

Article

Analysis of Road Safety Perception and Influencing Factors in a Complex Urban Environment—Taking Chaoyang District, Beijing, as an Example

Xinyu Hou and Peng Chen *

School of Information and Cyber Security, People's Public Security University of China, Beijing 100038, China; 2022211511@stu.ppsuc.edu.cn

* Correspondence: chenpeng@ppsuc.edu.cn

Abstract: Measuring human perception of environmental safety and quantifying the street view elements that affect human perception of environmental safety are of great significance for improving the urban environment and residents' safety perception. However, domestic large-scale quantitative research on the safety perception of Chinese local cities needs to be deepened. Therefore, this paper chooses Chaoyang District in Beijing as the research area. Firstly, the network safety perception distribution of Chaoyang District is calculated and presented through the CNN model trained based on the perception dataset constructed by Chinese local cities. Then, the street view elements are extracted from the street view images using image semantic segmentation and target detection technology. Finally, the street view elements that affect the road safety perception are identified and analyzed based on LightGBM and SHAP interpretation framework. The results show the following: (1) the overall safety perception level of Chaoyang District in Beijing is high; (2) the number of motor vehicles and the proportion of the area of roads, skies, and sidewalks are the four factors that have the greatest impact on environmental safety perception; (3) there is an interaction between different street view elements on safety perception, and the proportion and number of street view elements have interaction on safety perception; (4) in the sections with the lowest, moderate, and highest levels of safety perception, the influence of street view elements on safety perception is inconsistent. Finally, this paper summarizes the results and points out the shortcomings of the research.

Keywords: street view images; safety perception; semantic segmentation; object detection; LightGBM; SHAP



Citation: 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. <https://doi.org/10.3390/ijgi13080272>

Academic Editors: Wolfgang Kainz, Jiangfeng She, Min Yang and Jun Zhu

Received: 3 June 2024

Revised: 23 July 2024

Accepted: 30 July 2024

Published: 31 July 2024



Copyright: © 2024 by the authors. 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

For decades, with the rapid development of urbanization, different places in the city have been reshaped, and development in location, physical environment, and other aspects has become gradually unbalanced [1]. The original natural environment has been changed and replaced by a large number of artificial environments to varying degrees, such as roads, buildings, urban greening, etc. [2]. Environmental problems seriously affect people's psychological perception and behavior, and environmental factors also play an important role in shaping safety perception. Therefore, it is of great significance to objectively evaluate the urban environment, which can not only effectively adjust the urban functions and improve the urban environment but also improve the residents' safety perception level. Improving the level of safety perception is one of the prerequisites to enhance the vitality of public space and the effectiveness of social governance. It is essential to improve the impact effect and impact mechanism of environmental safety perception from the perspective of space users. As the skeleton of the city, urban roads are an important carrier of people's daily activities [3], and it is of great significance to carry out relevant research on roads.

Human perception refers to people's psychological feelings about a place, which reflect people's overall impression of a place [4,5]. As a level of perception, safety perception

first appeared in the theoretical research of Freud's psychoanalysis, which refers to the individual's prediction of potential physical or psychological hazards [6]. The Crime Prevention through Environmental Design (CPTED) theory summarizes the core elements that affect safety, such as territoriality, surveillance, access control, and image maintenance [7–9]. Different from actual safety, safety perception is a complex psychological feeling [10,11] which is not only related to the material environment [12] but also affected by many factors such as the social environment and individual characteristics [13,14]. Based on the above point of view, Li et al. [15] defined environmental safety perception as the degree of fear, anxiety, and other negative emotions caused by the surrounding physical environment when pedestrians are in the urban road space. Safety perception is not equal to the fear of crime; instead, it is usually related to the physical environment and the psychological feeling of people in the scene.

Traditional perceptual datasets are collected by social scientists through field surveys [16,17]. This method is time-consuming, high-cost and small in research scope. Street view objectively depicts the real urban landscape and was proved to be an effective and reliable data to measure the urban environment [18–20]. In recent years, more and more attention has been paid to perceptual prediction based on street view images. Salesses, Schechtner, and Hidalgo found that street view images can be used to assess the social and economic impact of the urban environment [21]. He et al. used GSV (Google Street View) to collect fine-scale quantitative data of the built environment and conducted research on the relationship between the built environment and violent crime [22]. Hwang et al. introduced a method of systematic social observation using GSV to detect visible cues of neighborhood change, and then conducted research on the uneven evolution of gentrification in cities [23]. Naik et al. [24] proposed a scene perception prediction method based on support vector regression. However, this method has the problem of relying on predefined feature mapping. With the popularity of deep learning, most scholars have begun to introduce it into the study of urban perception [25–27]. Porzi et al. [28] proved that deep learning is superior to traditional feature description in predicting human perception. Naik, Raskar, and Hidalgo [29] and Zhang et al. [1,19,20] used computer vision to simulate individuals to quantify urban perception scores.

At present, the dataset used in urban perception is mainly the place pulse dataset provided by MIT [4,28]. The city perception dataset collects street view images of many cities around the world but lacks training samples of cities in mainland China. There are obvious differences between Eastern and Western architectural styles and urban planning [30]. The city perception model derived from the global dataset may not be applicable to cities in mainland China. Moreover, the place pulse dataset provided by MIT contains streets of countries with different levels of development, and the model trained with this dataset may not be able to effectively identify the subtle differences in the safety perception of a specific city [31]. In addition, urban perception is a subjective assessment, which is influenced by people's social and cultural backgrounds. Although Dubey et al. [25] and Salesses et al. [21] claimed that the demographic statistics of survey respondents would not cause any deviation, their statement was only true in the sample area or cities similar to the sample area. When the dataset is directly used to predict and evaluate the perception of cities in other places, the deviation may still exist. Therefore, Yao et al. [12] proposed establishing a unique urban perception dataset in China. They constructed a human–computer confrontation scoring framework based on in-depth learning to support the perception research of cities and regions in mainland China and explained the results from the visual and urban functions. However, Yao's method is to use FCN (full revolutionary networks) [32] to extract features from street view images by semantic segmentation and then calculate safety perception based on the extracted features. Wang et al. [33] compared the method of Yao et al. with the model based on CNN; the results showed that the model based on CNN was more accurate and had higher correlation with urban functions. However, the framework of this “black box” nature lacks explicability [34]. Most relevant studies have confused the relationship between street view elements and urban perception, thus reducing their

effectiveness in urban planning practice requiring effective evidence [35]. It may be difficult for decision makers to answer questions such as how to improve the effectiveness of urban planning practices and how to assess the urban perception of specific streets through urban environment and infrastructure. At present, scholars have conducted interpretable analysis on urban perception. Xu et al. used Baidu Street View images in Shanghai to explore the influencing factors and degree of street safety perception evaluation and found that the degree of street management, green vision rate, and number of lanes all have a significant impact on safety perception, with the green vision rate and management degree having the greatest influence [36]. Fu et al. analyzed the composition and proportion of elements in street view images, studied the impact of different landscape designs and combinations on safety perception, and found that environmental design elements such as paving, people, outdoor facilities, and landscape architecture have a positive impact on safety perception [37]. Liu et al. [38] analyzed the influencing factors of urban perception based on street view images of the central urban area of Wuhan, and found that buildings, sky, and green space are the three types of land features that have the greatest impact on human perception. Li et al. [15] constructed a safety perception dataset based on the gathering places of universities in Lanzhou, and explained safety perception. They found that the proportions of images of the sky, sidewalks, roads, and trees were the four elements that had the greatest impact on environmental safety perception.

From the above research status, it can be seen that Chinese scholars at present have gradually carried out research based on the perception dataset of Chinese local cities, but there are some problems, such as the lack of in-depth research on the impact mechanism of street view safety perception, the lack of comprehensive analysis on the impact variables, and the relatively single research scenario, all of which lead to limitations in the research results. Therefore, it is of more universal significance to study and analyze the complex urban environment with larger scope and richer functions. Therefore, in view of the shortcomings of the current research, this paper plans to choose Chaoyang District in Beijing, which has a complex urban environment, as the research area and use the CNN model trained by the urban perception dataset constructed by Yao et al. to obtain the safety perception distribution of Chaoyang District. Then, a safety perception model was constructed using LightGBM and SHAP to conduct in-depth analysis of the distribution of safety perception, influencing factors, interaction of street view elements with safety perception, and the local environmental characteristics of urban areas. This paper helps to reveal the distribution of safety perception in complex environments and the corresponding influencing factors so as to provide guidance and suggestions for urban planning and residents' safety perception improvement.

2. Materials and Methods

2.1. Study Area and Data Source

2.1.1. Study Area

The study area is the Chaoyang District of Beijing (except the capital airport). Beijing is the political, economic, and cultural center of China and one of the regions with the strongest population flow in the country. Chaoyang District is located in the east of Beijing. It is the largest and most populous district in the main urban area of Beijing. It is an economically and culturally strong district in Beijing, with rich transportation services, living services, and other facilities. With a total area of 470.8 square kilometers, Chaoyang District governs 24 streets and 19 regions. By the end of 2020, Chaoyang District had a permanent population of 3.4525 million. Chaoyang District has rich urban functions and complex urban environment, and the number of crimes accounts for more than one third of the total number of crimes in Beijing. Research on the safety perception of Chaoyang District is of great significance to Beijing and all cities in China. The research unit of this paper is based on the road network division. There are 10,749 road sections in Chaoyang District. After data cleaning and removing the duplicate road sections, 9654 road sections are finally determined as the research unit. Figure 1 is research area and road network structure.



Figure 1. Research area and road network structure. The white line is the boundary of each street, and the gray line is the road network.

2.1.2. Data Source

The road network data of Beijing was obtained through OSM. A total of 420,000 sample points were set every 50 m on the road section. The 2020 street view image data were obtained and cut through Google Street View API. Finally, 1.68 million street view images were obtained. Each sample point corresponds to four street view pictures, two parallel to the road and two perpendicular to the road (the angles are $0-90^\circ$, $90-180^\circ$, $180-270^\circ$, and $270-360^\circ$, respectively). Finally, we selected the sample points within Chaoyang District and the corresponding street view pictures, with a total of 40,000 sample points and 160,000 street view pictures. Figure 2 shows the azimuth of four street view pictures at the sample point, two of which are parallel to the street and two of which are perpendicular to the street. Figure 3 shows the process of obtaining Google Street View images.

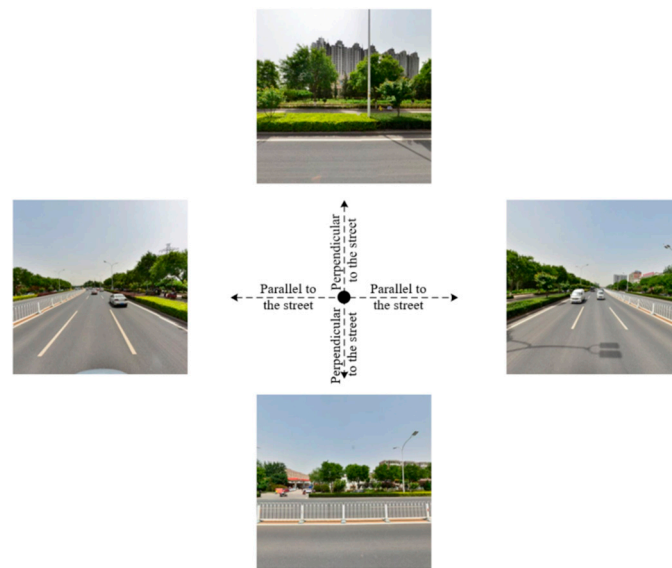


Figure 2. Azimuth angles of four street view images of sample points.

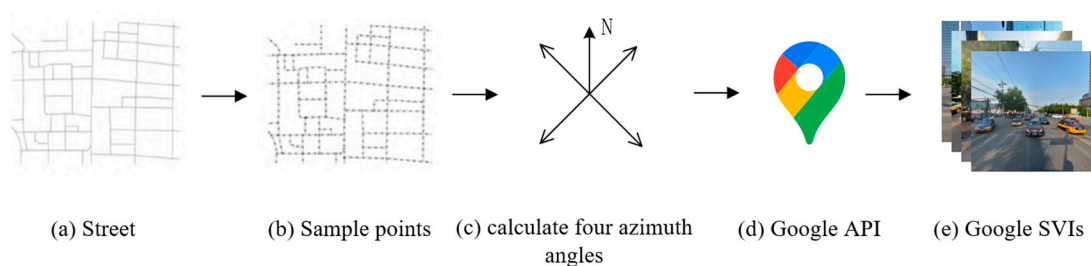


Figure 3. The process of obtaining Google Street View images.

2.2. Research Method

2.2.1. Research Framework

The main research ideas of this paper are as follows: firstly, the street view image is obtained from the sample points of the road network through Google map, and the CNN model based on the Chinese local city perception dataset is used to train the safety perception score of Beijing street view image data; then, LM-deeplabv3+ and yolov5 are used to extract the street view element indicators from the street view image, and the road safety perception dataset is constructed and combined with the safety perception score; secondly, the distribution of safety perception in the research area is analyzed; thirdly, the street view safety perception model is constructed based on the LightGBM algorithm, and SHAP is introduced to analyze the interpretability of environmental safety perception. The influencing factors of safety perception, the interaction of street view elements on safety perception, and local environmental characteristics are analyzed, respectively. Finally, this paper summarizes and puts forward the research deficiencies.

2.2.2. Safety Perception Score and Street View Element Extraction

(1) Safety perception score

Street view safety perception score is the dependent variable of this study. It mainly uses manual evaluation combined with deep learning to obtain large-scale safety perception score data. The selection of neural network is very important for the performance of the model. In order to extract sufficient image features, we selected the end-to-end CNN model constructed by Wang et al. [33] to train the street view safety perception score. The perception model refers to the structure of vgg16 [39], and its parameters and structure have been proven to be reliable in many aspects of image feature extraction [40–42]. More impor-

tantly, the model is trained based on the local dataset of China, which is more suitable for research on China than the place pulse 2.0 dataset of MIT. This local dataset included nearly 500 thousand street view images of major Chinese cities (Beijing, Shenzhen, Guangzhou, Shanghai, Wuhan, Hangzhou, etc.). To accurately produce safety perceptions in Chinese urban environments, local volunteers who are aware of the regional socioeconomic background are asked to work through the human–machine adversarial scoring framework. Volunteers score a displayed street-view image in terms of safety perception in a range of 1–100, with 0 being the lowest and 100 being the highest level of a perception. If an image is rated multiple times by several volunteers, the final score of the photograph will be set as the median value to avoid extreme scores. Figure 4 shows the processing flow of street view safety perception score.

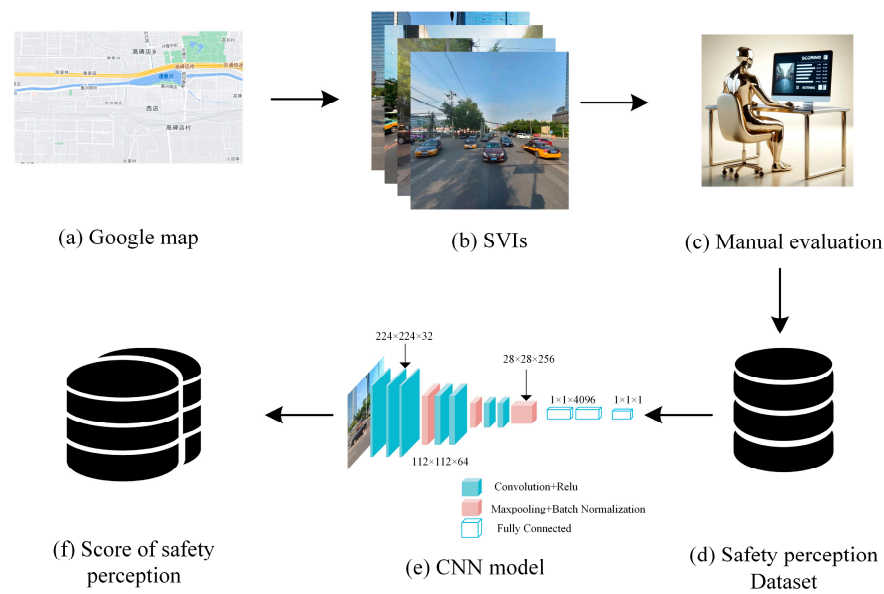


Figure 4. Process of safety perception scoring for street view images.

(2) Street view element extraction

Street view element is the explanatory variable of this study, which is extracted by image semantic segmentation and target detection. For street view elements (such as sky, roads, buildings, trees, etc.) with unclear boundaries and that are not easy to quantify, the proportion of the area of the elements in the street view image is extracted by image semantic segmentation only; for the street view elements with clear boundaries and that are easy to quantify (such as motor vehicles, pedestrians, bicycles, traffic lights, etc.), while using image semantic segmentation to extract the proportion of the area, target detection technology is used to extract the number of elements in the image.

We use image semantic segmentation to obtain the calculation formula of the proportion of various elements:

$$\text{Average pixel proportion of elements} = \frac{\sum_{i=1}^n \sum_{j=1}^4 \text{Pixel}_{\text{Image}_{ij}}(x)}{\sum_{i=1}^n \sum_{j=1}^4 \text{Pixel}_{\text{Image}_{ij}}} \quad (1)$$

where n represents that there are n sample points in a road section, $\text{Pixel}_{\text{Image}_{ij}}(x)$ represents the area of the street view element x in the j -th picture of the i -th sample point, and $\text{Pixel}_{\text{Image}_{ij}}$ represents the total area of the street view element in the j -th picture of the i -th sample point.

We use object detection to obtain the calculation formula for the number of various elements:

$$\text{Total number of elements} = \sum_{i=1}^n \sum_{j=1}^4 \text{Count}_{\text{Image}_{ij}(x)} \quad (2)$$

where $\text{Count}_{\text{Image}_{ij}(x)}$ indicates the number of street view element x in the j -th picture of the i -th sample point.

Image semantic segmentation is one of the important fields in computer vision technology. It takes the image as the processing object and classifies each point in the image; target detection mainly focuses on the objects that need to be recognized in the image and predicts the category and location of target objects by extracting features in the network. Considering the large amount of data in this paper, we need to take into account the effect and accuracy of segmentation, and we choose the LM-deeplabv3+ [43] lightweight model with mobilenetv2 [44] as the backbone network for street view image segmentation. The model has been pre-trained in the cityscapes dataset, which has a total of 5000 street view images, including 2975 training sets, 500 validation sets, and 1525 test sets. The average intersection union ratio (Miou) of the pre trained model reached 74.9%, and the average pixel accuracy (MPA) reached 83%, which was 1% lower than the average intersection union ratio of deeplabv3+ [45], while the average pixel accuracy was the same, but the parameter quantity was only one sixth of deeplabv3+. Yolov5 [46] target detection technology is selected to complete the training on the coco2017 dataset, and target detection is carried out on the street view image. This model has obvious advantages in processing time and accuracy. At present, it is widely used in vehicle detection, face detection, and many other fields [47,48]. The coco2017 dataset contains 80 categories of targets, including 118,287 training sets, 5000 validation sets and 40,670 test sets, with a mean average precision (map50) of 68.9%. Because yolov5 is a multi-objective classification model, the average detection accuracy is not high, but considering the actual elements in the road scene in this study, the number of motor vehicles, bicycles, pedestrians, and traffic lights is selected as the number index of street view elements. After testing, the number of each index can be correctly detected, so yolov5 can meet the needs of this study. Figure 5 shows the effect of extracting street view elements using LM-deeplabv3+ and yolov5.

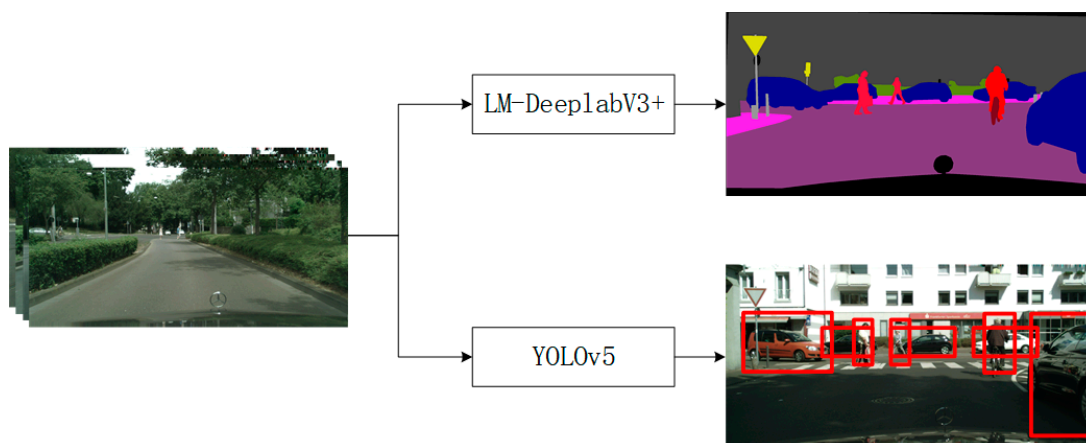


Figure 5. Extraction of street view elements. The red box indicates how to extract the number of street view elements through object detection.

Finally, we extracted the area proportion of 14 types of street view elements, namely road, sky, sidewalk, building, ground, motor vehicle, wall, fence, traffic signal board, trees, pedestrians, poles, traffic signal lights, and bicycles, and extracted the number of 4 types of street view elements: motor vehicle, pedestrian, bicycle, and traffic signal lights. The independent variable of regression analysis in this paper is the area proportion of

14 types of street view elements and the number of 4 types of street view elements, and the dependent variable is the safety perception score.

2.2.3. LightGBM and SHAP

Based on the safety perception score obtained through the safety perception model and the street view elements extracted by semantic segmentation, this paper selects the LightGBM algorithm to model and analyze the safety perception score and street view elements, and uses SHAP to interpret the results of the model output.

(1) LightGBM

Lightgbm (Light Gradient Boosting Machine) is an algorithm framework that implements gbdm (Gradient Boosting Decision Tree) [49]. It is a decision tree algorithm based on a histogram. By using Goss (Gradient based One-side Sample), it reduces a large amount of data with only small gradients and calculates information gain only using data with large gradients so as to ensure model accuracy and reduce the amount of training data. The algorithm also introduces a leaf-wise strategy with depth constraints, which finds the leaf with the largest splitting gain from all the current leaf nodes each time, splits, and cycles continuously. Under the same splitting times, leaf-wise can reduce more errors and obtain better accuracy than other strategies. Therefore, compared with traditional decision tree algorithms such as XGBoost 1.6.2 [50], LightGBM has faster training speed, lower memory consumption, and higher accuracy and is one of the best promotion methods at present.

The superparameter of a machine learning model is very important for training. There are two main methods to find the optimal parameters: one is the superparameter optimization based on grid search, and the other is the superparameter optimization based on Bayesian optimizer. Bayesian optimization method is the current state of the art in the field of superparameter optimization. It can be considered as the most advanced optimization framework at present. It can be applied to various fields of automatic machine learning (AutoML), not only in the field of hyperparameter optimization (HPO) but also in advanced fields such as neural network architecture search NAS and meta learning. The main idea of Bayesian hyperparametric optimization method is to update the posterior distribution of the objective function by adding samples until the posterior distribution matches the actual distribution. Almost all modern superparametric optimization methods that have achieved excellent results in efficiency and effect are based on the basic idea of Bayesian optimization, so this paper selects the Bayesian optimization method.

(2) SHAP

SHAP is a model result interpretation tool which originates from the concept of Shapley value in cooperative game theory, namely Shapley value [51]. As a classic post-attribution explanatory variable framework, the SHAP framework interprets the results of the machine learning model by calculating the contribution of each feature added to the model. It can not only quantify the impact of various factors in each sample on the predicted value of the model but also reflect the positive and negative impact and organically unify the global and local interpretation of the machine learning model [52]. SHAP is a kind of additive interpretation model which regards each feature as a contributor, calculates its specific contribution value, and adds the contribution values of all features to generate the final prediction. In this paper, the SHAP method is used to allocate the SHAP value for each element according to the impact of various street view elements on environmental safety perception.

The basic design idea of SHAP is first, to calculate the marginal contribution of a feature when it is added to the model, then calculate the different marginal contributions of the feature in all feature sequences, and finally calculate the SHAP value of the feature, that is, the mean value of all marginal contributions of the feature.

Suppose the i -th sample is x_i , the j -th feature of the i -th sample is x_{ij} , the marginal contribution of the feature is mc_{ij} , and the weight of the edge is w_i , where $f(x_{ij})$ is the

SHAP value of x_{ij} . For example, the SHAP value of the first feature of the i -th sample is calculated as follows:

$$f(x_{i1}) = mc_{i1}w_1 + \dots + mc_{i1}w_n \quad (3)$$

The predicted value of the model for this sample is y_i , and the baseline of the whole model (usually the average of the target variables of all samples) is y_{base} , so the SHAP value follows the following equation:

$$y_i = y_{base} + f(x_{i1}) + f(x_{i2}) + \dots + f(x_{ik}) \quad (4)$$

$f(x_{i1})$ is the contribution value of the first feature in the i -th sample to the final predicted value y_i . k means there are k features in total. The SHAP value of each feature represents the change predicted by the model when the feature is taken as a condition. For each feature, the SHAP value indicates its contribution. When $f(x_{i1}) > 0$, it represents that the feature improves the predicted value and has a positive effect; on the contrary, it shows that this feature reduces the predicted value and has a reverse effect. LightGBM can only reflect the importance of the feature by using the traditional feature importance, but it is not clear about the specific influence of the feature on the prediction results. The biggest advantage of SHAP value calculation is that it can not only reflect the influence of each feature in the sample on the prediction results but also show the positive and negative influence.

The redder the point color on the SHAP output graph, the greater the value of the feature itself; the bluer the color, the smaller the value of the feature itself. A positive SHAP value indicates that the feature has a positive impact on the safety perception score and a negative SHAP value indicates that the feature has a negative impact on the safety perception score.

3. Results

3.1. Spatial Distribution of Street View Safety Perception

We trained the safety perception score of street view images in Chaoyang District of Beijing through CNN model. The lower the score of safety perception, the more unsafe the street view image in this area is; the higher the safety perception score, the safer the street view image of the area. As shown in Figure 6, the overall street view safety perception level of Chaoyang District in Beijing is high. The area with the highest level of safety perception is the western central area with dense road networks, which has a high level of economic development, dense population, and perfect street construction. The overall level of safety perception in the eastern region is relatively low, and the road network in this region is relatively sparse, indicating that the level of economic development is relatively low compared with the western region, and the street construction is not as good as the western region as a whole.

3.2. Analysis of Influencing Factors of Street View Safety Perception

We build the dataset of all street view images and the corresponding safety perception score as the safety perception dataset, set 80% of the data as the training set and 20% of the data as the test set, and find the optimal superparameters of LightGBM through Bayesian optimization method. The optimal superparameters are shown in Table 1.

Table 1. Bayesian optimization for finding the optimal hyperparameters of LightGBM.

Name	Configuration
Num_estimators	100
Boosting_type	Gbdt
Objective	Regression
Metric	L1, L2
Num_leaves	46

Table 1. Cont.

Name	Configuration
Max_depth	7
Learning_rate	0.05
Feature_fraction	0.6
Bagging_fraction	0.9
Bagging_freq	2
Min_data_in_leaf	86

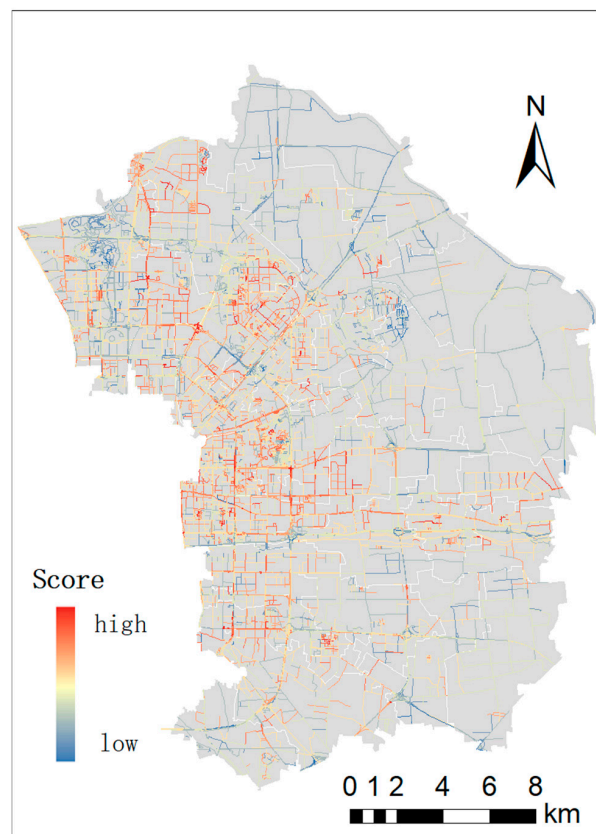
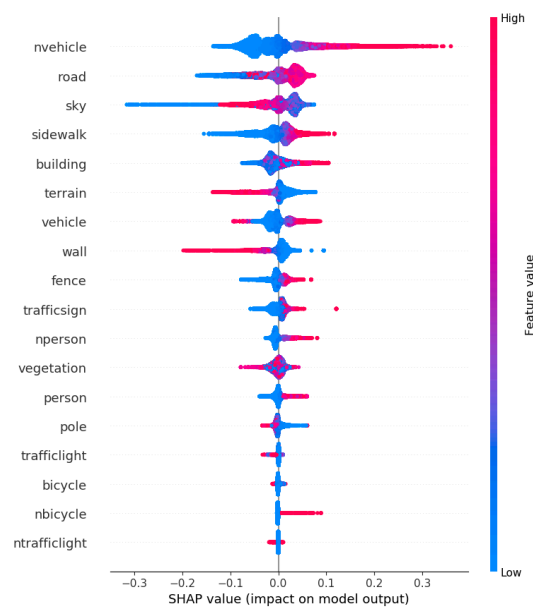


Figure 6. Distribution of safety perception.

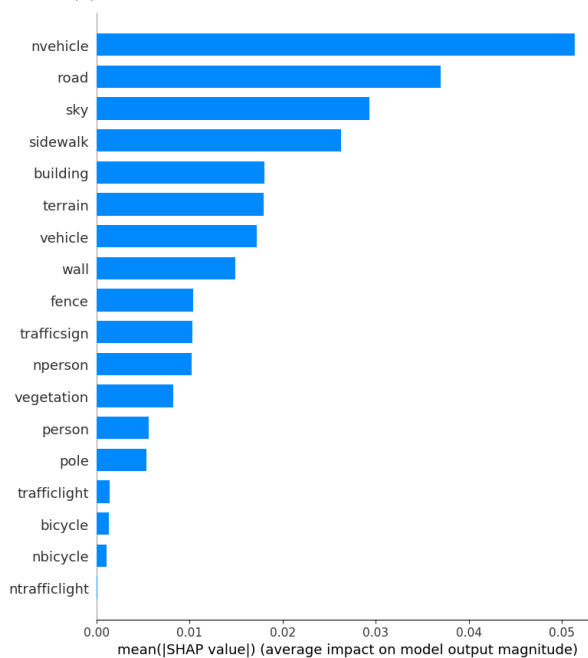
After setting the superparameter of LightGBM as the optimal superparameter, the LightGBM model is used to train the street view safety perception dataset, and the interpretability analysis is carried out using the SHAP method. The SHAP results are shown in Figure 7.

Figure 7a shows the SHAP values of each sample of all street view elements. The left label shows that the elements are sorted according to their importance, and the lower abscissa is the SHAP value of the element, that is, the weight of the impact. The wider the local distribution range, the greater the impact of the element. Each point in the figure represents a sample, and the color represents the numerical size of each element. The redder the point color on the SHAP value, the larger the value of the element; the bluer the color, the smaller the value of the element. It can be seen from the figure that the number of motor vehicles and the proportion of road, sky, and sidewalk area are the four most influential street view elements. Specifically, the greater the number of motor vehicles, the higher the proportion of roads and sidewalks, the greater the positive impact on safety perception, and the higher the degree of safety perception. Roads considered unsafe are mainly located outside the city. Roads considered safe are mainly distributed in the downtown area [53]. Because places with a large number of motor vehicles, roads,

and sidewalks are usually close to the central area of the city, a more lively and prosperous environment will make people feel safe. The higher the proportion of the sky, the greater the negative impact on safety perception and the lower the degree of safety perception. In addition, buildings, vehicles, fences, traffic signs, the proportion of pedestrian area, and the number of pedestrians and bicycles have a certain role in promoting image safety perception; the area proportion of open space, walls, trees, columns, and traffic lights and the number of traffic lights have an inhibitory effect on safety perception. The wall is similar to large obstacles, reducing the visibility and permeability of public space [54] and reducing informal supervision and “street eyes” [55], thus inhibiting safety perception. The area ratio of bicycles has no significant impact on safety perception. The results in Figure 7b are consistent with the impact in Figure 7a. The top four absolute values of the impact on safety perception are the number of motor vehicles, roads, skies, and sidewalks.



(a) SHAP value for each street view element



(b) Average of the absolute values of the SHAP values of street view elements

Figure 7. Impact of street view elements on environmental safety perception.

As shown in Figure 8a, as the number of motor vehicles increases, the SHAP value will increase significantly. As shown in Figure 8b, when the proportion of roads is less than 40%, the higher the proportion of roads, the higher the SHAP value; when the proportion of roads is greater than 40%, the higher the proportion of roads, the lower the SHAP value, because too many roads will make people feel empty, thus reducing safety perception. As shown in Figure 8c, when the proportion of the sky is less than 10%, the SHAP value increases with the increase in the proportion of the sky; when the sky proportion is greater than 20%, the SHAP value decreases with the increase in the sky proportion. As shown in Figure 8d, the more the proportion of sidewalks, the higher the SHAP value. This is because sidewalks are closely related to residents' daily life. Whether it is a nearby trip or a walk, sidewalks will provide people with a safe space isolated from motor vehicles, and the construction of sidewalks is related to the local economy and population level. Generally, the higher the economic and population level, the more perfect the construction of sidewalks.

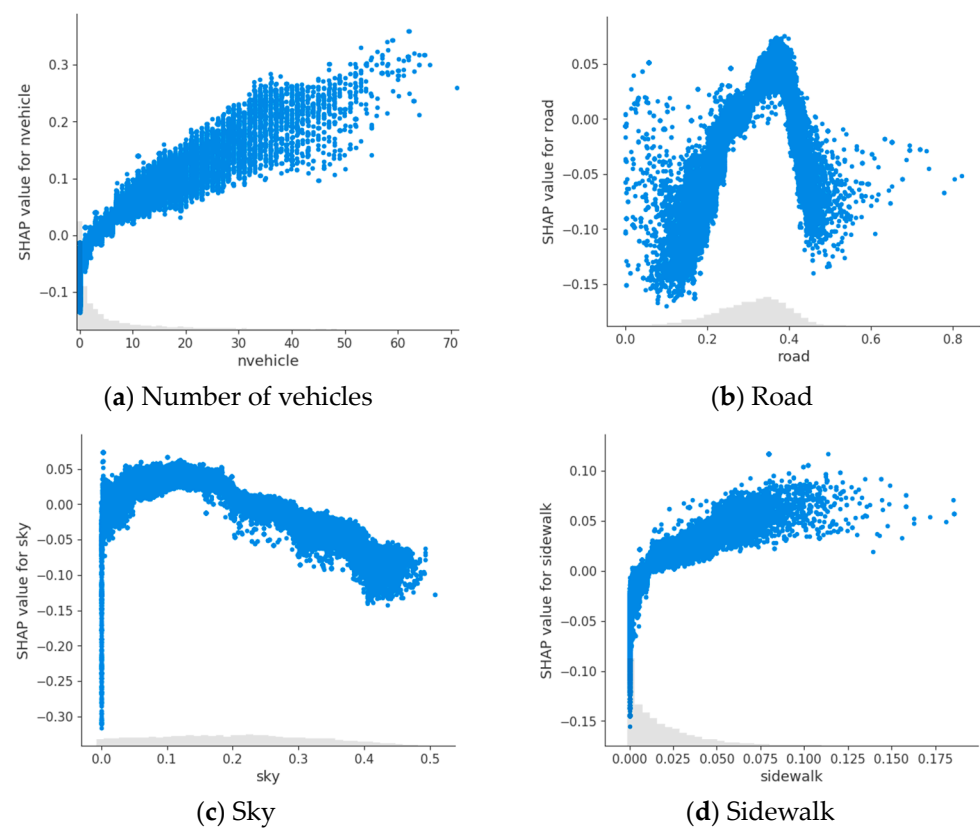


Figure 8. Impact of explanatory variables on safety perception.

3.3. Interaction of Street View Elements

(1) Interaction of different street view elements on safety perception

As shown in Figure 9, we conducted interaction analysis on the top seven street view elements of the SHAP value, including the number of motor vehicles, roads, sky, sidewalks, buildings, ground, and motor vehicles. We choose the street view elements with obvious interactions to analyze. We analyzed the street view elements with obvious interaction in Figure 9, as shown in Figure 10. In Figure 10a, when the number of motor vehicles is fixed, the more the proportion of road area is, the less conducive to the improvement of safety perception. Because the open road is relatively remote, it will make people feel uneasy. In Figure 10b, when the number of motor vehicles is fixed, the larger the proportion of building area, the more conducive to the improvement of safety perception. Where there are more buildings, human activities are more frequent, which will make people feel safe [31]. Moreover, as the proportion of building area increases, the positive interaction

effect gradually weakens. In Figure 10c, when the proportion of road area is less than 0.3, the greater the proportion of building area, the greater the positive effect of road on safety perception. In Figure 10d, when the area of the sky accounts for a certain proportion, the fewer the number of trees, the more conducive to the improvement of safety perception. This is because the high proportion of images of trees means that the trees are dense or the canopy is large, which reduces the visibility and visibility of pedestrians in the environment and makes people feel uneasy [56]. Moreover, when the proportion of sky area is less than 0.2, the increase is more obvious. With the increase in the proportion of sky area, the positive interaction effect gradually weakens. In Figure 10e, when the area ratio of motor vehicles is less than 0.1, the greater the area ratio of roads, the greater the positive impact of the area ratio of motor vehicles on safety perception; when the area ratio of motor vehicles is greater than 0.1, the smaller the area ratio of roads, the greater the positive impact of the area ratio of motor vehicles on safety perception. This is because in more remote areas, the presence of motor vehicles can indicate human activity and make people feel safe. When the proportion of motor vehicle area is relatively large, the smaller the proportion of road area, the more congested the road is, and congested roads are often more bustling, which will increase people's safety perception. In Figure 10f, when the area proportion of motor vehicles is fixed, the more the area proportion of buildings, the more conducive to the improvement of safety perception. Places with more buildings have more human activities, which makes people feel safe. In Figure 10g, when the proportion of sky area is less than 0.2, the larger the proportion of building area, the more conducive to the improvement of safety perception; when the area proportion of the sky is greater than 0.2, the smaller the area proportion of the building, the more conducive to the improvement of safety perception. The reason is that in areas where the sky area is relatively small, it is usually trees or buildings that obstruct the sky, and more trees can block pedestrians' line of sight, making people feel unsafe; the more buildings there are, the more human activities they can reflect, thereby enhancing people's safety perception. Street view images are collected along roads, and open areas with a high proportion of sky are mostly on urban main roads and highways. Vehicles in these areas tend to travel at high speeds, making it inconvenient to travel. Additionally, the presence of many buildings can make people feel unsafe. In Figure 10h, when the proportion of the ground area is fixed, the larger the proportion of the sky, the more conducive to the improvement of safety perception. Moreover, with the increase in the proportion of the ground area, the positive interaction effect is gradually weakened. In excessively open spaces, people may feel fear and unease.

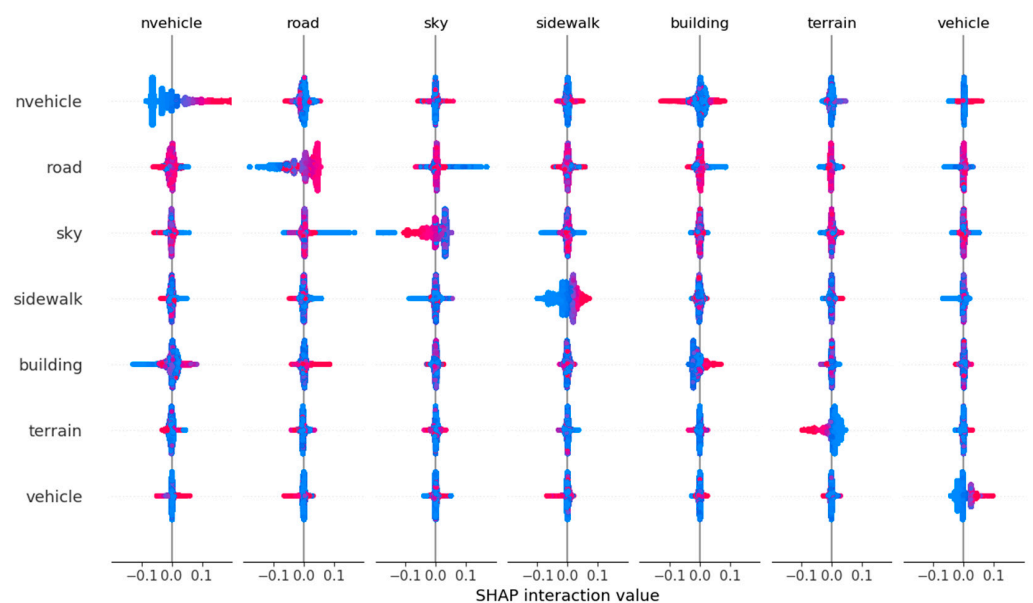


Figure 9. The interactive effect of bivariate on SHAP values.

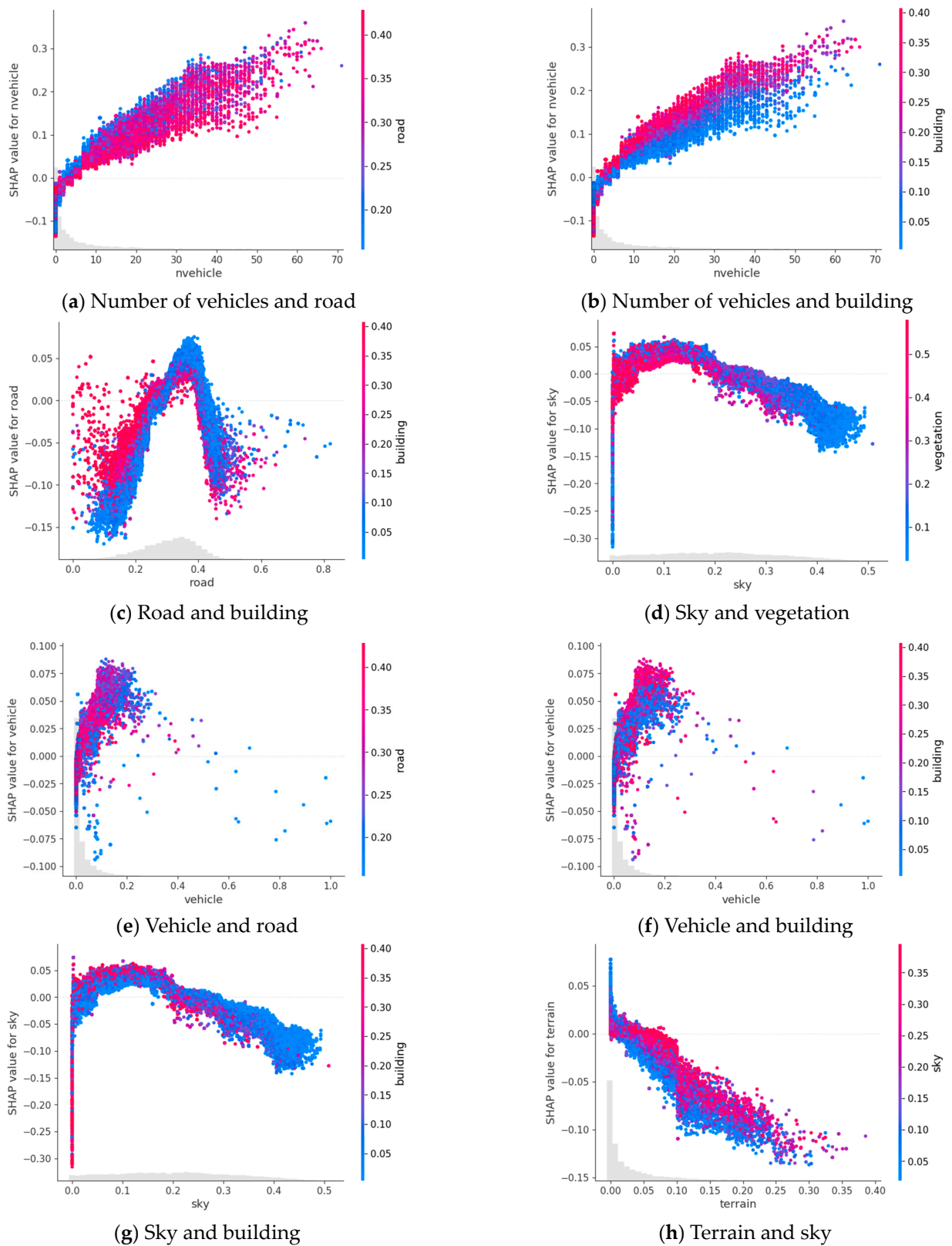


Figure 10. The interactive impact of street view elements on safety perception.

(2) Interaction between the proportion and quantity of street view elements

As Figure 11a shows, when the number of pedestrians is no more than 2, the area proportion of fewer pedestrians will make people feel safer. The reason for this phenomenon

may be that the large proportion of portraits means that pedestrians are closer to individuals, which breaks through the appropriate social distance and reduces safety perception. Hall, a famous American anthropologist, proposed four kinds of spatial distance, namely intimate distance, personal distance, social distance, and public distance [57]. Among them, “social distance” is generally 1.2~3.6 m, which refers to the communication distance between strangers. If it is below this range, people will feel uncomfortable. Jan Gehl further elaborated on this. According to the principle of arm length distance, that is, the minimum non-contact distance, he believed that human individuals tended to maintain this narrow but crucial distance so as to obtain a sense of safety and comfort in their situation [58]. Psychological research also shows that psychological space is also a sense of safety, and the pursuit of safety is human instinct. When this mental space is violated, it will cause anxiety and anxiety [59]. When the number of pedestrians is large, the proportion of pedestrian area has no obvious relationship with the number of pedestrians. This may be because when the number of pedestrians reaches a certain scale, people’s sense of safety will not be affected by the proportion of pedestrian area. In Figure 11b, when the number of motor vehicles is less than four, the relatively small proportion of the area of motor vehicles will make people feel safer, because when the number of motor vehicles is fixed, the larger the proportion of the area of motor vehicles, the closer the distance from pedestrians, which will bring the feeling of potential risk and make people feel unsafe. When the number of motor vehicles is large, the larger the proportion of motor vehicles, the safer people feel because human activities can generate informal monitoring, forming a “street eye”, thereby increasing residents’ safety perception [60].

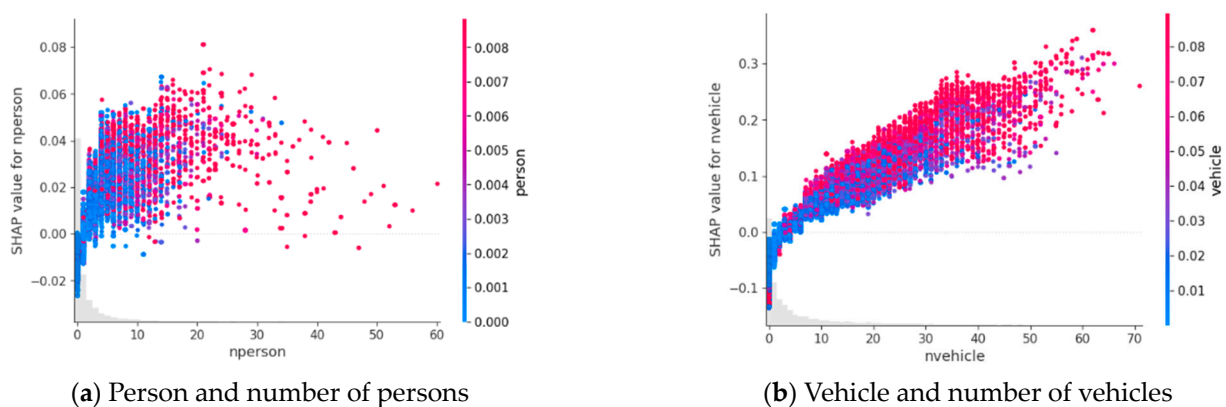


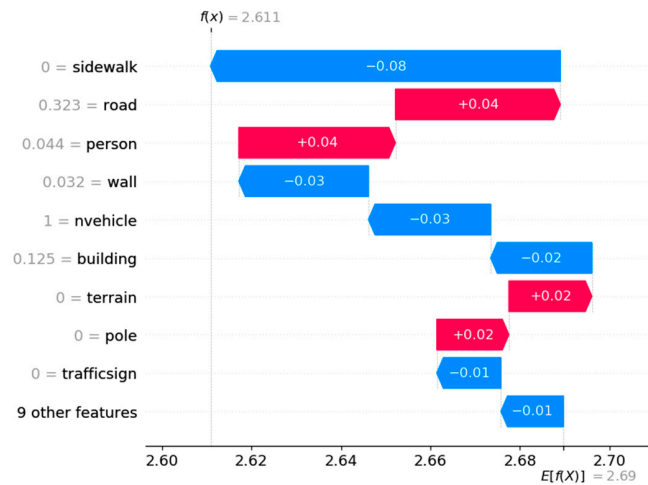
Figure 11. Interactive influence of the number and proportion of street view elements.

3.4. Local Interpretation of Street View Environment

The total SHAP value of the street is equal to the sum of the baseline value (the average of the total SHAP values of all street segments) plus the SHAP values of all street view elements. If the total SHAP value of the street segment is greater than or less than the baseline value, it indicates that the contribution of the characteristics of the street segment jointly increases or reduces the risk of crime. The same variable may have different SHAP values in different locations. The spatial change reveals the different effects of the local street landscape environmental characteristics on different street segments.

We select the lowest, moderate, and highest sample points of safety perception as examples to explain the local effect of street view elements. The SHAP value of each street view element is visualized as a waterfall in the following figure. As shown in the figure, in Figures 12a, 13a, and 14a, the red histogram represents the positive impact of street view elements and the blue histogram represents the negative impact of street view elements. The length of the histogram represents the absolute contribution of street view elements, and the corresponding value is the contribution value. Figures 12b, 13b, and 14b are street view images corresponding to the waterfall diagram.

As shown in Figure 12a, sidewalks have a significant negative impact on safety perception, while roads and pedestrians can significantly have a positive impact on safety perception. Figure 12b is the street view picture explained in Figure 12a. It can be seen from Figure 12b that such places are rural roads with a lack of natural monitoring. There are few pedestrians and social activities in the street view pictures, which will give people a feeling of unsafety. If the number of roads and pedestrians in such places increases significantly, people’s perception of safety in such places will be greatly improved. Therefore, the social environment and residents’ emotions should be considered to strengthen the construction of such sites [61].

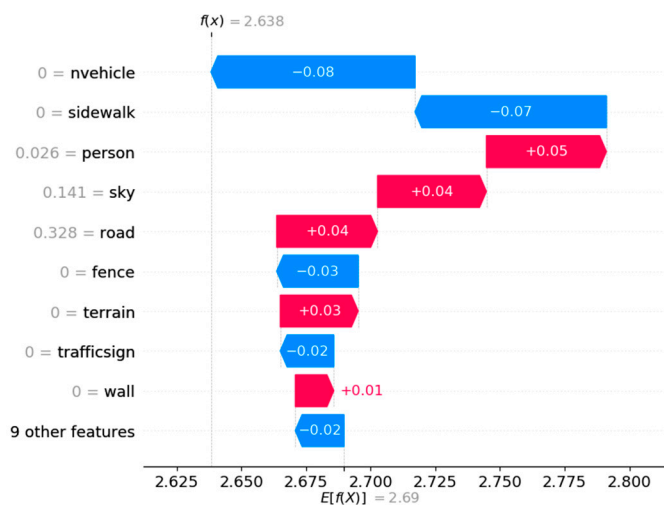


(a)



(b)

Figure 12. The place with the lowest environmental safety perception. (a) Local SHAP value. (b) The example Street View Image.



(a)

Figure 13. Cont.

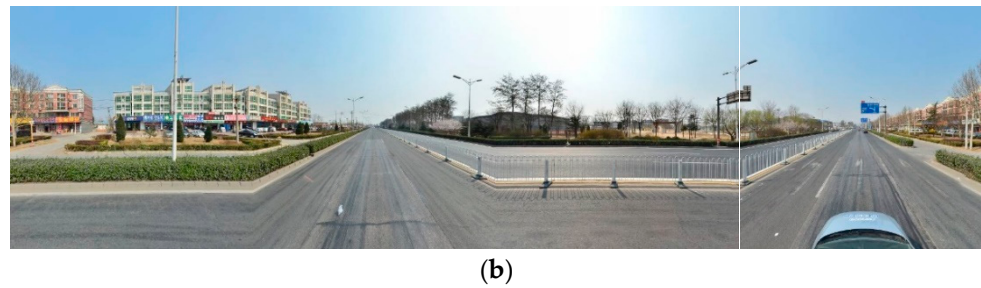


Figure 13. The place with the moderate environmental safety perception. (a) Local SHAP value. (b) The example Street View Image.

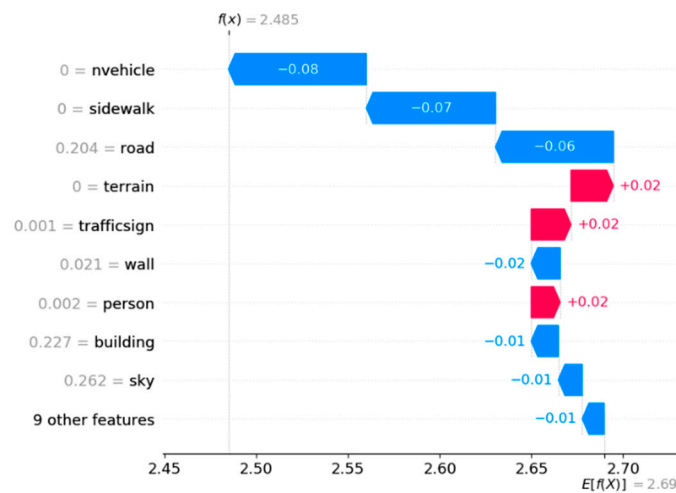


Figure 14. The place with the highest environmental safety perception. (a) Local SHAP value. (b) The example Street View Image.

As shown in Figure 13a, where the safety perception score is moderate, the number of motor vehicles and sidewalks have a significant negative impact on the safety perception score, while pedestrians, the sky, and roads have a significant positive impact. This is because such roads are not close to residential areas, there are fewer residents nearby, they are relatively open, and the speed of motor vehicles is usually fast, so the increase in the number of motor vehicles will make people feel unsafe. The moderate number of pedestrians in such places will also make people feel safe. The sky and road will also give people a feeling of openness. Combined with the street view image in Figure 13b, where pedestrians and social activities are relatively few, the presence of people can produce “informal supervision” [60], so as to improve people’s safety perception level.

As shown in Figure 14a, for places with the highest score of safety perception, the number of motor vehicles and the proportion of sidewalks and roads have the most significant impact, and the SHAP values are -0.08 , -0.07 , and -0.06 , respectively. Therefore, where the score of safety perception is the highest, the number of motor vehicles and the proportion of sidewalks and roads have a negative impact on safety perception. The

ground, traffic signs, and pedestrians have a certain positive impact. This is because the ground will make people feel comfortable. Traffic signs can restrict motor vehicles to a certain extent, making nearby residents feel more at ease. An appropriate increase in pedestrians will make people feel safer. The wall has a certain negative impact, although it also has a partition effect. However, it will also weaken the monitoring effect. Figure 14b is the real street view picture explained in Figure 14a. The environment of such places is clean and tidy, and the high land development and utilization rate can improve the comfort and safety of nearby residents [62].

4. Conclusions

4.1. Conclusions

In view of the shortcomings of the current research, this paper uses the Chinese local city perception dataset and selects Chaoyang District in Beijing as the representative city to analyze the road safety perception and influencing factors of the complex urban environment. This paper studies the distribution of road safety perception in Chaoyang District, Beijing, and evaluates and analyzes the influencing factors of street view safety perception based on LightGBM and the SHAP interpretation framework. Firstly, using CNN and machine learning technology, the domestic safety perception dataset is constructed to train the safety perception score of Chaoyang District in Beijing; then, LM-deep labv3+ and yolov5 are used to perform semantic segmentation and target detection on street view images, respectively, to obtain the area proportion and number of street view elements. Finally, a street view safety perception model is constructed based on LightGBM, and the key street view elements that affect the safety perception are obtained by interpretive analysis combined with the SHAP method, and the in-depth analysis is carried out. The main conclusions of this study are as follows:

- (1) The overall street view safety perception level of Chaoyang District in Beijing is relatively high. The region with the highest level of safety perception is the western central region with dense road networks, which has a high economic level and dense population. The overall level of safety perception in the eastern region is relatively low, and the road network in this region is relatively sparse, indicating that the level of economic development is relatively low compared with the western region, and the street construction is not as good as the western region as a whole. The rich and diverse functional layout increases accessibility and watchability, which has a positive impact on enhancing safety perception [63].
- (2) The four most influential street view elements are the number of motor vehicles and the proportion of the area of roads, skies, and sidewalks. Among them, the greater the number of motor vehicles and the higher the proportion of roads and sidewalks, the greater the positive impact on safety perception and the higher the degree of safety perception; the higher the proportion of sky area, the greater the negative impact on safety perception and the lower the degree of safety perception. Human activities can increase informal monitoring, thereby enhancing residents' safety perception.
- (3) There is interaction between street view elements on safety perception: when the number of motor vehicles is fixed, the greater the proportion of road area is, the less conducive to the improvement of safety perception; when the number of motor vehicles is fixed, the larger the proportion of building area, the more conducive to the improvement of safety perception. When there are fewer roads and more buildings, it is more conducive to the improvement of safety perception. When the proportion of sky area is fixed, the more trees, the less conducive to the improvement of safety perception. The proportion and number of street view elements interact with safety perception: when the number of pedestrians is less than two, the area proportion of fewer pedestrians will make people feel safer. When the number of motor vehicles is less than four, the area proportion of motor vehicles is relatively small, which makes people feel safer; when the number of motor vehicles is large, the larger the proportion of motor vehicles, the more people feel safe. This discovery is consistent with the

findings of psychological research, and the impact of human behavioral activities on environmental safety perception is also consistent with CPTED.

- (4) In the sections with the lowest, moderate, and highest level of safety perception, the influence of street view elements on street view safety perception is inconsistent. In the place with the lowest safety perception, sidewalks have a significant negative impact on safety perception, while roads and pedestrians can significantly have a positive impact on safety perception. In places with moderate safety perception scores, the number of motor vehicles and sidewalks have significant negative effects on safety perception scores, while pedestrians, the sky, and roads have significant positive effects. For the places with the highest score of safety perception, the number of motor vehicles and the proportion of sidewalks and roads have the most significant impact, which has a significant negative impact on safety perception. The ground, traffic signs, and pedestrians have a certain positive impact, while the wall has a certain negative impact. It can be seen that the contribution of various street view elements to safety perception varies in different regions.

4.2. Implications for Practice

The layout of the urban environment usually does not undergo significant changes in the short term. Therefore, based on the research results, we provide the following suggestions from the perspective of the police:

- (1) Appropriate monitoring can significantly enhance people's safety perception. In areas with low safety perception, emphasis should be placed on improving the infrastructure for security prevention and security control mechanisms and configuring security infrastructure such as installing surveillance cameras.
- (2) Strengthen the deployment of police resources and patrols in areas with low safety perception; invest in public security human resources and improve the level of public safety service management.
- (3) In addition, building a good community environment can enhance residents' safety perception. By promoting comprehensive management of spatial blind spots, making efficient use of the Internet to innovate police work, and strengthening cooperation between the police and the residents, residents' safety perception will be improved.

4.3. Research Deficiency

The research of this paper has the following shortcomings:

- (1) In view of the limited resources, the safety perception dataset of this study is the domestic dataset constructed by the existing research. Although they are all domestic cities, there will be some differences between different cities. Therefore, in the follow-up study, we can consider building a local safety perception dataset in the study area to obtain more objective and real safety perception data.
- (2) Due to the large amount of data and limited computing resources, the image processing in this paper selects a relatively lightweight semantic segmentation and target detection model, and the processing results have some errors. In the future, more accurate models and more powerful computing resources can be used to process data.
- (3) The research on the interaction of street view elements in this article is at a macro level, and the interaction of street view elements may vary in different regions. We will conduct more detailed research on this issue in the future. In addition, due to the limitations of street view image data, only daytime data can be obtained on a large scale. Therefore, we have not yet considered the performance differences in different time periods.
- (4) The research in this article revolves around street view images, and the analysis is also based on the information obtained from street view images. For variables such as the surrounding built environment and social variables that cannot be directly obtained from street view images, we will conduct further analysis in subsequent research.

- (5) Existing studies have shown that there is a deviation between safety perception and actual safety (crime risk), and this deviation will be further studied in the future.

Author Contributions: Conceptualization, Xinyu Hou; methodology, Xinyu Hou; software, Xinyu Hou; validation, Peng Chen; formal analysis, Xinyu Hou; investigation, Xinyu Hou; resources, Xinyu Hou; data curation, Xinyu Hou; writing—original draft preparation, Xinyu Hou; writing—review and editing, Peng Chen; visualization, Xinyu Hou; supervision, Peng Chen; project administration, Peng Chen; funding acquisition, Peng Chen. All authors have read and agreed to the published version of the manuscript.

Funding: This research received Funding for Discipline Innovation and Talent Introduction Bases in Higher Education Institutions (B20087).

Data Availability Statement: The original data presented in the study are openly available at <https://google.cn/streetview/>.

Acknowledgments: We thank anonymous reviewers for their comments on improving this manuscript.

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

References

- 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]
- Zhang, L.; Pei, T.; Chen, Y.; Song, C.; Liu, X. A review of urban environmental assessment based on street view images. *J. Geo-Inf. Sci.* **2019**, *21*, 46–58. [CrossRef]
- Kou, S.; Yao, Y.; Zheng, H.; Zhou, J.; Zhang, J.; Ren, S.; Wang, R.; Guan, Q. Delineating China’s urban traffic layout by integrating complex graph theory and road network data. *J. Geo-Inf. Sci.* **2021**, *23*, 812–824. [CrossRef]
- Ordonez, V.; Berg, T.L. Learning high-level judgments of urban perception. In Proceedings of the Computer Vision—ECCV 2014, Zurich, Switzerland, 6–12 September 2014; Springer International Publishing: Cham, Switzerland, 2014; pp. 494–510. [CrossRef]
- Tuan, Y. *Space and Place: The Perspective of Experience*; University of Minnesota Press: Minneapolis, MN, USA, 1977.
- An, L.J.; Cong, Z. A review of research. *Chin. J. Behav. Med. Brain Sci.* **2003**, *12*, 698–699. [CrossRef]
- Jeffery, C.R. *Crime Prevention through Environmental Design*; Sage: Beverly Hill, CA, USA, 1971.
- Crowe, T.D.; Fennelly, L.J. *Crime Prevention through Environmental Design*, 3rd ed.; China People’s Public Security University Press: Beijing, China, 2015.
- Marzbali, M.H.; Abdullah, A.; Ignatius, J.; Tilaki, M.J.M. Examining the effects of crime prevention through environmental design (CPTED) on residential burglary. *Int. J. Law Crime Justice* **2016**, *46*, 86–102. [CrossRef]
- Márquez, L. Safety perception in transportation choices: Progress and research lines. *Ing. Compet.* **2016**, *18*, 11–24.
- Li, X.; Zhang, C.; Li, W. Does the visibility of greenery increase perceived safety in urban areas? Evidence from the place pulse 1.0 dataset. *ISPRS Int. J. Geo-Inf.* **2015**, *4*, 1166–1183. [CrossRef]
- Yao, Y.; Liang, Z.; Yuan, Z.; Liu, P.; Bie, Y.; Zhang, J.; Wang, R.; Wang, J.; Guan, Q. A human-machine adversarial scoring framework for urban perception assessment using street-view images. *Int. J. Geogr. Inf. Sci.* **2019**, *33*, 2363–2384. [CrossRef]
- Van, D.W.A.; Van, S.L.; Stringer, P. Fear of crime in residential environments: Testing a social psychological model. *J. Soc. Psychol.* **1989**, *129*, 141–160. [CrossRef]
- Farrall, S.; Bannister, J.; Ditton, J.; Gilchrist, E. Social psychology and the fear of crime. *Br. J. Criminol.* **2000**, *40*, 399–413. [CrossRef]
- Li, X.; Yan, H.; Wang, Z. Evaluation of road environment safety perception and analysis of influencing factors combining street view imagery and machine learning. *J. Geo-Inf. Sci.* **2023**, *25*, 852–865. [CrossRef]
- Liu, Y.; Liu, X.; Gao, S.; Gong, L.; Kang, C.; Zhi, Y.; Chi, G.; Shi, L. Social sensing: A new approach to understanding our socioeconomic environments. *Ann. Assoc. Am. Geogr.* **2015**, *105*, 512–530. [CrossRef]
- Sampson, R.J. *Great American City: Chicago and the Enduring Neighborhood Effect*; University of Chicago Press: Chicago, IL, USA, 2012.
- Long, Y.; Liu, L. How green are the streets? An analysis for central areas of Chinese cities using Tencent street view. *PLoS ONE* **2017**, *12*, e0171110. [CrossRef] [PubMed]
- Zhang, F.; Zhang, D.; Liu, Y.; Lin, H. Representing place locales using scene elements. *Comput. Environ. Urban Syst.* **2018**, *71*, 153–164. [CrossRef]
- Zhang, F.; Hu, M.; Che, W.; Lin, H.; Fang, C. Framework for virtual cognitive experiment in virtual geographic environments. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 36. [CrossRef]
- Salesses, P.; Schechtner, K.; Hidalgo, C.E.S.A. The collaborative image of the city: Mapping the inequality of urban perception. *PLoS ONE* **2013**, *8*, e68400. [CrossRef] [PubMed]
- He, L.; Páez, 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]

23. Hwang, J.; Sampson, R.J. Divergent Pathways of Gentrification: Racial Inequality and the Social Order of Renewal in Chicago Neighborhoods. *Am. Sociol. Rev.* **2014**, *79*, 726–751. [[CrossRef](#)]
24. Naik, N.; Philipoom, J.; Raskar, R.; Hidalgo, C. Streetscore—Predicting the Perceived Safety of One Million Streetscapes. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops, Columbus, OH, USA, 23–28 June 2014; pp. 793–799. [[CrossRef](#)]
25. Dubey, A.; Naik, N.; Parikh, D.; Raskar, R.; Hidalgo, C. Deep Learning the City: Quantifying Urban Perception at a Global Scale. In *Computer Vision—ECCV 2016 (ECCV 2016)*; Leibe, B., Matas, J., Sebe, N., Welling, M., Eds.; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2016; Volume 9905. [[CrossRef](#)]
26. Liu, X.; Chen, Q.; Zhu, L.; Xu, Y.; Lin, L. Place-centric visual urban perception with deep multi-instance regression. In Proceedings of the 25th ACM International Conference on Multimedia, Mountain View, CA, USA, 23–27 October 2017; pp. 19–27. [[CrossRef](#)]
27. Liu, X.; He, J.; Yao, Y.; Zhang, J.; Liang, H.; Wang, H.; Hong, Y. Classifying urban land use by integrating remote sensing and social media data. *Int. J. Geogr. Inf. Sci.* **2017**, *31*, 1675–1696. [[CrossRef](#)]
28. Porzi, L.; Rota, B.S.; Lepri, B.; Ricci, E. Predicting and understanding urban perception with convolutional neural networks. In Proceedings of the 23rd ACM International Conference on Multimedia, Brisbane, Australia, 26–30 October 2015; pp. 139–148. [[CrossRef](#)]
29. Naik, N.; Raskar, R.; Hidalgo, C.E.S.A. Cities are physical too: Using computer vision to measure the quality and impact of urban appearance. *Am. Econ. Rev.* **2016**, *106*, 128–132. [[CrossRef](#)]
30. Ashihara, Y. *The Aesthetic Townscape*; MIT Press: Cambridge, UK, 1983.
31. Liu, Y.; Chen, M.; Wang, M.; Huang, J.; Thomas, F.; Rahimi, K.; Mamouei, M. An interpretable machine learning framework for measuring urban perceptions from panoramic street view images. *iScience* **2023**, *26*, 106132. [[CrossRef](#)] [[PubMed](#)]
32. Shelhamer, E.; Long, J.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 640–651. [[CrossRef](#)] [[PubMed](#)]
33. Wang, R.; Ren, S.; Zhang, J.; Yao, Y.; Wang, Y.; Guan, Q. A comparison of two deep-learning-based urban perception models: Which one is better? *Comput. Urban Sci.* **2021**, *1*, 1–13. [[CrossRef](#)]
34. Comber, S.; Arribas-Bel, D.; Singleton, A.; Dolega, L. Using convolutional autoencoders to extract visual features of leisure and retail environments. *Landsc. Urban Plan.* **2020**, *202*, 103887. [[CrossRef](#)]
35. Krizek, K.; Forysth, A.; Slotterback, C.S. Is there a role for evidence-based practice in urban planning and policy? *Plan. Theory Pract.* **2009**, *10*, 459–478. [[CrossRef](#)]
36. Xu, L.; Jiang, W.; Chen, Z. Research on Public Space Security Perception: Taking Shanghai Urban Street View Perception as an Example. *Landsc. Archit.* **2018**, *25*, 23–29.
37. Fu, X.; Yang, P.; Jiang, S.; Xing, J.; Kang, D. The impact of different landscape design elements and their combinations on landscape security. *Urban Probl.* **2019**, 37–44. [[CrossRef](#)]
38. Liu, Z.; Lv, J.; Yao, Y.; Zhang, J.; Kou, S.; Guan, Q. Research method of interpretable urban perception model based on street view imagery. *J. Geo-Inf. Sci.* **2022**, *24*, 2045–2057. [[CrossRef](#)]
39. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
40. Ha, I.; Kim, H.; Park, S.; Kim, H. Image retrieval using BIM and features from pretrained VGG network for indoor localization. *Build. Environ.* **2018**, *140*, 23–31. [[CrossRef](#)]
41. Liu, X.; Chi, M.; Zhang, Y.; Qin, Y. Classifying High Resolution Remote Sensing Images by Fine-Tuned VGG Deep Networks. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 7137–7140. [[CrossRef](#)]
42. Lu, X.; Duan, X.; Mao, X.; Li, Y.; Zhang, X. Feature extraction and fusion using deep convolutional neural networks for face detection. In *Problems in Engineering*; Wiley: Hoboken, NJ, USA, 2017; Volume 2017. [[CrossRef](#)]
43. Hou, X.; Chen, P.; Gu, H. LM-DeeplabV3+: A Lightweight Image Segmentation Algorithm Based on Multi-Scale Feature Interaction. *Appl. Sci.* **2024**, *14*, 1558. [[CrossRef](#)]
44. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.-C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520. [[CrossRef](#)]
45. Chen, L.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), 8–14 September 2018; pp. 801–818.
46. Jocher, G. Ultralytics YOLOv5[EB/OL]. Available online: <https://github.com/ultralytics/yolov5> (accessed on 22 November 2022).
47. Benjdira, B.; Khursheed, T.; Koubaa, A.; Ammar, A.; Ouni, K. Car detection using unmanned aerial vehicles: Comparison between faster R-CNN and YOLOv3. In Proceedings of the 2019 1st International Conference on Unmanned Vehicle Systems-Oman (UVS), Muscat, Oman, 5–7 February 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–6. [[CrossRef](#)]
48. Li, C.; Wang, R.; Li, J.; Fei, L. Face detection based on YOLOv3. In *Recent Trends in Intelligent Computing, Communication and Devices*; Springer: Singapore, 2019; pp. 277–284. [[CrossRef](#)]
49. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T. LightGBM: A highly efficient gradient boosting decision tree. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; ACM: New York, NY, USA, 2017; pp. 3149–3157. [[CrossRef](#)]

50. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016.
51. Lundberg, S.M.; Lee, S.I. A unified approach to interpreting model predictions. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; ACM: New York, NY, USA, 2017; pp. 4768–4777. [[CrossRef](#)]
52. Lundberg, S.M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, M.J.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2020**, *2*, 56–67. [[CrossRef](#)] [[PubMed](#)]
53. Yu, X.; Ma, J.; Tang, Y.; Yang, T.; Jiang, F. Can we trust our eyes? Interpreting the misperception of road safety from street view images and deep learning. *Accid. Anal. Prev.* **2024**, *197*, 107455. [[CrossRef](#)] [[PubMed](#)]
54. Navarrete-Hernandez, P.; Vetro, A.; Concha, P. Building safer public spaces: Exploring gender difference in the perception of safety in public space through urban design interventions. *Landsc. Urban Plan.* **2021**, *214*, 104180. [[CrossRef](#)]
55. Chiodi, S.I. Crime prevention through urban design and planning in the smart city era: The challenge of disseminating CP-UDP in Italy: Learning from Europe. *J. Place Manag. Dev.* **2016**, *9*, 137–152. [[CrossRef](#)]
56. Baran, P.K.; Tabrizian, P.; Zhai, Y.; Smith, J.W.; Floyd, M.F. An exploratory study of perceived safety in a neighborhood park using immersive virtual environments. *Urban For. Urban Green.* **2018**, *35*, 72–81. [[CrossRef](#)]
57. Hall, E.T. *The Hidden Dimension*, 1st ed.; Doubleday: Garden City, NY, USA, 1966.
58. Gehl, J. *Cities for People*; Island Press: Washington, DC, USA, 2010.
59. Stephan, E.; Liberman, N.; Trope, Y. Politeness and psychological distance: A construal level perspective. *J. Personal. Soc. Psychol.* **2010**, *98*, 268–280. [[CrossRef](#)] [[PubMed](#)]
60. Jacobs, J. *The Death and Life of Great American Cities*; Random House: New York, NY, USA, 1961.
61. Liu, X.; Xiao, H.; Wang, X.; Zhao, F.; Hu, Y. Research on the sense of safety in Geography: Comprehensive understanding, application and prospect based on location. *Hum. Geogr.* **2018**, *33*, 38–45. [[CrossRef](#)]
62. Aziz, H.M.A.; Nagle, N.N.; Morton, A.M.; Hilliard, M.R.; White, D.A.; Stewart, R.N. Exploring the impact of walk–bike infrastructure, safety perception, and built environment on active transportation mode choice: A random parameter model using New York City commuter data. *Transportation* **2018**, *45*, 1207–1229. [[CrossRef](#)]
63. Donder, L.D.; Buffel, T.; Dury, S.; Witte, N.D.; Verté, D. Perceptual quality of neighbourhood design and feelings of unsafety. *Ageing Soc.* **2013**, *33*, 917–937. [[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.