International Journal of Geo-Information

MDPI

Article

Modeling and Explaining Perceived Fear of Crime from Street View Imagery Using a GeoAI Framework

Somang Kim, Jaeyeon Choi and Youngok Kang *

Department of Social Studies (Geography), Ewha Womans University, Seoul 03760, Republic of Korea; ssomanghope@gmail.com (S.K.); jaeyeon9983@naver.com (J.C.)
* Correspondence: ykang@ewha.ac.kr; Tel.: +82-(2)3277-6646

Abstract

Understanding the spatial distribution and determinants of perceived fear of crime is essential for enhancing urban safety and promoting equitable city development. This study models and explains perceived fear of crime from street view imagery using a GeoAI framework that integrates deep learning, semantic segmentation, and explainable AI techniques. Focusing on Yeongdeungpo-gu in Seoul, South Korea—a district characterized by diverse urban morphologies—we collected 171,942 pairwise comparison responses through a large-scale crowdsourced survey designed to capture visual perceptions of crime-related fear. A Vision Transformer-based Siamese network (RSS-Swin) was employed to predict continuous fear-of-crime scores, while semantic segmentation (SegFormer-B5) and AutoML regression were applied to identify built-environment features influencing these perceptions. SHAP-based interpretability analysis was then used to quantify the importance and interactions of key visual elements. The results reveal that open and accessible streetscape components, such as roads and sidewalks, consistently mitigate perceived fear, whereas enclosed or unmanaged features, including walls, poles, and narrow alleys, heighten it. Moreover, the effects of vegetation, fences, and buildings vary across spatial contexts, emphasizing the need for place-sensitive interpretation. By integrating predictive modeling and explainable analysis, this study advances a transparent and scalable GeoAI framework for understanding the visual and environmental determinants of crime-related fear and supporting perception-aware CPTED strategies.

Keywords: perceived fear of crime; GeoAI; street view imagery; vision transformer; semantic segmentation; AutoML; SHAP; explainable AI

Check for updates

Academic Editors: Wolfgang Kainz, Lirong Yin, Shan Liu and Kenan Li

Received: 10 October 2025
Revised: 10 December 2025
Accepted: 30 December 2025
Published: 31 December 2025

Copyright: © 2025 by the authors. Published by MDPI on behalf of the International Society for Photogrammetry and Remote Sensing. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Residents' perceptions of the built environment play a critical role in shaping urban mobility, access to public spaces, and community well-being. Among these perceptions, the perceived fear of crime represents a spatially situated psychological response to the perceived risk of victimization. Unlike objective crime statistics, perceived fear of crime is shaped by subjective assessments and environmental cues [1–4]. Particularly in dense urban settings, fear perception often outweighs concerns about natural hazards or accidents. These perceptions not only affect mental health and quality of life but also challenge broader goals of sustainable and inclusive urban development [5–7].

The "Broken Windows Theory (BWT)" [8] posits that physical disorder signals and facilitates criminal behavior. A growing body of empirical evidence supports the link between disorderly environments and heightened levels of perceived fear and violent crime [9,10].

In response, many cities have adopted Crime Prevention Through Environmental Design (CPTED) principles, which emphasize spatial interventions—such as improved visibility, territoriality, and maintenance—to reduce both crime and fear [11]. However, existing studies and planning practices often rely on resident surveys or field audits, which are labor-intensive and limited in spatial coverage [12]. These methods lack the granularity and scalability needed to capture fine spatial variation or to generalize findings across diverse urban contexts [13–15].

Recent advances in computer vision and the increasing availability of street view imagery (SVI) have created new opportunities for large-scale, fine-grained visual analysis of urban safety [16–21]. Yet most existing applications have focused on generalized notions of "safety" rather than on the specific and measurable perception of fear of crime. Moreover, few studies have developed interpretable and context-sensitive models that can directly inform place-based policy and design.

To address these gaps, this study models and explains perceived fear of crime from street view imagery using a GeoAI framework that integrates deep learning, semantic segmentation, and explainable AI. Focusing on Yeongdeungpo-gu in Seoul, South Korea, we collected 171,942 pairwise comparison responses through a custom crowdsourcing platform explicitly targeting fear perception rather than general safety. The proposed framework consists of three analytical components: (1) Vision Transformer–based Siamese modeling for predicting fear scores from SVIs, (2) semantic segmentation and AutoML regression for identifying key built-environment features associated with perceived fear, and (3) SHAP-based explainability analysis for interpreting both consistent and context-dependent effects of these features.

This study makes three key contributions. First, it introduces a methodological innovation by incorporating a Vision Transformer into a Siamese ranking framework, achieving improved predictive performance and spatial generalizability over conventional CNN-based models. Second, it establishes a scalable perception modeling pipeline that combines semantic segmentation and AutoML-based regression to identify environmental correlates of fear. Third, it provides an interpretable GeoAI workflow linking predictive modeling with SHAP-based explanations, thereby revealing how specific streetscape elements—such as roads, sidewalks, walls, and vegetation—shape perceived fear of crime across spatial contexts. By integrating large-scale perception data, fine-grained visual inputs, and explainable modeling, this study advances a human-centered GeoAI approach for spatially mapping and interpreting perceived fear of crime.

The remainder of this paper is structured as follows: Section 2 reviews the related literature; Section 3 details the proposed methodology; Section 4 presents the results of predictive and interpretive analyses; Section 5 discusses theoretical and practical implications; and Section 6 concludes with a summary of the key findings.

2. Literature Review

2.1. Previous Studies on Fear of Crime Mapping

Perceived fear of crime is a complex, multifaceted phenomenon encompassing not only immediate personal risk but also emotional, social, and spatial dimensions. Prior research emphasizes that fear perception is not a single construct but a layered psychological response shaped by diverse factors including direct victimization, social trust, collective memory, and neighborhood context [1,22]. Recognizing this multidimensionality is crucial for avoiding deterministic interpretations and for ensuring that interventions are contextually appropriate.

Environmental criminology and urban studies have long shown that visual features of the built environment—such as disorder, poor lighting, or obstructed visibility—serve as

cues shaping perceptions of vulnerability. As perceived fear influences quality of life and spatial behavior, numerous studies have identified fear-prone urban areas through surveys and mapping techniques. For instance, Doran & Lees [14] and Kohm [23] demonstrated that visual and social disorder (e.g., graffiti, prostitution, poor maintenance) heighten fear in dark or isolated areas, while Fisher & Nasar [24] and Vrij & Winkel [25] associated fear with limited visibility and escape routes. In addition to these empirical findings, foundational theoretical perspectives—including Defensible Space Theory, CPTED, and the Broken Windows Theory—suggest that individuals interpret environmental cues through cognitive assessments of visibility, territoriality, order, and perceived guardianship. These theories collectively highlight that fear of crime arises not only from observable disorder but also from the broader socio-spatial meaning embedded in the built environment, providing an important conceptual basis for understanding how visual cues captured in SVI may shape perceptual responses.

With the advancement of geospatial technologies, studies have increasingly mapped fear of crime using participatory and digital tools such as mobile applications [26,27], online sketch maps [28,29], and web-based mapping platforms [30]. Solymosi et al. [27] and Pánek et al. [30] demonstrated that such participatory approaches can capture micro-scale fear patterns—such as dark alleys and poorly lit parks—often independent of actual crime statistics. Despite their value, these approaches remain constrained by participant subjectivity, uneven sampling, and limited generalizability across cities.

While these studies established an important conceptual foundation, most relied on subjective inputs or manually coded GIS variables. To enhance objectivity and scalability, the current research builds upon this foundation by employing SVI and deep learning-based GeoAI methods to detect and model visual indicators of perceived fear of crime in a systematic and reproducible manner.

2.2. Deep Learning and SVI-Based Perception Modeling

The proliferation of SVI platforms such as Google Street View, combined with the rise of deep learning, has transformed how researchers analyze visual perceptions of urban environments [31–34]. Unlike aerial imagery, SVIs provide ground-level, human-perspective visuals that capture the environmental cues influencing urban cognition and emotion.

A major milestone was the development of the Place Pulse datasets by the MIT Media Lab, which collected millions of pairwise comparisons of SVIs to assess perceptions such as safety, beauty, and liveliness [35,36]. These datasets catalyzed the creation of deep learning models capable of predicting urban perception directly from imagery. Two main data collection paradigms emerged: (1) absolute ratings using Likert scales [37,38], and (2) pairwise comparisons, where users select the more favorable image [39,40]. The latter proved more robust for modeling perceptual preferences, leading to models such as RSS-CNN [36].

Building on Place Pulse 2.0, subsequent studies introduced more sophisticated models [41–44]. Min et al. [41] proposed a multi-task deep relative attribute learning network (MTDRALN) to predict multiple perceptual attributes simultaneously. Xu et al. [42] integrated Siamese networks with semantic segmentation to connect visual objects with perceived attributes, while Guan et al. [43] developed city-adaptive models to account for contextual differences. Kang et al. [44] extended this paradigm by integrating global–local features for walkability perception prediction, achieving the highest accuracy (75.01%).

Despite these advances, most perception models have relied on convolutional neural networks (CNNs), which are limited in representing long-range dependencies critical for understanding complex urban scenes. Vision transformer architectures (e.g., Swin Transformer) offer improved contextual awareness and multiscale representation, making them better suited for modeling nuanced perceptual responses. However, their potential in fear-of-crime

analysis remains underexplored. Bridging this methodological gap is essential for improving both the predictive accuracy and interpretability of perception models in GeoAI research.

2.3. Modeling Built Environment Features Affecting Perceived Fear of Crime

A growing body of research examines how specific built-environment features influence perceived fear, informed by frameworks such as Defensible Space Theory (DST) [45], CPTED, and the BWT. DST emphasizes spatial design that enhances surveillance and territorial control, while CPTED extends these ideas through principles of visibility, access control, and maintenance. BWT highlights how signs of disorder (e.g., litter, graffiti, neglect) incite fear and weaken collective efficacy.

Recent studies increasingly combine these theories with computer vision and semantic segmentation to quantify how visual elements contribute to perceived safety or fear [16–19,21,46]. For example, Zhang et al. [16] used segmentation models trained on ADE20K to correlate sidewalks, grass, and cars with higher safety perceptions, while sky and walls correlated with increased fear. Jing et al. [17] and Ramírez et al. [18] found that greenery reduces fear by signaling maintenance and order, whereas poorly lit or cluttered features amplify fear. Wang et al. [19] and Hou & Chen [21] similarly demonstrated that open, accessible streets reduce fear, while visual barriers or neglected infrastructure elevate it.

Although general patterns emerge, feature effects vary contextually. For instance, trees may either reduce or increase fear depending on maintenance and visibility, and buildings may provide safety through surveillance but induce fear when abandoned or densely packed. Such inconsistencies underscore the need for context-sensitive, interpretable models capable of integrating both visual and spatial cues.

Most previous studies have used segmentation models such as DeepLab or PSPNet, which, while effective, struggle to capture the full semantic complexity of urban streetscapes. Transformer-based architectures, by contrast, enable fine-grained semantic interpretation and long-range feature dependency modeling. This study addresses these gaps by integrating vision transformers, semantic segmentation, AutoML, and SHAP-based explainability into a unified multi-stage GeoAI framework, thereby advancing the methodological frontier for analyzing how built environments shape the perception of fear of crime in urban spaces

3. Materials and Methods

This study focuses on Yeongdeungpo-gu, a district in Seoul, South Korea (Figure 1). With industrial, residential, and commercial functions coexisting within compact urban blocks, Yeongdeungpo-gu provides heterogeneous contexts that are well-suited for street-level, perception-based safety modeling. The study was structured as a three-step GeoAI workflow encompassing prediction, feature modeling, and explainable interpretation (Figure 2).



Figure 1. Study area.

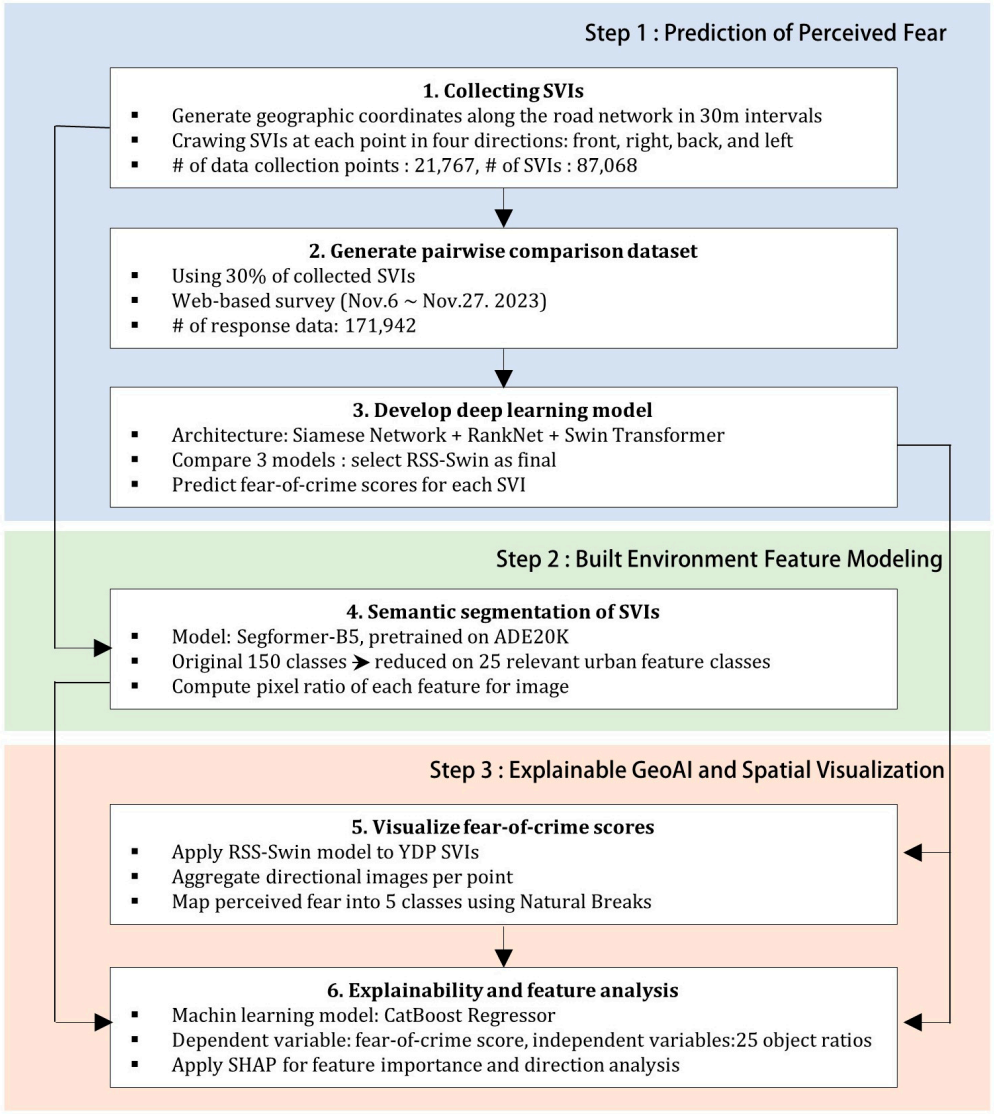


Figure 2. Research flow of this study. Arrows indicate the workflow sequence, and "#" represents the number of SVIs or responses.

3.1. SVI Acquisition and Preprocessing

In South Korea, SVIs can be accessed through major internet portals such as Google, Naver, and Kakao. For this study, SVIs were collected from Kakao Map, which provides both broader spatial coverage and more frequent updates than Google or Naver. SVIs were accessed via Kakao Map (https://map.kakao.com, accessed on 15 November 2023). Sampling locations for SVI collection were derived from road network data provided by the National Spatial Data Infrastructure (NSDI) portal. Using this dataset, we extracted coordinates at approximately 30 m intervals along road segments and retrieved SVIs corresponding to each location.

Since SVIs are panoramic images, visual perception varies by viewing direction. Therefore, four directional images (0°, 90°, 180°, 270°) were collected at each point. To emulate a pedestrian's viewpoint, the vertical angle (pitch) was fixed at 0°, aligning the camera's optical axis with the horizontal line of sight. The SVIs were crawled in 2023, although more than 90% of the images were originally captured between July 2021 and September 2022. These images were typically taken during daytime under clear weather conditions, which is standard for commercial street-view services. To ensure representation of diverse urban morphologies, additional SVIs were collected from Anyang City, which includes

both newly developed areas and older, traditional neighborhoods. In total, 41,648 images were collected from 10,412 locations in Yeongdeungpo-gu and 45,420 images from 11,355 locations in Anyang, resulting in a combined dataset of 87,068 images for analysis.

3.2. Perception Labeling Through Pairwise Comparisons

To train a model capable of predicting perceived fear of crime, a dataset comprising SVIs and corresponding perception based labels was constructed. Two key design decisions guided this process: (1) determining the sampling ratio of images to be included in the training dataset and (2) defining the method for collecting perceptual evaluations. To reduce the response burden while preserving spatial representativeness, we selected 30% of all SVI locations for the perception survey using stratified sampling. Image acquisition points were stratified according to road type (arterial, collector, and local roads) and land use classification (industrial, commercial/business, high density residential, low density residential, and others), excluding highway segments, and within each stratum the required number of points was randomly selected. The subsampling was applied to spatial points rather than to viewing directions; for every selected point all four headings (0°, 90°, 180°, and 270°) were retained so that no particular direction was systematically over or underrepresented. This stratified sampling yielded 2371 image points (9484 images) in Yeongdeungpo gu and 3107 points (12,428 images) in Anyang, resulting in a total of 20,886 usable SVIs after removing blurred or distorted content (Table 1).

Table 1. Summary of data.

Data	Yeongdeungpo-gu	Anyang	Total
Sample points	10,412	11,355	21,767
Number of SVIs	41,648	45,420	87,068
Number of stratified points for training	2371	3107	5478
Number of SVIs for training	9484	12,428	21,912
Final number of SVIs for pairwise comparison	9040	11,846	20,886

Fear-of-crime evaluations can generally be collected through one of two approaches: (1) an absolute rating method, in which participants assign numerical scores to individual images; or (2) a relative comparison method, where participants are shown two images and asked to select the one that appears safer. Previous research has demonstrated that the relative approach is typically more reliable and efficient for capturing subjective perceptions [39,40]. Accordingly, this study adopted a pairwise comparison method, in which participants were shown two SVIs side by side and asked to choose which scene appeared safer from crime-related risk, based solely on visual impressions. We clarified that "crime" in this study referred to interpersonal crime (such as assault, robbery, or harassment) rather than traffic accidents or natural hazards, and explicitly instructed participants to focus on visual cues related to vulnerability to such person-oriented crime.

A custom web-based survey platform was developed for this purpose (Figure 3). We recruited 65 participants representing diverse age and gender groups: 25 in their 20s, 20 in their 30s and 40s, and 20 aged 50 or older, including 35 female participants. The survey protocol was reviewed and approved by the Institutional Review Board (IRB) of Ewha Womans University on 17 October 2023. To minimize potential bias and maintain a perceptual focus, no information was collected regarding participants' familiarity with Yeongdeungpo-gu or their prior experiences with crime, ensuring that responses were based purely on visual interpretation. To further enhance data reliability, five trap question pairs with easily distinguishable differences were inserted to detect inattentive or malicious responses. Additionally, the question set was algorithmically structured to achieve statisti-

cal stability in preference scores even with a limited number of comparisons, minimizing variance caused by respondent heterogeneity [47]. The survey was conducted from 6 to 27 November 2023, producing 178,750 total responses. After excluding 6808 neutral responses (i.e., selections of "equal"), the final dataset contained 171,942 valid pairwise comparisons. These corresponded to 20,886 unique SVIs, with each image appearing in at least 16 comparisons on average. This high comparison frequency per image ensured that the derived preference scores were robust and less susceptible to individual-level bias, providing a solid foundation for subsequent model training.

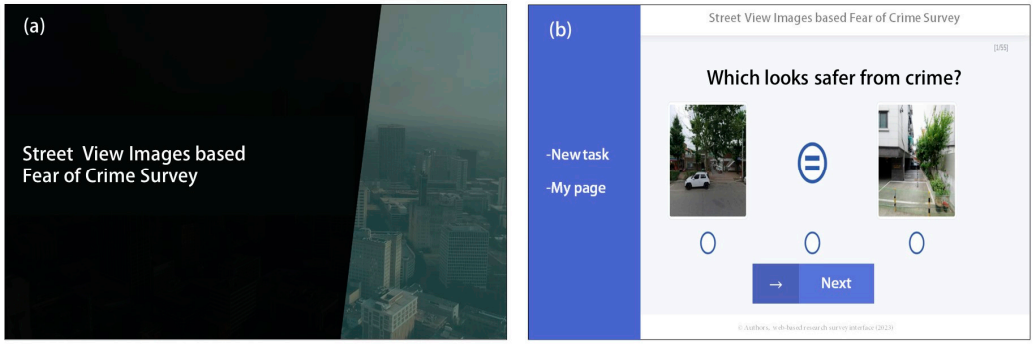


Figure 3. Survey website: (**a**) landing page; (**b**) response screen.

3.3. Step 1: Predicting Perceived Fear of Crime Using a Vision Transformer Model

To predict perceived fear-of-crime scores from SVI, a deep learning architecture was developed that strategically integrates a Siamese network, a RankNet-based ranking mechanism, and a Swin Transformer backbone for visual feature extraction (Figure 4). The proposed model comprises two main modules—a Feature Block and a Score Block—both embedded within the Siamese architecture (Koch et al. [48]). This design enables the network to process image pairs in parallel through identical subnetworks with shared weights, thereby learning subtle visual distinctions that influence comparative judgments. Each image is divided into fixed-size patches, which are embedded into high-dimensional vectors and passed through the Swin Transformer (Liu et al. [49]). The Transformer applies self-attention within local windows and shifts these windows across successive layers to capture both fine-grained object details and broader spatial configurations—an essential capability for modeling perceived fear, which often arises from spatial characteristics such as enclosure, visibility, and the clustering of disorderly elements. The feature vectors extracted from the two images are then jointly processed within the Score Block, which employs a RankNet-based probability ranking function [50]. This module learns to reflect relative fear judgments derived from pairwise comparisons. Using a Softmax-based cross-entropy loss, the model is optimized to predict which of two images evokes greater perceived fear, and simultaneously assigns each image a continuous, interpretable fear-of-crime score. By combining the hierarchical spatial reasoning of the Swin Transformer with pairwise perceptual learning through the Siamese–RankNet integration, the proposed model captures both object-level semantics and contextual cues across multiple scales. This architecture enables a fine-grained and scalable prediction of perceived fear, surpassing conventional CNN-based methods and offering a theoretically grounded, GeoAI-driven framework for urban safety analysis.

The dataset for model training consisted of pairwise comparisons, formally represented as $D = \{(x_i, x_j, y)\}$, where n denotes the total number of images and $i, j \in \{1, \ldots, n\}$ are image indices. Each pair (xi, xj) corresponds to a side-by-side presentation of two images, and the binary label $y \in \{0, 1\}$ indicates which image was perceived as safer from crime. Specifically, $y = 1$ denotes that the right-hand image x_j was chosen as safer (i.e., evoking

less fear of crime) than x_i, while y = 0 indicates the opposite. The model aims to learn a ranking function f_r: $\mathbb{R}^n \rightarrow \mathbb{R}$ that assigns each image a continuous safety score, where higher scores indicate lower perceived fear of crime. The RankNet-based framework is defined as follows:

$$\begin{cases} y = 1, \ f(x_i) < f(x_j) \\ y = 0, \ f(x_i) > f(x_j) \end{cases} \tag{1}$$

$$P_{ij} = \frac{e^{S_{ij}}}{1 + e^{S_{ij}}} \tag{2}$$

$$L_{r(i,j)} = -\overline{P}_{ij} log P_{ij} - (1 - \overline{P}_{ij}) log(1 - P_{ij}) \tag{3}$$

P_{ij}: the predicted probability that image x_i is perceived as more fearful than x_j.
$\overline{P}_{ij}$: the ground-truth label derived from pairwise comparisons.
S_{ij}: the difference between the safety scores of images x_i and x_j, defined as

$$f(x_i) - f(x_j)$$

$L_{r(i,j)}$: the RankNet loss for the image pair (x_i, x_j).

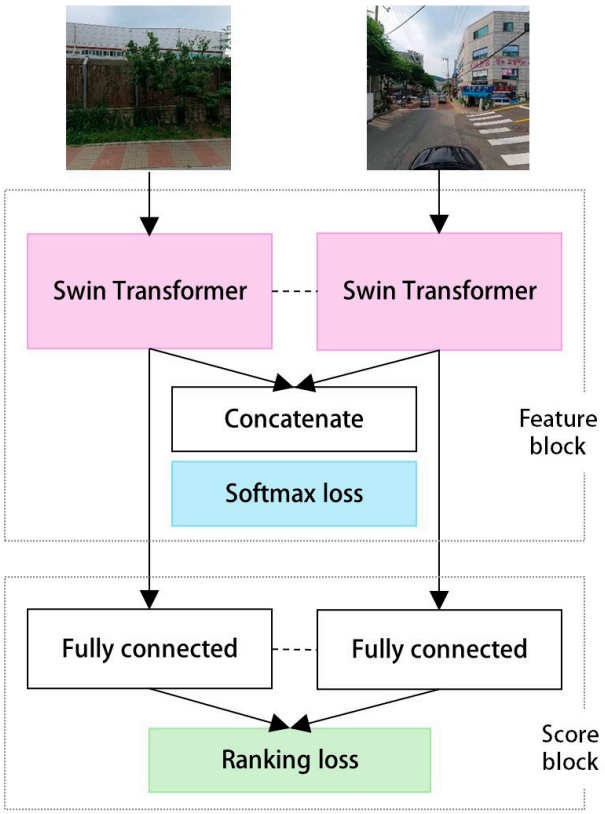


Figure 4. RSS-Swin model architecture.

To evaluate the model's performance, the proposed RSS-Swin was compared with existing models that predict perceptual scores from pairwise data. Three models were examined: (1) Global-Patch-RSS-CNN (Kang et al. [44])—the best-performing CNN-based baseline that integrates global and local visual patterns. It previously achieved an accuracy of 75.01% and outperformed earlier CNN variants such as RSS-CNN [36], semantic-enhanced CNN (Xu et al. [42]), and local patch models. (2) RSS-ViT—a Siamese + RankNet framework using a Vision Transformer (ViT) backbone (vit_base_patch16_224), pre-trained on ImageNet-1k via the TIMM library. ViT is recognized as the first Transformer-based architecture to suc-

cessfully apply self-attention to image classification. (3) RSS-Swin (proposed)—also built on the Siamese + RankNet framework but replacing the backbone with a Swin Transformer (swin_base_patch4_window7_224), pre-trained on ImageNet-1k. The Swin Transformer effectively handles high-resolution images with hierarchical and window-shifting attention, enabling robust spatial representation. Model training was performed on Amazon Web Services (AWS) using a g4dn.12xlarge instance (four NVIDIA T4 GPUs, 48 vCPUs, 900 GB NVMe SSD, Ubuntu 18.04.1). Of the total pairwise dataset, 80% (137,552 pairs) was used for training and 20% (34,390 pairs) for testing. Model accuracy was computed as the proportion of test pairs whose predicted ranking order matched the ground-truth label.

Table 2 summarizes the model comparison results. Among the three models, RSS-Swin achieved the highest accuracy (82.03%), outperforming all prior CNN-based frameworks. Although the numerical gain over previous models may appear moderate, the hierarchical attention mechanism of Swin Transformer provides a significant qualitative advance by effectively capturing spatial configurations relevant to fear perception. Compared with CNNs—often limited by scale variation and irregular layouts—the Swin Transformer exhibits enhanced generalizability across heterogeneous urban environments. Figure 5 illustrates the training and validation loss trajectories. Based on the validation loss curve, which reached its minimum without evidence of overfitting, the model parameters from epoch 3 were selected as the final version for deployment.

Table 2. Comparison of accuracy between models.

Model	Accuracy
Global-Patch-RSS-CNN	81.44%
RSS-ViT	81.87%
RSS-Swin	82.03%

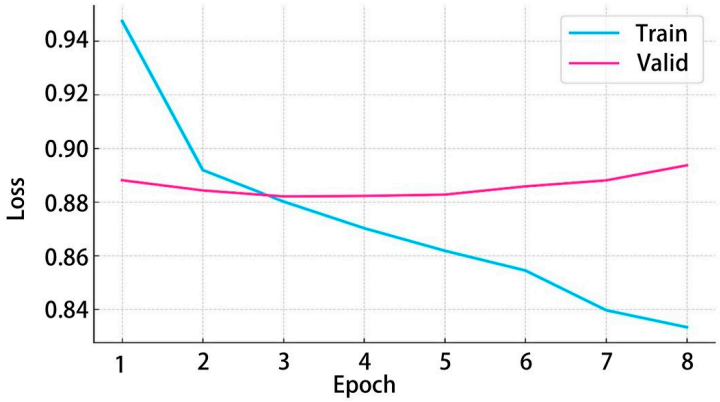


Figure 5. Train and validation loss over epochs.

3.4. Step 2: Modeling Built-Environment Features via Semantic Segmentation and AutoML

To further investigate how the built environment influences perceived fear of crime, we analyzed the relationship between the physical features extracted from SVIs and the predicted fear scores obtained from the RSS-Swin model in Stage 1. Semantic segmentation was employed to extract object-level features from each image, thereby enabling a detailed and quantitative representation of the urban streetscape. For semantic segmentation, we utilized the SegFormer-B5 model [51], trained on the ADE20K dataset [52]. SegFormer integrates the Vision Transformer architecture [53] with a hierarchical CNN-based encoder, enabling efficient feature learning across multiple scales. Compared with conventional CNN-based segmentation models, SegFormer achieves lower segmentation error and higher parameter efficiency, particularly in terms of mean Intersection over Union

(mIoU) performance. Figure 6 compares the segmentation outputs of DeepLabV3 and SegFormer-B5, both trained on ADE20K. The SegFormer-B5 model demonstrated clearer object boundaries, finer segmentation of small-scale elements, and fewer misclassifications across complex urban scenes.

To examine the influence of streetscape features on perceived fear of crime, we defined both dependent and independent variables. The dependent variable was the perceived fear-of-crime score predicted for each SVI in Stage 1. The independent variables were derived from the semantic segmentation outputs of SVIs collected in Yeongdeungpo-gu. To ensure analytical relevance, segmentation classes in which more than 90% of pixel values across all images were zero were excluded. Consequently, 25 classes out of the original 150 ADE20K categories were retained as independent variables for subsequent modeling (Table 3).



Figure 6. Comparison of semantic segmentation results of SegFormer-B5 and DeepLabV3 models for SVIs.

To explore the relationships between these 25 independent variables and the dependent fear-of-crime scores, we employed machine-learning-based regression. Selecting an appropriate model and tuning its hyperparameters typically requires extensive experimentation; therefore, to streamline this process, we adopted an automated machine learning (AutoML) approach using the open-source PyCaret library. The dataset was divided into training (80%) and testing (20%) subsets, and model performance was evaluated using 5-fold cross-validation. A variety of candidate algorithms were compared using standard

regression metrics, including Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2) (Equations (4)–(7)):

$$MAE = \frac{\sum|y - \hat{y}|}{n} \tag{4}$$

$$MSE = \frac{\sum(y - \hat{y})^2}{n} \tag{5}$$

$$RMSE = \sqrt{\frac{\sum(y - \hat{y})^2}{n}} \tag{6}$$

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \tag{7}$$

y = actual value, $\overline{y}$ = mean of observed value, $\hat{y}$ = predicted value, n = # of samples.

Table 3. Descriptive statistics of variables.

Category	Name	Min	Max	Mean	SD
Dependent variables	Fear of crime score	−7.04	3.85	−0.38	1.51
Independent variables	Wall	0	100	2.39	8.01
	Building	0	100	33.12	24.13
	Sky	0	89.04	15.77	12.51
	Tree	0	79.28	8.83	12.16
	Road	0	87.23	21.00	13.06
	Grass	0	61.57	0.74	3.96
	Sidewalk	0	46.00	4.46	6.01
	Person	0	17.12	0.12	0.47
	Earth	0	93.61	0.50	3.00
	Plant	0	87.55	1.59	4.31
	Car	0	46.71	5.76	7.01
	Fence	0	63.41	0.88	2.44
	Railing	0	33.79	0.21	1.14
	Box	0	38.86	0.15	0.96
	Signboard	0	29.75	0.62	1.54
	Stairs	0	29.58	0.13	0.82
	Truck	0	61.73	0.22	1.50
	Awning	0	43.56	0.23	1.19
	Streetlight	0	7.11	0.08	0.20
	Pole	0	20.58	0.36	1.14
	Minibike	0	20.92	0.13	0.64
	Trade name	0	22.38	0.09	0.54
	Pot	0	9.56	0.05	0.27
	Traffic light	0	9.61	0.04	0.26
	Ashcan	0	18.29	0.07	0.45

Among all tested models, the CatBoost Regressor [54] demonstrated the highest explanatory power (Table 4). CatBoost is a gradient-boosting algorithm that employs ordered boosting and ordered target statistics to mitigate overfitting and prevent target leakage during model training. After hyperparameter optimization, the final CatBoost model achieved an R^2 of 0.7964 (Table 5). This optimized model was subsequently used to assess the relative importance of built-environment features and their contributions to perceived fear of crime.

Table 4. Comparison of model performance. Sorted in R^2 order.

Model	MAE	MSE	RMSE	R^2
CatBoost Regressor	0.5319	0.4847	0.6962	0.7874
Light Gradient Boosting Machine	0.5561	0.5276	0.7263	0.7686
Extra Trees Regressor	0.5530	0.5384	0.7338	0.7639
Extreme Gradient Boosting	0.5619	0.5399	0.7348	0.7632
Random Forest Regressor	0.5583	0.5444	0.7378	0.7612
Gradient Boosting Regressor	0.5980	0.6066	0.7788	0.7340
K Neighbors Regressor	0.6205	0.6710	0.8191	0.7057
AdaBoost Regressor	0.7012	0.7893	0.8883	0.6539
Linear Regression	0.7246	0.8798	0.9380	0.6142
Ridge Regression	0.7246	0.8798	0.9380	0.6142
Bayesian Ridge	0.7246	0.8798	0.9380	0.6142
Elastic Net	0.7520	0.9541	0.9767	0.5817
Lasso Least Angle Regression	0.7794	1.0219	1.0109	0.5519
Lasso Regression	0.7794	1.0219	1.0109	0.5519
Huber Regressor	0.7559	1.0430	1.0213	0.5427
Decision Tree Regressor	0.7996	1.1063	1.0518	0.5149
Least Angle Regression	0.7996	1.2223	1.0998	0.4637
Orthogonal Matching Pursuit	0.9211	1.4074	1.1863	0.3829
Passive Aggressive Regressor	1.1041	1.8925	1.3522	0.1716
Dummy Regressor	1.1985	2.2819	1.5105	−0.0004
CatBoost Regressor	0.5319	0.4847	0.6962	0.7874
Light Gradient Boosting Machine	0.5561	0.5276	0.7263	0.7686
Extra Trees Regressor	0.5530	0.5384	0.7338	0.7639
Extreme Gradient Boosting	0.5619	0.5399	0.7348	0.7632
Random Forest Regressor	0.5583	0.5444	0.7378	0.7612

Table 5. Analysis of the test dataset of CatBoost Regressor tuned to optimal hyperparameters.

Model	MAE	MSE	RMSE	R^2
CatBoost Regressor	0.5254	0.4715	0.6867	0.7964

3.5. Step 3: Explaining Model Outputs Using SHAP

To enhance the interpretability of the AutoML-selected regression model, SHAP (SHapley Additive exPlanations) analysis was employed to quantify both global feature importance and instance-level contributions, as well as to explore interactions among built-environment variables. This approach is particularly effective for interpreting complex ensemble models such as the CatBoost Regressor, whose internal decision structure is not easily explained by model parameters alone. SHAP is grounded in cooperative game theory, where the prediction process is regarded as a game in which each feature contributes a portion of the final output. The SHAP value, therefore, represents the marginal contribution of an individual feature to a specific prediction, averaged over all possible feature coalitions [55]. This enables a unified interpretation of both local and global behavior in non-linear machine-learning models. Formally, the SHAP value for an independent variable i with respect to a prediction $f(x)$ is defined as follows:

$$\varphi_i(f,x) = \sum_{z' \subseteq x'} \frac{|z'|!(M-|z'|-1)!}{M!}\left[f_x(z') - f_x(z' \setminus i)\right] \quad (8)$$

φ_i: SHAP value of i th feature;
f: CatBoost Regressor;
x: input data;

M: total number of features (25);

z': a binary vector indicating whether each feature is included ($z'_i = 1$ if feature i is present).

By computing φ_i for all features and samples, the SHAP framework provides a transparent decomposition of the model output into additive feature effects. This allows for both global interpretation (identifying features with consistently strong influence) and local analysis (explaining why specific locations or street scenes yield higher or lower perceived fear scores). The subsequent section (Section 4) presents the empirical SHAP-based results and visual interpretations derived from this analysis.

4. Results

4.1. Visualization of Street-Level Perceived Fear

After removing duplicate entries from the initial dataset of 41,648 SVIs in Yeongdeungpo-gu, a total of 39,412 unique street-view images were used to predict perceived fear-of-crime scores using the RSS-Swin model. The distribution of the predicted scores is shown in Figure 7. The scores follow an approximately symmetric distribution, centered around a mean of −0.3806, with a minimum of −7.0358 and a maximum of 3.8459. Figure 8 presents representative examples of individual SVIs along with their corresponding predicted fear-of-crime scores. These examples visually demonstrate how environmental characteristics—such as narrow alleys, walls, or the presence of greenery—affect the model's perception of fear.

Figure 9 visualizes the predicted fear-of-crime scores at each image acquisition point across Yeongdeungpo-gu. For each location, the final score was computed as the average of the four directional images (0°, 90°, 180°, and 270°). The values were subsequently classified into five categories using the Natural Breaks (Jenks) method. Lower scores (shown in red) indicate areas associated with higher perceived fear of crime, whereas higher scores (blue) correspond to lower perceived fear and a stronger sense of safety. The spatial distribution reveals clear and interpretable patterns across the district. Areas with elevated perceived fear are relatively concentrated in the southern part of Yeongdeungpo-gu, whereas low-fear areas cluster around Yeouido and the northern commercial zones.

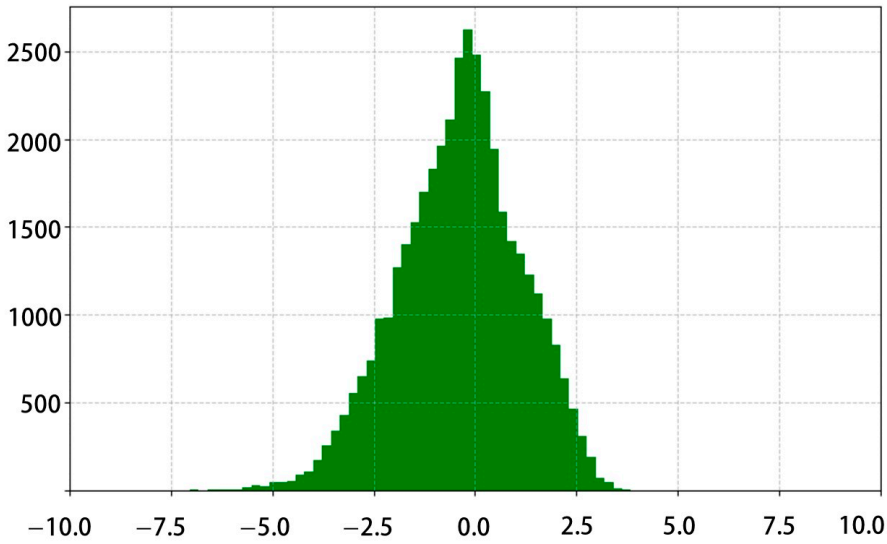


Figure 7. Distribution of fear of crime score for 39,412 SVIs: the x-axis represents fear of crime score, and the y-axis presents the number of SVIs.

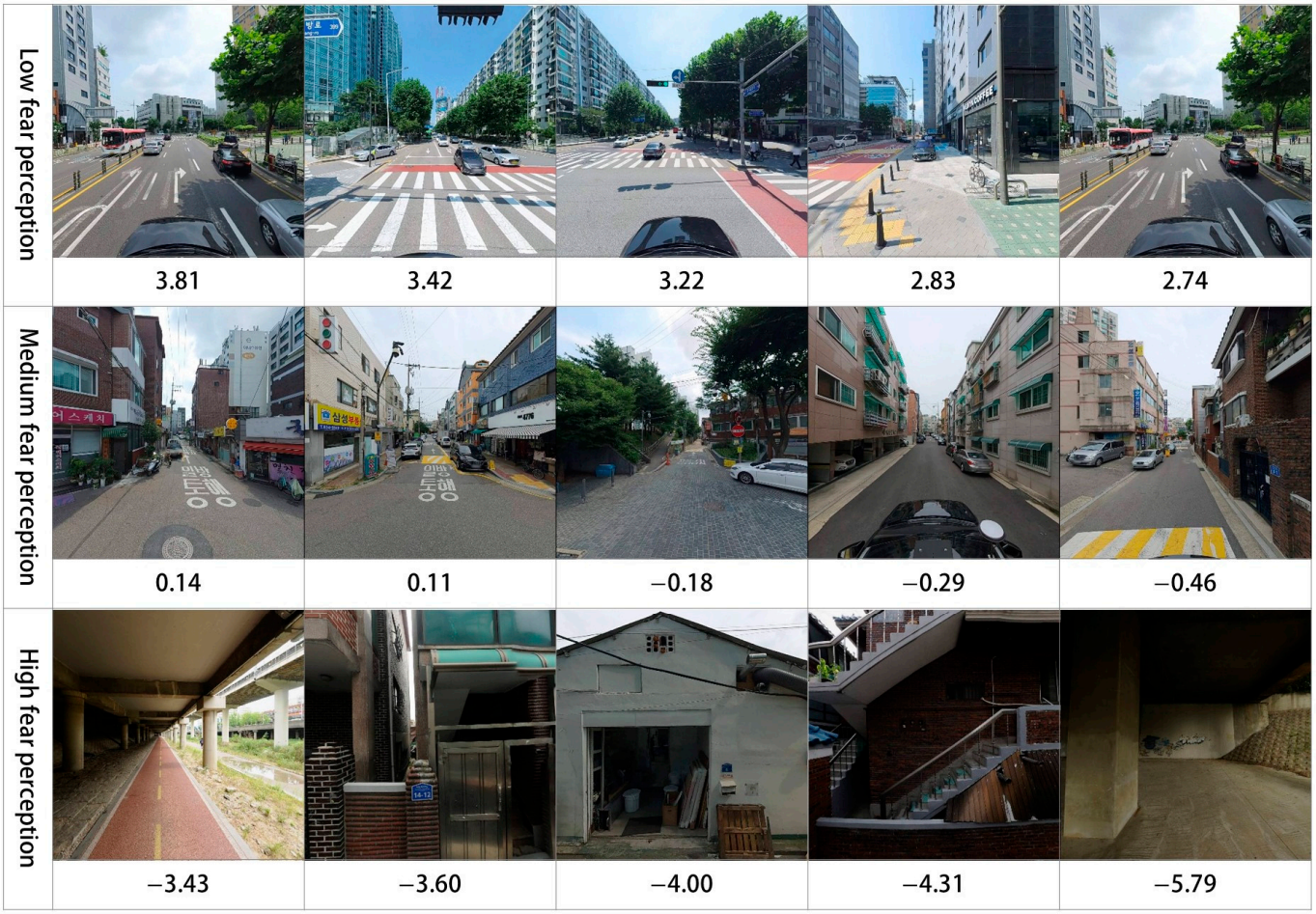


Figure 8. Examples of the fear of crime score of SVIs.

A closer examination of the high-fear clusters labeled "a" and "b" shows that, although these regions generally exhibit elevated fear levels, the major arterial roads within them tend to appear blue, suggesting a relatively greater sense of safety. Area "a" corresponds to an industrial district characterized by aging factory buildings and narrow alleyways, which limit visibility and create a sense of enclosure, while area "b" consists mainly of densely packed single-family houses and villas arranged along confined streets that intensify spatial closure.

In contrast, areas labeled "c" and "d," which represent low-fear zones, exhibit distinctly different physical characteristics. Area "c" is a commercial and business district with broad open streets and enhanced visibility, providing a clear sense of spatial openness, whereas area "d" is dominated by large apartment complexes featuring wide, well-maintained roads and organized layouts, both of which contribute to a stronger perception of safety.

Overall, the spatial patterns suggest that traffic activity, lighting, and openness mitigate perceived fear even within neighborhoods generally regarded as unsafe. These findings highlight the importance of micro-level variations in urban morphology, which can produce heterogeneous fear perceptions within short spatial distances—an important consideration for perception-aware urban design and CPTED-based interventions.

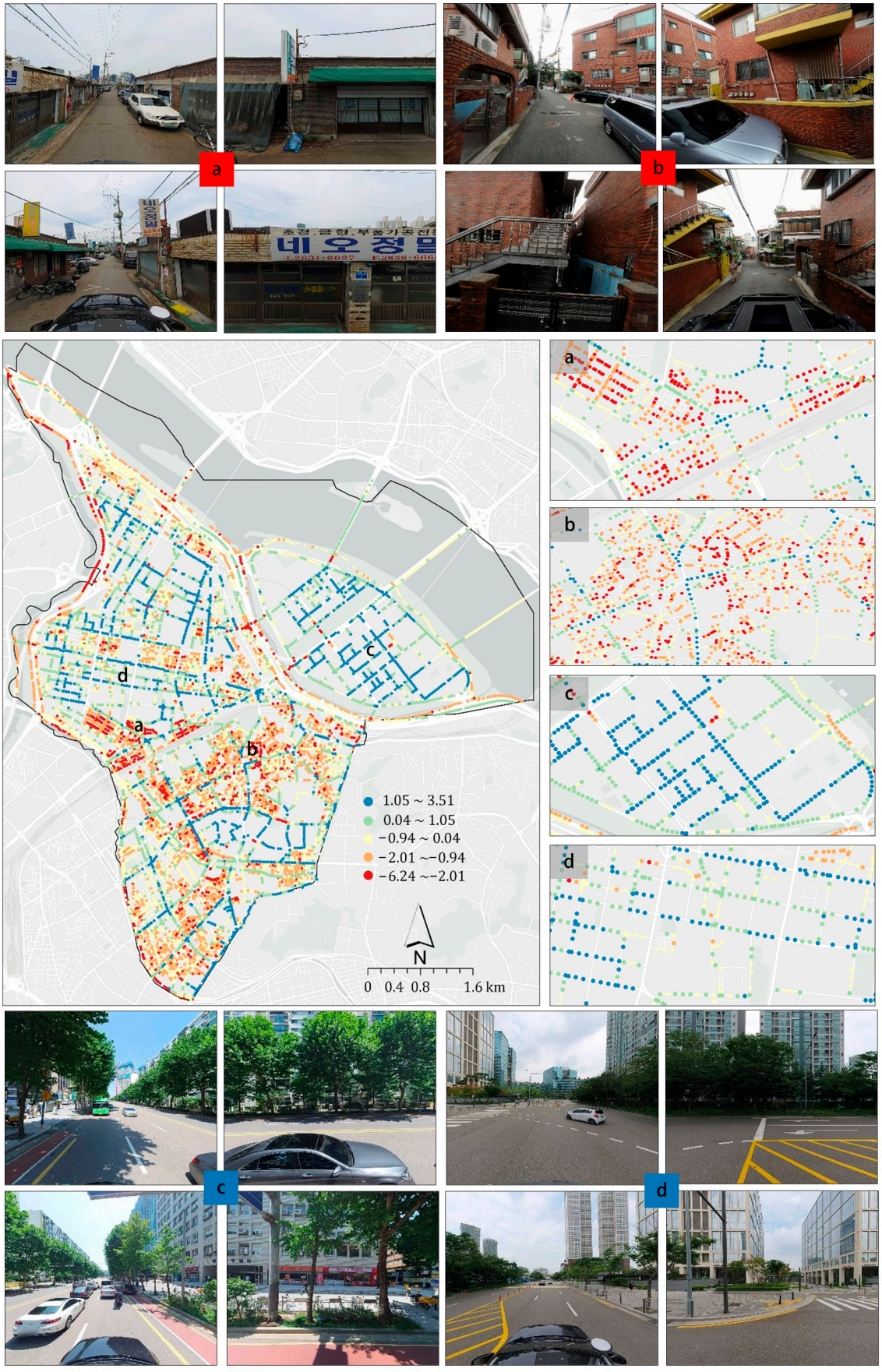


Figure 9. Visualization of perceived fear of crime at the street level. Lowercase letters (**a**–**d**) denote representative locations on the map and their corresponding street view image examples.

4.2. Feature Importance and Directional Effects

Using the previously trained CatBoost regressor model, we analyzed the relative importance and directional influence of built environment features on the perceived fear of crime. Figure 10 presents the results of the feature importance analysis, which quantifies the average impact of each variable on the predicted values. Feature importance was computed using Equation (9), defined as the expected absolute difference between the predicted values when the feature was included and when it was excluded. The results indicated that road had the strongest influence, followed by sidewalk, tree, building, and car.

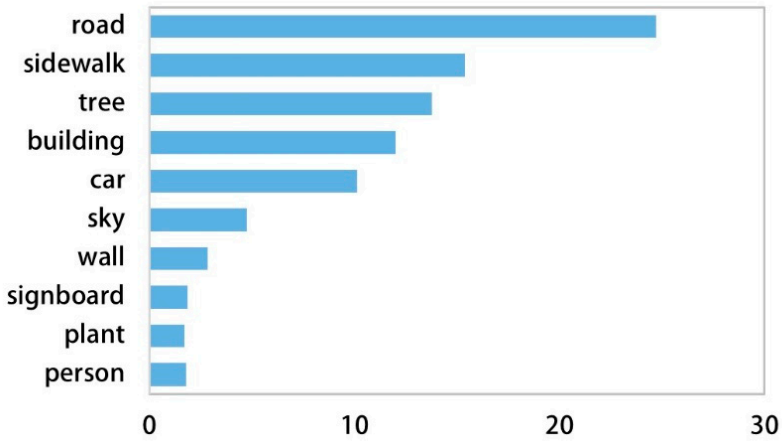


Figure 10. The 10 major features of CatBoost regressor and the importance of each feature.

These features jointly represent the key structural and functional elements of street environments that contribute most significantly to perceived safety.

$$\text{Importance}(f_i) = \mathbb{E}\left[\left|\hat{y}_{with\ f_i} - \hat{y}_{without\ f_i}\right|\right] \quad (9)$$

f_i: ith independent variable;

$\hat{y}$: predicted score of perceived fear of crime.

To further interpret the directional effects of each variable, Figure 11 shows the SHAP summary plot, which visualizes the distribution of SHAP values across all data points. Features are ordered by their overall importance, while the color of each point indicates the relative pixel proportion of that object in the SVI—red representing higher ratios and blue indicating lower ones. The horizontal position of each point shows both the magnitude and direction of influence: points located to the right of the centerline contribute to lower perceived fear of crime (i.e., greater perceived safety), whereas points to the left are associated with higher perceived fear. This interpretation aligns with the directionality of the model output, where higher predicted scores correspond to a reduced sense of fear.

Among all variables, road exhibited a strong positive association, suggesting that a higher proportion of visible roadway in an image is linked to a lower level of perceived fear. Similar effects were observed for sidewalk and car, indicating that the presence of well-developed pedestrian and vehicular infrastructure enhances the sense of safety. In contrast, wall, earth, and ashcan features, which often appear in enclosed or poorly maintained environments, were associated with higher perceived fear when their proportions increased. Features such as pole and awning followed a comparable trend, likely reflecting their frequent occurrence along narrow alleyways, which can evoke a sense of spatial confinement. Interestingly, features such as plant, fence, and grass exhibited SHAP values distributed on both sides of the axis, indicating context-dependent effects.

In some environments, these features contribute positively by improving visual appeal and natural visibility, whereas in others—particularly in isolated or poorly lit settings—they

may increase perceived fear by reducing sightlines or signaling neglect. These mixed patterns underscore the importance of contextual and compositional relationships among urban elements in shaping perceptual safety, suggesting that identical objects can generate contrasting impressions depending on their spatial arrangement and surrounding conditions.

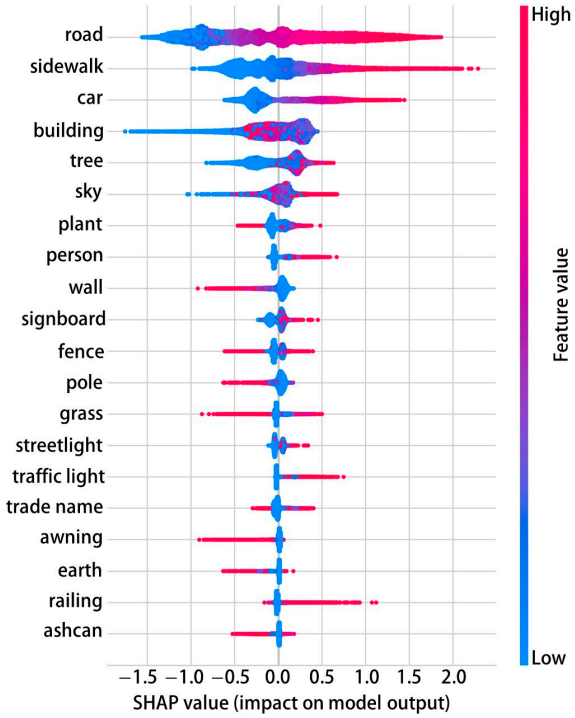


Figure 11. SHAP summary plot of total data.

4.3. Interaction and Non-Linear Relationships

To further investigate how built-environment features jointly shape perceptions of fear, SHAP dependence plots were employed to analyze the interactions and non-linear relationships among key variables. These plots compare each major feature's SHAP value with that of its most strongly correlated secondary feature, thereby illustrating how their joint variations influence the predicted fear-of-crime score. Figure 12 presents the results for the top eight influential variables, visualizing their pairwise interactions. To capture the underlying non-linear trends, we applied LOWESS (Locally Weighted Scatterplot Smoothing), which performs localized regression by assigning greater weight to nearby data points. In this study, 30% of the data was used within each local regression window to generate smoothed interaction curves.

The results reveal clear patterns for several key features. Road and sidewalk both exhibited a consistent tendency where higher proportions corresponded to lower predicted fear, confirming their collective role in enhancing perceived safety. A particularly strong interaction was observed between road and building. When road coverage was low and building coverage was high, perceived fear increased substantially, suggesting that narrow streets with dense building walls heighten spatial confinement. In contrast, when road coverage was high and building coverage remained moderate, perceived fear decreased, reflecting the fear-reducing effect of openness and visibility. A similar relationship was found for sidewalk and road; at comparable levels of sidewalk presence, higher road coverage was consistently associated with lower fear scores. Together, these results imply that the coexistence of well-developed roads and sidewalks contributes significantly to reducing perceived fear by fostering greater accessibility and visibility within the streetscape.

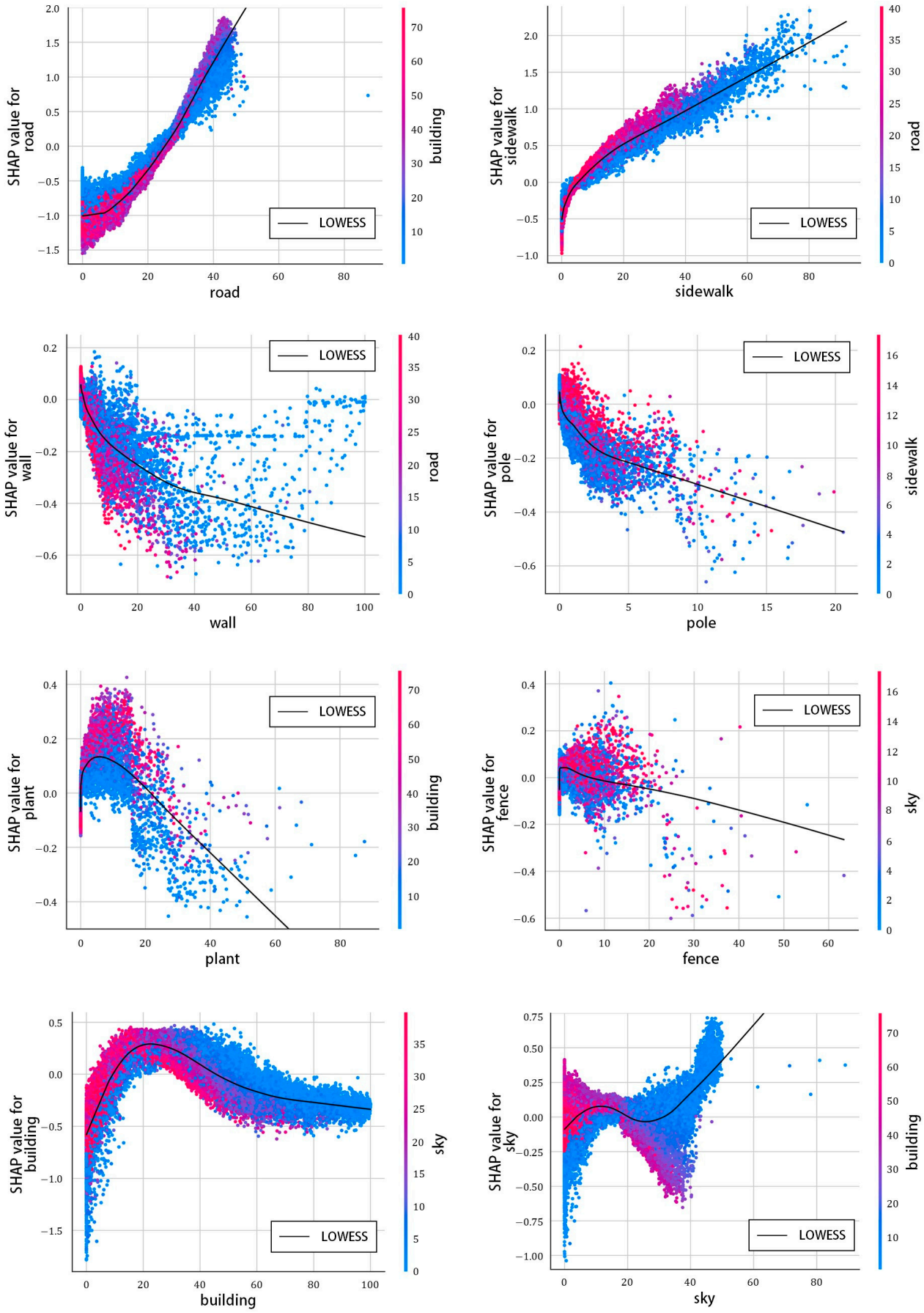


Figure 12. SHAP value of 8 factors and interaction between variables.

Negative associations were also observed for wall, pole, plant, and fence, which tended to increase perceived fear when their proportions were high. For example, wall exhibited a strong interaction with road, and fear levels were highest in areas where both features were prominent—such as underpasses, tunnels, or beneath elevated highways—where visibility is restricted and enclosure is accentuated. Similarly, pole was closely related to sidewalk, and fear perception intensified when pole density increased alongside low sidewalk coverage, implying that cluttered or obstructed pedestrian paths can evoke unease. Plant showed a partially non-linear trend: its presence appeared to reduce fear up to approximately 20% coverage, particularly when combined with visible buildings, but beyond that threshold, the effect plateaued. For fence, a fear-reducing tendency emerged primarily when the proportion of sky was high, suggesting that transparent or low-height boundaries in open spaces may promote a sense of order and safety.

The relationship between building and sky coverage revealed more complex, non-linear dynamics. When building coverage was below roughly 20%, incremental increases in building proportion were generally associated with lower fear—especially when a large portion of the sky was visible—indicating that moderate urban density with open views may be perceived as safer. However, as building coverage exceeded this threshold, the relationship became less stable. Interestingly, sky alone did not exhibit a clear directional influence; rather, its interaction with buildings proved decisive. In areas with low sky visibility but high building density, fear tended to decrease, possibly due to a sense of familiarity or enclosure common in well-populated neighborhoods. Conversely, in open areas characterized by high sky visibility and low building density, perceived fear was also lower, likely due to improved sightlines and unobstructed spatial awareness.

Overall, these findings highlight the non-linear and context-dependent nature of environmental perception. Both dense yet structured urban settings and open, visually connected spaces can mitigate perceived fear—albeit through distinct psychological and spatial mechanisms. This emphasizes the importance of context-aware design strategies that balance enclosure, visibility, and accessibility when addressing fear of crime in urban environments.

4.4. Contrasting Contexts: Top and Bottom Deciles with Visual Evidence

To further explore contextual ambiguities in how built-environment elements shape fear perception, we conducted a decile-based comparative analysis. Specifically, the top 10% (low-fear) and bottom 10% (high-fear) areas were selected according to their predicted fear-of-crime scores to examine contrasting feature patterns. Figure 13a,b present the corresponding SHAP value distributions for both groups, highlighting the relative contribution of each streetscape feature. Across both deciles, road, sidewalk, car, and person consistently exhibited negative SHAP values for fear (i.e., safety-enhancing effects), whereas wall, earth, pole, awning, and ashcan were associated with elevated fear when their proportions increased. These findings reaffirm the dominant role of openness, accessibility, and human activity in mitigating perceived fear.

For building, no consistent pattern was observed in low-fear areas; however, in high-fear areas, lower building coverage corresponded to a marked rise in fear levels. This pattern likely reflects the enclosed spatial morphology of narrow residential blocks, where densely packed buildings create limited visibility and confinement—conditions visually evident in Figure 14b. Similarly, sky coverage showed no clear trend in low-fear zones, but in high-fear zones, greater sky visibility was linked to lower perceived fear, suggesting that openness and unobstructed sightlines mitigate anxiety even in otherwise vulnerable areas.

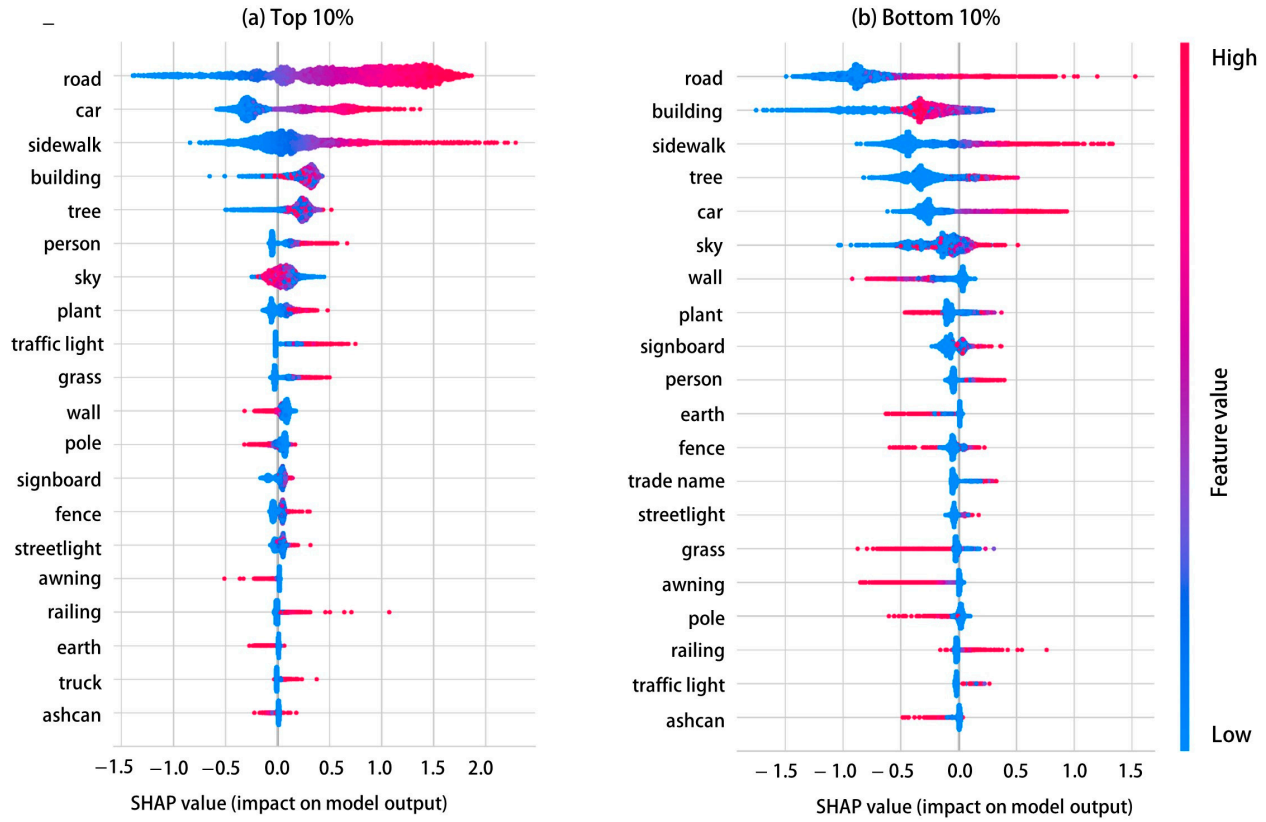


Figure 13. SHAP summary plots of top 10% data and bottom 10% data and their sampled SVIs.

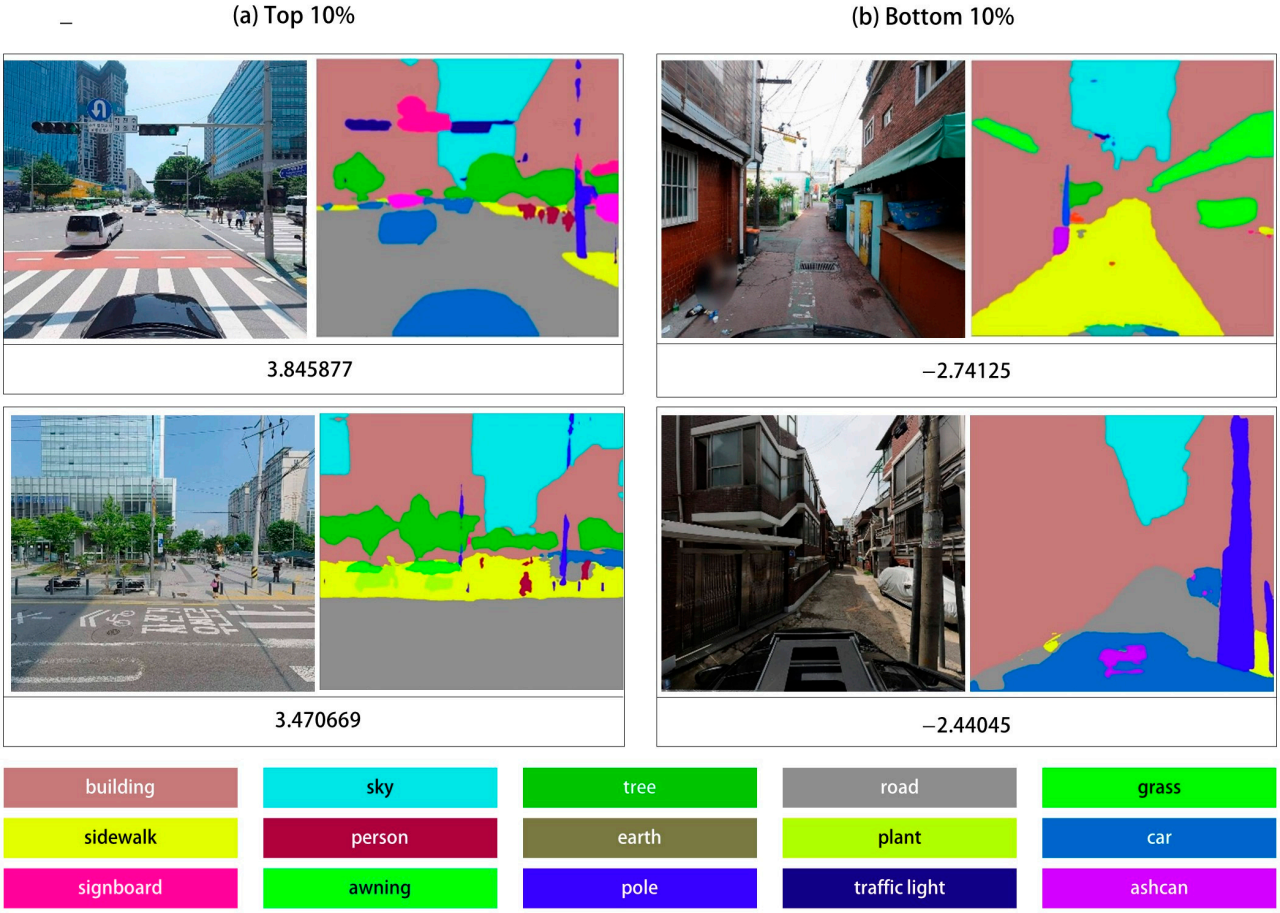


Figure 14. Examples of semantic segmentation of top 10% data and bottom 10% data.

Interestingly, fence, grass, and plant features displayed reversed effects between low- and high-fear contexts. In low-fear areas, a higher presence of these features corresponded to reduced fear, reinforcing their role in promoting order, aesthetic quality, and environmental comfort. Conversely, in high-fear areas, greater proportions of the same features correlated with increased fear perception. A closer inspection of imagery indicates that grass and plant elements in high-fear areas often appeared in riverside parks, vacant lots, or semi-isolated spaces away from dense pedestrian flows, as well as in the form of potted plants or climbing vegetation along aged residential and commercial façades. When situated in such poorly monitored or dimly lit settings, these features may inadvertently signal neglect or isolation, amplifying fear rather than alleviating it. The fence class, as defined in the ADE20K-based segmentation, encompasses a broad range of structures including pedestrian barriers, road dividers, railings, and metal grates. These objects exhibited context-dependent meanings across the fear spectrum. In low-fear environments, fences generally served as pedestrian-friendly boundaries or traffic separators, contributing to perceived order and safety. In contrast, fences in high-fear areas were often linked to construction sites, vacant properties, or residential security installations—such as barred windows or wire mesh—that visually communicate restriction or socio-spatial tension. Such representations may evoke feelings of danger or abandonment, thereby reinforcing fear rather than diminishing it.

Collectively, these results underscore that identical urban features can convey opposite perceptual signals depending on their spatial and social context. Thus, understanding the dual semantics of built-environment elements is critical for developing place-sensitive CPTED strategies and context-aware visual analytics that account for both physical form and situational meaning in urban fear mapping.

5. Discussion

5.1. Mapping Fear of Crime

The spatial distribution of perceived fear in Yeongdeungpo-gu, derived through the proposed GeoAI-based framework, reveals clear morphological and contextual patterns that align with previous studies on environmental criminology and urban perception [14,23,27]. Physically or socially disordered areas, including deteriorated residential blocks and poorly lit alleyways, were associated with higher fear levels. By applying a GeoAI-based analytical framework, this study extends prior findings by situating fear perception within the fine-grained spatial and morphological structure of a dense East Asian urban environment.

Traditional approaches to measuring fear of crime such as cross-sectional surveys and ecological momentary assessment (EMA) offer valuable subjective information but have clear limitations. Surveys cannot capture fine spatial detail, and EMA requires continuous participation and is difficult to apply at a citywide scale. Our SVI based GeoAI method falls between these two approaches. It depends on daytime images that are updated periodically, but it allows for consistent and detailed mapping of fear perception across large areas with minimal cost. This highlights both the strengths of GeoAI such as scalability, spatial granularity, and reproducibility, and its limitations related to image recency and lack of time-of-day variation.

Spatial visualization results showed that high-fear zones were concentrated in the southern part of the district, characterized by aging housing stock and irregular street networks, whereas low-fear zones were mainly distributed around Yeouido and major arterial corridors with wide visibility and active pedestrian flows. This pattern empirically supports the principles of CPTED and DST [45], emphasizing the roles of natural surveillance, territorial reinforcement, and spatial legibility in mitigating fear.

The intra-zonal heterogeneity captured by the model demonstrates the importance of local spatial morphology in shaping perceptual outcomes. Fear perception was not uniformly distributed, even within short spatial intervals, but dynamically influenced by variations in enclosure, visibility, and accessibility. These findings underscore the potential of fine-resolution GeoAI models in detecting subtle perceptual gradients that are often overlooked in traditional survey-based analyses.

Cultural and social context further mediates how residents interpret urban form. In Korean cities, collective vigilance within apartment complexes and sensitivity to secluded or poorly monitored spaces contribute to the cognitive construction of safety. The proposed framework integrates these socio-spatial dynamics into a quantifiable spatial layer of fear, bridging subjective perception and objective urban form. In summary, this GeoAI-based spatial analysis advances understanding of how visual, morphological, and cultural factors jointly shape perceived fear, providing actionable insights for evidence-based and context-sensitive urban safety planning.

5.2. Contextualizing the Effects of Built Environment Features on Perceived Fear of Crime

The findings reaffirm existing evidence on the environmental determinants of perceived fear while contributing new insights into how spatial context moderates these effects. Features such as roads, sidewalks, and parked vehicles consistently acted as fear-reducing elements, reinforcing their role in promoting natural surveillance, walkability, and urban order—core principles of CPTED. Conversely, features such as walls, poles, and awnings, which often signify spatial enclosure or physical disorder, were associated with higher fear levels, in line with the propositions of the BWT.

However, SHAP-based model interpretation revealed that several features, including sky, vegetation, fences, and buildings, exhibit nonlinear and context-dependent effects. Visible sky reduced fear primarily in dense and enclosed settings, indicating that its influence depends on spatial configuration rather than openness alone. Similarly, fences demonstrated opposite effects depending on the surrounding environment—acting as protective boundaries in well-maintained areas but as visual signals of neglect when associated with temporary or deteriorating structures.

These results show that the perceptual meaning of urban features is not fixed but evolves depending on their form, co-occurrence, and surrounding context. Simplified assumptions such as “green equals safe” or “openness equals unsafe” fail to capture these nuances. Vegetation, for example, can enhance visual comfort in managed spaces but increase fear in overgrown or poorly lit conditions. Overall, the findings highlight that fear of crime emerges from the dynamic interaction between visual semantics, spatial configuration, and sociocultural interpretation. For urban planners, this emphasizes the need for locally calibrated CPTED strategies that reflect how built-environment cues are differently perceived across spatial and cultural contexts.

5.3. Policy Implications for Crime-Fear Mapping and CPTED Planning

Perceived fear of crime represents an essential dimension of urban safety planning and is increasingly recognized as a critical diagnostic layer in CPTED. In South Korea, local governments are encouraged to produce crime safety maps that identify locations where residents feel vulnerable. These maps are used to prioritize interventions such as improved lighting, enhanced visibility, and maintenance of public spaces. Conventional approaches relying on field audits and resident surveys, however, remain time-consuming and spatially limited.

The methodology proposed in this study provides a scalable and data-driven alternative. By combining pairwise perception data with Transformer-based deep learning and

explainable AI techniques, the proposed GeoAI framework enables fine-scale mapping of perceived fear directly from SVI. Once SVI data are available, planners can generate high-resolution fear maps that reflect visual perceptions of crime-related risk with minimal additional cost. This approach facilitates efficient and timely decision-making, supporting local governments in allocating urban safety resources based on data-driven, perception-aware evidence.

By focusing explicitly on crime-related fear rather than general safety, this framework offers actionable insights into how the built environment influences public anxiety. It strengthens the diagnostic capacity of CPTED planning by identifying not only where fear is concentrated but also which environmental features contribute to that perception. Such evidence-based mapping enhances transparency and enables targeted interventions aligned with the visual and emotional experiences of residents.

Beyond producing high resolution fear maps, the proposed framework can directly inform concrete policy actions. Local governments could use the identified fear prone street segments to prioritize street lighting upgrades, sidewalk and crossing improvements, façade maintenance, activation of underused public spaces, and the redesign of alleys or blind spots that reduce natural surveillance. Future extensions of this work could incorporate additional datasets such as police incident records, land use and zoning information, pedestrian and traffic flow data, socio economic indicators, and the locations of CCTV cameras and public facilities. Integrating SVI based fear scores with these layers would support more comprehensive vulnerability assessments and help planners design place specific CPTED strategies that respond both to perceived risk and to objective environmental and social conditions. In addition, structured fieldwork-based safety inventories can be used to ground-truth and validate the model outputs at selected locations, so that the proposed GeoAI framework complements rather than replaces on-site assessments.

5.4. Limitations and Future Research

While this study proposes a novel framework for predicting and mapping perceived fear of crime using SVI and GeoAI, several limitations should be noted. First, all analyses were based on daytime SVI, which may not capture the full range of temporal variation in fear perception. As fear often intensifies after dark, future studies should incorporate nighttime imagery or employ generative models such as GAN based day to night translation to analyze nocturnal safety perceptions. In particular, subsequent research should explicitly consider nighttime environmental factors such as streetlight illumination, nighttime visibility and visual range, and the presence or absence of pedestrians after dark, as the relationships between streetscape features and fear of crime may change or even reverse under low light conditions.

Second, the segmentation model (SegFormer B5 trained on ADE20K) may not fully capture region specific urban characteristics, such as localized signage, informal fencing, or cultural design features common in Korean cities. Future work could enhance model accuracy by fine tuning segmentation models on custom datasets that better represent regional urban textures and safety related visual cues.

Third, the current study focuses solely on perceptual fear without integrating other spatial indicators such as crime incidents, pedestrian flows, or land use patterns. Expanding the framework into a multi-layered urban safety model would improve its practical applicability and enable more comprehensive risk assessments. Integrating fear scores with other spatial datasets could support the development of composite vulnerability indices that inform holistic and proactive safety interventions.

Fourth, while the pairwise comparison experiment yielded a large number of image level labels, the pool of respondents was relatively small (N = 65) and drawn from a limited

demographic and cultural context. Consequently, the generalizability of the findings to other populations or cities may be constrained. Future research would benefit from recruiting larger and more diverse samples to validate and extend the proposed framework and to examine whether the identified relationships between streetscape features and perceived fear hold across different socio demographic groups.

Finally, we did not collect information about how familiar each respondent was with the locations shown in the SVIs. Although participants were instructed to base their judgments solely on visual cues, prior research suggests that place familiarity, routine activity patterns, and personal experiences with an area can attenuate or amplify fear perceptions. These familiarity effects are also likely to interact with time of day, as residents may feel more comfortable in well-known streets after dark but more anxious in unfamiliar areas. Because SVI-based models rely on static daytime images, they may not fully reflect these familiarity- or time-dependent perceptual processes. Future studies should combine SVI-based tasks with data on participants' activity spaces, home locations, or in situ EMA-style reports to more explicitly examine how familiarity and temporal context jointly shape perceived fear of crime.

6. Conclusions

This study proposed a GeoAI-based framework for modeling and explaining perceived fear of crime at the street level using street view imagery and explainable deep learning. By integrating a Swin Transformer–based Siamese-RankNet model with semantic segmentation and SHAP-based interpretation, the framework provides both high predictive accuracy and transparent insights into how built-environment features influence fear perception.

Empirical results from Yeongdeungpo-gu, Seoul, demonstrate that the framework effectively captures spatial heterogeneity in perceived fear, revealing how open and accessible elements such as roads and sidewalks mitigate fear, whereas enclosed or unmanaged features like walls and poles intensify it, depending on spatial context. The explainable AI component enhances interpretability, allowing planners and researchers to better understand the mechanisms underlying spatial variations in fear perception.

Methodologically, this study advances GeoAI research by linking computer vision, spatial analytics, and human-centered perception modeling within a single interpretable framework. Unlike traditional approaches relying on surveys or CNN-based classifiers, the proposed model leverages pairwise human judgments to generate scalable, image-level predictions of perceived fear. For spatial information science, it demonstrates how subjective visual experiences can be quantified and analyzed spatially; for urban planning, it provides a practical diagnostic tool to identify fear-prone microspaces and inform targeted CPTED interventions.

In summary, this research contributes to the emerging domain of perception-based GeoAI by bridging human spatial cognition and data-driven modeling. The proposed framework offers a replicable and scalable pathway for understanding how visual environments shape emotional responses in cities, ultimately supporting the design of safer, more perceptually inclusive urban environments.

Author Contributions: Conceptualization, Youngok Kang; Methodology, Software, Validation, Formal Analysis, Somang Kim and Jaeyeon Choi; Data Curation, Somang Kim and Jaeyeon Choi; Writing—Original Draft Preparation, Somang Kim and Jaeyeon Choi; Writing—Review and Editing, Youngok Kang; Visualization, Somang Kim and Jaeyeon Choi; Supervision, Youngok Kang; Project Administration, Youngok Kang; Funding Acquisition, Youngok Kang. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (Grant 2023R1A2C1005819) and Korea Agency for Infrastructure Technology Advancement (KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (Grant RS-2022-00143782).

Data Availability Statement: Data can be provided upon request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Ferraro, K.F.; LaGrange, R.L. The measurement of fear of crime. *Sociol. Inq.* **1987**, *57*, 70–101. [CrossRef]
2. Hale, C. Fear of crime: A review of the literature. *Int. Rev. Vict.* **1996**, *4*, 79–150. [CrossRef]
3. Rader, N.E.; May, D.C.; Goodrum, S. An empirical assessment of the threat of victimization: Considering fear of crime, perceived risk, avoidance, and defensive behaviors. *Sociol. Spectr.* **2007**, *27*, 475–505. [CrossRef]
4. Wyant, B.R. Multilevel impacts of perceived incivilities and perceptions of crime risk on fear of crime: Isolating endogenous impacts. *J. Res. Crime Delinq.* **2008**, *45*, 39–64. [CrossRef]
5. Anderson, J.M.; MacDonald, J.M.; Bluthenthal, R.; Ashwood, J.S. Reducing crime by shaping the built environment with zoning: An empirical study of Los Angeles. *Univ. Penn. Law Rev.* **2012**, *161*, 699–757. [CrossRef]
6. De Nadai, M.; Vieriu, R.L.; Zen, G.; Dragicevic, S.; Naik, N.; Caraviello, M.; Lepri, B. Are safer-looking neighborhoods more lively? A multimodal investigation into urban life. In Proceedings of the 24th ACM International Conference on Multimedia, Amsterdam, The Netherlands, 15–19 October 2016; pp. 1127–1135. [CrossRef]
7. Zhou, H.; Liu, L.; Lan, M.; Zhu, W.; Song, G.; Jing, F.; Zhong, Y.; Su, Z.; Gu, X. Using Google Street View imagery to capture micro built environment characteristics in drug places, compared with street robbery. *Comput. Environ. Urban Syst.* **2021**, *88*, 101631. [CrossRef]
8. Wilson, J.Q.; Kelling, G.L. Broken windows. *Atl. Mon.* **1982**, *249*, 29–38.
9. He, L.; Paez, A.; Liu, D. Built environment and violent crime: An environmental audit approach using Google Street View. *Comput. Environ. Urban Syst.* **2017**, *66*, 83–95. [CrossRef]
10. Lee, I.; Jung, S.; Lee, J.; Macdonald, E. Street crime prediction model based on the physical characteristics of a streetscape: Analysis of streets in low-rise housing areas in South Korea. *Environ. Plan. B Urban Anal. City Sci.* **2019**, *46*, 862–879. [CrossRef]
11. Jeffery, C.R. *Crime Prevention Through Environmental Design*; Sage Publications: Beverly Hills, CA, USA, 1971.
12. Abraham, J.; Kang, Y.; Ceccato, V.; Näsman, P.; Duarte, F.; Gao, S.; Ljungqvist, L.; Zhang, F.; Ratti, C. Crime and visually perceived safety of the built environment: A deep learning approach. *Ann. Am. Assoc. Geogr.* **2025**, *115*, 1613–1633. [CrossRef]
13. Bilen, Ö.; Aşkın, Ö.E.; Büyüklü, A.H.; Ökten, A.; Gür, M. How the fear of crime spatially differs among the districts of Istanbul. *Soc. Sci.* **2013**, *8*, 153–164. [CrossRef]
14. Doran, B.J.; Lees, B.G. Investigating the spatiotemporal links between disorder, crime, and the fear of crime. *Prof. Geogr.* **2005**, *57*, 1–12. [CrossRef]
15. Nasar, J.L.; Fisher, B.S. 'Hot spots' of fear and crime: A multi-method investigation. *J. Environ. Psychol.* **1993**, *13*, 187–206. [CrossRef]
16. Zhang, F.; Zhou, B.; Liu, L.; Liu, Y.; Fung, H.H.; Lin, H.; Ratti, C. Measuring human perceptions of a large-scale urban region using machine learning. *Landsc. Urban Plan.* **2018**, *180*, 148–160. [CrossRef]
17. Jing, F.; Liu, L.; Zhou, S.; Song, J.; Wang, L.; Zhou, H.; Wang, Y.; Ma, R. Assessing the impact of street-view greenery on fear of neighborhood crime in Guangzhou, China. *Int. J. Environ. Res. Public Health* **2021**, *18*, 311. [CrossRef] [PubMed]
18. Ramírez, T.; Hurtubia, R.; Lobel, H.; Rossetti, T. Measuring heterogeneous perception of urban space with massive data and machine learning: An application to safety. *Landsc. Urban Plan.* **2021**, *208*, 104002. [CrossRef]
19. Wang, Y.; Zeng, Z.; Li, Q.; Deng, Y. A complete reinforcement-learning-based framework for urban-safety perception. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 465. [CrossRef]
20. Kang, Y.; Abraham, J.; Ceccato, V.; Duarte, F.; Gao, S.; Ljungqvist, L.; Zhang, F.; Näsman, P.; Ratti, C. Assessing differences in safety perceptions using GeoAI and survey across neighbourhoods in Stockholm, Sweden. *Landsc. Urban Plan.* **2023**, *236*, 104768. [CrossRef]
21. Hou, X.; Chen, P. Analysis of road safety perception and influencing factors in a complex urban environment—Taking Chaoyang District, Beijing, as an example. *ISPRS Int. J. Geo-Inf.* **2024**, *13*, 272. [CrossRef]
22. Pain, R. Place, social relations and the fear of crime: A review. *Prog. Hum. Geogr.* **2000**, *24*, 365–387. [CrossRef]
23. Kohm, S.A. Spatial dimensions of fear in a high-crime community: Fear of crime or fear of disorder? *Can. J. Criminol. Crim. Justice* **2009**, *51*, 1–30. [CrossRef]
24. Fisher, B.S.; Nasar, J.L. Fear of crime in relation to three exterior site features: Prospect, refuge, and escape. *Environ. Behav.* **1992**, *24*, 35–65. [CrossRef]

25. Vrij, A.; Winkel, F.W. Characteristics of the built environment and fear of crime: A research note on interventions in unsafe locations. *Deviant Behav.* **1990**, *11*, 363–375. [CrossRef]
26. Chataway, M.L.; Hart, T.C.; Coomber, R.; Bond, C. The geography of crime fear: A pilot study exploring event-based perceptions of risk using mobile technology. *Appl. Geogr.* **2017**, *86*, 300–307. [CrossRef]
27. Solymosi, R.; Bowers, K.J.; Fujiyama, T. Mapping fear of crime as a context-dependent everyday experience that varies in space and time. *Leg. Criminol. Psychol.* **2015**, *20*, 193–211. [CrossRef]
28. Jakobi, Á.; Pődör, A. GIS-based statistical analysis of detecting fear of crime with digital sketch maps: A Hungarian multicity study. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 229. [CrossRef]
29. Ogneva-Himmelberger, Y.; Ross, L.; Caywood, T.; Khananayev, M.; Starr, C. Analyzing the relationship between perception of safety and reported crime in an urban neighborhood using GIS and sketch maps. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 531. [CrossRef]
30. Pánek, J.; Ivan, I.; Macková, L. Comparing residents' fear of crime with recorded crime data: Case study of Ostrava, Czech Republic. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 401. [CrossRef]
31. Biljecki, F.; Ito, K. Street view imagery in urban analytics and GIS: A review. *Landsc. Urban Plan.* **2021**, *215*, 104217. [CrossRef]
32. Kang, Y.; Zhang, F.; Gao, S.; Lin, H.; Liu, Y. A review of urban physical environment sensing using street view imagery in public health studies. *Ann. GIS* **2020**, *26*, 261–275. [CrossRef]
33. Kang, Y. GeoAI application areas and research trends. *J. Korean Geogr. Soc.* **2023**, *58*, 395–418.
34. Kang, Y. Human-centered geospatial data science. In *The International Encyclopedia of Geography: People, the Earth, Environment, and Technology*; Wiley-Blackwell: Hoboken, NJ, USA; American Association of Geographers: Washington, DC, USA, 2025. [CrossRef]
35. Salesses, P.; Schechtner, K.; Hidalgo, C.A. The collaborative image of the city: Mapping the inequality of urban perception. *PLoS ONE* **2013**, *8*, e68400. [CrossRef] [PubMed]
36. Dubey, A.; Naik, N.; Parikh, D.; Raskar, R.; Hidalgo, C.A. Deep learning the city: Quantifying urban perception at a global scale. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; pp. 196–212. [CrossRef]
37. Blečić, I.; Cecchini, A.; Trunfio, G.A. Towards automatic assessment of perceived walkability. In Proceedings of the International Conference on Computational Science and Its Applications, Melbourne, Australia, 2–5 May 2018; Springer: Cham, Switzerland, 2018; pp. 351–365. [CrossRef]
38. Santani, D.; Ruiz-Correa, S.; Gatica-Perez, D. Looking south: Learning urban perception in developing cities. *ACM Trans. Soc. Comput.* **2018**, *1*, 13. [CrossRef]
39. Bijmolt, T.H.; Wedel, M. The effects of alternative methods of collecting similarity data for multidimensional scaling. *Int. J. Res. Mark.* **1995**, *12*, 363–371. [CrossRef]
40. Stewart, N.; Brown, G.D.; Chater, N. Absolute identification by relative judgment. *Psychol. Rev.* **2005**, *112*, 881–911. [CrossRef]
41. Min, W.; Mei, S.; Liu, L.; Wang, Y.; Jinag, S. Multi-task deep relative attribute learning for visual urban perception. *IEEE Trans. Image Process.* **2020**, *29*, 657–669. [CrossRef]
42. Xu, Y.; Yang, Q.; Cui, C.; Shi, C.; Song, G.; Han, X.; Yin, Y. Visual urban perception with deep semantic-aware network. In Proceedings of the International Conference on Multimedia Modeling, Thessaloniki, Greece, 8–11 January 2019; Springer: Cham, Switzerland, 2019; pp. 28–40. [CrossRef]
43. Guan, W.; Chen, Z.; Feng, F.; Liu, W.; Nie, L. Urban perception: Sensing cities via a deep interactive multi-task learning framework. *ACM Trans. Multimed. Comput. Commun. Appl.* **2021**, *17*, 13. [CrossRef]
44. Kang, Y.; Kim, J.; Park, J.; Lee, J. Assessment of perceived and physical walkability using street view images and deep learning technology. *ISPRS Int. J. Geo-Inf.* **2023**, *12*, 186. [CrossRef]
45. Newman, O. *Defensible Space: Crime Prevention Through Urban Design*; Collier Books: New York, NY, USA, 1973.
46. Ye, X.; Wang, Y.; Dai, J.; Qiu, W. Generated nighttime street view image to inform perceived safety divergence between day and night in high density cities: A case study in Hong Kong. *J. Urban Manag.* **2024**, *13*, 11. [CrossRef]
47. Yoo, G.; Lee, D.; Lee, C.; Nam, K. Generating pairwise comparison set for crowd-sourcing based deep learning. *J. Korea Ind. Inf. Syst. Res.* **2022**, *27*, 1–11. [CrossRef]
48. Koch, G.; Zemel, R.; Salakhutdinov, R. Siamese neural networks for one-shot image recognition. In Proceedings of the 32nd International Conference on Machine Learning, Lille, France, 6–11 July 2015; pp. 1–10.
49. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin Transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, Canada, 28 February 2022; IEEE: Piscataway, NJ, USA, 2021; pp. 9992–10002. [CrossRef]
50. Burges, C.J.; Shaked, T.; Renshaw, E.; Lazier, A.; Deeds, M.; Hamilton, N.; Hullender, G.N. Learning to rank using gradient descent. In Proceedings of the 22nd International Conference on Machine Learning, Bonn, Germany, 7–11 August 2005; ACM: New York, NY, USA, 2005. [CrossRef]
51. Xie, E.; Wang, W.; Yu, Z.; Anandkumar, A.; Alvarez, J.M.; Luo, P. SegFormer: Simple and efficient design for semantic segmentation with transformers. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 12077–12090.

52. Zhou, B.; Zhao, H.; Puig, X.; Fidler, S.; Barriuso, A.; Torralba, A. Scene parsing through ADE20K dataset. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 5122–5130. [CrossRef]
53. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale. In Proceedings of the International Conference on Learning Representations (ICLR 2021), Vienna, Austria, 4 May 2021.
54. Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; Gulin, A. CatBoost: Unbiased boosting with categorical features. In Proceedings of the Advances in Neural Information Processing Systems 31 (NeurIPS 2018), Montreal, Canada, 3–8 December 2018; Volume 31, pp. 1–11. [CrossRef]
55. Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. In Proceedings of the Advances in Neural Information Processing Systems 30 (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017; Volume 30, pp. 4765–4774. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.