applied sciences

MDPI

Article

Spatial Image-Based Walkability Evaluation Using Regression Model

Jiyeon Hwang, Kwangwoo Nam and Changwoo Lee *

Department of Computer Information Engineering, Kunsan National University, Gunsan 54150, Republic of Korea; ghkdwldus2@naver.com (J.H.); kwnam@kunsan.ac.kr (K.N.)
* Correspondence: leecw@kunsan.ac.kr

Abstract: Governments worldwide have invested considerable money and time into creating pedestrian-oriented urban environments. However, generalizing arbitrary standards for walking environments is challenging. Therefore, this study presents a method for predicting walkability scores of evaluations using five regression models, including Multiple linear, Ridge, LASSO regression, SVR, and XGBoost. The models were trained using semantic segmentation, walkability evaluations based on crowdsourcing, and image scores obtained using the TrueSkill algorithm, and their performances were compared. Feature selection was employed to improve the accuracies of the models, which were retrained using the importance of extracted features. Among the five regression models, XGBoost, a tree-based regression model, exhibited the lowest error rate, high accuracy, and greatest performance improvement after retraining. This study is expected to generalize the walking environments preferred by various people and demonstrate that objective walkability evaluations are possible through a computer system rather than through subjective human judgment.

Keywords: crowdsourcing; walkability evaluation; semantic segmentation; TrueSkill; regression model; feature importance

check for updates

Citation: Hwang, J.; Nam, K.; Lee, C. Spatial Image-Based Walkability Evaluation Using Regression Model. *Appl. Sci.* **2024**, *14*, 4079. https://doi.org/10.3390/app14104079

Academic Editor: Douglas O'Shaughnessy

Received: 26 March 2024
Revised: 26 April 2024
Accepted: 8 May 2024
Published: 11 May 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

To enhance pedestrian safety and convenience, numerous projects aimed at improving living environments are actively promoted worldwide [1]. The Architecture and Urban Research Institute of the Republic of Korea regularly holds international seminars on the theme of "Creating a Good and Safe Walking City" to share excellent precedents for improving walking environments and to discuss strategies for creating safe walking environments [2]. Deep learning and computer vision are emphasized for their potential roles in understanding cities, and various computer vision algorithms are being diversely applied in urban systems [3,4]. Jeong et al. [5] calculated importance based on the correlation of direct and indirect factors that should be considered when designing a pedestrian environment using an analytic network process.

Chen et al. [6] extracted and integrated features of inner London using geotagged Flickr photographs, and explored the characteristic differences and the dynamics of areas where more people assemble [7]. Verma et al. [8,9] developed a mobile application to collect visual and audio datasets including the characteristics of the urban street fluctuating with time, and extracted the attributes of urban areas using various computer algorithms. Kim et al. [10] used a Tobit regression model to analyze the physical environmental factors affecting walking. Their results showed that the probability of walking increased as the population density, diversity of land use, street connectivity index, and four-way crossroad density increased. Lee et al. [11] surveyed the perceptions of walking environments around neighborhood parks and analyzed the factors affecting pedestrian environment satisfaction through Stepwise regression analysis.

Shao et al. [12] integrated objective image analysis and subjective visual comfort evaluations, and assumed that the street comfort of four cities in China was related to city development and urban regeneration. Kang et al. [13] used a double-column CNN to reflect

human visual perception and accurate safety scores, and compared the performance of support vector regression and CNN models. Seresinhe et al. [14] explored the components of beautiful outdoor places using the Places365 CNN and trained a neural network to automatically identify scenic places. They found that built locations, such as buildings and bridges, as well as natural locations were positively related to the walking environment. Liu et al. [15] collected street-view images of Beijing, China, through Baidu Maps, and proposed machine learning techniques to automatically evaluate the qualities of urban environments using Scale Invariant Feature Transform histograms (SIFT), AlexNet, and GoogLeNet.

Zhang et al. [16] proposed a deep learning model that predicts human perceptions of large-scale urban regions and argued that it could help map the distribution of human perceptions for an entire city. Dubey et al. [17] created a website to collect pairwise comparisons of six perceptual attributes and trained a Siamese-like CNN on the Place Pulse 2.0 dataset to predicts pairwise comparisons. They argued that their method could help make data-driven decisions to improve urban appearance and aid computational studies in areas such as architecture and economics. Chen et al. [18] conducted a study to predict the effect of the street environment on the mood of citizens in Guangzhou, China, based on various factors of the street environment using stepwise, ridge, and Least Absolute Shrinkage and Selection Operator (LASSO) regressions. Additionally, studies have shown that streets in urban fringe areas are more likely to make residents motivated, happy, relaxed, and focused than those in city centers.

Sung et al. [19] constructed a multi-level regression model to investigate the relationship between the residential environments and walking activities in Seoul, and supported Jacobs's urban vitality theory that "Urban vitality is exerted by people's walking activities, and these activities are related to the built environment". Rossetti et al. [20] proposed a method for identifying the characteristics of public spaces using the SegNet model and quantifying the perceptions of landscapes through discrete choice models. Huang et al. [21] used a transformer network to extract semantic features and proposed a human perception model that enhances feature representations to integrate visual elements. The correlation between environmental perceptions and resident activities were explored using the SlowFast network. Areas perceived as "wealthy", which showed a high active index, comprised high-rise buildings and considerable traffic flow, whereas those perceived as "beautiful" comprised greenery and landscapes.

Park et al. [22,23] extracted the street environment from the perspective of pedestrians using Google Street View images and semantic segmentation techniques, trained a model to predict pedestrian satisfaction with street environments, and compared its performance with that of other models. By analyzing the relationship between street environmental factors and walking satisfaction, they determined that walking satisfaction improved as visual complexity increased. Lieu et al. [24] analyzed the relationship between street environmental factors and pedestrian satisfaction through semantic segmentation using Deeplabv3+, object detection using YOLOv3, and Multilevel ordered logistic regression analysis. They found that pedestrian satisfaction was low in street environments with many cars or signs, whereas more green spaces increased their satisfaction. Additionally, they emphasized that the visual factors considered important in street environments vary because the subjective criteria differ from person to person based on their social status or personal experiences.

According to these studies, the walking environment and human walking activities are closely related, and the components of the walking environment are related to human emotions. This study proposes a method for predicting walkability scores of evaluations by using regression models and synthesizing the walking environments preferred by individuals. The main contributions of this study are as follows:

(1) It proposes regression models that predict image scores based on the walking environment, trained on the results of a walkability evaluation conducted through crowdsourcing.
(2) The regression models, trained on the results of the walkability evaluation, analyze the features influencing the image score and extract feature importance accordingly.

(3) It verifies and compares the accuracy of the proposed models after retraining it using the extracted feature importance.

The remainder of this paper is organized as follows: Section 2 presents the methods for conducting the walkability evaluation and analyzing environmental features. Section 3 presents the results of the experiment conducted using the proposed methods. Finally, Section 4 concludes the study.

2. Proposed Method

This study proposes a walkability evaluation method using road images and Figure 1 shows the flowchart of study. The proposed method consists of data acquisition and feature analysis.

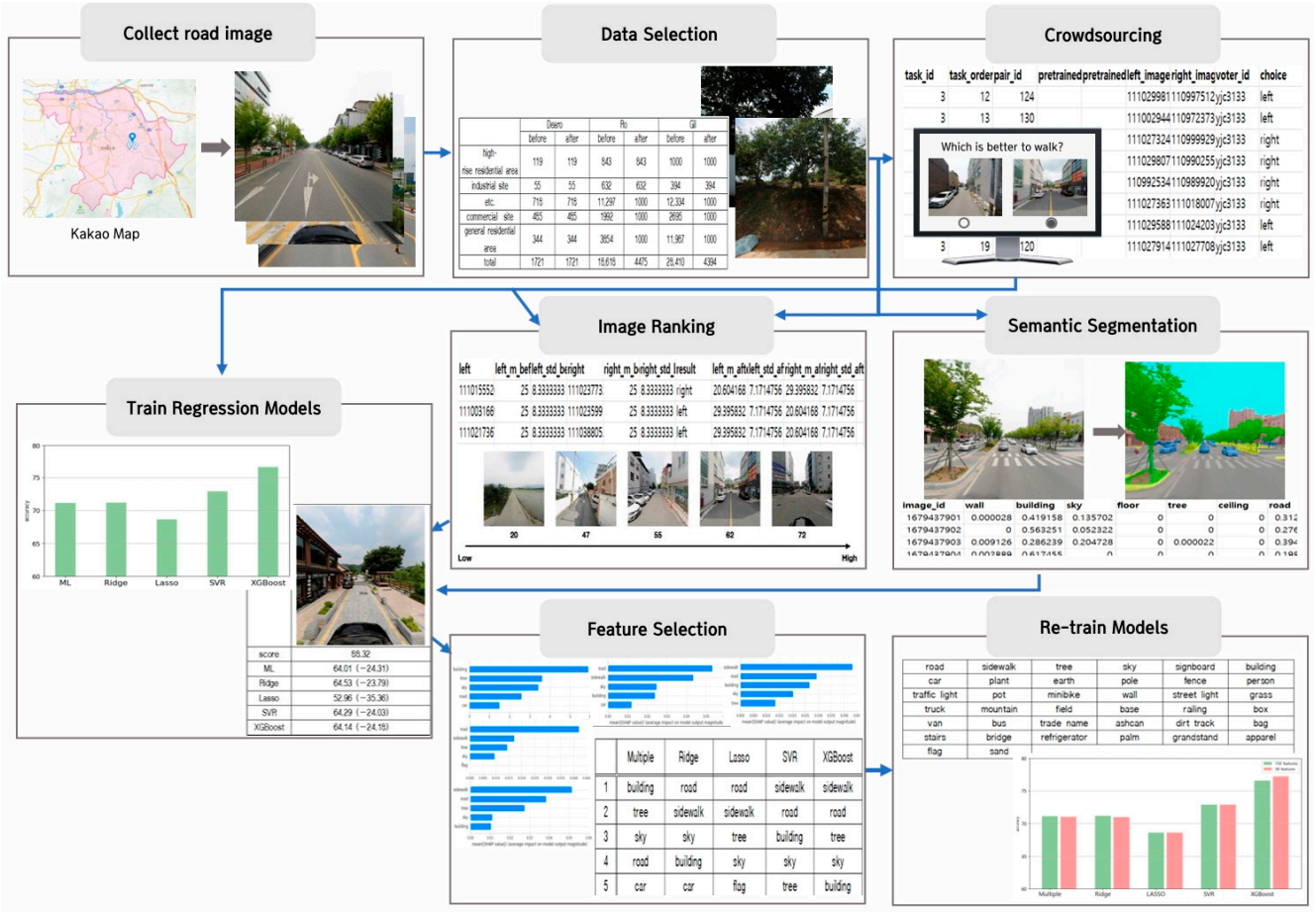


Figure 1. Flowchart of this study.

For data acquisition, images of roads required for the walkability evaluation are collected, and walkability is evaluated through crowdsourcing. Semantic segmentation is employed to extract the walking environment features, and the TrueSkill algorithm is used to calculate scores for the images used in the walkability evaluations, thereby ranking them. Additionally, a web application is built to visually confirm the results of the walkability evaluation.

Regarding feature analysis, a dataset for training the regression models is constructed by the 150 features obtained through semantic segmentation and the image scores calculated using the TrueSkill algorithm. The training dataset comprises a data identifier, image name, image score, and 150 features. The five regression models are trained on the dataset, and their performances in predicting walkability scores are compared. The regression models analyze features that significantly contribute to predicting walkability scores for evaluation and extract feature importance based on their contribution. To improve accuracy, the

five regression models are retrained using the extracted feature importance. Through this process, we conduct this study to find the spatial correlations existing in road images.

2.1. Data Acquisition

2.1.1. Image Collection

Road images for the walkability evaluation were exclusively collected in Jeonju-si, Republic of Korea, using Kakao Maps. Data was extracted from a total of 49,156 points by dividing the road network of Jeonju-si into 30 m intervals. The panoramic images of each point were divided into left, front, right, and back, as shown in Figure 2, resulting in 196,624 images. Thereafter, images containing large portions of tunnels, buildings, and trees were manually removed, as it is difficult to determine whether these environments are suitable for walking. Stratified extraction was performed based on road widths (daero, ro, gil) and land types (industrial, commercial, high-rise housing, detached house, etc.). "Daero" are roads with eight or more lanes, "ro" are roads with two–seven lanes, and "gil" are roads that are narrower than "ro." Land types were classified according to their use cases. After stratified extraction, 10,590 points out of 49,156 points were extracted, and 42,360 images were used for the walkability evaluation [25,26].



Figure 2. Road views of Jeonju-si, Republic of Korea (**a**) left, (**b**) front, (**c**) right, (**d**) back.

2.1.2. Crowdsourcing

Walkability evaluation is conducted by having crowdsourcing participants [27,28] select the image of the environment they consider more suitable for walking from two randomly presented road images. The pairwise image comparison employs the comparison set generation algorithm proposed by Yoo et al. [29], and a total of 193,600 pairwise comparison datasets are constructed using 21,168 images out of 42,360 road images. The image pairwise comparison dataset is represented as Equation (1).

$$P = \{i_x, j_x, s_x\}_{x=1}^{N},\ i,j \in \{1,\ldots,N\},\ s \in \{+1,-1\}, \quad (1)$$

where i_x and j_x denote the images and N denotes the total number of images. Additionally, $s = +1$ indicates that i_x was selected, whereas $s = -1$ indicates that j_x was selected.

The TrueSkill algorithm is used to calculate the walkability scores (mean, standard deviation) of road images used in the walkability evaluation. TrueSkill obtains stable scores by reducing noise in the binary predictions for the pairwise image comparisons [30]. The algorithm calculates the scores $N(\mu_i, {\sigma_i}^2)$ and $N\left(\mu_j, {\sigma_j}^2\right)$ for images i and j, respectively, and normalizes their values to between 0 and 100. These scores are updated each time an evaluation is conducted and are used to determine the final ranking. Table 1 presents the image information dataset, which includes the number of times an image was compared with other images in the walkability evaluation, the number of wins in the comparison, and the image score calculated through the TrueSkill algorithm. Table 2 presents the pairwise image score dataset containing the image scores before and after the comparison and the results of the comparisons.

Table 1. Image information dataset for two images.

Image Name		
Number of comparisons	16	16
Number of wins	3	15
Trueskill_mean	17.226043	34.964508
Trueskill_std	2.8507967	3.0705123

Table 2. Image scores dataset of pairwise comparison.

	Left	**Right**
Image name		
Mean_before	36.153447	23.979166
Std_before	3.4737805	2.7657728
Result	left	
Mean_after	36.362659	23.979166
Std_after	3.4737805	2.7657728
Image name		
Mean_before	28.317252	30.239095
Std_before	3.5223532	3.4375636
Result	right	
Mean_after	27.181376	31.320974
Std_after	3.292593	3.2244403

2.1.3. Semantic Segmentation

Semantic segmentation is used to classify pixel areas of images, identifying features that constitute the walking environment. The more detailed the pixel areas, the easier it is to distinguish the components of the walking environment in road images. Therefore, this study employs the Unified Perceptual Parsing network (UPerNet) model, trained on the ADE20K dataset, which provides 150 classes. The model utilizes the Swin Transformer [31] as its backbone network [32]. The UPerNet (Swin-B) model, which has demonstrated state-of-the-art semantic segmentation performance for the ADE20K dataset, divides the pixel

areas of the image into their corresponding features from the 150 features. Figure 3 shows a visualization of the semantic segmentation results, a road image of Jeonju-si. Table 3 presents the semantic segmentation results of Figure 3, showing the ratio of the number of pixels for the extracted features. The feature numbers are displayed according to the order presented in the ADE20K dataset. From the image shown in Figure 3, the following 11 features were extracted: road, sky, tree, plant, sidewalk, car, building, grass, mountain, house, and streetlights.

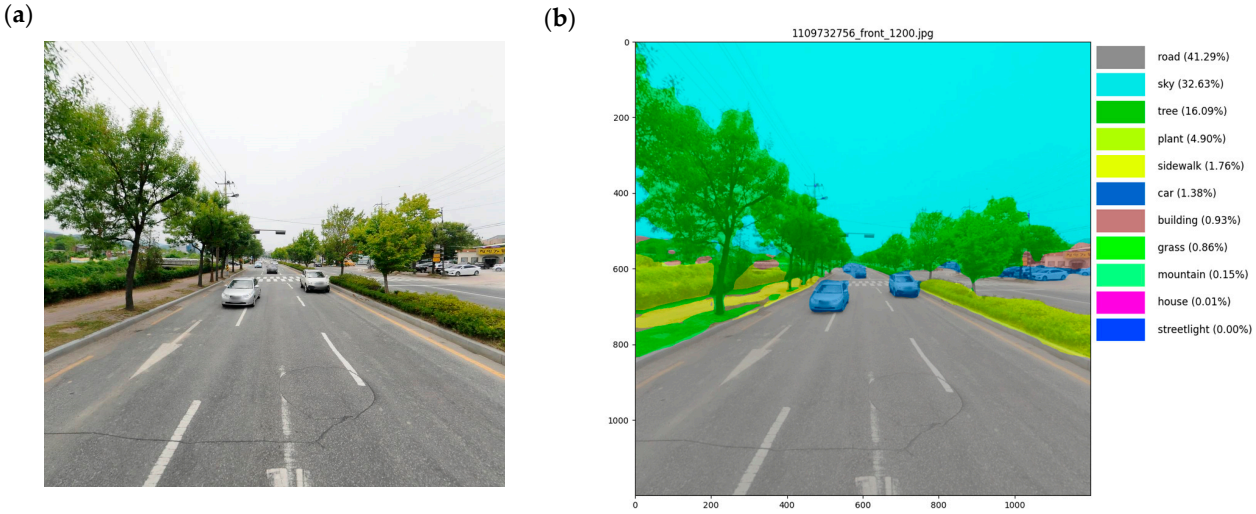


Figure 3. (**a**) Road image (**b**) semantic segmentation results.

Table 3. Semantic segmentation results for the image shown in Figure 3.

Number	2	3	5	7	10	12	17	18	21	26	88
Class	building	sky	tree	road	grass	sidewalk	mountain	plant	car	house	streetlight
Ratio (%)	0.009258	0.326276	0.160924	0.41294	0.008617	0.017583	0.001495	0.048962	0.013838	0.000083	0.000024

However, expressing the crowdsourcing results in numerical form, along with semantic segmentation results and image scores calculated by the TrueSkill algorithm, presents difficulties in generalizing the criteria of the walkability evaluators or visually understanding the walking environment preferred by pedestrians. To address these issues, a web application is built to visually represent the data.

2.2. Feature Analysis

2.2.1. Regression Model Training

To predict the walkability scores for the evaluation, linear (multiple linear, ridge, LASSO [33]), nonlinear (SVR [34]), and tree-based (XGBoost [35]) regression models are built. The following environment is used for model training: Windows 10, AMD RYZEN 9 5950X CPU, NVIDIA RTX 3080 TI GPU, and 64 GB RAM. Scikit-learn is used to build the multiple linear regression, ridge regression, LASSO regression, and SVR, whereas the XGBoost library 2.0.3 is used to build eXtreme Gradient Boosting (XGBoost). These five regression models are trained on the results of the walkability evaluation and the model exhibiting the best performance is identified by comparing the accuracies of all models. The following performance metrics are used for the evaluations: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Square Error (RMSE), and coefficient of determination (R^2) [32].

2.2.2. Feature Selection

The features constituting the walking environment are extracted through semantic segmentation. The feature importance is extracted based on their degree of reflection in the

model. These extracted features are considered feature importance for the regression model to evaluate walkability and can be seen as pedestrian-friendly environmental elements. In the feature analysis process, the SHapley Additive exPlanations (SHAP) [36] is used to numerically calculate how much the model depends on certain features, as can be seen in Equation (2).

$$\varnothing_i(f, x) = \sum_{z' \subseteq x'} \frac{|z'|!(M - |z'| - 1)!}{M!} \left[f_{(x)}(z') - f_{(x)}(z' \setminus i) \right] \quad (2)$$

Based on the degree to which the feature contributes to the model, the feature importance of the five regression models is visualized. The feature importance by removing duplication from the feature importance of five models is extracted. Using the feature importance without duplicates, we conduct a study to compare the accuracy of the model as the number of features increases. The exhaustive search algorithm is used to find the number of features resulting in the best accuracy by adding a specific number of features. The method of adding features represents in Algorithm 1. We represent feature importance of the highest accuracy, and the five models are retrained using this feature importance. The accuracy of retrained models is compared with that of the models trained on 150 features.

Algorithm 1: exhaustive search

Input: particular set F $= \{x_1, x_2, \cdots, x_d\}$, size of subset $\left[\hat{d_1}, \hat{d_2}\right]$

Output: subset $\hat{F}$
score = 0;
while (TRUE) {
  S = next_subset ($F, \hat{d_1}, \hat{d_2}$);
  if ($S \neq$NULL) {
    $s = J(S)$;
    if ($s >$ *score*) {$\hat{F} = S$; *score* = s;}
  }
  else break;
}
return $\hat{F}$;

3. Experimental Results

3.1. Training the Regression Models

The five models are trained on image scores obtained by the TrueSkill algorithm and semantic segmentation results containing 150 feature distribution. Table 4 shows the performances of the regression models. The multiple linear regression exhibited accuracy of 71.15%, and its dependent variable (image score) for the independent variable (150 classes) was unstable; this model was considered unsuitable for image scores prediction. Ridge, LASSO regression and SVR exhibited accuracies of 71.19, 68.64 and 72.91%, respectively. XGBoost exhibited accuracy of 76.60%, which was higher than the accuracy levels of linear and nonlinear regressions.

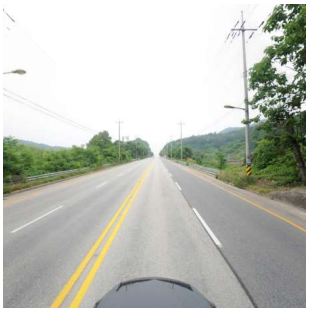
Table 4. Model results for walkability evaluations.

Model	MAE	MSE	RSME	R^2	Accuracy
Multiple	1,893,901.1099	1.1760	108,443,675.1976	−5.8665	71.15%
Ridge	0.0804	0.0104	0.1024	0.4768	71.19%
LASSO	0.0900	0.0127	0.1128	0.3650	68.64%
SVR	0.0931	0.0087	0.0953	0.5635	72.91%
XGBoost	0.0743	0.0089	0.0947	0.5522	76.60%
Dubey et al. [17]					73.50%
Kim et al. [25]					75.01%

Dubey et al. [17] conducted crowdsourcing evaluation to select urban landscape suitable for six perceptions (Safe, Lively, Beautiful, Wealthy, Depressing, Boring), whereas this study evaluated only one perception (walkability). Despite the differences in evaluation indicators, it was judged feasible to compare as both studies conducted qualitative human evaluation of the walking environment. Ranking StreetScore (RSS)-CNN [17], which predicts scores for pairwise comparisons, exhibited an accuracy of 73.5%. The deep learning model based on the RSS-CNN architecture was built by Kim et al. [25], achieving an accuracy of 75.01% when trained on the same data as in this study. This was higher than those of ridge, LASSO regression, and SVR built in this study but lower than that of XGBoost. According to reference [37] about XGBoost, a tree boosting system, generates individual trees and provides an algorithm for weighted quantile sketch for appropriate tree learning. It gives a weight to the weakly trained samples which are then reflected in the next learning model to reinforce the error. That is, in this study, it can be seen that XGBoost, a gradient boosting system, exhibited higher performance than other linear and nonlinear regression models.

Table 5 compares the walkability scores predicted by the regression models with those obtained using the TrueSkill algorithm. For a first image with score of 88.32 points calculated by TrueSkill algorithm, ridge regression most closely predicted with 64.53 points. For the second image with a score of 49.47 points by TrueSkill algorithm, XGBoost predicted the closest score of 49.84 with a difference of 0.37. Therefore, it was determined that XGBoost, which demonstrated the best performance among the five regression models and provided score predictions similar to those generated by the TrueSkill algorithm, is suitable for predicting image scores in the walkability evaluation.

Table 5. Image score predictions of the five models.

Image		
Score	88.32	49.47
Multiple	64.01 (−24.31)	61.00 (+11.53)
Ridge	64.53 (−23.79)	60.85 (+11.38)
LASSO	52.96 (−35.36)	60.50 (+11.03)
SVR	64.29 (−24.03)	56.01 (+6.54)
XGBoost	64.14 (−24.18)	49.84 (−0.37)

Various regression models have been trained using semantic segmentation results to evaluate walkability; however, their accuracies did not reach 76%. It was determined that the crowdsourcing results were unreliable because walkability evaluation was not conducted properly by Lee et al. [31]. Therefore, in this study, a web application was built to verify the data structure of the evaluation results. The web application visually provided the image information dataset and the pairwise image score dataset constructed in Section 2.1.2. Each image was compared 16 times with other images in the walkability evaluation, and was classified in the web application based on the number of times it was selected by the evaluator.

In Figure 4a, the 21,168 road images were classified from 0 to 16 based on the number of times they were selected in the walkability evaluation, and displayed in the web application. If the user selected an image at a specific number of times, the web application showed the image compared to other images, the image scores before and after the comparisons, and the results of the walkability evaluation, as shown in Figure 4b. When the user clicked the "image_segmentation" link in Figure 5b, the semantic segmentation results of the corresponding image were visually shown as can be seen in Figure 4c. Additionally, when the user clicked the "segmentation" link in Figure 4b, the semantic segmentation results for the compared images were shown in a graph, as shown in Figure 4d, and the walking environment components of the two images. In the future, the prediction performance is expected to improve by identifying and removing evaluation data that degrade the accuracy through this web application.

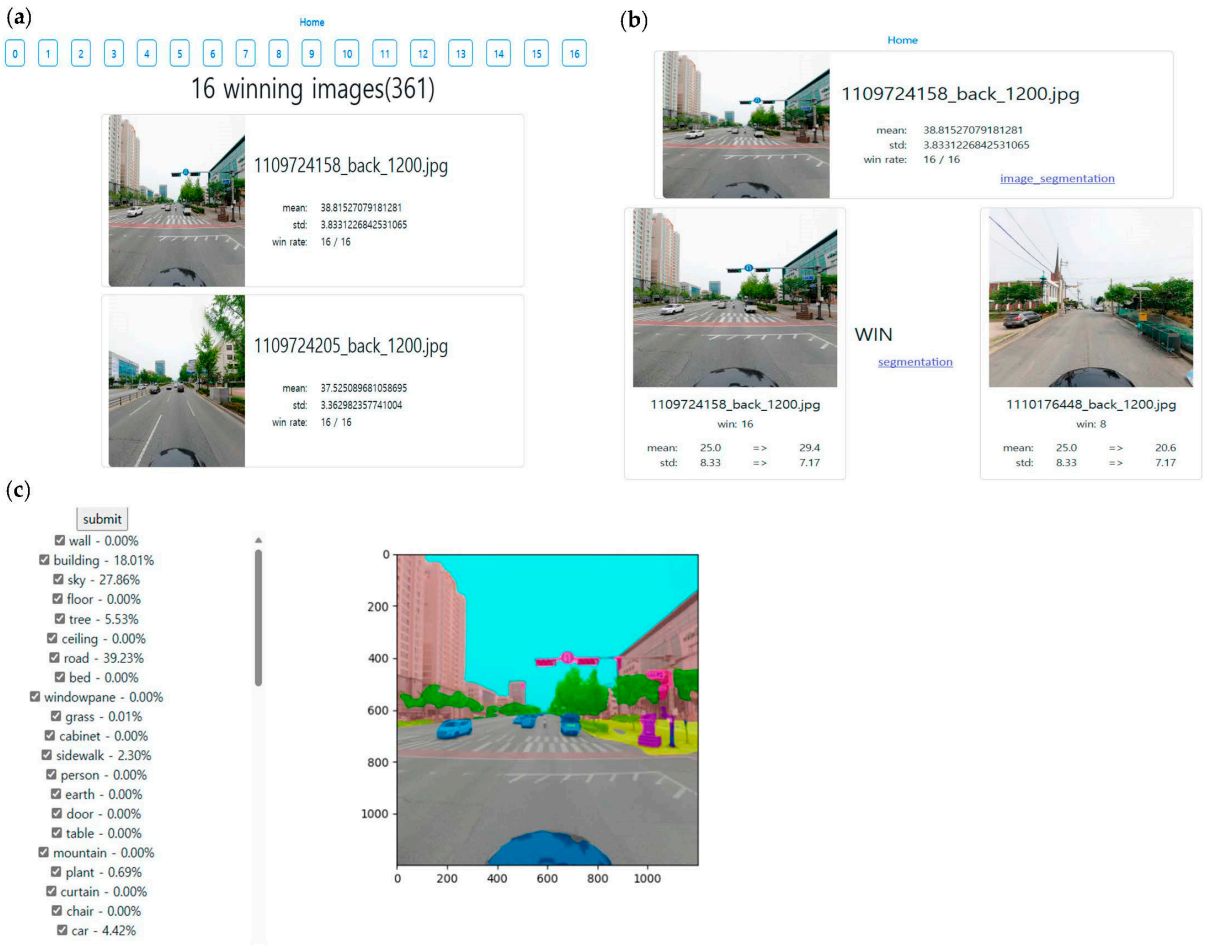


Figure 4. *Cont.*

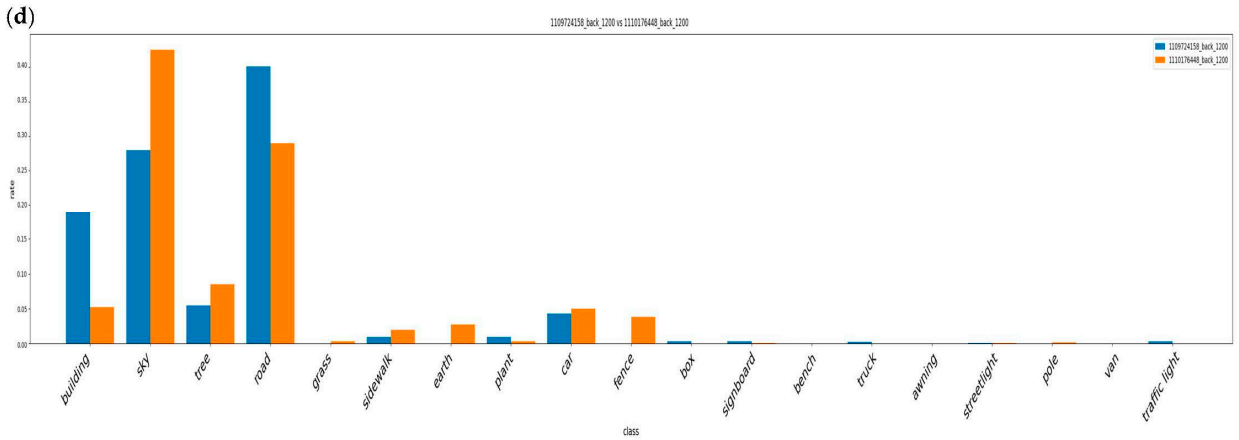


Figure 4. Web pages showing the (**a**) image information, (**b**) walkability evaluation results, (**c**) semantic segmentation results, (**d**) semantic segmentation results for pairwise comparisons.

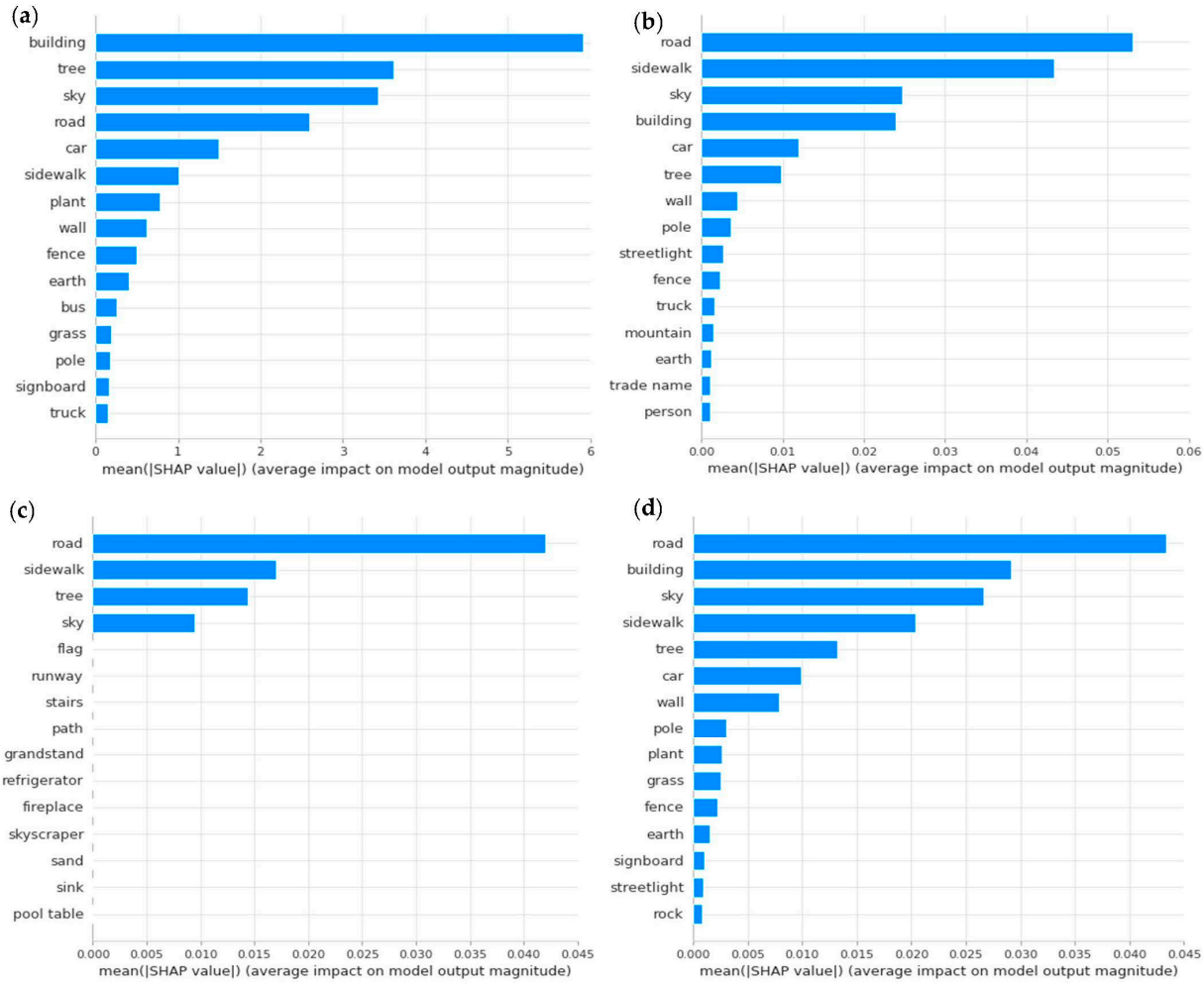


Figure 5. *Cont.*

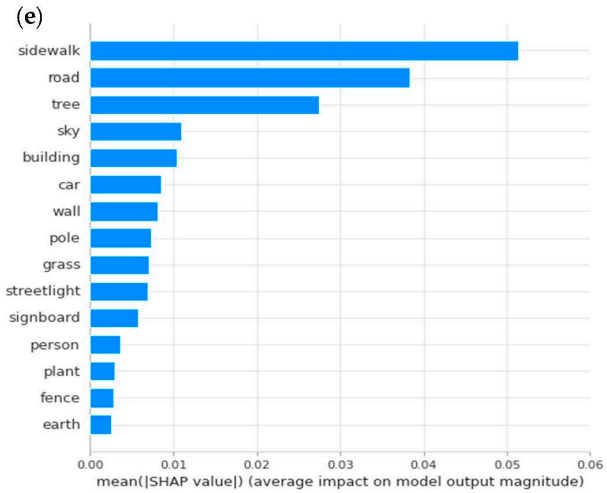


Figure 5. Feature importances of the five regression models. (**a**) multiple linear, (**b**) ridge, (**c**) LASSO, (**d**) SVR, (**e**) XGBoost.

3.2. Feature Importance

We used the SHAP technique to extract the important features considered by the five regression models during the evaluation of walkability. The 15 most important features out of 150 were visualized in Figure 5. In Figure 5a, for multiple regression, it was determined the higher the proportion of features such as "building", "tree", "sky", "road", and "car", the higher score was predicted. Figure 5b shows that ridge regression was affected by the "road", "sidewalk", "sky", and "car" features, in that order, and Figure 5c shows that LASSO regression was affected by the "road", "sidewalk", "tree", and "sky", in that order. Figure 5d shows that the high prediction scores of SVR were related to "sidewalk", "road", "building", "sky", and "tree". Additionally, Figure 5e shows that "road", "building", "sky", "sidewalk", and "tree" are significant features on XGBoost. Table 6 contains 30 features that are highly dependent on the model's training process. Table 7 shows the feature importance obtained 66 features by removing common features from the feature importance of each model. After examining the top 5 feature importance in five regression models, "road" and "sky" were extracted, and upon examining the top 10, "tree" and "sidewalk" were also extracted. The fact that the following features were significantly extracted can be observed because road images were used in the walkability evaluation.

Table 6. Feature importance by each model (ranked 1–30).

	Multiple	Ridge	LASSO	SVR	XGBoost
1	building	road	road	sidewalk	road
2	tree	sidewalk	sidewalk	road	building
3	sky	sky	tree	building	sky
4	road	building	sky	sky	sidewalk
5	car	car	flag	tree	tree
6	sidewalk	tree	runway	car	car
7	plant	wall	stairs	pole	wall
8	wall	pole	path	earth	pole
9	fence	streetlight	grandstand	wall	plant
10	earth	fence	refrigerator	grass	grass
11	bus	truck	fireplace	signboard	fence
12	grass	mountain	skyscraper	field	earth
13	pole	earth	sand	ceiling	signboard
14	signboard	trade name	sink	person	streetlight

Table 6. *Cont.*

	Multiple	Ridge	LASSO	SVR	XGBoost
15	truck	person	pool table	base	rock
16	dirt track	ashcan	counter	plant	truck
17	trade name	traffic light	chest of drawers	box	railing
18	van	railing	signboard	van	bag
19	railing	dirt track	column	pot	ashcan
20	field	plant	box	house	traffic light
21	person	bag	case	traffic light	bus
22	traffic light	pot	stairway	bicycle	field
23	house	field	pillow	path	minibike
24	pot	awning	flower	water	box
25	streetlight	minibike	palm	bridge	van
26	base	signboard	stove	bench	sculpture
27	minibike	bridge	counter top	stairway	trade name
28	mountain	base	bench	windowpane	pot
29	ashcan	water	hill	apparel	awning
30	rock	ship	book	minibike	mountain

Table 7. Feature importance of each model after removing duplicates.

road	sidewalk	tree	sky	signboard	building
car	plant	earth	pole	fence	person
traffic light	pot	minibike	wall	streetlight	grass
truck	mountain	field	base	railing	box
van	bus	trade name	ashcan	dirt track	bag
stairway	path	awning	water	bridge	bench
sculpture	rock	flag	runway	stairs	grandstand
refrigerator	fireplace	skyscraper	sand	sink	pool table
ceiling	counter	chest of drawers	column	case	bicycle
house	pillow	flower	palm	stove	counter top
windowpane	hill	apparel	ship	book	door

3.3. Retraining the Regression Models

Using the 66 features listed in Table 7, a study was conducted to enhance the accuracy of XGBoost, which exhibited the best performance among the five regression models. The model was trained incrementally, from 30 to 66 features, using an exhaustive search algorithm, and its performance was showed through graphs. XGBoost exhibited the highest accuracy (77.24%) when trained on 38 features, an increase of 0.64% compared to when it was trained on 150 features (76.60%). The performance of the model by increasing the number of features in increments of 2 (30, 32, 34, . . ., 62, 64, 66) is shown in Figure 6, and the 38 features resulting in the highest XGBoost accuracy are listed in Table 8. It is worth noting that three features (“rock”, “stroke”, and “awning”) were excluded from the XGBoost feature importance listed in Table 6, and the features “person”, “base”, “steps”, “bridge”, “refrigerator”, “palm”, “grandstand”, “apparel”, “flag”, and “sand” were newly extracted.

Other models were retrained using 38 features of importance optimized for XGBoost. In Figure 7, the accuracies of the models trained on 150 features in Section 3.1 was compared with those of the models trained on the 38 features of importance. The accuracy of multiple linear regression decreased by 0.11% to 71.04% and that of ridge regression decreased by 0.17% to 71.02%. However, the accuracies of the LASSO regression and SVR trained on 150 or 38 features remained unchanged. Moreover, the accuracy of XGBoost increased by 0.64%, showing the greatest performance improvement. It was determined that 38 features are the optimal number for XGBoost, while the number of features suitable for other regression models varies.

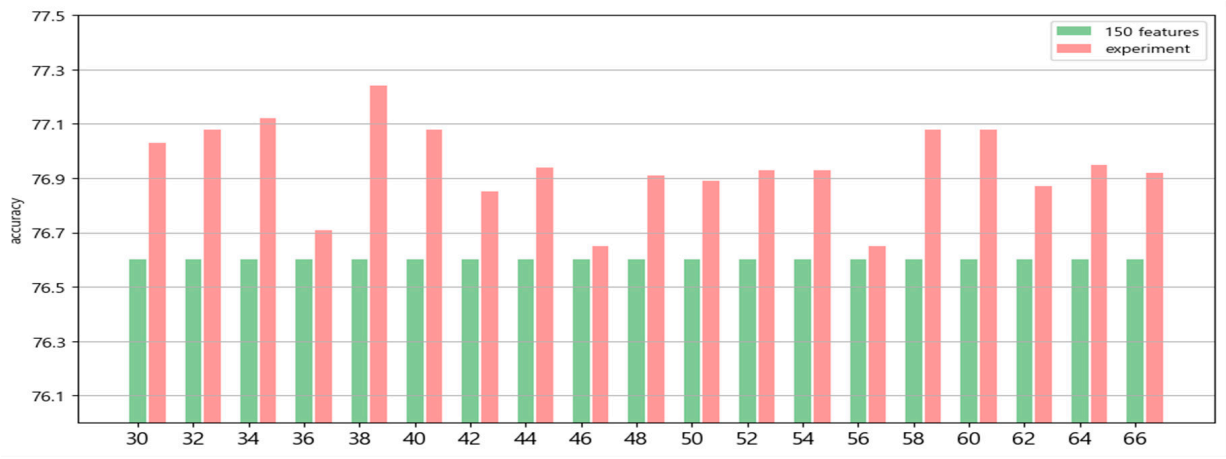


Figure 6. Performance results of XGBoost according to the number of features.

Table 8. Importance of the 38 features in the optimized XGBoost model.

road	sidewalk	tree	sky	signboard	building
car	plant	earth	pole	fence	person
traffic light	pot	minibike	wall	streetlight	grass
truck	mountain	field	base	railing	box
van	bus	trade name	ashcan	dirt track	bag
stairs	bridge	refrigerator	palm	grandstand	apparel
flag	sand				

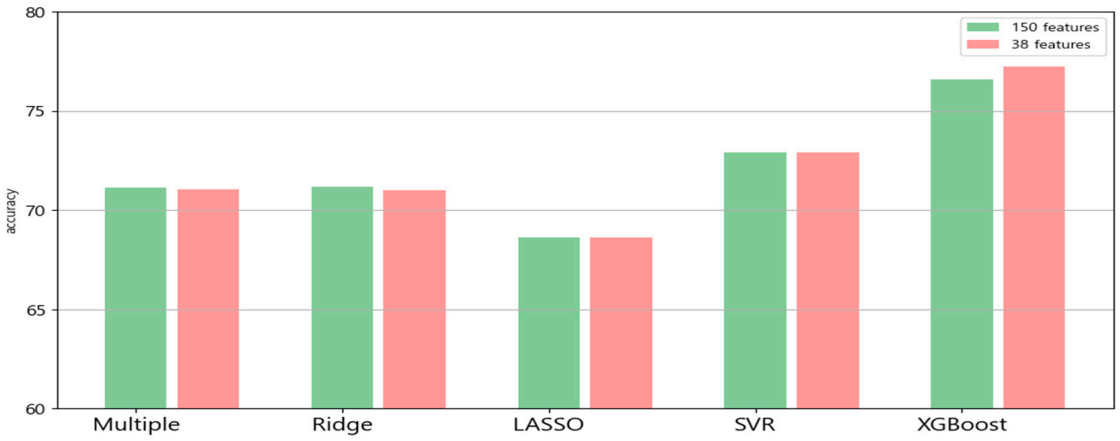


Figure 7. Performance of regression models using the feature importance presented in Table 8.

4. Conclusions and Future Work

This paper proposed a method for evaluating the preferences for walkability environments based on road images. The proposed method aimed to understand the correlation between people's preferences for walkability environments and the components of road images. To assess people's preferences for walkability environments, several regression models were trained using the semantic segmentation results of the road images, and feature selection was conducted based on the analysis of feature importance from the trained results. Furthermore, a retraining process using only selected features was performed to improve the accuracy of walkability preference evaluations. The proposed method is an automated walkability evaluation method that can replace traditional methods using various GIS data manually.

The proposed study needs many improvements: Firstly, due to the pedestrian preference evaluations being obtained from an unspecified number of individuals through crowdsourcing, the walkability preference dataset is too diverse to be captured by generalized deep learning models. Therefore, there is a need for the development of deep learning models capable of representing the diversity of pedestrian preferences. Secondly, when people evaluate walkability preferences, they tend to make judgments based on quickly observable holistic features rather than detailed specifics. Therefore, it is necessary to identify the correlations between more specified semantic objects during the feature analysis process.

Author Contributions: Conceptualization, software, validation, investigation, data curation, writing—original draft preparation, visualization, J.H.; methodology, writing—review and editing, C.L.; supervision, project administration, K.N. and C.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the Korea Agency for Infrastructure Technology Advancement (KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (Grant RS-2022-00143336).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

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

References

1. Dragović, D.; Krklješ, M.; Slavković, B.; Aleksić, J.; Radaković, A.; Zećirović, L.; Alcan, M.; Hasanbegović, E. A Literature Review of Parameter-Based Models for Walkability Evaluation. *Appl. Sci.* **2023**, *13*, 4408. [CrossRef]
2. Y.S. Landscape Architecture Korea. Available online: https://www.lak.co.kr/news/boardview.php?id=14571 (accessed on 19 October 2023).
3. Ibrahim, M.R.; Haworth, J.; Cheng, T. Understanding cities with machine eyes: A review of deep computer vision in urban analytics. *Cities* **2020**, *96*, 102481. [CrossRef]
4. Hsieh, I.-H.; Cheng, H.-C.; Ke, H.-H.; Chen, H.-C.; Wang, W.-J. A CNN-based wearable assistive system for visually impaired people walking outdoors. *Appl. Sci.* **2021**, *11*, 10026. [CrossRef]
5. Jeong, G.S.; Roh, H.C.; Hwang, J.H. Analysis of Priority of Direct and Indirect Factor for the Pedestrian Environment Design. *Urban Des.* **2010**, *11*, 5–18.
6. Chen, M.; Arribas-Bel, D.; Singleton, A. Quantifying the Characteristics of the Local Urban Environment through Geotagged Flickr Photographs and Image Recognition. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 264. [CrossRef]
7. Chainikov, D.; Zakharov, D.; Kozin, E.; Pistsov, A. Studying Spatial Unevenness of Transport Demand in Cities Using Machine Learning Methods. *Appl. Sci.* **2024**, *14*, 3220. [CrossRef]
8. Verma, D.; Jana, A.; Ramamritham, K. Quantifying urban surroundings using deep learning techniques: A new proposal. *Urban Sci.* **2018**, *2*, 78. [CrossRef]
9. Verma, D.; Jana, A.; Ramamritham, K. Machine-based understanding of manually collected visual and auditory datasets for urban perception studies. *Landsc. Urban Plan.* **2019**, *190*, 103604. [CrossRef]
10. Kim, H.C.; Ahn, K.H.; Kwon, Y.S. The effects of residential environmental factors on personal walking probability: Focused on Seoul. *J. Urban Des. Inst. Korea* **2014**, *15*, 5–18.
11. Lee, G.-M.; Lee, W.-S.; Jung, S.-G.; Jang, C.-K. The influence of pedestrian environment perception on pedestrian environment satisfaction and expected health promotion effects-focused on park user for health promotion. *J. Korean Inst. Landsc. Arch.* **2016**, *44*, 137–147. [CrossRef]
12. Shao, Y.; Yin, Y.; Xue, Z.; Ma, D. Assessing and Comparing the Visual Comfort of Streets across Four Chinese Megacities Using AI-Based Image Analysis and the Perceptive Evaluation Method. *Land* **2023**, *12*, 834. [CrossRef]
13. Kang, H.-W.; Kang, H.-B. A Safety Score Prediction Model in Urban Environment Using Convolutional Neural Network. *KIPS Trans. Softw. Data Eng.* **2016**, *5*, 393–400. [CrossRef]
14. Seresinhe, C.I.; Preis, T.; Moat, H.S. Using deep learning to quantify the beauty of outdoor places. *R. Soc. Open Sci.* **2017**, *4*, 170170. [CrossRef] [PubMed]
15. Liu, L.; Silva, E.A.; Wu, C.; Wang, H. A machine learning-based method for the large-scale evaluation of the qualities of the urban environment. *Comput. Environ. Urban Syst.* **2017**, *65*, 113–125. [CrossRef]
16. Zhang, F.; Zhou, B.; Liu, L.; Liu, Y.; Fung, H.H.; Lin, H.; Ratti, C. Measuring human perceptions of a large-scale urban region using machine learning. *Landsc. Urban Plan.* **2018**, *180*, 148–160. [CrossRef]

17. Dubey, A.; Naik, N.; Parikh, D.; Raskar, R.; Hidalgo, C.A. Deep learning the city: Quantifying urban perception at a global scale. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 196–212.
18. Chen, C.; Li, H.; Luo, W.; Xie, J.; Yao, J.; Wu, L.; Xia, Y. Predicting the effect of street environment on residents' mood states in large urban areas using machine learning and street view images. *Sci. Total. Environ.* **2022**, *816*, 151605. [CrossRef] [PubMed]
19. Sung, H.; Lee, S. Residential built environment and walking activity: Empirical evidence of Jane Jacobs' urban vitality. *Transp. Res. Part D Transp. Environ.* **2015**, *41*, 318–329. [CrossRef]
20. Rossetti, T.; Lobel, H.; Rocco, V.; Hurtubia, R. Explaining subjective perceptions of public spaces as a func-tion of the built environment: A massive data approach. *Landsc. Urban Plan.* **2019**, *181*, 169–178. [CrossRef]
21. Huang, J.; Qing, L.; Han, L.; Liao, J.; Guo, L.; Peng, Y. A collaborative perception method of human-urban environment based on machine learning and its application to the case area. *Eng. Appl. Artif. Intell.* **2023**, *119*, 105746. [CrossRef]
22. Park, K.; Lee, S. Application and validation of a deep learning model to predict the walking satisfaction on street level. *J. Urban Des. Inst. Korea Urban Des.* **2018**, *19*, 19–34. [CrossRef]
23. Park, K.D.; Ki, D.H.; Lee, S.G. Analysis of visual characteristics of urban street elements on walking satisfaction in Seoul, Korea—Application of Google Street View and deep learning technique of semantic segmentation. *J. Urban Des. Inst. Korea Urban Des.* **2021**, *22*, 55–72. [CrossRef]
24. Lieu, S.; Ha, J.; Kim, H.; Ki, D.; Lee, S. Analysis of street environmental factors affecting subjective perceptions of streetscape image in Seoul, Korea: Application of deep learning semantic segmentation and YOLOv3 object detection. *J. Korea Plan. Assoc.* **2021**, *56*, 79–93. [CrossRef]
25. Kim, J.Y.; Kang, Y.O. Development of a Deep Learning Model to Predict the Qualitative Evaluation of a Walking Environment based on Street View Images. *J. Korean Soc. Geospat. Inf. Syst.* **2022**, *30*, 45–56.
26. Park, J.; Kang, Y.; Kim, J. Development of Walkability Evaluation Index Using StreetView Image and Semantic Segmentation. *J. Korean Cartogr. Assoc.* **2022**, *22*, 53–68. [CrossRef]
27. Lee, D.; Lee, C. A Study on the Applicability of Crowdsourcing for Cadastral Reform. *J. Korea Soc. Cadastre* **2012**, *28*, 55–70.
28. Kim, H.; Lee, S.; Jin, U.K. Improvement of Crowdsourcing based Software Development Process. *Korean Inst. Inf. Sci. Eng.* **2016**, *6*, 654–656.
29. Yoo, K.; Lee, D.; Lee, C.; Nam, K. Generating Pairwise Comparison Set for Crowed Sourcing based Deep Learning. *J. Korea Ind. Inf. Syst. Res.* **2022**, *27*, 1–11.
30. Herbrich, R.; Minka, T.; Graepel, T. TrueSkill™: A Bayesian skill rating system. *Adv. Neural Inf. Process. Syst.* **2006**, *19*, 569–576.
31. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021.
32. Lee, K.; Nam, K.; Lee, C. A Study on the Walkability Scores in Jeonju City Using Multiple Regression Models. *J. Korea Ind. Inf. Syst. Res.* **2022**, *27*, 1–10.
33. Muthukrishnan, R.; Rohini, R. LASSO: A feature selection technique in predictive modeling for machine learning. In Proceedings of the 2016 IEEE International Conference on Advances in Computer Applications (ICACA), Coimbatore, India, 24 October 2016.
34. Awad, M.; Khanna, R. *Efficient Learning Machinces: Theories, Concepts, and Applications for Engineers and System Designers*; Apress: New York, NY, USA, 2015; pp. 70–83.
35. 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.
36. Lundberg, S.M.; Lee, S.I. A unified approach to interpreting model predictions. In Proceedings of the NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017.
37. Ramraj, S.; Uzir, N.; Sunil, R.; Banerjee, S. Experimenting XGBoost algorithm for prediction and classification of different datasets. *Int. J. Control Theory Appl.* **2016**, *9*, 651–662.

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.