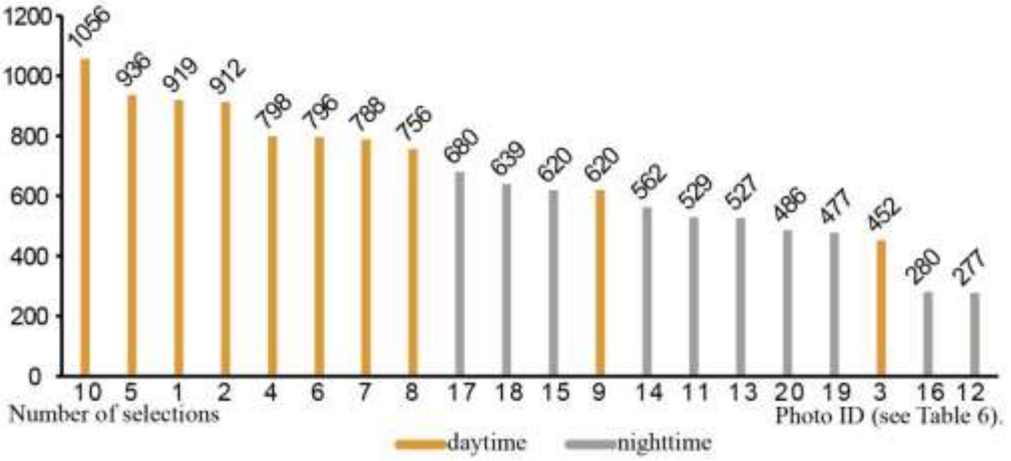


Figure 4. Ranking of photos by selection count (n=69 participants; total=13,110 pairwise choices).

Conversely, images with the highest selection frequencies were daytime photographs. Considering the area-ratio analysis of visual elements obtained with the trained Semantic Segmentation model from Section 2 (**Table 7**), high-rated photographs commonly exhibited sufficient lighting, clean and well-maintained environments, a moderate presence of people, and relatively low vehicle traffic (hence lower perceived danger). In contrast, low-rated photographs typically had insufficient lighting and environmental issues such as visible graffiti and rubbish. Specifically, graffiti was notable in Photo 12 (graffiti area ratio 1.89%), Photo 16 (1.06%), and Photo 3 (0.76%), while rubbish was prominent in Photo 19 (rubbish area ratio 3.68%) (**Table 11**), suggesting negative effects of these elements on perceived safety.

Table 11. Area ratios (%) of visual elements in the photos used for the pairwise-comparison perceived-safety experiment.

Rank	Photo ID	Shooting time (D/N)	Bicycle (a)	Car (b)	Graffiti (c)	Greenery (d)	Light (e)	People (f)	Rubbish (g)	Building (h)	parking area(i)	Road (j)	Sidewalk (k)	Road signs(l)	Sky (m)
1	10	D	0.15	0.07	-	13.06	-	0.32	-	-	-	6.85	16.43	0.04	6.40
2	5	D	0.08	0.46	-	1.85	0.14	0.85	-	-	-	2.44	18.20	-	5.41
3	1	D	0.66	4.68	-	2.62	2.54	0.34	-	39.34	-	3.43	20.60	0.74	4.50
4	2	D	0.16	5.82	-	3.08	-	0.07	0.05	11.98	6.59	4.73	14.96	-	12.36
5	4	D	1.57	0.65	-	1.41	-	-	0.23	25.42	-	5.07	20.18	-	6.04
6	6	D	-	-	-	4.82	0.07	3.18	5.12	-	-	6.10	15.64	0.85	12.43
7	7	D	0.25	3.55	-	0.26	-	0.50	9.14	20.35	-	3.00	16.43	-	2.92
8	8	D	-	14.59	0.26	-	-	-	-	-	6.04	15.52	1.24	-	9.81
9	17	N	0.35	0.11	-	0.83	0.23	4.85	0.73	14.35	-	13.91	12.52	-	3.36
10	18	N	-	1.02	-	1.69	4.19	1.69	0.04	-	-	2.16	26.91	-	3.68
11	9	D	-	10.42	0.64	-	-	-	-	20.49	6.57	29.35	-	-	-
12	15	N	-	4.74	0.12	0.59	0.48	-	-	34.91	-	4.32	15.92	-	5.72
13	14	N	-	0.91	-	1.53	2.18	0.12	-	48.91	-	2.07	17.48	-	-
14	11	N	0.36	4.77	-	0.96	2.12	-	-	27.72	1.96	19.55	5.26	-	8.66
15	13	N	0.05	5.14	0.94	0.03	0.16	-	-	-	1.57	28.39	4.95	-	5.67
16	20	N	-	9.51	0.07	-	0.67	0.06	-	-	1.65	5.39	18.71	-	9.58
17	19	N	-	2.37	-	2.91	0.52	0.24	3.68	1.69	-	2.00	16.01	1.47	11.76

Rank	Photo ID	Shooting time (D/N)	Bicycle (a)	Car (b)	Graffiti (c)	Greenery (d)	Light (e)	People (f)	Rubbish (g)	Building (h)	parking area(i)	Road (j)	Sidewalk (k)	Road signs(l)	Sky (m)
18	3	D	-	4.08	0.76	-	-	-	-	19.07	26.16	-	-	-	0.63
19	16	N	-	4.92	1.06	-	0.20	-	-	-	23.05	-	-	-	-
20	12	N	-	19.54	1.89	-	-	-	-	-	10.05	9.90	-	-	-

Table 11. (*Continued*)

Note. *"D" = daytime; "N" = nighttime. Only the subset of classes used in the scoring model (a–g) are shown here; the full per-image feature table (including building, sky, road, sidewalk, etc.) is unchanged and available in the supplement.*

(3) Ranking of Influential Factors in Perceived Safety

In the final questionnaire item, participants ranked the seven elements considered influential—"lighting," "people," "cars," "cleanliness," "types of shops," "graffiti," and "greenery"—by perceived importance. Overall, "lighting" (1.7) was ranked highest, followed by "people" (3.1) and "cleanliness" (3.6). "Greenery" (4.9) and "graffiti" (5.8) tended to receive lower importance rankings (**Table 12**).

Table 12. Rankings of perceived-safety influence factors and mean ranks by gender, nationality, and age group.

Rank	Male	Male(mean)	Female	Female(mean)	Japanese	Japanese(mean)	Chinese	Chinese(mean)
1	Lighting	1.8	Lighting	1.5	Lighting	1.7	Lighting	1.6
2	People	3.1	People	3.0	Cleanliness	2.7	People	3.0
3	Cleanliness	3.6	Cleanliness	3.6	People	3.1	Cleanliness	3.8
4	Car	4.1	Types of shops	4.4	Cars	4.0	Cars	4.3
5	Types of shops	4.8	Greenery	4.5	Graffiti	4.7	Types of shops	4.4
6	Greenery	5.1	Cars	4.6	Types of shops	5.6	Greenery	4.6
7	Graffiti	5.4	Graffiti	6.3	Greenery	5.9	Graffiti	6.0
Rank	< 20	< 20(mean)	20s	20s(mean)	≥ 30	≥ 30(mean)	Overall	Overall(mean)
1	Lighting	1.4	Lighting	1.8	Lighting	1.3	Lighting	1.7
2	People	3.0	People	3.0	People	3.1	People	3.1
3	Cars	3.2	Cleanliness	3.6	Cleanliness	3.2	Cleanliness	3.6
4	Greenery	4.6	cars	4.3	Types of shops	4.5	Cars	4.3
5	Types of shops	4.8	Types of shops	4.6	Greenery	4.6	Types of shops	4.6
6	Cleanliness	5.4	Greenery	5.0	Cars	4.7	Greenery	4.9
7	Graffiti	5.6	Graffiti	5.5	Graffiti	6.5	Graffiti	5.8

Notes: *Lower mean rank indicates stronger perceived influence (1 = most, 7 = least).*

Weighted score = ∑(rank n × count at rank n); total appearances = ∑ counts across all ranks;

mean rank = weighted score / total appearances.

(4) Comparisons by Gender, Nationality, and Age

(i) Gender

Average ranks of influential factors by gender are summarized in **Table 12**. As shown by the curves in **Figure 5**, overall selection tendencies were broadly similar between men and women. However, some images showed gender differences in selection frequency. For women, “lighting” (1.5) exerted the strongest influence; for men, the rank of “cars” (4.1) was somewhat higher (**Table 12**: “cars” ranked 4th for men vs. 6th for women), suggesting partially different emphases in perceived safety.

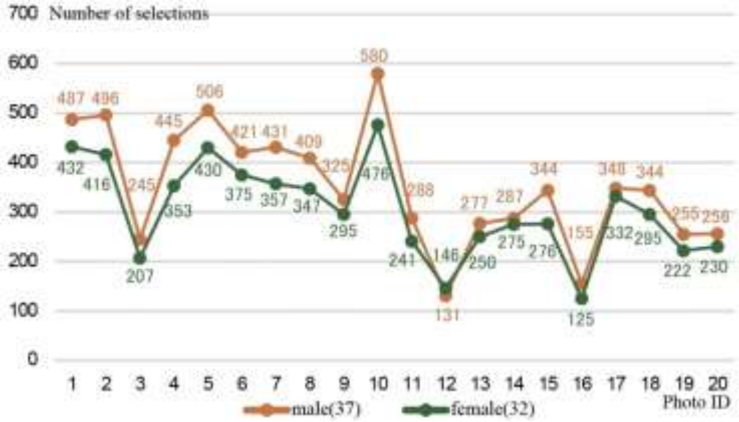


Figure 5. Comparison of photo-selection counts by gender (male vs. female).

Using participant data, we compiled and analyzed rankings of the photographs judged as safer (**Table 13**). Comparing Photo 12 (vehicle area ratio 19.54%) and Photo 16 (4.92%), the male group ranked Photo 16 19th and Photo 12 20th, whereas the female group ranked Photo 12 19th and Photo 16 20th, indicating possible gender differences in how vehicle area influences perceived safety.

Table 13. Selection rankings by gender, nationality, and age group (photo IDs; see **Table 7**).

Rank	Male	Female	Japanese	Chinese	< 20	20s	≥ 30	Overall
1	10	10	10	10	10	10	10	10
2	5	1	2	5	6	5	1	5
3	2	5	5	1	2	1	2	1
4	1	2	1	2	5	2	5	2
5	4	6	7	6	7	7	6	4
6	7	7	8	4	17	4	8	6
7	6	4	4	7	8	6	4	7
8	8	8	6	8	1	8	7	8
9	17	17	9	17	4	17	9	17
10	15	18	15	18	9	18	17	18
11	18	9	18	15	15	15	18	9
12	9	15	17	9	19	9	15	15
13	11	14	3	14	18	14	11	14
14	14	13	13	11	14	11	13	11
15	13	11	14	13	20	13	14	13
16	20	20	11	20	13	19	20	20
17	19	19	19	19	11	3	19	19
18	3	3	20	3	3	20	3	3
19	16	12	16	12	12	16	12	16
20	12	16	12	16	16	12	16	12

Both genders appeared to value a sense of openness. Photos 12 and 13 were taken at the same location but with different angles; due to better visibility and openness, Photo 13 received higher evaluations from both men and women. Similarly, at the same site, Photo 2 (with a clearer view) was rated higher than Photo 3 (**Figure 3**).

For women, lighting had the strongest impact (ranked 1st), and cleanliness was also important (ranked 3rd). rubbish reduced perceived safety; for example, Photo 19—with a rubbish area ratio of 3.68%—ranked 17th among women **(Tables 11** and **12**). Women tended to place greater emphasis on atmosphere, especially openness and visibility. Photo 12 (female rank 19th) and Photo 13 (female rank 14th) illustrate this: although taken at the same location, Photo 12's composition features cars (19.54% area) obstructing sightlines, whereas Photo 13 offers an open view and was rated higher by women. The lighting area ratio was also larger in Photo 13 (2.18%) than in Photo 12 (0%), indicating that brighter, more open spaces contribute to a greater sense of safety among women.

(ii) Nationality

By nationality, **Table 12** shows that Japanese participants emphasized "cleanliness" (2.7), whereas Chinese participants tended to prioritize "lighting" (1.6) and "people" (3.0). For instance, images with larger lighting areas and the presence of people—such as Photo 18 (lighting 4.19%, people 1.69%) and Photo 14 (lighting 2.18%, people 0.12%)—Were more highly rated by Chinese participants. Japanese participants were more sensitive to cleanliness; locations with visible dirt or rubbish tended to receive lower evaluations from them. For example, Photo 6, which contained a relatively large amount of rubbish (5.12%), received low ratings among Japanese respondents. In addition, photographs where shop facades were clearly visible—Such as Photo 4 and Photo 15 (building area ratios 25.42% and 34.91%, respectively)—Were rated higher by the Chinese group.

Overall selection tendencies were broadly similar between the two nationalities (**Figure 6**), but some images received different evaluations. Focusing on cases with rank differences ≥3: Photo 3 was rated lower by Chinese participants (18th) than by Japanese participants (13th). Despite the relatively large rubbish area in Photo 6 (5.12%), Chinese participants ranked it 5th, higher than the Japanese group (8th). For Photo 17, the Chinese group ranked it 9th vs. 12th among Japanese. We infer that Photo 3's constrained visibility might evoke greater unease among Chinese participants; Photo 6 suggests cross-national differences in sensitivity to cleanliness (greater tolerance for rubbish among Chinese participants); and Photo 17 contains Chinese-style storefronts/signage that may feel culturally familiar and thus safer to Chinese participants.

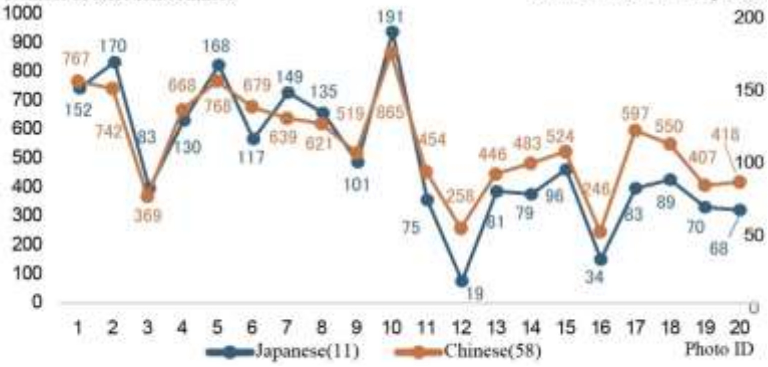


Figure 6. Comparison of photo-selection counts by nationality.

In summary, nationality-based comparisons suggest that Chinese participants emphasize visibility and the presence of people—Aligning with “types of shops” and “people”—Whereas Japanese participants are more attuned to “cleanliness” and “graffiti” in urban street environments.

(iii) Age

By age group (**Table 12**), participants aged 30 and over assigned the highest importance to “lighting” (1.3), indicating strong sensitivity to illumination. Participants under 20 showed notable concern regarding “cars” (3.0), suggesting greater influence of vehicle presence on perceived safety among younger respondents.

Overall selection tendencies by age were broadly similar (**Figure 7**), but some images differed by ≥3 ranks across groups. Photos 1, 6, and 19 are representative. For Photo 1 (vehicle area ratio 4.68%), ranks were 8th (<20), 3rd (20s), and 2nd (≥30), indicating that higher age was associated with higher perceived safety; thus, participants under 20 may respond more sensitively to vehicle presence and more strongly equate “more cars = less safe.” For Photo 6 (rubbish 5.12%), the ranks were 2nd (<20), 7th (20s), and 5th (≥30), implying relatively higher acceptance among the youngest group; participants in their 20s and ≥30s may rely more on cleanliness as a safety criterion. A similar trend appeared for Photo 19 (rubbish 3.68%): ranks were 12th (<20), 16th (20s), and 17th (≥30), suggesting generational differences in sensitivity to rubbish/dirt, with younger participants showing greater tolerance.

Note 3. A 24-mm-equivalent focal length approximates the human field of view.

Note 4. Such settings reproduce a natural perspective in street photography.

Note 5. The suitability of 24-mm framing for perceptual tasks has been discussed in prior work; detailed parameters are summarized in **Table 11**.

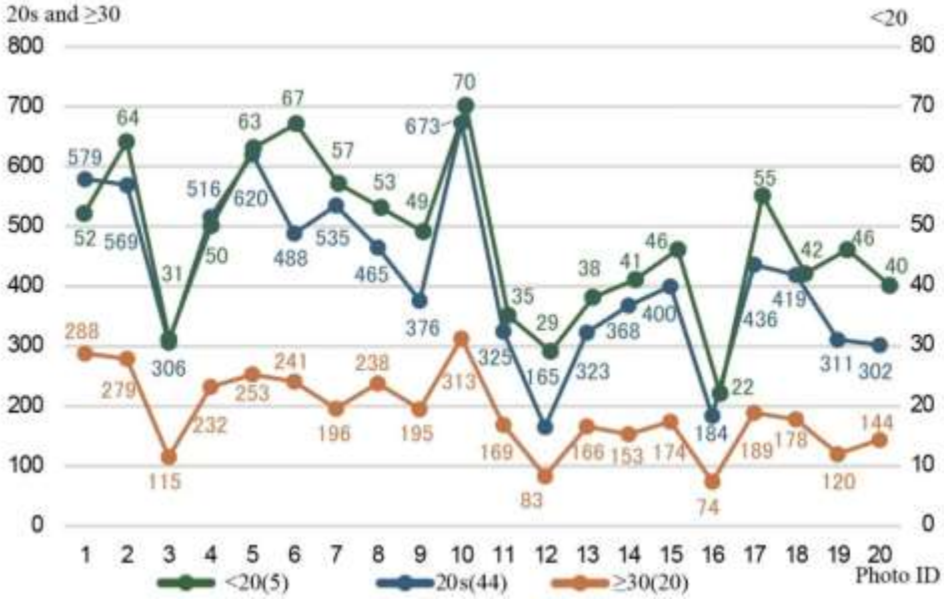


Figure 7. Comparison of photo-selection counts by age group.

Robustness to day/night. Because our “lighting” feature captures the presence of fixtures rather than scene luminance, day–night conditions can shift perceived safety independent of fixture area. Descriptively, the average rank for daytime photos is 6.5 versus 14.5 for nighttime (Δ = 8 rank positions across the 20 images), and matched locations (e.g., 1 vs. 14, 4 vs. 17, 8 vs. 11, 3 vs. 16) consistently show lower perceived safety at night. We therefore add a daytime dummy to the score model (§4.1) to absorb baseline visibility differences and reduce confounding in the element weights.

4. Construction of the AI-based evaluation system

We develop an AI evaluation system that computes perceived-safety scores by leveraging Semantic Segmentation outputs on street images. First, using the segmentation model trained via transfer learning in Section 2, we analyze the 20 street photographs used in the pairwise-comparison experiment and compute the area ratios of evaluation elements relevant to perceived safety (e.g., lighting fixtures, greenery, rubbish, people, automobiles, bicycles, graffiti). Next, we align these area ratios with the photograph rankings obtained in Section 3 to estimate each element's influence weight (area weight) on perceived safety. This yields a scoring function that, given any street image, can rapidly and automatically compute and display a perceived-safety score based on element area ratios and learned weights.

(1) Image Analysis and Feature Extraction

Using the trained Semantic Segmentation model, we processed all 20 images from the pairwise-comparison experiment. For each pixel, the model predicts a class label and outputs a segmentation mask image (with a unique color for each class) (**Tables 2** and **11**). From these outputs we computed, for every image, the occupancy rate of each class (i.e., the proportion of pixels in the full image) and derived an Semantic Segmentation-based estimate of the perceived-safety score. Thus, each image is represented by a vector of visual features; the primary features (area ratios) are summarized in **Table 11** and **Figure 8**.

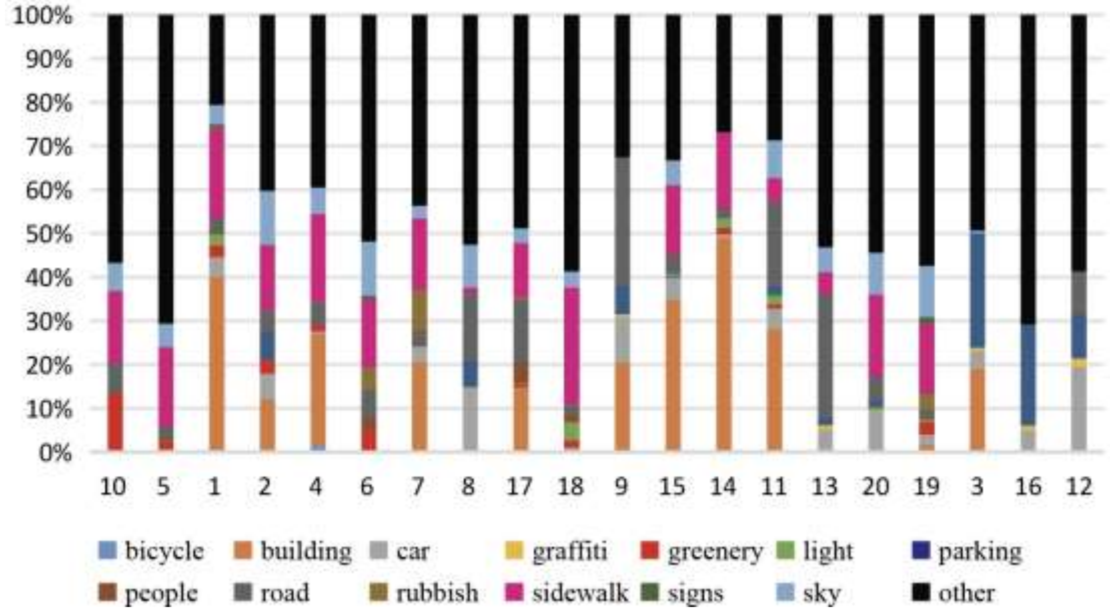


Figure 8. Area ratios of influential elements in the experimental photos (descending order).

(2) Inference Method

We adopt a scoring approach that aggregates Semantic Segmentation-derived area ratios of elements (e.g., lighting, greenery/cleanliness, people, cars, graffiti, advertisements) with their influence weights. Concretely, the perceived-safety score for an image is the weighted sum of element area ratios, where the weights are statistically estimated from the pairwise-comparison results. This procedure provides a basis for the objectivity and practical validity of the AI-generated scores.

4.1. Evaluation method

We computed element-wise area ratios from Semantic Segmentation outputs and formed a linear scoring function

$$S_n = \sum_{i=1}^{n} w_i X_i + \gamma D_{\text{day},n} + b$$

where X collects area ratios (e.g., lighting, greenery, pedestrians, automobiles, bicycles, litter, graffiti)[8]. We estimated the weights w by optimizing agreement between the score-implied ordering of images and the aggregate pairwise ranking (Section 3), subject to L2 regularization. This procedure is descriptive and correlation-based; it does not claim causality. To mitigate overfitting, we tuned the regularization strength via five-fold cross-validation on pairwise comparisons, holding out disjoint sets of image pairs. Agreement with human judgments is summarized in **Figure 9**.

Formula:

$$S_n = w_a X_a + w_b X_b + w_c X_c + \cdots + w_n X_n + \gamma D_{\text{day},n} + b$$

$$=$$

$$S_n = \sum_{i=1}^{n} w_i X_i + \gamma D_{\text{day},n} + b$$

S_n: perceived-safety score of photograph n;

X_n: area ratio of a given class;

w_n: weight coefficient;

$\gamma D_{day,n}$: 1 for daytime and 0 otherwise, γ is the day/night coefficient

b: intercept;

(Abbreviations correspond to **Table 11**.)

Example: a higher occupancy of greenery (X_d) increases the score, whereas a higher occupancy of graffiti (X_c) decreases it. The pairwise-comparison ranking of photographs is:

$$S_{10} > S_5 > S_1 > S_2 > S_4 > S_6 > S_7 > S_8 > S_{17} > S_{18} >$$

$$S_9 > S_{15} > S_{14} > S_{11} > S_{13} > S_{20} > S_{19} > S_3 > S_{16} > S_{12}$$

Illustration:

$S_{10} > S_{18}$:

$$0.15\%\cdot w_a + 0.07\%\cdot w_b + 13.06\%\cdot w_d + 0.32\%\cdot w_f + \gamma D_{day,n} + b$$

$$>$$

$$1.02\%\cdot w_b + 1.69\%\cdot w_d + 4.19\%\cdot w_e + 1.69\%\cdot w_f + 0.04\%\cdot w_g + b$$

Complete computation results are provided in **Table 14**; validation is shown in **Figure 9**. Although the AI-generated scores do not perfectly match the pairwise-comparison curve, both exhibit similar rising trends overall. Discrepancies may reflect day–night differences, measurement precision of lighting, and the need for finer class definitions. These issues will be addressed to further improve accuracy in future work.

Table 14. Estimated weights for each segmentation class.

Class	Abbrev.	Weight coefficient	Multiplier	Scaled weight(for computation)
bicycle	a	3.09	2.5	7.73
car	b	1.00	2.5	2.50
graffiti	c	-8.02	2.5	-20.05
greenery	d	1.00	2.5	2.50
lighting	e	1.00	2.5	2.50
people	f	2.58	2.5	6.45
rubbish	g	-1.00	2.5	-2.50

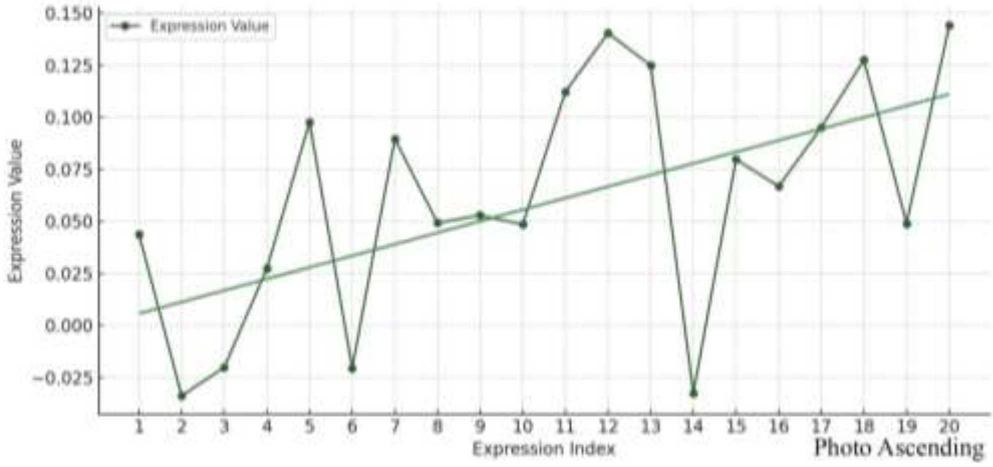


Figure 9. AI-derived perceived-safety scores (ascending order). n = 69 participants; 13,110 pairwise comparisons.

4.2. Implementation of the AI perceived-safety program

(1) Visualization and Information Panel

The program overlays Semantic Segmentation results on the original image for visualization. We set the original image opacity to 0.7 and the segmentation mask to 0.3 to balance visibility and information. An information panel on the right displays the image's perceived-safety score at the top (Safety Score). Below it, for each class (e.g., people, cars, sky, greenery, sidewalk), three items are listed (**Table 15**):

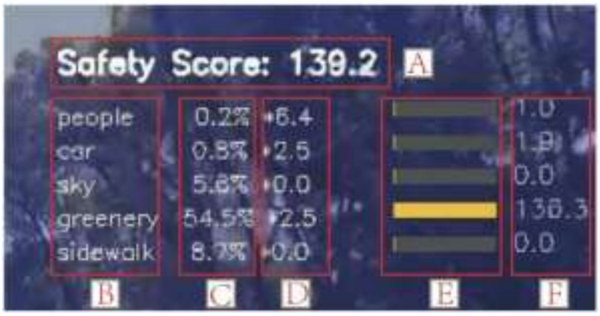
B/C: the element name and its area ratio (% of the image),

D: the pre-estimated weight coefficient,

E/F: the weighted score (contribution to the perceived-safety score), i.e., (area ratio×weight).

A horizontal bar is also shown per class to convey contribution magnitude intuitively.

Table 15. Information panel of the AI evaluation program.



```
Safety Score: 139.2  A
people     0.2%  ×6.4     1.0
car        0.8%  ×2.5     1.9
sky        5.6%  ×0.0     0.0
greenery  54.5%  ×2.5   136.3
sidewalk   8.7%  ×0.0     0.0
   B         C     D    E    F
```

Code	Item
A	Overall perceived-safety score
B	Elements
C	Area ratio (%)
D	Weight coefficient
E	Bar visualization
F	Element score

(2) Three Modes of Operation

The current program provides three modes:

① Batch scoring for still images.

For a set of pre-captured still images, Mode ① runs Semantic Segmentation in batch and automatically computes perceived-safety scores. Area ratios of influential elements (lighting, people, cars, greenery, etc.) are also exported (e.g., to Excel) for downstream analysis.

② Frame-wise scoring for videos.

For pre-recorded videos, Mode ② performs Semantic Segmentation on each frame and computes frame-level perceived-safety scores (**Figures 10** and **11**). Area ratios and scores are recorded in tabular form to visualize temporal changes.

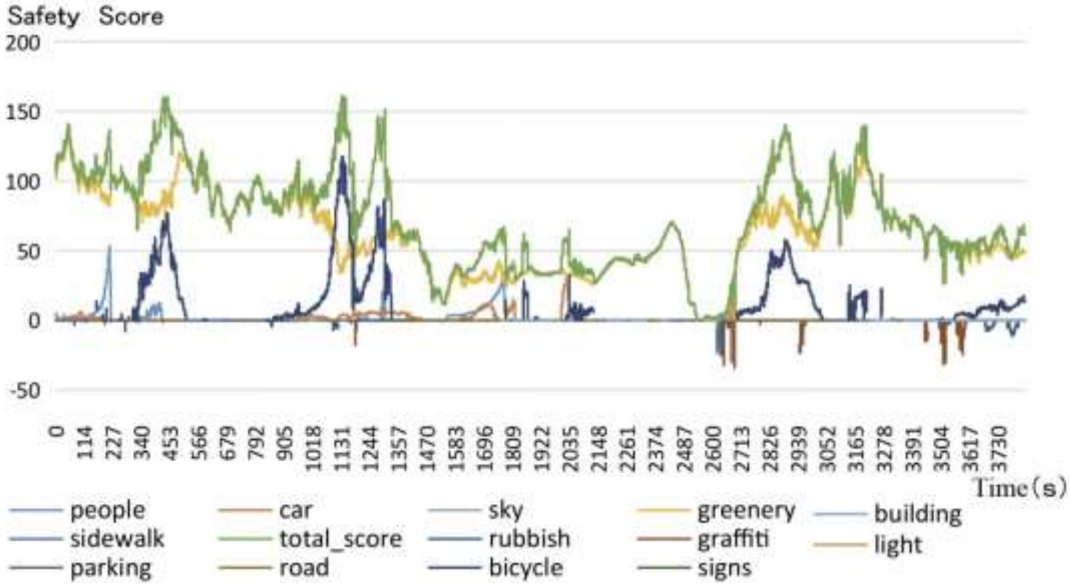


Figure 10. Per-element scores in the video analysis for AI-based perceived-safety evaluation.

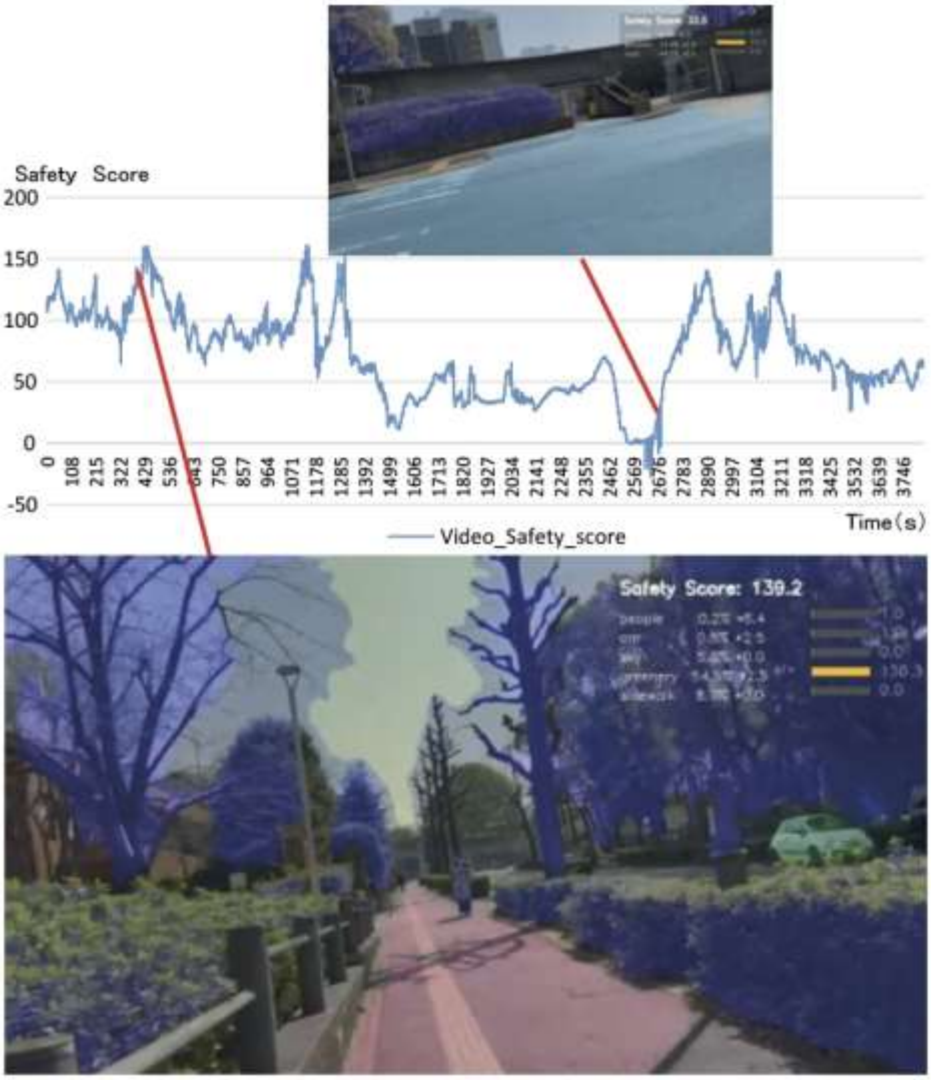


Figure 11. AI perceived-safety score — Example from Mode ② (video analysis).

③ Real-time scoring via a PC camera.

Mode ③ analyzes real-time video captured by a PC-mounted camera and outputs instantaneous perceived-safety scores (**Figure 12**). Because Semantic Segmentation is computationally demanding, this mode currently requires a high-performance PC environment and is not yet practical on mobile devices.

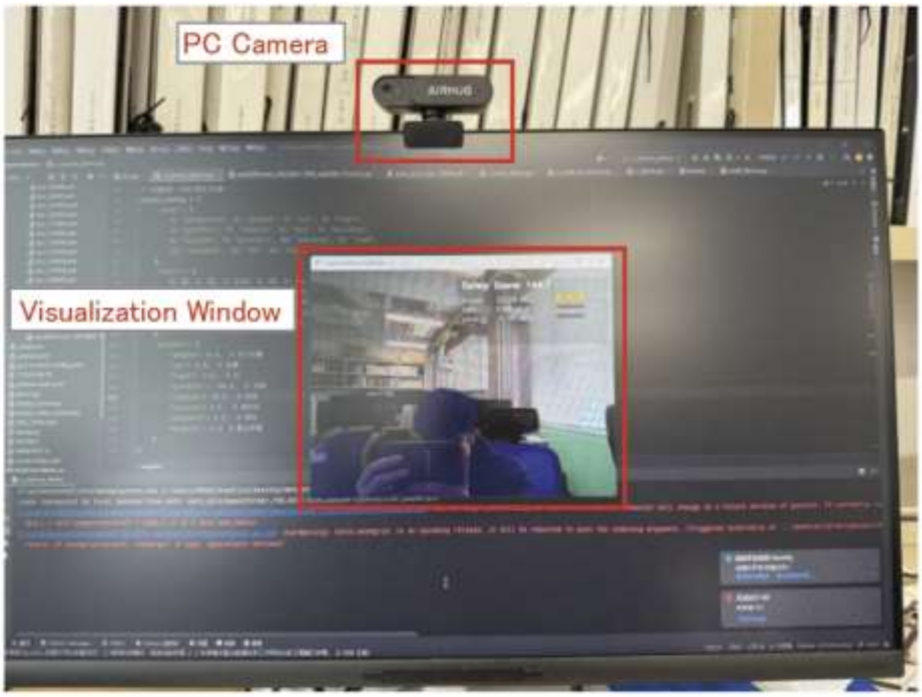


Figure 12. Example of Mode ③ (real-time perceived-safety scoring via PC camera).

5. Results and discussion of the ai perceived-safety program

5.1. Results of the pairwise-comparison experiment

This chapter reports the results of the pairwise-comparison experiment and discusses the AI perceived-safety evaluation program. We analyze how lighting, cleanliness, and graffiti affected participants' evaluations of perceived safety, and we examine differences by gender, age, and nationality. We also assess the program's validity in light of the four CPTED principles—Natural surveillance, access control, territorial reinforcement, and target hardening—And propose a safety-evaluation method.

(1) Computed Scores and Overall Tendencies

We aggregated the pairwise-comparison outcomes and computed a perceived-safety score for each image. Based on these data, we analyzed selection frequencies and produced a ranking of perceived safety for urban streets in Nagoya.

① Effect of lighting. Photographs with sufficiently installed streetlights and relatively larger lighting area ratios—such as Photo 14 (2.18%) and Photo 18 (4.19%)—tended to be rated higher. By contrast, Photo 12, which lacked lighting (0%), ranked at the bottom, indicating a strong influence of lighting on perceived safety (**Table 11**).

② Effect of cleanliness. Streets with little rubbish or graffiti and a well-maintained appearance were rated more favorably. For example, Photos 10 and 5—both clean—ranked near the top, whereas images with recorded graffiti or rubbish, such as Photo 12 (graffiti area ratio 1.89%) and Photo 19 (rubbish area ratio 3.68%), tended to receive lower evaluations.

③ Effect of graffiti. Walls and building edges exhibiting graffiti were more likely to be perceived as "poorly managed" or indicative of "unsafe areas." In practice, images with relatively larger graffiti ratios—Photo 12 (1.89%), Photo 16 (1.06%), and Photo 13 (0.94%)—all remained at or below the middle tier of perceived-safety scores, suggesting that visual signs of physical disorder influenced evaluations (**Table 11**).

Some participants implicitly associated graffiti with higher crime risk, indicating that it functions as an unconscious symbol of "poor public safety."

Notably, the learned weights for automobiles and bicycles are positive (**Table 14**). We hypothesize that, in our photos, these elements co-occur with broader carriageways and active storefronts—correlates of openness and natural surveillance—so the net association appears positive. This interpretation is descriptive; future work will separate parked vs. moving vehicles and replace area ratios with instance counts or perspective-weighted measures to reduce confounding.

5.2. A CPTED–AI integrated method for safety evaluation

Building on CPTED theory, we propose a more practical crime-prevention and urban-design workflow by integrating AI techniques with CPTED principles. Comparing the properties of our AI evaluations with CPTED's principles, we propose the following integrated model:

① Simulation with the AI model. Use the AI model during design to forecast perceived safety for new block plans. For example, simulate "How much would perceived safety improve if additional CCTV were installed?" and quantify the impact.

② Real-time monitoring and urban improvement. Leverage sensor data and AI analytics in real time to continuously monitor urban safety conditions. Automatically detect, for instance, "dark paths with no people flow for a certain period" or "areas requiring cleaning," thereby supporting maintenance and management workflows.

③ Integration of CPTED and AI evaluation. Combine AI analysis with CPTED principles to operationalize urban-safety design. From the perspectives of natural surveillance, access control, and territorial reinforcement, optimize the urban environment. By employing this method, AI augments and strengthens conventional, subjective evaluations of perceived safety, enabling more efficient and accurate assessment of urban environments (**Table 16**).

Table 16. Relationship between CPTED principles and the AI evaluation.

CPTED principle	Link to AI evaluation	Practical implication
Target hardening	AI-based analysis of human flow, gaze direction, and building features	A comprehensive, integrated evaluation method; optimize the placement of CCTV cameras and late-night (24-hour) businesses.
Access control	Analysis of road/sidewalk layout	Identify low-traffic, high-risk areas and plan countermeasures.
Natural surveillance	AI-based analysis of lighting facilities	Automatically detect areas with poor visibility/surveillance and optimize street-light placement.
Territorial reinforcement	Detection of graffiti and rubbish	Propose environmental-management improvements (regular cleaning, streetscape-maintenance measures).

5.3. Social-psychological implications

Our findings reinforce that quick safety judgments are shaped by readable environmental cues: illumination, visible order, and signs of guardianship. In design and maintenance terms, interventions that increase perceived visibility (e.g., light placement that improves face/ground contrast and reduces occlusion) and restore order (routine removal of rubbish/graffiti hot spots) should raise perceived safety even without changing crime incidence. The subgroup differences suggest targeting: routes for women and older adults may benefit disproportionately from lighting quality, whereas areas serving Japanese commuters may gain more from cleanliness-first maintenance. Because our AI system quantifies these cues at scale, it can support diagnostics (what to fix) and evaluation (did perceived safety improve) in CPTED-aligned programs.

6. Conclusion

6.1. Summary of the study

This paper proposed and validated a new method that integrates subjective assessment with objective analysis by combining the pairwise-comparison method with semantic-segmentation deep learning. As a first step, we constructed a region-specific Semantic Segmentation model by fine-tuning a Mask2Former model—Pretrained on ADE20K—using street photographs actually taken in the Shinsakae district of Nagoya. This enabled high-precision extraction of visual information tailored to the target streetscape.

Next, we conducted a pairwise-comparison experiment on 20 street photographs. Participants selected, for each presented pair, the scene they perceived as safer; from these responses, we derived relative rankings of perceived safety for all images. We then analyzed the same 20 photographs with the Semantic Segmentation model and extracted thirteen visual elements—such as lighting, cleanliness, greenery, people, automobiles, rubbish, and graffiti—computing each element's area ratio. These elements correspond to factors commonly discussed in CPTED as related to perceived safety.

Finally, we statistically examined relationships between the subjective rankings and the Semantic Segmentation-derived area ratios and estimated the influence (weight coefficients) of each element on perceived safety. Based on these results, we implemented an AI system that detects the proportion of each visual element in an input image and outputs a numerical perceived-safety score.

The contributions are threefold. First, we demonstrate the effectiveness of a fusion approach that combines subjective evaluation with AI analysis. Whereas subjective assessments have traditionally been costly to collect and limited to case-by-case discussion, pairing an efficient pairwise-comparison design with AI techniques shows potential to overcome these difficulties. Second, we quantitatively clarify how concrete urban-design elements relate to perceived safety. By showing, with data, the effects of adjustable design features—such as openness, green coverage, and visibility—this work informs future design guidelines and improvements. Third, we highlight directions to improve AI evaluation, pointing out information not fully captured by current image-segmentation methods and indicating the need for more advanced multimodal analysis and additional data.

6.2. Future work

We identify four key directions:

① Expanding data collection and improving generalizability.

This study fine-tuned on 263 images from a limited area in Shinsakae, and participant attributes tended to be urban-resident biased. Direct application to other areas or cities is therefore limited. Future work should gather image and participant data from more diverse urban environments (residential, commercial, suburban new towns, etc.) and different cultural contexts to improve model generalization and reliability.

② Incorporating dynamic elements.

Although the present system can compute real-time perceived-safety scores from still imagery, it does not yet consider dynamics such as the motion of people and cars (e.g., approach/retreat, speed). Because actual perceived safety depends strongly on time of day and moving objects, future models should incorporate video data and sequential-frame analysis to capture dynamic features.

7. Limitations

First, "lighting" captures the presence of fixtures rather than scene illumination; future work will add luminance/contrast measures. Second, day–night conditions were not explicitly modeled and may confound results. Third, sample imbalance (58 Chinese vs. 11 Japanese participants) limits subgroup generalization; stratified analyses are warranted. Finally, the segmentation model was fine-tuned on a single district, constraining external validity.

③ Integrating visual and non-visual information and broadening environmental factors.

Our current model relies on Semantic Segmentation from ADE20K and Shinsakae images; while visual elements such as lighting, graffiti, and greenery are handled, environmental conditions—weather (clear/rainy/cloudy) and time of day (day/night)—are not fully accounted for. Non-visual factors (sound, smell) and bodily sensations (temperature, humidity) may also influence perceived safety. Future work will integrate multimodal inputs—e.g., acoustic and climate sensors—beyond camera images to enable more comprehensive urban-safety evaluation.

④ Personalizing scores by individual characteristics.

Perceived safety varies across individuals. We plan to administer brief profiling questionnaires (e.g., personality tendencies, risk sensitivity) and adjust element weights per user to present personalized perceived-safety scores.

By improving model accuracy, realizing real-time evaluation, and further integrating CPTED, this approach can evolve into a more practical urban-safety evaluation system. The developed AI model can rapidly and quantitatively surface issues and improvement items for city areas, contributing to enhanced perceived safety and comfort for residents and visitors and, ultimately, to reduced crime risk.

Looking ahead, we also envisage linkage with future AR devices (e.g., smart glasses) to present real-time perceived-safety scores at the observation point and, for example, support safer walking-route suggestions for women returning home at night. Integration with public monitoring systems could enable real-time analysis and visualization of changes in perceived safety, facilitating early detection and prediction of potential crime risks and supporting preemptive warnings and countermeasures.

We aim to translate these findings into an AI-based perceived-safety evaluation system that can be deployed across urban planning, crime prevention, and safety assurance during disasters. Our long-term goal is societal implementation that contributes to safer and more comfortable urban environments.

Beyond vision, affective state and trait risk sensitivity likely modulate cue weighting. Incorporating brief scales (e.g., state anxiety, risk perception) will enable mediation/moderation tests linking cues → appraisal → choice. This will clarify how environmental fixes translate into psychological relief across populations.

We now control for day vs. night via a dummy variable, which reduces baseline differences in perceived visibility; however, absolute luminance and contrast are still not explicitly measured. Future versions will incorporate photometric features (e.g., mean luminance, face/ground contrast) so that fixture presence and illumination intensity can be disentangled more precisely.

8. Ethics approval and consent

The study involved minimal-risk behavioral judgments by adult volunteers, with anonymized responses and the option to discontinue at any time. Before participation, an information page described the study purpose, procedures, voluntary nature, and data handling; by proceeding, participants provided implied consent. The authors acknowledge this limitation and have taken additional safeguards.

Privacy and safeguards. All faces and license plates in street photos have been blurred. Only de-identified, aggregate results are reported; identifiable images will not be shared. Future data collection will be conducted under prospective ethics approval.

9. Data and code availability

De-identified pairwise responses and code are available at [DOI/link] upon publication. Due to privacy, raw photos are shared only in anonymized form (faces and license plates blurred); semantic-segmentation masks and per-image feature tables are provided.

10. Notes

1. Intersection over Union (IoU). An accuracy metric for image segmentation that measures the overlap between
   the predicted region and the ground-truth (GT) region for a given class:
   IoU = |Predicted ∩ Ground Truth| / |Predicted ∪ Ground Truth|.
2. Mean Intersection over Union (mIoU). The average of IoU values over all classes, serving as a comprehensive
   indicator of overall accuracy in multi-class segmentation.
3. Wandell, B. A. Foundations of Vision. Sinauer Associates, 1995. This book notes that the human horizontal
   field of view extends to roughly 200°, while high-acuity vision mediated by the fovea spans about 2°, and the range generally attended to is approximately 30–40°.
4. Palmer, S. E. Vision Science: Photons to Phenomenology. MIT Press, 1999. This book provides detailed
   discussions of human visual processing, the span of focused attention, and physiological mechanisms underlying
   blur in peripheral vision.
5. Lens distortion. Owing to lens characteristics, slight distortion may occur near the edges of the field of view;
   however, for the purposes of this study—recording the overall street environment—this distortion is considered acceptable.

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Hino, K. , Koide, O . : Effect of Adopt-a-park-program with Regard to Fear of Crime in Parks, journal of Architecture and Planning, Vol. 70, No. 592, pp. 117-122, 2005. 6(in Japanese)
2. Hino, K. , Ishii, N. , Hijikata, T. , Hino, A. , Amemiya, M. : Natural Surveillance on a Pedestrian Street in Tsukuba Science City, Japan: Utilization of the survey for the amount of "eyes on the street", Reports of the City Planning Institute of Japan, Vol. 9, No. 2, pp. 64-68, 2010. 8(in Japanese)
3. Jeffery, C. R.: Crime Prevention Through Environmental Design. Beverly Hills, CA: Sage Publications, 1971
4. Newman, O. Defensible Space: Crime Prevention through Urban Design. Macmillan Publishing Company, 1973

5. Cozens, P. M., Hillier, D., & Prescott, G.: Crime and the Design of Residential Property: Exploring the Theoretical Background, Property Management, Vol. 19, No. 4, pp. 222-248, 2001.5
6. Cozens, P. , Love, T. : A Review and Current Status of Crime Prevention through Environmental Design (CPTED), Journal of Planning Literature, Vol. 30, No. 4, pp. 393-412, 2015.8
7. Naik, N . , Philipoom, J. , Raskar, R. , Hidalgo, C. : Streetscore–Predicting the Perceived Safety of One Million Streetscapes, 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2014.7
8. Xiao, H. , Natsume, Y. : Research on factors affecting road environment safety perception and evaluation based on CPTED theory, 14th International Symposium on Architectural Interchanges in Asia, 2024. 9
9. Takase, D. , Nambu, S. , Hino, K. , Tanaka, Y. : ESTIMATING ANXIETY CAUSED BY FEAR OF NIGHTTIME CRIME FROM PHYSICAL ENVIRONMENTAL CHARACTERISTICS, journal of Architecture and Planning, Vol. 87, No. 792, pp. 329-336, 2022. 2(in Japanese)
10. Dubey, A., Naik, N., Parikh, D., Raskar, R., Hidalgo, C. : Deep Learning the City: Quantifying Urban Perception at a Global Scale, European Conference on Computer Vision (ECCV 2016), PP. 196-212, 2016.9
11. Li, X., Zhang, C., Li, W., Ricard, R., Meng, Q., Zhang, W. : Assessing Street-Level Urban Greenery Using Google Street View and a Modified Green View Index, Urban Forestry & Urban Greening, Vol. 14, No. 3, pp. 675-685, 2015.6
12. Ge, X., Shen, X., Zhou, Y.: DDC-Net: Semantic Segmentation for Urban Roads Based on Improved Capsule Networks. Journal of Applied Science and Engineering (JASE), 28(10), 2163–2176, 2025. https://doi.org/10.6180/jase.202510_28(10).0009
13. Yin, S., Wang, L., Chen, T., et al.: LKAFormer: A Lightweight Kolmogorov–Arnold Transformer Model for Image Semantic Segmentation. ACM Transactions on Intelligent Systems and Technology (TIST), 2025. https://doi.org/10.1145/3759254
14. Wang, Y.: MRCNNAM: Mask Region Convolutional Neural Network Model Based on Attention Mechanism and Gabor Feature for Pedestrian Detection. Journal of Applied Science and Engineering, 26(11), 1555–1561, 2023. (Indexed in DOAJ.)
15. Yin, S., Wang, L., Teng, L.: Threshold Segmentation Based on Information Fusion for Object Shadow Detection in Remote Sensing Images. Computer Science and Information Systems, 21(4), 1221–1241, 2024. https://doi.org/10.2298/CSIS231230023Y
16. Teng, L., et al.: FLPK-BiSeNet: Federated Learning Based on Prior Knowledge and Bilateral Segmentation Network for Image Edge Extraction. IEEE Transactions on Network and Service Management, 20(2), 1529–1542, 2023. https://doi.org/10.1109/TNSM.2023.3273991
17. Jiang, Y.: A Novel DABU-Net Model Based on Principal Component Analysis for Intelligent Collaborative Robot Design. Journal of Applied Science and Engineering, 27(11), 3533–3541, 2024. https://doi.org/10.6180/jase.202411_27(11).0010