

Article

Assessing Urban Safety Perception Through Street View Imagery and Transfer Learning: A Case Study of Wuhan, China

Yanhua Chen ^{1,2} and Zhi-Ri Tang ^{1,*}¹ School of Intelligent Systems Science and Engineering, Jinan University, Zhuhai 519070, China² Division of Science, Engineering and Health Studies, School of Professional Education and Executive Development, The Hong Kong Polytechnic University, Hong Kong 999077, China

* Correspondence: tangzhiri@jnu.edu.cn

Abstract

Human perception of urban streetscapes plays a crucial role in shaping human-centered urban planning and policymaking. Traditional studies on safety perception often rely on labor-intensive field surveys with limited spatial coverage, hindering large-scale assessments. To address this gap, this study constructs a street safety perception dataset for Wuhan, classifying street scenes into three perception levels. A convolutional neural network model based on transfer learning is developed, achieving a classification accuracy of 78.3%. By integrating image-based prediction with spatial clustering and correlation analysis, this study demonstrates that safety perception displays a distinctly clustered and uneven spatial distribution, primarily concentrated along major arterial roads and rail transit corridors by high safety levels. Correlation analysis indicates that higher safety perception is moderately associated with greater road grade, increased road width, and lower functional level while showing a weak negative correlation with housing prices. By presenting a framework that integrates transfer learning and geospatial analysis to connect urban street imagery with human perception, this study advances the assessment of spatialized safety perception and offers practical insights for urban planners and policymakers striving to create safer, more inclusive, and sustainable urban environments.



Academic Editor: Anna Mazzi

Received: 10 June 2025

Revised: 15 July 2025

Accepted: 1 August 2025

Published: 25 August 2025

Citation: Chen, Y.; Tang, Z.-R. Assessing Urban Safety Perception Through Street View Imagery and Transfer Learning: A Case Study of Wuhan, China. *Sustainability* **2025**, *17*, 7641. <https://doi.org/10.3390/su17177641>

Copyright: © 2025 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/>).

Keywords: safety perception; street-view imagery; transfer learning; spatial statistics

1. Introduction

Human perception, referring to urban inhabitants' cognitive appraisal of urban scenes, has established links with the physical appearance of cities [1–4]. High-level perception like the urban environment's safety and beauty would coexist with the spatial arrangement and social structure of urban regions [5]. For example, urban scenes generate favorable emotions as vegetation grows [6], while urban disorder might cause psychological discomfort [7]. As a result, many stakeholders, including urban planners and governments, need to look into the relationship between human perceptions and the urban physical environment, which could strengthen residents' feeling of identity and belonging [8–12]. Many studies have been conducted to study the relationship between the physical appearance structure of cities and other higher-level characteristics of perception [13].

Crime Prevention Through Environmental Design (CPTED) provides a framework for understanding how built environments influence perceived safety, highlighting factors such as visibility, lighting, territoriality, and maintenance [14]. Grounded in Place Theory [15], the interaction of place and space plays an important role in understanding the

relationship between humans and place. The perception of safety, as one of the key aspects of psychological feeling, is understood as a socially constructed, subjective experience shaped by one's emotional connection to place [16] and not solely identified by the physical infrastructure of a place and how individuals interpret their surroundings [17]. Traditional methods collect data on human perceptions of safety through offline or online questionnaires [18–20]. Generally, related studies mainly use survey data as ground truth data and then quantitatively analyze the human perceptions with influencing factors in local areas [21–23]. However, such approaches are often costly and time consuming, as they rely heavily on manual data collection and organization and are typically limited in spatial scale. With the advancement of machine learning, approaches that integrate traditional survey data with computational models have gained increasing attention. These methods leverage the knowledge embedded in human-labeled data to enable machines to learn and predict human perceptions [24]. However, they still face significant challenges, particularly the limited availability and uneven spatial distribution of labeled data, which constrains their generalizability and applicability at large scales.

There is a great need to utilize large-scale urban data, processed through advanced analytical methods, to gain a more comprehensive understanding of how high-level characteristics of human perception respond to a place [11,25–29]. Recently, large-volume street view imagery provides a panoramic function of 360° horizontally as well as 180° vertically at a sampling place of the scene, simulating the urban residents' eye-view perspective and providing the large-volume visual features as well as the surrounding scenes [30,31]. The semantic and scene information extracted from large-volume street imagery reflects human social sensing of the urban settings, a place in human-scale environments [32–34]. Additionally, a state-of-the-art deep-learning approach enables the demonstration of human perception by processing big data at a large scale [1,35]. Deep-learning algorithms, integrated with street-view imagery, have been successfully employed to extract street-level image features and predict human perceptions in urban street scenes [1,3,11]. However, convolutional neural network (CNN) models [36] that are based on deep learning methods generally require extensive training with labeled data to achieve strong classification performance [37], challenging the direct application of existing frameworks to urban safety-sensitive classification tasks, combined with insufficient labeled data for safety-sensitive urban classification tasks [38,39]. Furthermore, most studies focus on the improvement of perception recognition performance, seldom considering the estimation of human perception in terms of practical policymaking through the lens of streetscapes.

This study views urban safety as a perception of place and proposes a spatial framework that quantifies it using street-view imagery and deep learning. As such, the objective of this study is to map spatial patterns of urban safety perception at a large scale and gain nuanced insights into scene visualization and correlation analysis with street attributes in the context of safety perception. The specific aims are (1) to advance the empirical research of measuring safety perception based on the transfer learning method, (2) to interpret the scene visualization of relative safety levels, and (3) to generate evidence-based policy recommendations by analyzing the spatial patterns, transportation accessibility, road infrastructure, and housing prices.

2. Methods

2.1. Study Area

The selected study area includes the seven central administrative districts of Wuhan in China, namely, Jiangan District, Jianghan District, Qiaokou District, Hanyang District, Wuchang District, Qingshan District, and Hongshan District, as shown in Figure 1. Wuhan is located in the central region of China, the eastern part of Hubei Province, between 30°52' N latitude and

114°31' E longitude, and its center is located at the junction of Yangtze and Han rivers, with a total land area of 8569.15 square kilometers and a resident population of approximately 10.89 million in 2017. The average population density was 1271 persons per square kilometer. Within the urban area, Jiangnan District had a resident population of approximately 741,784, Jiangnan District—496,289, and Qiaokou District—528,604. Hanyang District recorded a population of 631,185, Wuchang District—1,044,072, Qingshan District—426,289, and Hongshan District—1,073,545. Hongshan District has the largest land area of 573.28 square kilometers among the seven administrative districts. However, its urbanization rate of permanent residents is relatively low at about 85%, the lowest among the seven districts. The other six districts all have a 100% urbanization rate. Although Jiangnan District covers only 28.29 square kilometers, it has the highest population density at 25,790 people per square kilometer, ranking first among the seven districts. Qiaokou District follows closely. Wuhan's urban pattern is primarily ring-shaped, with the central city located on the Second Ring Road, which has rapid economic development, complete infrastructure, and a compact building layout, primarily providing basic human life services and financial services to the city. The second to third ring road is a new urban area, rich in human resources such as educational resources, a natural environment with many lakes, mountains, and other natural resources, as well as convenient transportation, primarily to provide a better living environment for human lives, as well as to carry out tourism and ecological functions. The seven economically developed and developing administrative districts of Wuhan, as well as the annotation area (red circles in Figure 1), were chosen with a substantial amount of data and variability to ensure the data's representativeness and analyzability.

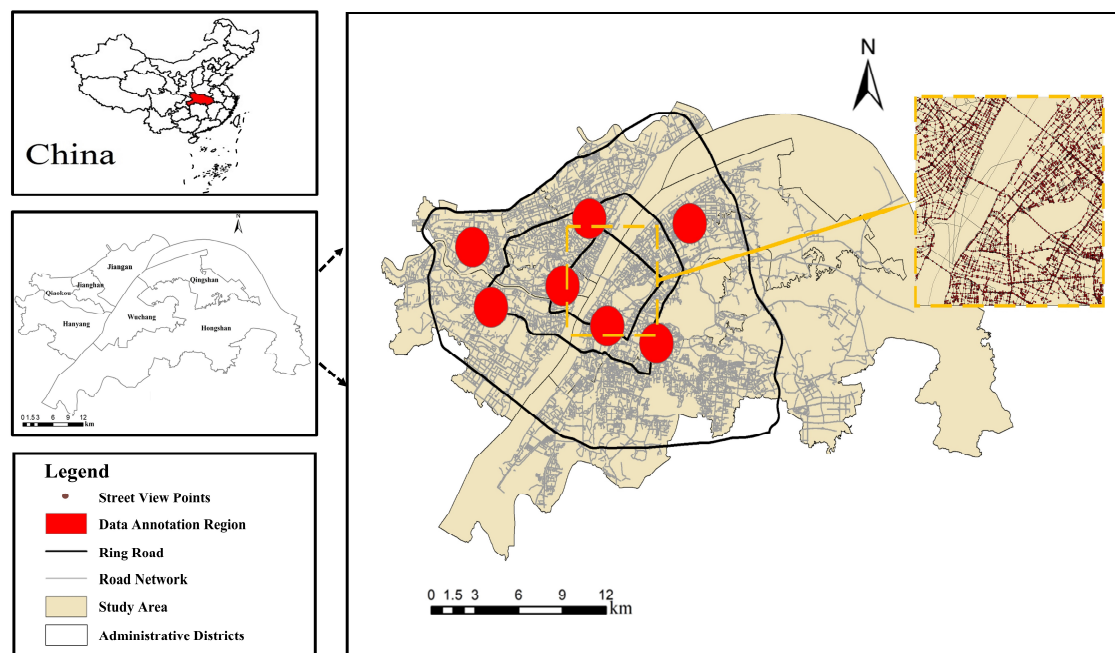


Figure 1. The location of the study area with a base map of Gaode road network data.

2.2. Dataset

2.2.1. Street View Image Data

Baidu Panorama, launched in 2013 and regularly updated, offers a detailed streetscape road network that covers both urban and high-traffic areas, providing comprehensive city streetscape data for this research (<https://lbsyun.baidu.com>). The Baidu Map Open Platform user manual was used to guide the data crawling process. Street view images were captured of the previously described road network, with specifications including image size, orientation, camera angle, and other relevant parameters. After data collection, a total of 44,683 street points corresponding to the seven administrative districts of Wuhan

city were obtained. The street view images corresponding to 8 orientations were obtained at each location, with a picture size of 480×360 and a camera shooting angle of 0° ; i.e., the panoramic view of the sampling device on the collection vehicle was obtained. This constructs the street view image dataset required in this paper; i.e., a point is taken every 200 m on a road (the road terminal is taken as a street viewpoint if the length is less than 200), each point is taken every 45° direction, and each image contains a unique image name as well as pixel and location information, etc.

As shown in Figure 1, the labeled dataset was constructed based on an analysis of community streets and commercial centers in representative areas of Wuhan, supported by detailed field research on the city's urban living streetscapes. This study selects seven sample areas, including business districts in various administrative regions, old industrial residential areas in Qingshan District, historic urban zones in Wuchang, and urban village redevelopment zones in Qiaokou District. These areas collectively reflect the diversity of Wuhan's urban living environment in recent years, encompassing affluent, middle-income, and low-income neighborhoods, as well as both older and newly developed communities. As such, they serve as representative areas with distinct land-use attributes.

2.2.2. Safety Perception Survey

The concept of safety in this study is interpreted as a subjective human perception rather than an objective risk assessment. It draws upon both contemporary frameworks, such as CPTED, and foundational urban theories. CPTED emphasizes how discomfort may arise from antisocial behavior, environmental neglect, poor lighting, or a lack of vitality and legibility in urban design, which can contribute to a sense of insecurity. This sense of safety can vary depending on individual cognition and can be interpreted as a lack of fear or as the perception that walking on that street would not pose a threat. In parallel, this study was also informed by "a well-used city street is a safe street," which underscores the importance of active street life and natural surveillance in shaping the perception of urban safety [40]. Together, these perspectives reinforce the idea that safety is not solely determined by measurable risks but also by how individuals cognitively and emotionally respond to the urban environment.

In this study, ten annotators, selected for their GIS-related expertise or familiarity with the city of Wuhan (with at least one year of residence), were tasked with labeling a dataset of street view images. The annotation was conducted anonymously with a gender-balanced composition (5 males and 5 females). Among them, three annotators (2 males, 1 female) were aged between 40 and 60, and the remaining seven (3 males, 4 females) were aged between 18 and 25. Each annotator applied their perceptions to evaluate and categorize the safety of each scene into three distinct levels: high, neutral, and low (Figure 2). Each image was randomly assigned to annotators after shuffling the dataset from the data annotation region shown in Figure 1, and the images did not contain any spatial or geographic indicators (e.g., street names, district names). To reduce potential cognitive bias and geographic preconceptions, each annotator labeled images independently using an interface that displayed one image at a time, with all spatial indicators (e.g., street names, district names) removed, as illustrated in Figure 2. They assessed whether a given street scene appeared safe, neutral, or unsafe, based on their holistic impression of walking through that environment. Rather than isolating specific visual elements (e.g., greenery, traffic, lighting), annotators were guided to provide an overall perceptual judgment while being mindful not to overreact to features like highways, whose scale may induce fear unrelated to actual street-level threat. Additionally, they were reminded that dense urban centers or sparse suburban areas, as well as varying levels of greenery, are not inherently linked to perceived safety or danger. As a result, a comprehensive set of 4456 unique street

locations and 35,648 annotated street scene images was obtained. The post-processing of the annotated street scene data is detailed in Section 2.3, including the removal of unclear images and the handling of sample imbalance and confusing issues.

This is an example



Which level of safety this place?

Please select *one* of the following options

Unsafe Neutral Safe

Figure 2. Snapshot of form for safety perception survey.

2.2.3. Other Datasets

There are types of datasets for spatial analysis and correlation analysis. Firstly, through preprocessing, the road data of seven urban areas in Wuhan city contains the following attribute information: the map width number where the road is located, road user number, connection start and end node user number, connection end node user number, road name, road alias, road grades, functional class, road composition, road type, number of lanes, road width, road all and administrative division code where the road is located. Road grades range from 1 to 11, where 1 represents minor roads, 2 represents the county and rural internal roads, 3 represents township roads, 4 represents county roads, 5 represents provincial roads, 6 represents the urban ordinary roads, 7 represents the urban secondary roads, 8 represents urban trunk roads, 9 represents urban expressway, 10 represents national highway, and 11 represents expressway. Road function is classified into six levels: Level 1 is Expressway, Level 2 is National Road, Level 3 is Provincial Road, Level 4 is County Road, Level 5 is Township Road, and Level 6 is Village or Local Road. Road composition ranges from 1 to 16, 1 for up and down lane separation, 2 for within the intersection, 3 for JCT (highway connector), 4 for traffic circle, 5 for service area, 6 for approach road, 7 for auxiliary road, 8 for approach road plus JCT, 9 for the exit, 10 for entrance, 11 for right turn lane A, 12 for right turn lane B, 13 for left turn lane A, 14 for left turn lane B, 15 for general road, 16 for left and right turn lanes, and 17 is the non-motorized lane borrowing lane. Secondly, transportation data, such as railways and subway lines, was extracted from OpenStreetMap website. Thirdly, this study collected housing price data using web-scraping techniques from <http://lianjia.com/> and <http://ke.com/>, including various types of data such as neighborhood housing prices, second-hand housing prices, rental housing prices, and new housing prices, containing attributes of price levels and latitude and longitude coordinates. The collected housing price data for Wuhan was transformed to approximate a normal distribution, revealing a trend of higher housing prices in the city center and lower prices in the surrounding areas. Ordinary Kriging interpolation was selected for spatial interpolation of this data.

2.3. Transfer Learning Models

The fine-tuning method in transfer learning as depicted in Figure 3 shows the architecture of the transfer learning-based safety perception prediction model. First, due to the influence of shooting time (sunrise, sunset), weather (sunny, cloudy), and change in viewpoint, the street scene dataset can cause blurred objects, no details of objects, shaded objects, or no obvious scenes, etc. At the same time, it is difficult to use the labeled samples directly for deep learning training because of the unbalanced and confused sample categories, so it is necessary for data enhancement on the labeled street scene images [41,42]. Therefore, data enhancement is needed on the labeled street scene images, including undersampling, increasing or decreasing brightness, cropping, etc. Then, we fine-tuned the VGG (Visual Geometry Group) CNN model pre-training on the Places365 database (VGG-PLACE365) [43] by changing the hyperparameter setting of the last fully connected layer by fixing all layers before the fully connected layer 7 in the VGG-PLACE model. It outputs the model with fixed hyperparameters and weights and reads the original image data through the network as the features of the bottleneck layer corresponding to the model for data input. Then the following occurs: outputting the picture data as the data from the bottleneck layer to the next reinitialization layer, designing the network structure of the last layer, and finally calculating the loss through the Softmax function and outputting the prediction probability. This study used the TensorFlow deep learning framework, which provides a large number of tools for deep learning training, evaluation, and prediction [44] and implements conversion between CPU/GPU, providing Python 3.6, C++ 11, and Matlab R2018a interfaces. The hardware configuration of the training process is Intel CPU E3-1240 v5, 3.5 GHz, CPU training.

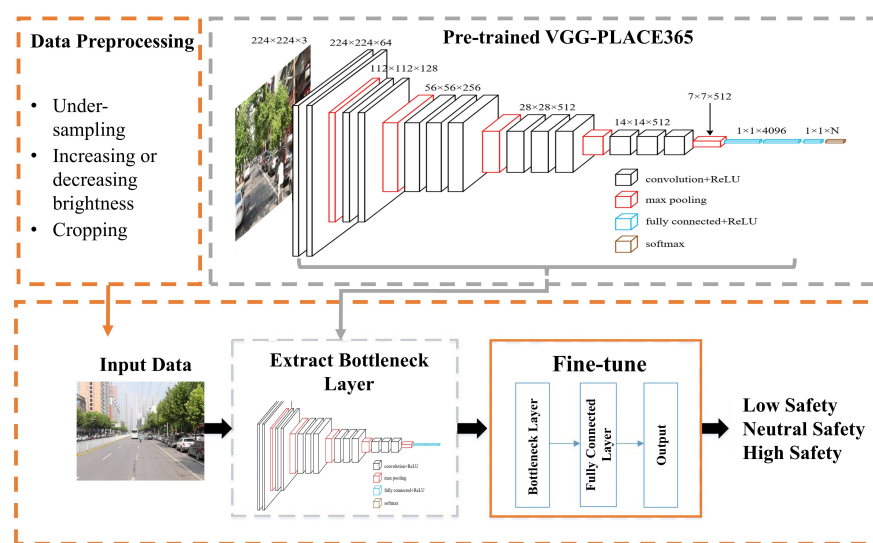


Figure 3. Architecture of transfer learning-based safety perception prediction model.

The dataset was randomly selected and divided into three categories: training set, validation set, and test set, accounting for 70%, 10%, and 20%, respectively, comprising 4456 street locations with a total of 35,648 street images. After data pre-processing, the dataset consisted of 23,582 street images, of which 8343 images represented low safety, 7471 images were classified as neutral, and 7768 images depicted high safety. The training, validation, and test sets were independent of each other, with no duplicates.

The model's accuracy was evaluated as an optimization index or performance evaluation metric. Predictions were made using the Softmax classifier on the test set, where higher accuracy indicated better classification performance. Overfitting during model training

was assessed through loss analysis; if the loss rate leveled off and approached zero, the model was considered well-converged and reached a local optimal solution.

Based on the analysis of the transfer learning strategy, learning rate and batch size critically affect training speed and ultimate accuracy. Various learning rates and sample configurations were tested to optimize training time for fine tuning. A subset of 4000 samples was selected for initial training, with a balanced ratio of 1:1:1 among low-safety, neutral, and high-safety samples. During tuning, the training, validation, and test set ratios were adjusted to refine parameter settings.

After extensive experimentation and parameter tuning, the final split ratio for the training, validation, and test sets was determined to be 70%, 10%, and 20%, respectively. In this study, fine tuning was applied to the task-specific dataset of the street view safety perception dataset. Although this dataset focuses on social perception, it shares a similar feature distribution with the Places dataset due to its use of outdoor street imagery. As a result, only a small portion was allocated to the validation set to monitor accuracy differences between training and validation phases, enabling timely adjustments to mitigate overfitting. Given the high inter-class similarity within the street safety perception data, a larger training set was employed to enhance the model's capacity to effectively learn task-specific features.

Once the training, validation, and test set ratios were finalized, the learning rate and batch size were optimized, resulting in a final training configuration of 4000 steps using a smaller sample dataset containing 4000 images. Table 1 presents the accuracy results obtained under different combinations of learning rates and batch sizes.

Table 1. Training effects of small samples with different hyperparameter combinations.

Rank	Learning Rate	Batch Size	Accuracy	Overfitting
1	0.0005	100	69.4%	No
2	0.00001	100	65.3%	Yes
3	0.0001	100	64.5%	No
4	0.001	100	60.1%	No
5	0.01	100	55%	No

Table 1 shows results at a batch size of 100, with the last column assessing overfitting by comparing training and validation accuracy curves. Overfitting is indicated by a large accuracy gap or abnormal patterns in either curve. A learning rate of 0.0005 was identified as optimal and used for further training and evaluation of the full street safety rating dataset.

2.4. Statistical Methods

The probability of scene identification for the classified dataset is based on the scene classification in Place 365 [45,46]. The predicted scenes for each image were extracted. The model predicted and processed 28,173 streets with sampling points, and a total of 43,232 street points were assigned a predicted safety level across eight different orientations. Based on this, the final prediction dataset comprises 345,856 images. To ensure reliability, any scene with a predicted class probability below 0.9 was excluded, as low-confidence predictions may be inaccurate. The best-recognized image is selected as the statistical object. Finally, the street scene with the highest probability score will have a series of scene attribute descriptions. The text description based on this scene attribute surface is the Sun attribute dataset to determine which location of the image will make people feel safe in the

place (the street scene image corresponds to the safety level label 3) [47]. The feature heat maps were generated using the class activation map CAM technique [48].

Spatial distribution pattern analysis and analysis of the influence of related factors were conducted on the perception of street safety in seven administrative districts. Tools from ArcGIS 10.2 exploratory analysis, spatial statistical methods, spatial analysis methods, and metric spatial distribution methods were used to conduct the spatial distribution analysis of street safety perceptions in the seven central urban areas of Wuhan City. Geoda software v1.1 was used to analyze spatial autocorrelation, clustering, and regression for each urban area in the seven central urban areas of Wuhan. The tools are applied including Getis-Ord G_i^* statistics, global and local Moran I, and buffer analysis. IBM SPSS Statistics 25.0 was used to analyze different influencing factors within the seven central urban areas of Wuhan City in the whole experimental area.

3. Results and Discussion

3.1. Performance of the Transfer Learning Model

The street images with street safety level were trained with a learning rate of 0.0005, training steps of 4000, training batch size of 100, validation batch size of 100, and validation interval of 100 steps. The final test set accuracy reached 78.3%. As shown in Figure 4, the training accuracy (orange line) and validation accuracy (blue line) gradually increase and stabilize, with a small gap indicating good model performance. Similarly, the loss curves for both sets decrease and stabilize without significant divergence or upward trends, suggesting that the model is not overfitting and effectively learns features for street safety rating.

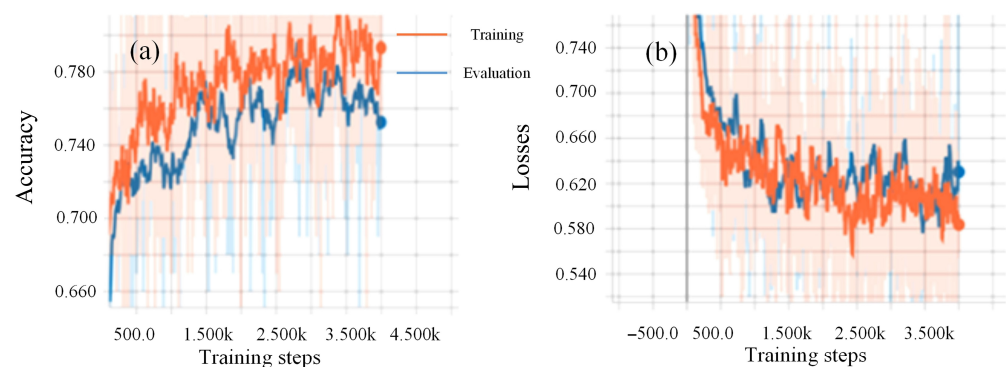


Figure 4. (a) Accuracy and (b) losses.

Regarding the test set of 4717 sheets for the confusion matrix representation, it is shown in Figure 5. The diagonal line indicates the producer accuracy for each safety perception category. The values outside the diagonal line indicate the proportion of a category predicted to be that category. Each value in the horizontal row except the diagonal is classified into other categories; the vertical row indicates that a different category was predicted to be this category. Figure 5 presents the confusion matrix, where street safety categories are arranged from top to bottom as neutral, high, and low. The model performs well in distinguishing high and low-safety categories, with mutual misclassifications under 200. Most misclassifications occur in the neutral category, which is frequently confused with both high and low safety with 200 to 400 such errors. This suggests that the perceptual similarities among the three classes, especially involving neutral cases, contribute to reduced classification accuracy. Given the subjective nature of safety perception, clear boundaries between categories are inherently difficult to define.

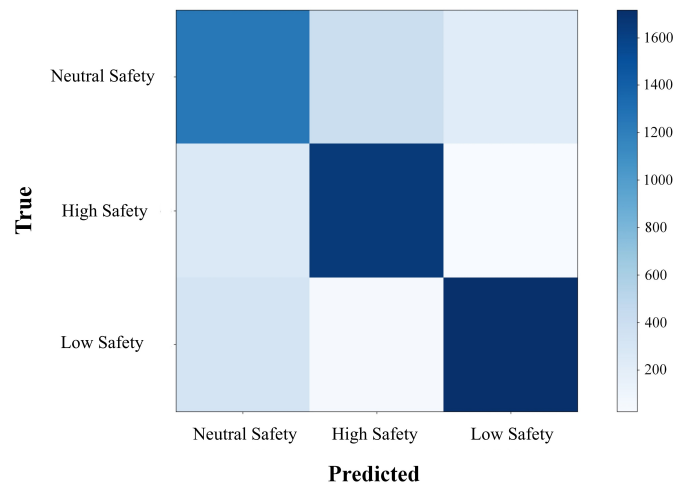


Figure 5. Confusion matrix, where neutral safety, high safety, and low safety are represented.

3.2. Explanation of Urban Safety Prediction

Sample images with the highest probability scores for each safety category are shown in Figure 6. By directly observing images that represent the highest, neutral, and lowest levels of perceived safety, one can discern the variations in the sense of safety across different scenarios. On wide and well-maintained roads, the level of street safety is generally high. Such streets are typically lined with residential buildings, featuring a clean and orderly environment with smooth traffic flow, which evokes a sense of security and comfort. Streets with a neutral level of safety are predominantly located in residential areas, characterized by lush greenery and significant shading from trees. Although these areas are relatively quiet, the limited visibility due to the dense foliage may create a sense of seclusion, thereby influencing the perception of safety. Conversely, streets with a low level of perceived safety are often found in dilapidated corners, marked by abandoned buildings, accumulated garbage, and inadequate lighting. These factors significantly diminish people's sense of safety and may even evoke feelings of unease.

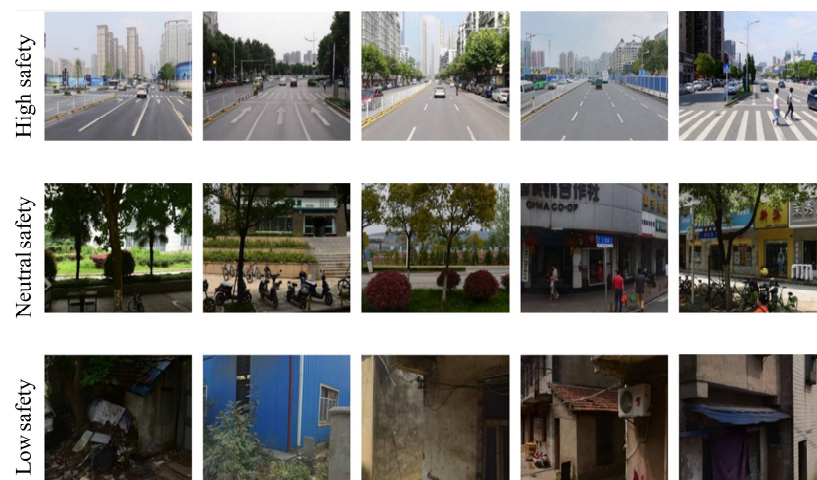


Figure 6. Sample street view images of urban safety prediction.

Statistics of scenarios associated with different safety levels and CAM visualizations reveal how the model internally distinguishes between perceived safety levels by scene category and the spatial and visual structure within each scene. As can be seen from Table 2, the ratio of different scenes is calculated according to the ratio of the number of predicted scenes to the images of that type of scene. The different scenes are ranked according to the

calculated ratio. Only the top three scenes that are predicted are listed in Table 2. As shown in Table 2, different scene types are associated with different predicted safety levels. For low safety, the most frequent scenes include gas stations, slum-like areas, and bus stations. These scenes often feature disordered open spaces, unpaved roads, temporary structures, or poor visibility elements that tend to reduce perceived safety [49]. Figure 7 further illustrates that such environments lack greenery, pedestrians, or clear road infrastructure, which are key indicators of spatial neglect and reduced social surveillance [50]. CAM heatmaps show that the model attends to these degraded or ambiguous features, such as broken fences, construction zones, or empty foregrounds, suggesting its sensitivity to environmental disorder.

Table 2. Scenarios associated with different safety levels.

Low Safety		Neutral Safety		High Safety	
Scene	Proportion	Scene	Proportion	Scene	Proportion
gas_station	27%	gas_station	37%	crosswalk	50%
slum	18%	bus_station	23%	highway	27%
bus_station	16%	highway	10%	bus_station	11%

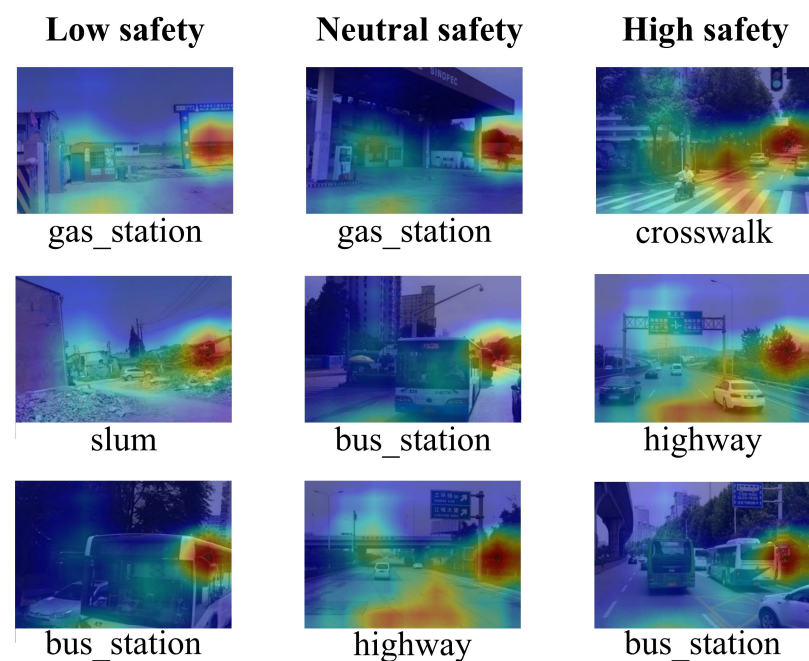


Figure 7. Example of CAM with images of different categories of safety.

In contrast, scenes associated with high safety commonly include crosswalks and well-maintained urban roads. These images typically contain clear road markings, visible traffic signals, moving vehicles, and human activity. The CAM visualizations confirm that the model focuses on these structured and socially active elements. Such attention patterns align with the urban design literature, which emphasizes order, visibility, and social presence as positive predictors of safety perception [51].

Some categories such as bus stations and highways appear in both low or neutral and high-safety predictions. This indicates that scene type alone does not determine perception; instead, the quality and context of the visual environment—including lighting, cleanliness, openness, and activity—play a crucial role. For example, highways with greenery and clear spatial orientation may be judged as safe [52], while those lacking context or environmental support are often considered neutral or ambiguous.

3.3. Spatial Analysis of Urban Safety

First, the model trained with transfer learning techniques predicted and processed safety perception based on 43,232 street viewpoints. The average safety score at the viewpoint level was 1.94 (standard deviation = 0.58) and 1.82 (standard deviation = 0.55) when averaged by street. Overall, safety perception in Wuhan's seven central districts is moderate, reflecting uneven urban development. Each street was assigned a score based on at least one viewpoint, with most streets linked to a single viewpoint. For streets with more than one, the average of directional scores was used. Due to short road lengths, averaging had little effect on the overall safety evaluation.

Secondly, the spatial distribution of street safety based on the standard deviation ellipse obtained is shown in Figure 8. For the mapping of street-level perceived safety, the safety scores were classified into five distinct categories using the Natural Breaks (Jenks) method. The classification ranges are as follows: Safety Level 1 (Low): scores between 1.000 and 1.333; Safety Level 2 (Sub-Low): scores between 1.333 and 1.775; Safety Level 3 (Neutral): scores between 1.775 and 2.187; Safety Level 4 (Sub-High): scores between 2.187 and 2.550; and Safety Level 5 (High): scores between 2.550 and 3.000.

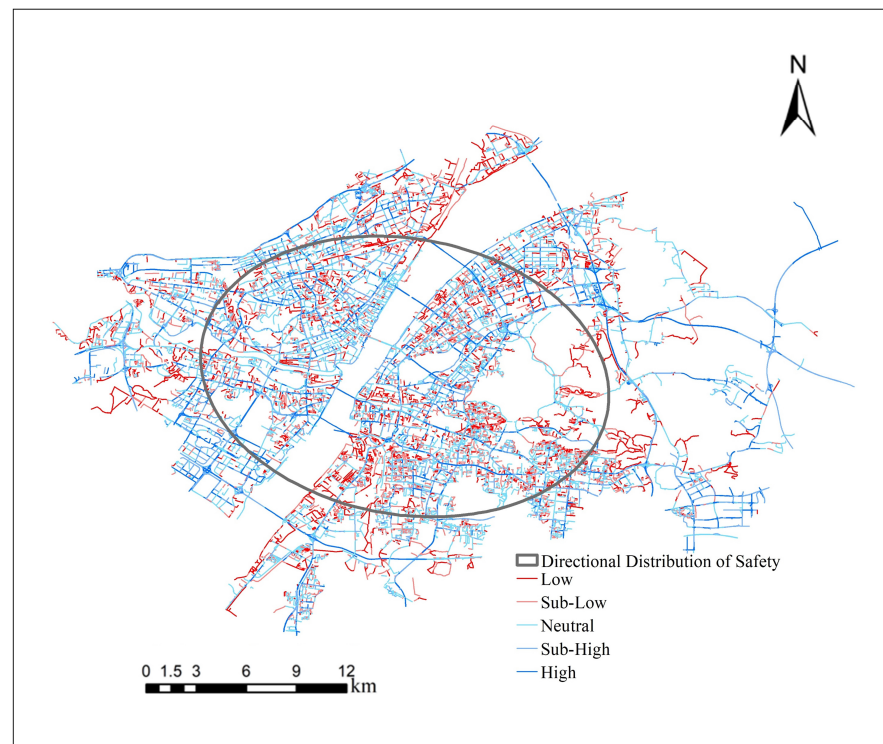


Figure 8. Spatial distribution of street safety perception based on standard deviation ellipse.

As shown in Figure 8, the spatial distribution of perceived street safety in Wuhan is dispersed and primarily aligned along a northwest–southeast direction. Streets located west of the Yangtze River generally exhibit higher safety perception compared to those on the eastern side. In Wuchang, Hongshan, and Qingshan districts east of the Yangtze River, higher safety perception is primarily concentrated along specific streets. In contrast, Jiangan District, located west of the Yangtze, shows a more scattered distribution, unlike the more clustered patterns observed in Qiaokou, Jiangnan, and Hanyang districts. Figure 8 illustrates that the distribution of perceived street safety in Wuhan's seven central urban districts primarily follows the city's major transportation corridors. Key thoroughfares crossing the Yangtze River, as well as arterial roads connecting urban districts, exhibit notably high predicted safety scores.

Figure 9a presents a hotspot analysis of perceived urban safety based on statistical confidence levels. Deep red areas indicate significant hotspots at the 99% confidence level, highlighting regions with a dense concentration of high safety perception. These hotspots are not randomly scattered but tend to cluster in specific urban zones—such as the Central Business District in Jianghan, Xin-Si in Hanyang, Optics Valley in Hongshan, traditional residential areas in Qingshan, and riverside areas in Wuchang. These zones often share features like strong economic activity, urban renewal, or stable residential populations, which may contribute to their higher perceived safety.

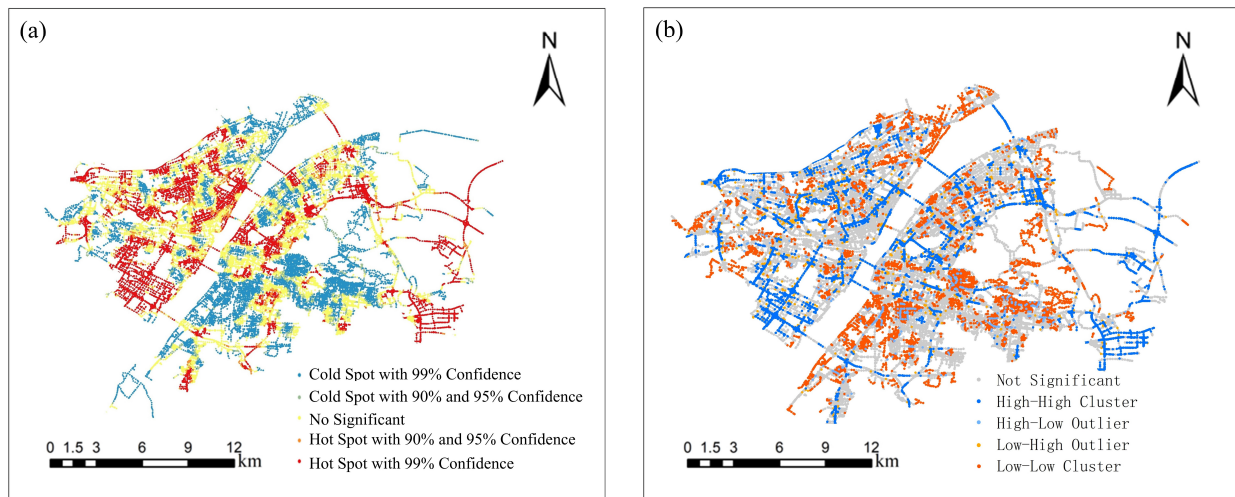


Figure 9. (a) Hotspot analysis map and (b) LISA map of safety perception.

Figure 9b shows the local spatial correlation index (LISA) as the local Moran's I index. The global Moran's I of 0.465 ($p \approx 0$, $z = 111.8$) indicates significant positive spatial autocorrelation, showing that street safety perceptions in Wuhan cluster geographically rather than occurring randomly. The spatial clustering patterns are generally consistent with the results of hotspot analysis, reinforcing the spatial heterogeneity of urban safety in Wuhan. In the LISA map, blue-shaded areas denote high–high clusters and high–low outliers, while red-shaded areas indicate low–low clusters and low–high outliers. Gray areas are statistically insignificant and do not exhibit any meaningful spatial association. Notably, the Xin-Si area in Hanyang District, identified as a hotspot, does not display consistent high-value clustering in the LISA results as some road segments appear statistically insignificant. This divergence may be attributed to the transitional urban structure of the Xin-Si area, which represents an emergent zone of Wuhan's new urbanization strategy. Similar spatial patterns can be observed in other districts, including most parts of Jianghan District, the border region of Qiaokou and Jianghan, newly developed areas in Hanyang, parts of Jiang'an near Jianghan, educational zones in Qingshan, riverside areas in Wuchang, and urban centers of Hongshan District. These areas exhibit statistically significant high–high or low–low clusters of safety perception, reflecting socio-spatial polarization.

Both the hotspot and LISA analyses indicate a clear and spatially segmented pattern of urban safety, with high-safety clusters forming cohesive, regionally concentrated zones. This suggests that, while urban safety perception tends to cluster in particular regions, there is considerable variation in local perceptions within these broader zones, revealing underlying socio-spatial dynamics and uneven urban development.

The spatial distribution of these hotspots of safety streets suggests that both historical urban cores (e.g., Jianghan district) and newly urbanized districts (e.g., Optics Valley area in Hongshan District) can serve as focal points of safety perception, depending on their level of infrastructure development, economic activity, and community stability. In

contrast, areas with lower levels of perceived safety tend to be less concentrated and more fragmented, often located in peripheral or transitional zones with weaker infrastructure or social cohesion. Specifically, high–low and low–high safety spatial outliers suggest the presence of internal spatial heterogeneity and potential transitional areas.

Finally, Figure 10 illustrates the road attributes with perceived safety. Figure 10a–d show that street safety perception varies with road grade, function, composition, and width. The horizontal coordinates in Figure 10a indicate that there are differences in the street safety perception levels of different levels of roads. Overall, the average value of street safety perception of expressways (road grades 11), national highways (road grades 10), urban expressways (road grades 9), urban trunk roads (road grades 8), and provincial roads (road grades 5) is relatively high, and the average value of street safety perception of minor roads (road grades 1) and urban ordinary roads (road grades 6) exhibit relatively lower perceived safety values. This suggests that roads with higher administrative or functional importance tend to be perceived as safer. Figure 10b presents the relationship between street safety perception and road functional level, where a higher numeric value indicates a lower functional class. It can be observed that roads classified as provincial, county, township, and village roads (functional levels 3 to 6) generally correspond to higher perceived safety values. This implies that, while expressways and national highways show high perceived safety by grade, in terms of functional hierarchy, moderately ranked roads also exhibit a strong association with perceived safety. The horizontal coordinates in Figure 10c are the road composition. The statistics conclude that the additional roadway intersections or general intersections constructed through different roadway compositions do not significantly impact the perception of street safety. Figure 10d shows the road width, and, overall, the mean value of the perception of street safety increases against the increase in road width, possibly due to better traffic separation and more organized lane structures.

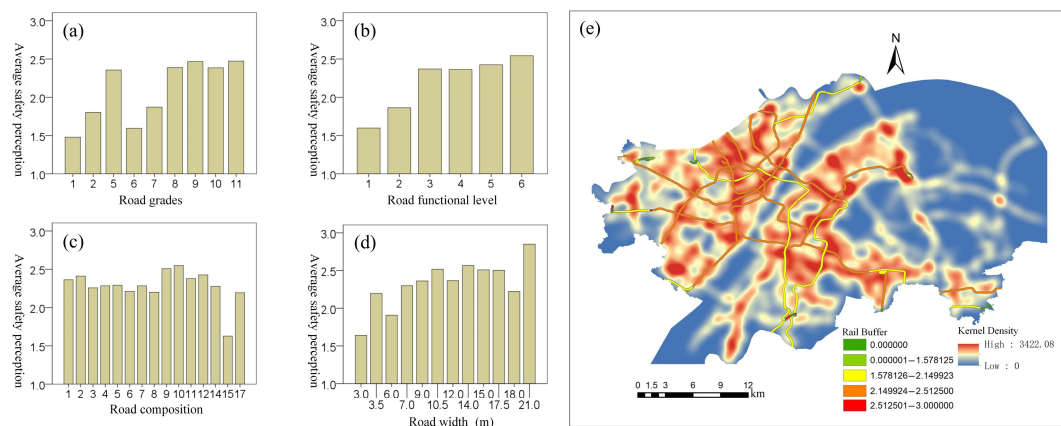


Figure 10. (a–d) Average safety perception level among road attributes. (e) Safety kernel density map with different safety values on the rail buffer zone.

To further explore the relationship between rail transit and perceived street safety, a buffer zone analysis was conducted. Buffer areas were created at 200 m intervals around rail transit lines, and the average street safety perception values within these buffers were calculated, as shown in Figure 10e. The analysis reveals that, in areas with high street density, the mean street safety perception values within the rail transit buffer zones are generally higher than the average for the entire study area. Specifically, the average perception scores in these areas mostly fall within the range of 2.15 to 2.5. Although the buffers surrounding Rail Transit Line 7 and Line 8 fall within a slightly lower range of 1.58 to 2.14, their specific scores of 2.07 and 2.14, respectively, are still above the overall mean value for the study area.

The findings suggest that the presence of rail transit infrastructure with higher-grade roadways and well-designed road function might exert a positive influence on how safe people feel in the surrounding street environment. Based on the kernel density map, the spatial distribution of street safety perception across the seven administrative districts of Wuhan is primarily concentrated along the city's major arterial roads. Notably, most rail transit lines are located within high-density areas of street safety perception. These regions of general traffic convenience typically correspond to areas with dense populations, robust economic development, and well-established infrastructure. These findings suggest that the streets adjacent to rail transit lines are associated with a higher level of perceived safety. These findings suggest that the streets adjacent to rail transit lines are associated with a higher level of perceived safety.

3.4. Estimation of Urban Safety

Table 3 presents the results of the Pearson correlation analysis between various variables and perceived street safety. The findings reveal that road grade ($r = 0.425$, $p < 0.01$), road functional level ($r = 0.530$, $p < 0.01$), and road width ($r = 0.459$, $p < 0.01$) are all positively correlated with perceived street safety, with correlations significant at the 0.01 level. Among them, the correlation between road functional level and perceived safety is the highest, suggesting that streets classified with lower functional classes are generally perceived as safer by the public.

Table 3. Correlation analysis of variables and perception of safety.

Variables	Pearson Correlation	Number
Road grade	0.425 **	
Road function level	0.530 **	28,173
Road width	0.459 **	
House price	−0.019 **	43,232

** Significant correlation at the 0.01 level (two-tailed).

Road grade and road width also demonstrate moderate positive correlations with safety perception, indicating that wider roads and roads of higher classification tend to enhance the public's sense of safety. However, it is important to note that these correlations, while significant, are not particularly strong (r values are all below 0.6), implying that additional factors beyond road attributes also influence safety perception.

In contrast, housing price shows a very weak negative correlation ($r = -0.019$) with perceived safety. Given its small effect size, this association is not robust and likely reflects the complexity of socioeconomic perception rather than a direct relationship.

3.5. Discussion and Policy Implications

Although the concept of sense of safety originated from psychological studies, it has gradually evolved into a sociological concern, especially during the rapid urbanization process of the 20th century [5]. As cities continue to grow in both population and complexity, the relationship between urban physical development and psychological well-being has become increasingly evident [10]. How to capture residents' perceptions of safety in a scale and cost-efficient way aiming to inform urban planning and policy formulation has been challenging, and, thus, there exists great room for further development of perception measurement and analysis. This paper proposes an automated framework to analyze perceived safety in urban districts by integrating street-level imagery, transfer learning techniques, and spatial analysis tools, which allows for a macro-level view of how safety

perception correlates with urban factors such as transportation networks, road hierarchies, and housing prices.

Grounded in Place Theory [15], this research emphasizes the role of physical urban elements related to sense responses to place using advanced technologies. Based on this, a sense of place and sense of safety are thus closely linked, as perception evaluations become part of how a location is defined [17]. The sense of urban safety is not merely a reaction to crime or disorder [53] but also a reflection of how people interpret and experience their environments [10] to enrich Place Theory. The spatial patterning underscores how targeted urban design interventions at the street level can significantly influence human perception [54]. Recognizing these perceptions as integral to place making supports a human-centered approach to urban design, where human perception contributes to spatial quality and sustainable urban development.

This study also offers actionable policy insights. First, the integration of big data and artificial intelligence, particularly transfer learning-based models, can support urban governance. These technologies allow for large-scale, efficient, and dynamic assessment of urban safety perception without relying solely on costly or time-consuming field surveys. By combining street imagery, spatial data, and machine learning, urban managers can gain a more holistic understanding of safety conditions and make evidence-based policy decisions. Beyond perception prediction, such automated approaches offer practical support for decision making in urban planning processes. At this stage, attention must also be given to the robustness and reliability of the models to ensure trustworthy outcomes.

Second, to support inclusive and livable cities, safety perception should be integrated into urban planning. Policy actions should include perception-based indicators in planning systems [26], prioritize road design improvements, and transportation accessibility. Creating safe places should address both technical and perception needs, supporting a human-centered approach aligned with sustainable urban development goals. Spatial statistical analysis reveals that perceived safety levels in Wuhan's seven central administrative districts exhibit a clear pattern of piecewise spatial clustering, with high-value and low-value zones forming distinct and contiguous regions rather than being evenly or randomly distributed throughout the city shown in Figure 9. From a sustainability standpoint, addressing such unevenness in perceptual safety requires targeted, street-level data analysis tailored to neighborhood-specific conditions [55,56]. This ensures that urban development enhances perceived safety equitably across space, avoiding the deepening of spatial unevenness.

The proposed method holds both potential and limitations for broader applications in other urban contexts. Although this study focuses on Wuhan, the proposed framework is transferable to other urban contexts to a certain extent. The model relies on widely available data sources (e.g., street-level imagery, urban spatial data) and can be adapted through fine tuning using local perception data. By fine tuning the model with locally sourced imagery and integrating city-specific urban features, the framework can be adapted to different urban contexts while maintaining methodological consistency. Some case studies have explored perception-based analysis in diverse urban settings [57,58], reinforcing the importance of incorporating human perception into urban planning with advanced methods. These studies demonstrate that safety perception varies across different cultural and spatial contexts but can still be systematically analyzed through similar methodological frameworks. However, this study has several limitations. First, the findings may be influenced by Wuhan's unique spatial structure, road hierarchy, and cultural perceptions of public safety. Differences in urban morphology, data availability (e.g., street-level imagery coverage), and local interpretations of safety could affect the model's performance and generalizability. Second, the annotators' judgments reflect perceptions rooted in a specific

sociocultural background, and their applicability to cities with different spatial structures or social environments may be limited. To address this, future work should apply the framework to multiple cities with diverse urban forms and cultural settings to evaluate its generalizability by expanding the annotation process via crowdsourcing platforms. Third, while the observed correlations (e.g., with road infrastructure or housing prices) suggest meaningful spatial patterns, they do not imply causality. Further studies could incorporate additional modeling techniques and control for confounding variables to improve causal inference.

4. Conclusions

In the context of human-centered sustainable cities, this study constructs a street safety perception dataset for Wuhan by labeling street scenes into three safety levels. A CNN based on the VGG-PLACES model was trained through transfer learning to predict safety perception, achieving an accuracy of 78.3%. By mapping the spatial distribution of safety perception, this study highlights how physical characteristics of places like a street aligned with social and emotional labels can identify the street-level safety distribution, thus promoting environments that are not only functional but also perceived as safe. Spatial statistical analysis revealed that perceived safety exhibits spatial clustering, with high-value and low-value areas interwoven. Correlation analysis showed that road attributes were moderately positively associated with perceived safety, while housing prices showed a weak negative correlation. This research demonstrates the feasibility of linking urban spatial features with psychological perceptions through computer vision. Future work should expand sample diversity, optimize model performance by multi-cities studies, and incorporate more influencing factors with non-linear relationship analysis to better understand the interaction between human perception and the urban environment.

Author Contributions: Conceptualization, Z.-R.T.; Methodology, Y.C.; Software, Y.C.; Validation, Y.C.; Writing—original draft, Y.C.; Funding acquisition, Z.-R.T. All authors have read and agreed to the published version of the manuscript.

Funding: This work is partly support by the National Natural Science Foundation of China (Grant No. 62401226); partly supported by the Fundamental Research Funds for the Central Universities (Grant No. 21624357); partly supported by Jinan University Special Project for Quality Enhancement and Upgrading of Experimental Teaching Reform (Grant No. 82625033); partly supported by Jinan University 2025 Annual “Artificial Intelligence +” Educational Reform Research Project (Grant No. 82625693).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

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

References

1. Dubey, A.; Naik, N.; Parikh, D.; Raskar, R.; Hidalgo, C.A. *Deep Learning the City: Quantifying Urban Perception at a Global Scale*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 196–212.
2. 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](#)]
3. Naik, N.; Philipoom, J.; Raskar, R.; Hidalgo, C. Streetscore-predicting the perceived safety of one million streetscapes. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Columbus, OH, USA, 23–28 June 2014; pp. 779–785.

4. Porzi, L.; Rota Bulò, S.; Lepri, B.; Ricci, E. Predicting and understanding urban perception with convolutional neural networks. In Proceedings of the 23rd ACM International Conference on Multimedia, New York, NY, USA, 26–30 October 2015; pp. 139–148.
5. Ordonez, V.; Berg, T.L. Learning high-level judgments of urban perception. In Proceedings of the Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, 6–12 September 2014; Proceedings, Part VI 13; Springer: Berlin/Heidelberg, Germany, 2014; pp. 494–510.
6. Lindal, P.J.; Hartig, T. Architectural variation, building height, and the restorative quality of urban residential streetscapes. *J. Environ. Psychol.* **2013**, *33*, 26–36. [[CrossRef](#)]
7. Kelling, G.L.; Coles, C.M. *Fixing Broken Windows: Restoring Order and Reducing Crime in our Communities*; Simon and Schuster: New York, NY, USA, 1997.
8. Cheng, Y.; Zhang, J.; Wei, W.; Zhao, B. Effects of urban parks on residents' expressed happiness before and during the COVID-19 pandemic. *Landsc. Urban Plan.* **2021**, *212*, 104118. [[CrossRef](#)]
9. Doran, D.; Severin, K.; Gokhale, S.; Dagnino, A. Social media enabled human sensing for smart cities. *AI Commun.* **2016**, *29*, 57–75. [[CrossRef](#)]
10. Giannico, V.; Spano, G.; Elia, M.; D'Este, M.; Sanesi, G.; Laforteza, R. Green spaces, quality of life, and citizen perception in European cities. *Environ. Res.* **2021**, *196*, 110922. [[CrossRef](#)] [[PubMed](#)]
11. Glaeser, E. Cities, productivity, and quality of life. *Science* **2011**, *333*, 592–594. [[CrossRef](#)] [[PubMed](#)]
12. Ulrich, R.S. Visual landscapes and psychological well-being. *Landsc. Res.* **1979**, *4*, 17–23. [[CrossRef](#)]
13. Ito, K.; Kang, Y.; Zhang, Y.; Zhang, F.; Biljecki, F. Understanding urban perception with visual data: A systematic review. *Cities* **2024**, *152*, 105169. [[CrossRef](#)]
14. Jeffery, C.R. Crime prevention through environmental design. *Am. Behav. Sci.* **1971**, *14*, 589. [[CrossRef](#)]
15. Relph, E. *Place and Placelessness*; Pion: London, UK, 1976.
16. Gómez, F.; Torres, A.; Galvis, J.; Camargo, J.; Martínez, O. Hotspot mapping for perception of security. In Proceedings of the 2016 IEEE International Smart Cities Conference (ISC2), Trento, Italy, 12–15 September 2016; pp. 1–6.
17. Cresswell, T. *Place: An Introduction*; John Wiley & Sons: New York, NY, USA, 2014.
18. Cresswell, T.J. In *Place/Out of Place: Geography, Ideology and Transgression*; The University of Wisconsin-Madison: Madison, WI, USA, 1992.
19. Nasar, J.L. The evaluative image of the city. *J. Am. Plan. Assoc.* **1990**, *56*, 41–53. [[CrossRef](#)]
20. Schroeder, H.W.; Anderson, L.M. Perception of personal safety in urban recreation sites. *J. Leis. Res.* **1984**, *16*, 178–194. [[CrossRef](#)]
21. Goodchild, M.F. Citizens as sensors: The world of volunteered geography. *GeoJournal* **2007**, *69*, 211–221. [[CrossRef](#)]
22. Goodchild, M.F. Formalizing place in geographic information systems. In *Communities Neighborhoods, and Health*; Springer: New York, NY, USA, 2010; pp. 21–33.
23. Porta, S.; Renne, J.L. Linking urban design to sustainability: Formal indicators of social urban sustainability field research in Perth, Western Australia. *Urban Des. Int.* **2005**, *10*, 51–64. [[CrossRef](#)]
24. 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](#)]
25. Hofman, J.M.; Watts, D.J.; Athey, S.; Garip, F.; Griffiths, T.L.; Kleinberg, J.; Margetts, H.; Mullainathan, S.; Salganik, M.J.; Vazire, S. Integrating explanation and prediction in computational social science. *Nature* **2021**, *595*, 181–188. [[CrossRef](#)]
26. Ji, T.; Chen, J.-H.; Wei, H.-H.; Su, Y.-C. Towards people-centric smart city development: Investigating the citizens' preferences and perceptions about smart-city services in Taiwan. *Sustain. Cities Soc.* **2021**, *67*, 102691. [[CrossRef](#)]
27. Kong, S.; Shen, X.; Lin, Z.; Mech, R.; Fowlkes, C. Photo aesthetics ranking network with attributes and content adaptation. In Proceedings of the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Proceedings, Part I 14. Springer: Berlin/Heidelberg, Germany, 2016; pp. 662–679.
28. Molina, M.; Garip, F. Machine learning for sociology. *Annu. Rev. Sociol.* **2019**, *45*, 27–45. [[CrossRef](#)]
29. Moustafa, K. Make good use of big data: A home for everyone. *Cities* **2020**, *107*, 102903. [[CrossRef](#)] [[PubMed](#)]
30. Anguelov, D.; Dulong, C.; Filip, D.; Frueh, C.; Lafon, S.; Lyon, R.; Ogale, A.; Vincent, L.; Weaver, J. Google street view: Capturing the world at street level. *Computer* **2010**, *43*, 32–38. [[CrossRef](#)]
31. Less, E.L.; McKee, P.; Toomey, T.; Nelson, T.; Erickson, D.; Xiong, S.; Jones-Webb, R. Matching study areas using Google Street View: A new application for an emerging technology. *Eval. Program Plan.* **2015**, *53*, 72–79. [[CrossRef](#)]
32. Jia, J.; Zhang, X.; Huang, C.; Luan, H. Multiscale analysis of human social sensing of urban appearance and its effects on house price appreciation in Wuhan, China. *Sustain. Cities Soc.* **2022**, *81*, 103844. [[CrossRef](#)]
33. Koo, B.W.; Guhathakurta, S.; Botchwey, N. How are neighborhood and street-level walkability factors associated with walking behaviors? a big data approach using street view images. *Environ. Behav.* **2022**, *54*, 211–241. [[CrossRef](#)]
34. Wan, J.; Wang, D.; Hoi, S.C.H.; Wu, P.; Zhu, J.; Zhang, Y.; Li, J. Deep learning for content-based image retrieval: A comprehensive study. In Proceedings of the 22nd ACM International Conference on Multimedia, New York, NY, USA, 3–7 November 2014; pp. 157–166.

35. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)]
36. Donahue, J.; Jia, Y.; Vinyals, O.; Hoffman, J.; Zhang, N.; Tzeng, E.; Darrell, T. Decaf: A deep convolutional activation feature for generic visual recognition. In Proceedings of the International Conference on Machine Learning, PMLR, Beijing, China, 21–26 June 2014; pp. 647–655.
37. Oquab, M.; Bottou, L.; Laptev, I.; Sivic, J. Learning and transferring mid-level image representations using convolutional neural networks. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1717–1724.
38. Deng, L.; Yang, M.; Qian, Y.; Wang, C.; Wang, B. CNN based semantic segmentation for urban traffic scenes using fisheye camera. In Proceedings of the 2017 IEEE Intelligent Vehicles Symposium (IV), Los Angeles, CA, USA, 11–14 June 2017; pp. 231–236.
39. Wang, S.-Y.; Wang, O.; Zhang, R.; Owens, A.; Efros, A.A. CNN-generated images are surprisingly easy to spot. for now. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 8695–8704.
40. Jacobs, J. *The Death and Life of Great American Cities*; Random House: New York, NY, USA, 1961; Volume 21, pp. 13–25.
41. Pan, S.J.; Yang, Q. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **2009**, *22*, 1345–1359. [[CrossRef](#)]
42. Torrey, L.; Shavlik, J. Transfer Learning. In *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*; IGI global: Hershey, PA, USA, 2010; pp. 242–264.
43. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. In Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015), San Diego, CA, USA, 7–9 May 2015.
44. Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M. TensorFlow: A System for Large-Scale Machine Learning. In Proceedings of the 12th USENIX symposium on operating systems design and implementation (OSDI 16), Savannah, GA, USA, 2–4 November 2016; pp. 265–283.
45. Zhou, B.; Lapedriza, A.; Torralba, A.; Oliva, A. Places: An image database for deep scene understanding. *J. Vis.* **2017**, *17*, 296. [[CrossRef](#)]
46. Zhou, B.; Lapedriza, A.; Xiao, J.; Torralba, A.; Oliva, A. Learning deep features for scene recognition using places database. *Adv. Neural Inf. Process. Syst.* **2014**, *27*, 487–495.
47. Patterson, G.; Hays, J. Sun attribute database: Discovering, annotating, and recognizing scene attributes. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 2751–2758.
48. Zhou, B.; Khosla, A.; Lapedriza, A.; Oliva, A.; Torralba, A. Learning deep features for discriminative localization. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2921–2929.
49. Riggs, W. Perception of safety and cycling behaviour on varying street typologies: Opportunities for behavioural economics and design. *Transp. Res. Procedia* **2019**, *41*, 204–218. [[CrossRef](#)]
50. Gong, F.-Y.; Zeng, Z.-C.; Zhang, F.; Li, X.; Ng, E.; Norford, L.K. Mapping sky, tree, and building view factors of street canyons in a high-density urban environment. *Build. Environ.* **2018**, *134*, 155–167. [[CrossRef](#)]
51. Harvey, C.; Aultman-Hall, L.; Hurley, S.E.; Troy, A. Effects of skeletal streetscape design on perceived safety. *Landsc. Urban Plan.* **2015**, *142*, 18–28. [[CrossRef](#)]
52. 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](#)]
53. Keizer, K.; Lindenberg, S.; Steg, L. The spreading of disorder. *Science* **2008**, *322*, 1681–1685. [[CrossRef](#)]
54. 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](#)]
55. Zhang, F.; Wu, L.; Zhu, D.; Liu, Y. Social sensing from street-level imagery: A case study in learning spatio-temporal urban mobility patterns. *ISPRS J. Photogramm. Remote Sens.* **2019**, *153*, 48–58. [[CrossRef](#)]
56. Zhang, F.; Zhang, D.; Liu, Y.; Lin, H. Representing place locales using scene elements. *Comput. Environ. Urban Syst.* **2018**, *71*, 153–164. [[CrossRef](#)]
57. Su, L.; Chen, W.; Zhou, Y.; Fan, L. Exploring city image perception in social media big data through deep learning: A case study of Zhongshan City. *Sustainability* **2023**, *15*, 3311. [[CrossRef](#)]
58. Wei, J.; Yue, W.; Li, M.; Gao, J. Mapping human perception of urban landscape from street-view images: A deep-learning approach. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *112*, 102886. [[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.