Assessing visual quality of the environment: Evidence extracted from the image segmentation of street views

Abstract

The visual quality of the environment influences human perceptions of safety, aesthetics, vibrancy, and more. Assessing visual quality is crucial in design practice, whether for pre-design site analysis or for assessing the performance of design proposals. However, existing computational visual assessment models depend heavily on real-world data and can not evaluate design proposals when projects are not built. This study addresses the **question**: How can the visual quality of the built environment be computationally assessed for design purposes? It tries to demonstrate the potential as well as limitations of computational assessment in the design stage. **Methodologically**, it employs image segmentation to extract the proportions and numbers of elements in street images, followed by multiple tests using various models and methods to model the relationship between these data and human ratings of the images. A final linear regression model is developed, and it is further incorporated into agent-based simulation to assess visual quality in the case of Rotterdam Central Station. The **results** reveal that factors such as the number of pedestrians and the proportions of trees, buildings, and water positively influence perceptions of street 'liveness.' The developed linear regression model offers a practical tool for assessing the visual quality of new scenes in design practice. However, this study also highlights significant limitations of this computational assessment, including its focus on a single environment type (streets), the restricted range of analyzable visual features, and the low accuracy of the mathematical and computational models used.

Keywords: Urban analytics, scientific evidence, perception, agent-based simulation, urban design

1 Introduction

Visual quality is an important aspect of the built environment. For humans, visual perception contributes to as much as 80% of the whole perception of the environment (Haupt and Huber, 2008). Improving visual quality is an important task of designers' daily practices. Architects are typically enthusiastic about the aesthetics of buildings; urban designers emphasize many visual qualities of the outdoor environments, such as imageability, complexity, enclosure, human scale, transparency, legibility, etc (Ewing et al., 2013). Studies reveal that visual qualities correlate with the intensity of people's activities (Gehl, 1987, Whyte, 2001, De Nadai et al., 2016). Visual features affect peoples' perception of the environment's

safety, comfort, etc., and affect peoples' willingness to walk (Ewing and Handy, 2009, Paydar et al., 2023).

Visual quality is commonly considered highly subjective to people's individual tastes, making it hard to assess objectively and deliver replicable results. Thanks to some computational means, it is possible to conduct more objective and replicable evaluations, as explained below.

1.1 From visibility to content in views

In the spatial design field, visibility analysis, as a simple analysis of whether things are visible, has long been developed using computational means. The computational measurement of visibility dates back to Benedikt (1979)'s 'isovist' concept, which represents the range that is visible to people at a given point in space. This concept was later applied in 2D to 3D building or environment analysis (Chen et al., 2021, Xiang et al., 2021, Zare et al., 2022). Various software programs have been developed for visibility analysis, such as Depthmap, the 'view studies' component in Ladybug, Isovist app (`https://isovists.org`). It is also relatively easy to conduct visibility analysis by coding Python scripts and implementing them on modeling platforms such as Rhino and ArcGIS.

Beyond visibility, the nuanced content in people's views can also be analyzed (which this study suggests naming as 'vision analysis') thanks to the development of image processing in the computer vision field in recent years. Many machine learning *algorithms* have been developed to do tasks such as image segmentation, object detection, and pattern recognition (He et al., 2018, Redmon et al., 2016, Ronneberger et al., 2015, Simonyan and Zisserman, 2015). Some open-source datasets laid the foundation to train new *models* for vision analysis (Term explanation: 'Models' are trained on datasets using 'algorithms,' and models have weights while algorithms do not have, as illustrated in Fig. 1. Algorithms and models can have the same names, such as the 'linear regression' algorithm and 'linear regression' models). These datasets include CityScape, Place Pulse, and so on. Some software *libraries* such as MMSegmentation, PixelLib, and Scikit-image (Chen et al., 2019, India, 2023, van der Walt et al., 2014), which integrate such algorithms or models, provide more user-friendly options for doing the vision analysis with less coding effort.

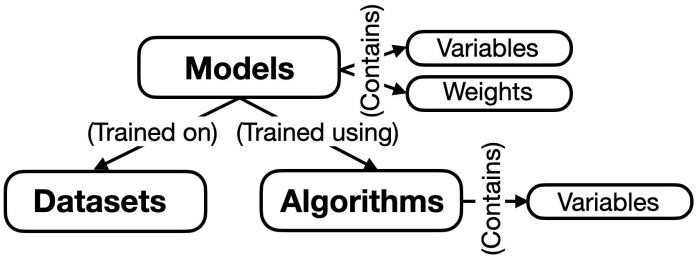

Figure 1: Differences between the terms 'algorithms' and 'models'

1.2 Visual quality assessment based on computer vision

In the spatial design field, many recent studies have done visual quality assessments based on computer vision, using machine models to analyze the content in views. The assessed visual qualities include safety (Naik et al., 2014), greenery,

openness, enclosure (Tang and Long, 2019, Li et al., 2017, Lee et al., 2022, Wang et al., 2021), complexity (Florio et al., 2023), imageability, legibility (Filomena et al., 2019), and so on. The visual quality analytics using computer vision is so extensive that it leads to literature reviews (Ibrahim et al., 2020), or even reviews of reviews (Liu and Sevtsuk, 2024).

Studies also commonly combined visual quality with other types of qualities (e.g., human scale urban form, functions mix) to analyze cities. These studies combine visual data (data extracted from street images) with other types of data, including point of interest (POI), mobile phone or GPS data, demographic data, social-economic data, and so on (Long and Ye, 2019, Li et al., 2022).

1.3 Knowledge gap and research question

The developed models in the aforementioned studies, however, heavily rely on the projects' real-world data, making them not applicable for projects during the design stage, when projects are not yet built. For example, extracting information from street images, some of the aforementioned studies analyze detailed street elements like fences, street furniture, facade texture, and advertisement billboards, or analyze advanced image features like color, hue, luminous, edges, patterns (Florio et al., 2023, Naik et al., 2014, Tang and Long, 2019). These detailed street elements and advanced image features are only available in street photos of the built projects, while not available in the typical 3D representation models during the design stage (Fig. 2). How to assess the visual quality of design proposals remains a knowledge gap. Given this knowledge gap, this study asks the following research question: *How to computationally assess the visual quality of the built environment for design use?*

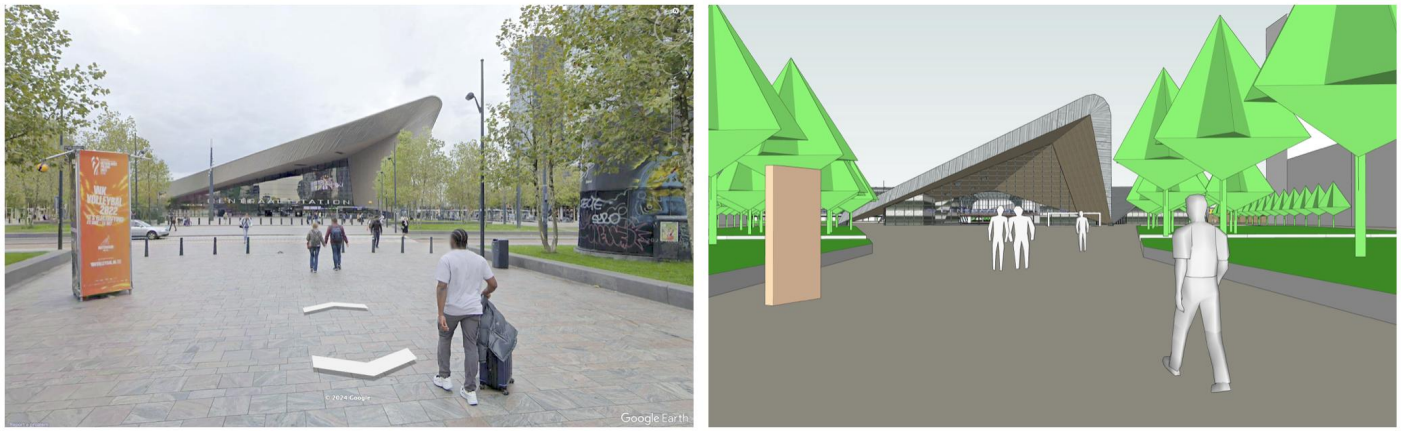

Figure 2: A photo of a real-world scene vs a 3D representation model of a design

To answer this question, this study builds a computational model that assesses visual quality based on limited features that are available in typical 3D design representation models. This study also intends to show the potential as well as the limitation of computational means in assessing visual quality during the design stage. The final developed model has its broader general relevance as it can be used in design simulation. After this introduction section, the following sections are outlined as follows: Section 2 describes the methodology, Section 3 presents the results, and Section 4 discusses.

2 Methodology

This study follows a structured methodology consisting of multiple components, including manipulating a dataset that has been rated by humans, extracting image features from the dataset, testing various models to model the relationship between the image features and visual perception, and incorporating the final chosen model into a simulation to do design assessment (Fig. 3). Many factors influence the modeling accuracy, so many tests are made on each component of the modeling. The full methodology is explained in detail below.

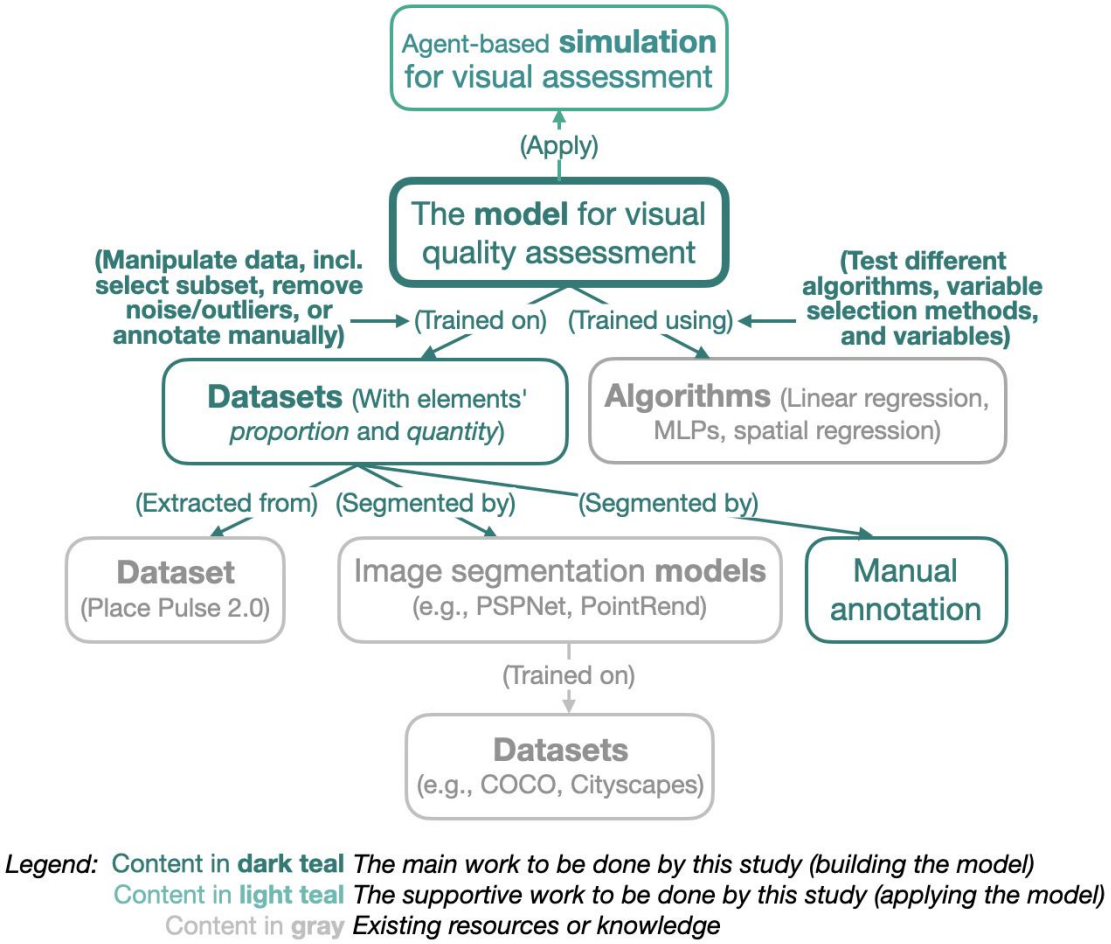

Figure 3: Research methodology

2.1 Data and manipulation

This study uses the Place Pulse 2.0 (PP2) dataset, which comprises images from 56 cities across the world, each evaluated based on six urban attributes: safety, liveliness, boredom, wealth, depression, and beauty. These images are initially rated by humans in a simple manner – pairwise comparison, then the numbers of clicks are transformed into rating scores of each image using the 'true' skill algorithm (Salesses, 2012) (Fig. 4). This study manipulates the dataset by removing noises and outliers and then comparing model accuracy when using the whole dataset versus using its subsets for modeling.

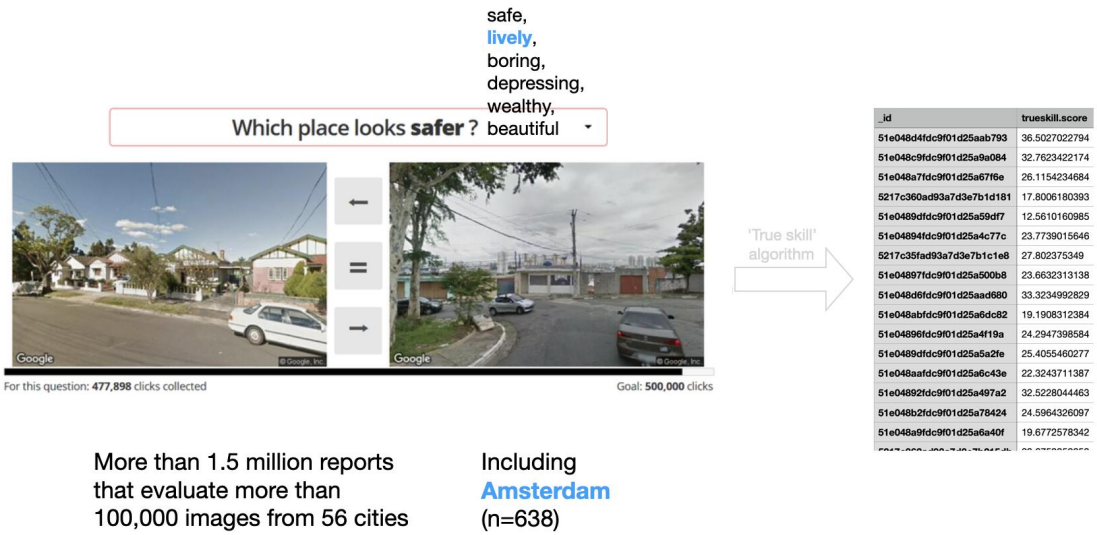

Figure 4: How the visual quality is rated in the Place Pulse 2.0 dataset (Salesses, 2012)

2.2 Image features to extract

This study extracts two types of image features in street view images – the proportion of uncountable elements in the view (e.g., how many percentages of trees) and the quantity/number of countable elements (e.g., how many people) (Fig. 5). These two features are chosen from many others because they are extractable from typical 3D design representation models of design proposals.

Figure 5: Extracting two types of image features: the proportion and the number of elements

The proportion of elements is extracted first by machine and then by manual annotation. As the input of the regression model, the images should be segmented as accurately as possible. This study first tried to use semantic segmentation models (e.g., the PSPNet models, embedded in the MMSegmentation library) to automatically extract the different proportions of elements in the images. However, the accuracy of using semantic segmentation models is visually problematical (e.g., Fig. 6, left). Also, since the element categories are predefined in the datasets that these models trained on, it is not

possible to investigate new types of elements (e.g., water bodies). This study finally decided to use the most accurate method – manually annotating images (by the first author, using the platform Roboflow. Roboflow provides the option to utilize the embedded Segment Anything algorithm to pre-segment images to save manual efforts) (Fig. 6, left).

The number of elements (or 'instances') in each picture is extracted using the instance segmentation model PointRend ResNet50 ((which is pre-trained on the COCO dataset using Mask R-CNN algorithm) embedded in PixelLib Python library. The instance segmentation applied is shown to be relatively accurate (Fig. 5, right), so it is conducted entirely by the machine in this study.

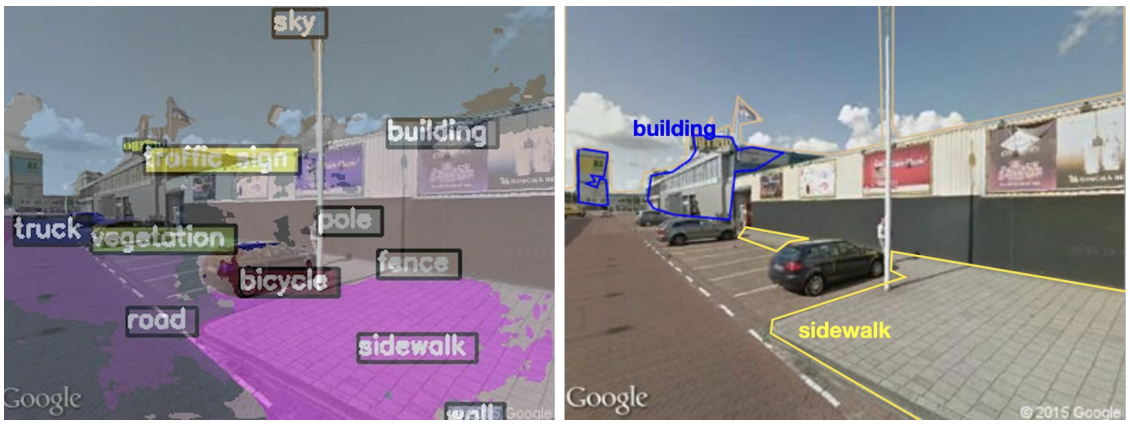

Figure 6: Compare the (semantic) segmentation of images by machine and by human

2.3 Algorithms for testing

This study tests different mathematical and computational algorithms to model the relationship between the human-rated 'liveness' scores and the image features. The tested algorithms include linear regression, Multilayer Perceptrons (MLPs), and spatial regression. Linear regression is one of the most common and simple mathematical models, assuming linear relationships between independent and dependent variables. The tests of linear regression are conducted in the software program SPSS. SPSS also provides an option – Automatic Linear Modeling, which is similar to linear regression, so this study also tests it. Beyond linear algorithms, non-linear algorithms are also tested. This study chooses Multilayer Perceptrons (MLPs) (one type of neural network model typically used for classification and regression tasks) as an example of non-linear models to test. When testing the MLPs algorithm, the modeling is approached as both regression and classification tasks [In the classification test using MLPs, this study classified the image 'lively' scores into 10 categories and trained the model to predict each image's category] (Table 1). Street images may have spatial auto-correlations (Cliff and Ord, 1981) that decrease the modeling accuracy. Therefore, spatial regression models, which can resolve auto-correlations, are also tried. The spatial regression is conducted in the software program GeoDa.

Table 1: Different models for testing

	Linear	**Non-linear**
(Regression)	linear regression; automatic linear modeling	MLPs; spatial regression
(Classification)	–	MLPs

Note: MLPs (Multilayer Perceptrons) are a type of neural network used for both regression and classification tasks. Spatial regression can be either linear or nonlinear, depending on the model specification.

2.4 Variable selection

The extracted data from image segmentation (i.e., the proportions and quantities of different elements) contains information about many elements, which are potential variables for models; however, not all the elements' information is essential or meaningful, necessitating variable selection. This study carries out variable selection in two steps. First, it utilizes multiple (mathematical or computational) variable selection methods and compares the outcomes. The aim is to determine which combination of variables yields high model accuracy. Considering variable selection methods are also dependent on the algorithms/models chosen, in the research design stage, no specific methods are determined, keeping the choices open. Second, the author makes manual selections to eliminate variables (elements) that have little design relevance.

2.5 Incorporating the model into agent-based simulation

To show the usefulness of the linear regression model for design use, this study incorporates the model into agent-based simulation, and applies the simulation to the case of Rotterdam Central Station. In the simulation, the 3D model of the scene is modeled regarding the proportion and number of elements. (This modeling process can be seen as a reverse process to the image feature extraction process) (Fig. 7); Each agent (representing each pedestrian) has a view of the scene, and the scene's visual quality is calculated using the linear regression model. Technically, this simulation is first run in MassMotion, a movement simulation software program, to get the movement trajectories of all the agents; then, it is executed using Python script (coded by the author) in Rhino to get a mapping of the visual quality.

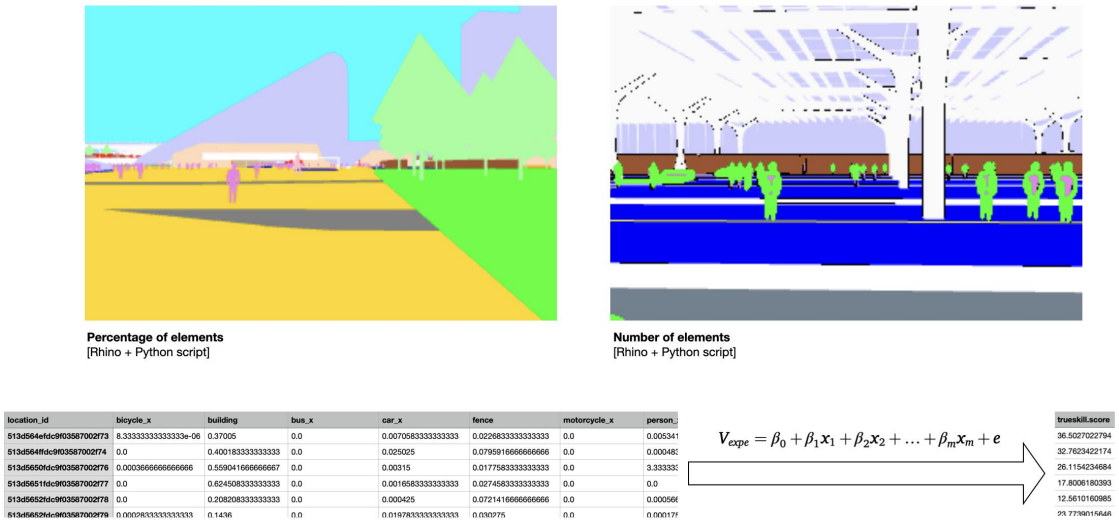

Figure 7: Modeling a design scene regarding the proportion and the number of elements

3 Results

This section presents the results. Subsection 3.1 presents the relationship between humans' perception and view features in a linear regression model. Despite linear regression being the final chosen model in this study, it is one among many choices of models, with many factors affecting the accuracy of models (Subsection 3.2). Subsection 3.3 further integrates the linear regression model into an agent-based simulation approach, for assessing the visual experience in Rotterdam Central Station, demonstrating the practical usage of this linear regression model in design practice.

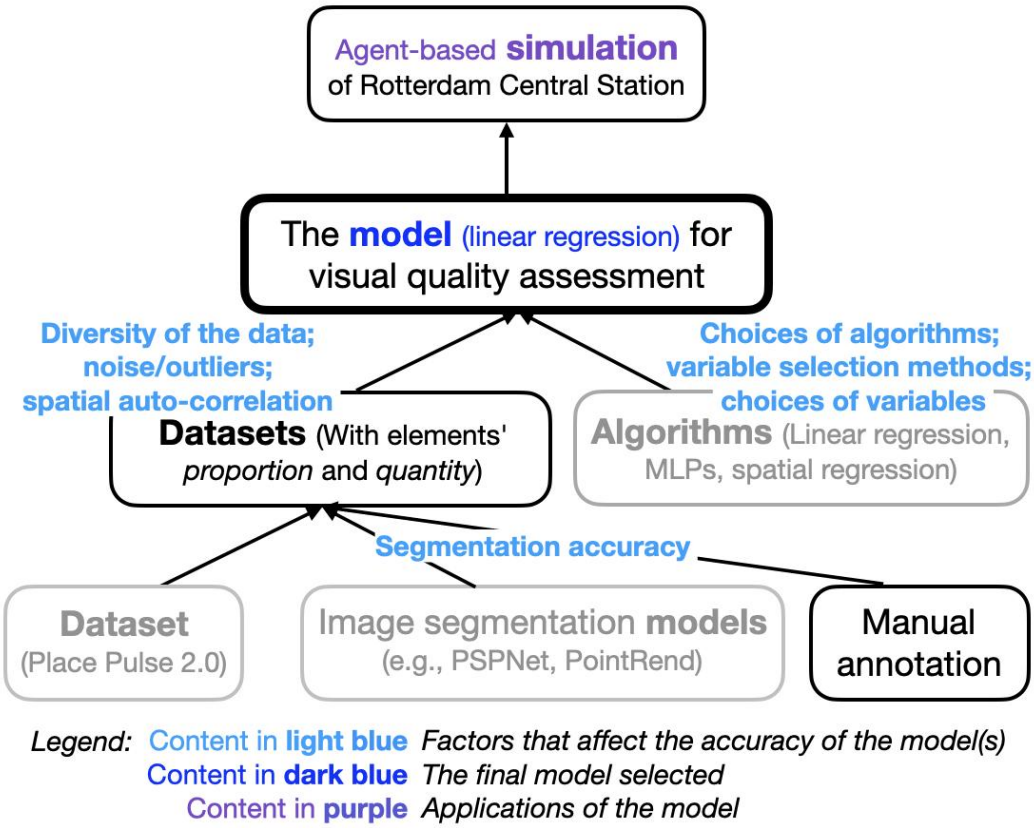

Figure 8: Research results (in colored texts)

3.1 A linear regression model that models the relationship between view elements and 'liveness' perception

Using the linear regression algorithm (Eq. 1), a linear regression model is developed to model the relationship between human perception of 'lively' and street image features (Eq. 2. It is the same model that used in Test 8 in Table 3). Table 2 displays the varying importance of different visual elements contributing to the 'lively' score. It can be seen from the standardized coefficients that the building proportion in the view contributes the most to the 'lively' quality, followed by the person numbers, tree proportion, and water proportion. Sky, although it is an element that people generally like, surprisingly contributes negatively. This may be because when there is more sky in the scene, there are fewer other elements, hence less visual diversity, leading to a declining perception of 'lively.'

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_m x_m + e \tag{1}$$

$$\begin{aligned} Y_{\text{LivelyScore}} = {} & 22.388 - 6.554 \cdot x_{\text{Sky}} + 8.457 \cdot x_{\text{Building}} + 0.539 \cdot x_{\text{Person\_No}} \\ & + 39.143 \cdot x_{\text{Water}} + 4.596 \cdot x_{\text{Vegetation}} + 2.202 \cdot x_{\text{ShannonIndex}} + e \end{aligned} \tag{2}$$

Table 2: Importance of different facilities regarding comfort

Varibles	Unstandardized Coefficients (B)	95% Confidence Interval Lower Bound	Upper Bound	Standardized Coefficients	P-value
(Constant)	22.388	20.457	24.320		<.001
Sky	-6.554	-11.685	-1.422	-0.127	0.012
Building	8.457	5.135	11.780	0.257	<.001
Person_No.	0.539	0.272	0.807	0.174	<.001
Water	39.143	16.241	62.044	0.141	<.001
Vegetation	4.596	1.421	7.771	0.145	0.005
ShannonIndex	2.202	0.656	3.747	0.124	0.005

Note: a) Dependent variable: 'lively' score. b) $R^2 = 0.167$, adjusted $R^2 = 0.156$. c) The Shannon index is calculated based on the proportions of buildings, sky, vegetation, and sidewalks.

3.2 The accuracy of modeling affected by various factors

The final model developed above is built upon a series of works and operations (Fig. 8); Various factors affect the accuracy of modeling, including factors related to data (data diversity, data noise and outliers, spatial auto-correlation, and image segmentation accuracy), the choices of different models, the choices of variable selection methods, and the choice of variables. This study runs multiple tests with different factors, with the results listed in Table 3 (The final selected model is the one used in Test 8), and more details are explained below.

Table 3: Tests on different factors affecting model accuracy

Test	Data	Data manipulation	Algorithm	Variable selection method	Variables selected	Accuracy (adjusted R^2)
1	Whole set [without element quantity] (n=92692)	–	Linear regression	[Auto-selection] 'Backward' method	[14 out of 19 variables][1] car, vegetation, fence, sky, wall, person, terrain, bicycle, pole, traffic.sign, traffic.light, building, truck, train	0.065
2	Whole set (n=92692)	–	Linear regression	[Auto-selection] 'Backward' method	[39 out of 97 variables] car, $car_{No.}$, potted.plant, road, bicycle, $bicycle_{No.}$, pole, vegetation, $person_{No.}$, sky, wall, fence, $fire.hydrant_{No.}$, $traffic.light_{No.}$, ⋯	0.095

(Continued on next page)

(Continued from previous page)

Test	Data	Data modification	Model	Variable selection method	Variables selected	Accuracy (adjusted R^2)
3	Subset Amsterdam (n=506)	–	Linear regression	[Auto-selection] 'Backward' method	[21 out of 97 variables]	0.210
4	Subset Amsterdam (n=490)	Noise and outliers excluded	Linear regression	[Auto-selection] 'Backward' method	[17 out of 97 variables]	0.230
5	Subset Amsterdam (n=490)	Noise and outliers excluded; manually annotated	Linear regression	[Auto-selection] 'Backward' method	[26 out of 135 variables] sidewalk, train, parking.meter, building.wall^{annotated}, traffic.light, kite, road.bricks.red^{annotated}, water, stop.sign, ^{annotated} ...	0.265
6	Subset Amsterdam (n=490)	Noise and outliers excluded; manually annotated	Linear regression	[Manual selection]	[5 out of 135 variables] sky, building$_{annotated}$, person$_{No.}$, water$_{annotated}$, vegetation	0.144
7	Subset Amsterdam [include interaction terms] (n=490)	Noise and outliers excluded; manually annotated	Linear regression	[Auto-selection] 'Backward' method	[6 out of 15 variables][4] sky, building$_{annotated}$, sky × person$_{No.}$, sky × water$_{annotated}$, building$_{annotated}$ × vegetation, person$_{No.}$ × water$_{annotated}$	0.162
8	Subset Amsterdam (n=490)	Noise and outliers excluded; manually annotated	Linear regression	[Manual selection]	[6 out of 136 variables][5] sky, building$_{annotated}$, person$_{No.}$, water$_{annotated}$, vegetation, ShannonIndex	**0.156**

(Continued on next page)

(Continued from previous page)

Test	Data	Data modification	Model	Variable selection method	Variables selected	Accuracy (adjusted R^2)
9	Subset Amsterdam (n=490)	Noise and outliers excluded; manually annotated	Automatic Linear Modeling	[Auto-selection by SPSS]	[10 out of 135 variables] $bicycle_{No.}$, $car_{No.}$, $person_{No.}$, $potted.plant_{No.}$, fence, rider, sky, vegetation, $buliding_{annotated}$, $water_{annotated}$	0.278
10	Subset Amsterdam (n=490)	Noise and outliers excluded; manually annotated	Spatial regression[6]	[Manual selection]	[5 out of 135 variables] sky, $building_{annotated}$, $person_{No.}$, $water_{annotated}$, vegetation	0.161
11	Subset Amsterdam (n=490)[7]	Noise and outliers excluded; manually annotated	MLPs [classification]	[Manual selection]	[5 out of 135 variables] sky, $building_{annotated}$, $person_{No.}$, $water_{annotated}$, vegetation	[Not applicable][8]
12	Subset Amsterdam (n=490)[7]	Noise and outliers excluded; manually annotated	MLPs [regression][9]	[Manual selection]	[5 out of 135 variables] sky, $building_{annotated}$, $person_{No.}$, $water_{annotated}$, vegetation	0.101
13	Subset Amsterdam (n=490)[7]	Noise and outliers excluded; manually annotated	MLPs [regression][9]	[Semi-automatic selection][10]	[5 out of 135 variables] car, $building_{annotated}$, $bicycle_{No.}$, $car_{No.}$, $person_{No.}$	0.137

Note: [1] This data has no element quantity/number information and, therefore, has fewer variables;

[2] The '$_{annotated}$' subscript represents that the corresponding element is manually annotated;

[3] The '$_{No.}$' subscript represents that the variable is the quantity/number of the corresponding element.

[4] This data includes five variables (that were selected in test 6) and their ten interaction terms.

[5] This data has the Shannon index added.

[6] The spatial weight matrix was created using Euclidean distance between the geometric centroids of the spatial units. The spatial regression model used is the spatial lag model, which has a higher accuracy than the spatial error model in this case (0.161 vs 0.153).

[7] The NaN values are imputed with zeros.

[8] In this classification task, the data is divided into ten categories. The model achieves an accuracy of 0.173, meaning it correctly classifies 17.3% of the instances. While this is slightly better than random guessing (which would yield 10.0% accuracy for ten categories), the performance is still poor.

[9] Hyperparameters: solver: 'adam', learning rate init: 0.001, hidden layer sizes: (100, 50), alpha: 0.0001, activation: 'relu.'

[10] In this test, The variables are selected by machine, based on ANOVA F-test. However, the number of top variables to select is decided manually, as five.

3.2.1 (Geographical) Diversity of data samples

The diversity of data samples affects the accuracy of modeling, as shown by the comparison between tests 2 and 3 in Table 3. It is a common understanding in data analysis that the more diverse the dataset is, the harder it becomes to identify clear patterns, as the variability in the data increases. This study first tried the whole Place Pulse 2.0 dataset (data sample size $n = 92692$, after exclusion of images with missing score values), which contains street view images from 56 cities across the world; then it tried to use a subset of the whole dataset, which contains data of only one city – Amsterdam ($n = 506$). The latter test resulted in significantly higher accuracy ($R^2 = 0.210$) than the former ($R^2 = 0.095$). The difference in accuracy when modeling using the whole dataset and the subset, also indicates that geographically, different environments vary significantly. [Therefore, it is important to note that this study's final model has a limited context, and its generalization ability in other cities or geographical areas needs further examination.]

3.2.2 Data noise and outliers

Data noise and outliers affect the accuracy of modeling, as shown by the comparison between tests 3 and 4 in Table 3. When analyzing the relationship between 'building proportion' and the 'lively' score, it first seems no statistically reliable relationship exists (with a high p-value, > 0.05). This is counter-intuitive because common sense suggests that a scene with more buildings would appear livelier (e.g., Fig. 9, a)). Presumptively, a strong, reliable relationship should exist (p-value < 0.05). After examining the data, the author realized that in the dataset, while most images have buildings as positive elements, there are some images where the buildings look really negative to pedestrians' perception, as they basically are walls with little or no windows and cover a large proportion of the scene (e.g., Fig. 9, b)). After removing these 'outlier' images, the p-value dramatically reduces to a value less than 0.05. This study also removed images that are totally unrecognizable (e.g., Fig. 9, c)) and images of tunnel environments which are significantly different from normal

street scenes (e.g., Fig. 9, d)). After noise removal (data sample size from $n = 506$ to $n = 490$), the accuracy of modeling increased (from $R^2 = 0.210$ to $R^2 = 0.230$). This accuracy difference also indicates that different scene environments will have impacts on the model accuracy. [Therefore, the developed model is limited in urban street environments, and the generalization ability needs further examination when applied to other environments such as indoor spaces and tunneling environments]

Figure 9: The data noise or outliers in the data

3.2.3 Spatial autocorrelation of data samples

Spatial autocorrelation of data samples affects the accuracy of modeling, as shown by the comparison between tests 10 and 6 in Table 3. Given the same data, when spatial autocorrelation is addressed (by applying the spatial weight matrix), the accuracy value increases (from $R^2 = 0.144$ to $R^2 = 0.161$). Spatial regression takes the assumption that visual quality is associated with spatial autocorrelation, while linear regression takes the assumption that visual quality depends on the image features and is independent of locations. Despite being more accurate in these tests, the spatial regression model is a less usable model than the linear regression model. This is because spatial regression models are dependent on weight matrices, and each weight matrix is unique (linked with the corresponding geographical area), making the associated spatial regression not transferrable to other geographical contexts.

3.2.4 The accuracy of image segmentation

The accuracy of image segmentation affects the accuracy of modeling, as shown by the comparison between Tests 4 and 5 in Table 3. With manually annotated variables, the model's accuracy increased (from $R^2 = 0.230$ to $R^2 = 0.265$). This study's final selected model includes manually annotated variables (Eq. 2, Test 8 in Table 3). More accurate semantic segmentation of images can be achieved with the fast development of image segmentation in the computer field (e.g. see (PaperswithCode)).

3.2.5 The choice of algorithms

The choice of (mathematical or machine learning) algorithms affects the accuracy of modeling, as shown by the comparison between tests 5 and 9; and between tests 6, 10, 11, and 12 in Table 3. Auto-linear regression in SPSS software has slightly higher accuracy than linear regression ($R^2 = 0.278$ vs $R^2 = 0.265$). Spatial regression has a higher accuracy than linear regression ($R^2 = 0.161$ vs $R^2 = 0.144$). Typically, MLP (Multilayer Perceptron) regression is able to capture

more complex relationships than linear regression hence more accurate. However, surprisingly, the MLP regression is no better than the linear regression in this study's tests ($R^2 = 0.101$ to $R^2 = 0.144$). This may be due to the sample size being small ($n = 490$), while MLP is usually effective on large data. When changing the task from a regression task to a classification task, the MLP's performance is still not impressive (accuracy as 0.17 vs 0.10, see table note [8] in Table 3).

This study adopted linear regression as the final algorithm for the training model, considering the normal linear regression algorithm is clearer for interpretation, and none of the other algorithms yielded dramatically improved accuracy. Linear regression is usually an oversimplification of reality, but statistically, nevertheless, it shows the net effects of independent variables on the dependent variable.

3.2.6 The choice of variable selection methods

When using the machine to auto-select variables, different variable selection methods can lead to different final sets of selected variables, resulting in models with different accuracy. For linear regression, in SPSS software, there are 'forward,' 'backward,' and 'stepwise' as variable selection methods. When testing the whole set of data ($n = 92692$), these three methods resulted in similar accuracy. When testing with the subset data of Amsterdam ($n = 490$), the 'backward' method resulted in the highest accuracy (Table 4). For MLP regression, this study tested various methods, including KBest F-test, KBest Mutual Info, Lasso, Tree Importance, and PCA. The highest accuracy resulted from the KBest F-test method (Table 4). Since automatic selections by the machine usually include variables that hardly have relevance to spatial design during the design stage, or can not be modeled reliably (e.g., the proportion of parking meters and traffic lights and the number of bicycles in Test 5), this study has to drop them and make manual selections further. After several tests, a combination of variables that balance the accuracy and design relevance was chosen (Test 6 in Table 3).

Table 4: Model accuracy affected by variable selection methods

Models	**Selection Methods**	**Accuracy (adjusted R^2)**
Linear regression (Based on Test 5 in Table 3)	Stepwise	0.220
	Forward	0.220
	Backward	0.265
MLP regression (Based on Test 13 in Table 3)	KBest (F-test)	0.201
	KBest (Mutual Info)	- 0.130
	Lasso	0.051
	Tree Importance	- 0.102
	PCA	0.151

3.2.7 The choice of variables

Models' accuracy is determined by the variables selected. Typically, more variables enable more accurate models. For example, in comparisons between Tests 1 and 2 in Table 3, when the variables of element quantity were added, the model's accuracy improved (from 0.065 to 0.095). Many variables automatically selected by the machine have little relevance to the spatial design or can not be modeled reliably and, hence, are dropped. The manual selection finally includes five variables: sky, $\text{building}_{\text{annotated}}$, $\text{person}_{\text{No.}}$, $\text{water}_{\text{annotated}}$, and vegetation (see Test 6 in Table 3). Beyond these selected variables, two types of artificial variables calculated based on them, interaction terms and the Shannon index, are also tested.

Interaction terms represent the combined effect of two or more independent variables on the dependent variable (e.g., exercise and diet may have combined effects on weight loss). Including interaction terms slightly increase the model accuracy, as shown by the comparison between Tests 7 and 6 in Table 3. However, in this study, the interaction terms selected by the machine have no evident real-world meaning (e.g., sky × $\text{person}_{\text{No.}}$).

Shannon index has been used as an indicator for measuring diversity in the built environment (Zhong et al., 2020, Zachary and Dobson, 2021); In urban design research, visual diversity or complexity are qualities commonly mentioned to be important to environmental perception (Ewing and Handy, 2009, Jacobs, 1961); This study achieved a statistically reliable Shannon index (p-value < 0.05), calculated based on the visual proportions of several elements, including buildings, sky, trees, grass, and sidewalks. The model's accuracy increased with this Shannon index being added as an extra variable (See the comparison between Tests 8 and 6).

3.2.8 Comparing the accuracy of this study's final model with other studies'

The above tests explained how the various factors affect models' accuracy; with some manipulations of these factors, relevant studies in the literature have achieved much higher accuracy values than this study's (This study's final model's $R2 = 0.167$, and the adjusted $R2 = 0.156$). For example, $R^2 = 0.54$ in (Naik et al., 2014), $R^2 = 0.51$ in (Nagata et al., 2020), and $R^2 = 0.238$ in (Liang, 2020)). This study's approach only includes the proportions and quantities of visual elements (to make it usable for design renderings). In comparison, these other papers incorporated more advanced visual features as variables, or non-visual variables. In the study by Naik et al. (2014), all variables are features that only exist in photos, and do not exist in typical cartoon-like design renderings. In the paper by Nagata et al. (2020), the width of roads and interactive terms are important variables; also, that paper includes many statistically unreliable variables (p-value > 0.05). In (Liang, 2020)'s paper, variables of urban form and road network are included.

3.3 Agent-based simulation of visual experience at Rotterdam Central Station

Fig. 10 shows the result of the final linear regression model being incorporated into an agent-based simulation and applied to the case of Rotterdam Central Station. The mappings show that the new station square, compared to the old station square, generally has higher visual qualities. This is because, in the old scenario, when people stand in the

plaza area (location 1), their sight lines towards positive environmental elements (including buildings, trees, and humans) are pretty much blocked by vehicles. The mapping results correspond with the users' real-world perceptions. These mappings demonstrate the effectiveness of the final model, and illustrate the potential of computational assessment of visual quality in design practice.

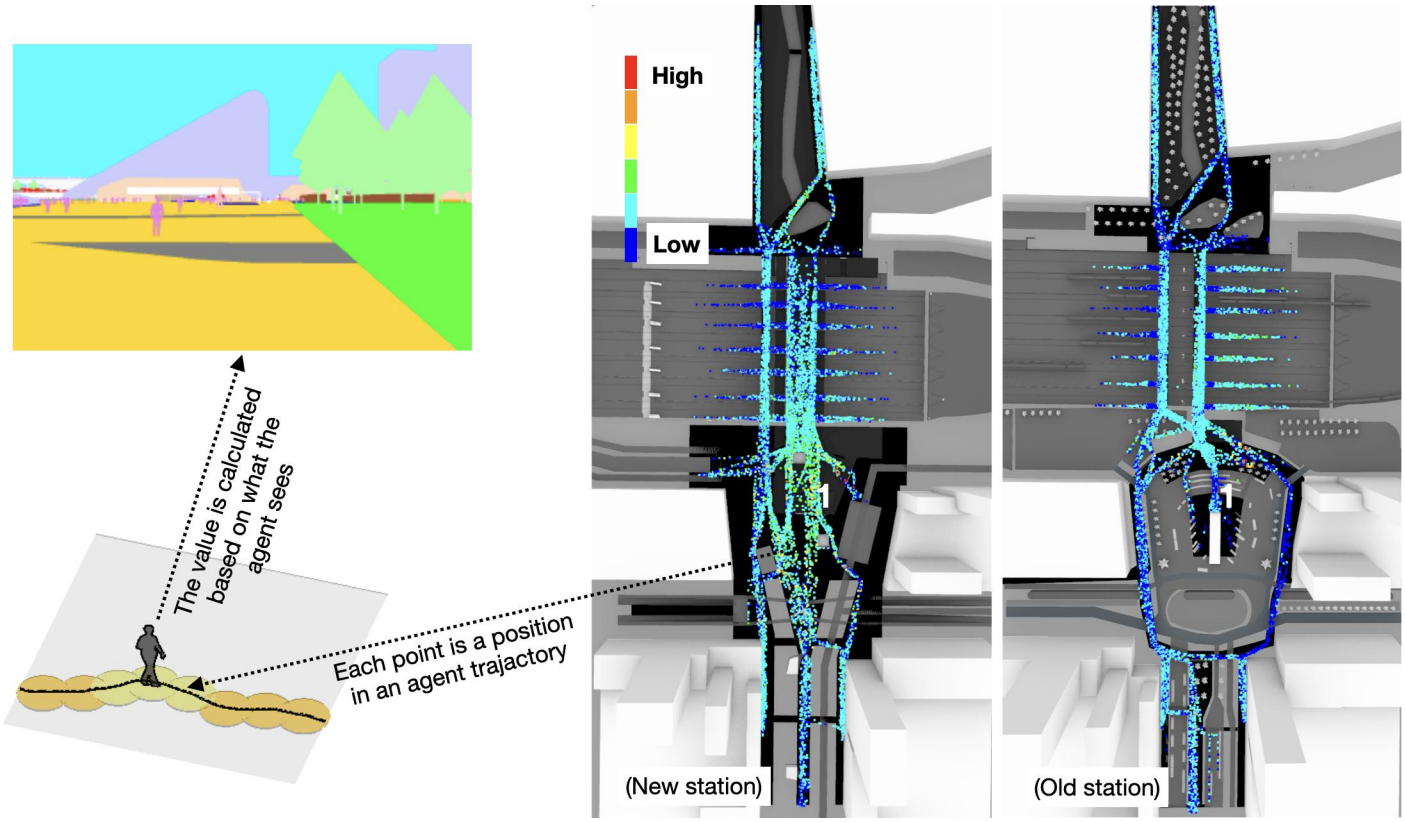

Figure 10: Visual quality mapping of the case of Rotterdam Central Station area

It requires significant technical efforts to conduct the simulation. Firstly, this study optimized computational accuracy and the speed of image processing. E.g., while exporting .jpg pictures results in huge errors in the calculation of element proportion in the view, exporting as .png format resolves this issue. The processing of all agent views by Python script using a pixel-by-pixel approach takes hours, while using PIL (Python Image Library) only takes a few minutes. Secondly, this study addressed the differences between different digital tools. E.g, while an agent position is recorded with coordinates as X, Y, Z in MassMotion (the movement simulation software), the same position is denoted as X, -Z, Y in Rhino (in 3D spatial modeling software).

Of notice, in this agent-based simulation, the visual quality mapping relies on the movement trajectories; and the movement trajectories are generated based on the origin-destination (OD) matrix using movement algorithms, which potentially brings bias to the visual mapping.

4 Discussions

4.1 Limitations of assessing visual quality using computational means

This study exhibits the limitations of computational assessment of visual quality. As shown by the research results, models are typically built on many primary works and correspondingly affected by various associated factors. The final linear regression model is trained on data with limited variability regarding geography (Amsterdam) and environment type (outdoor street environment). Therefore, the model should be limitedly used for new cases with similar characteristics [NB: Some research suggests different environments can result in significant variations of perception (Lou et al., 2024)]. Training a model requires huge efforts, and yet the model's generalizability and accuracy are often limited. This highlights the necessity of traditional means of visual quality assessment – designers' (who are experts with aesthetics to some degree) manual judgments.

4.2 Future research directions

There are several future research directions toward more generalizable and accurate models (Fig 11). Future research can verify the training data' variations, by including new data of more types of geography, environments (outdoor), or buildings (indoor); These new data can either be collected in similar ways to what this study did – extracting image features from datasets like Place Pulse 2.0, or be collected from VR experiments – where digital virtual environments provide directly-extractable image features. Of notice, In this study, manual labeling/annotation was used to ensure the segmentation accuracy and to investigate some elements that are not included in segmentation models' variable categories. However, manual labeling, as a laborious task, is challenging to apply to large datasets. Thus, image segmentation by machine is still promising and requires more accurate segmentation models trained on more properly labeled datasets, using future new algorithms.

Future research can also do city-level mapping using the developed model, and compare the effectiveness of mappings using this model versus using existing other models (Fig. 11). If the developed model has a similar effectiveness, then it provides a new way of city-level visual quality assessment that does not rely on street images but on 3D models. (This would be meaningful for areas that have no street images but have rough 3D spatial information).

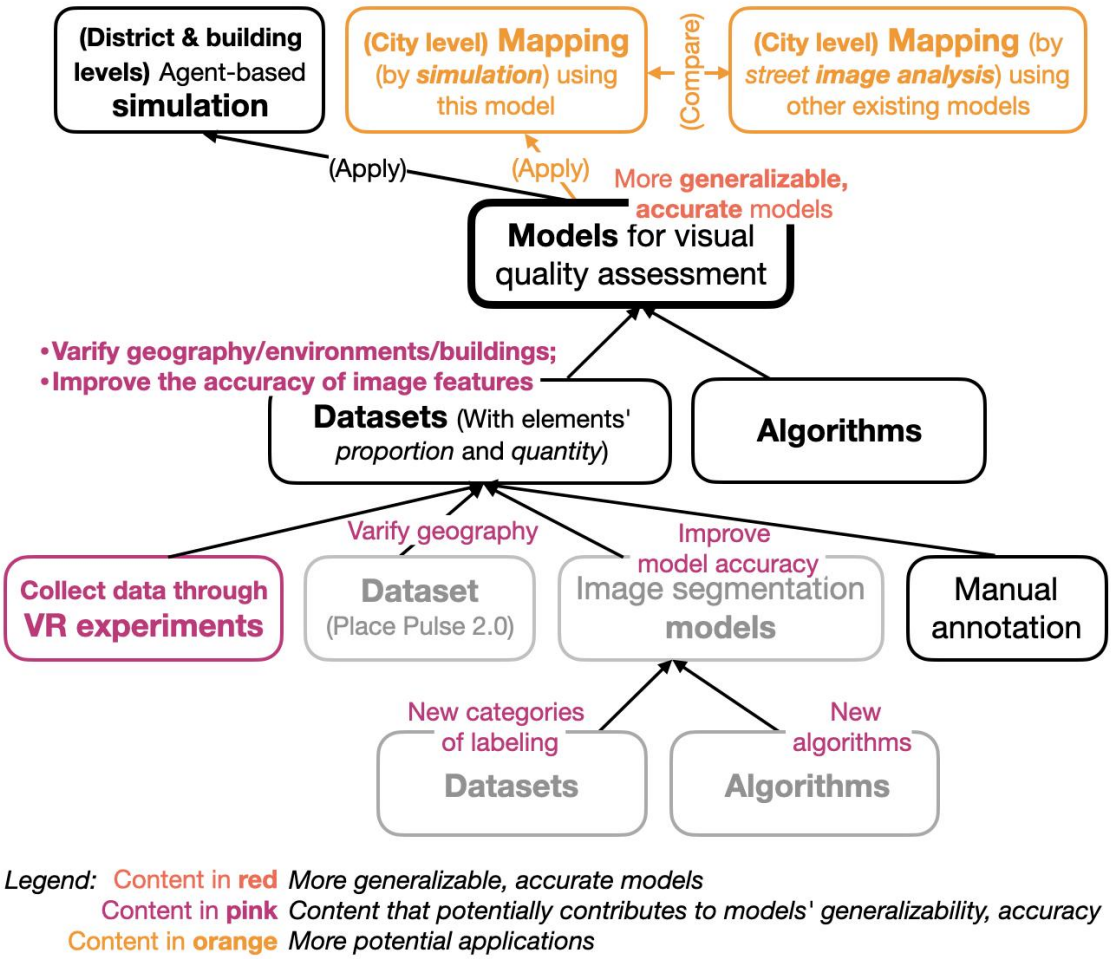

Figure 11: Future research directions (in colored texts)

4.3 Research contributions

This study contributes to the knowledge body in the following ways: Firstly, by building a linear regression model that calculates visual quality ('lively') based on the environment's elements, it extends the quantitative understanding of visual quality. Secondly, this study shows the possibility, as well as outlines the limitations and future directions of visual quality assessment through computational means.

5 Conclusions

In the introduction section, a research question is proposed: How to computationally assess the visual quality of the built environment for design use? This question is answered by this study's final linear regression model. This model is developed based on extensive tests on a series of relevant factors. The usage of this model is further demonstrated through an agent-based simulation, of the Rotterdam Central Station area.

References

Benedikt, M.L., 1979. To take hold of space: Isovists and isovist fields. Environment and Planning B: Planning and Design 6, 47–65. doi:10.1068/b060047.

Chen, E., Yang, S., Zhuang, Y., 2021. Analysis of Urban Security Based on Visibility: Taking Utrecht Railway Station Area as an Example. Architecture Technique 27, 121–123. doi:10.19953/j.at.2021.04.031.

Chen, K., Wang, J., Pang, J., Cao, Y., Xiong, Y., Li, X., Sun, S., Feng, W., Liu, Z., Xu, J., Zhang, Z., Cheng, D., Zhu, C., Cheng, T., Zhao, Q., Li, B., Lu, X., Zhu, R., Wu, Y., Dai, J., Wang, J., Shi, J., Ouyang, W., Loy, C.C., Lin, D., 2019. MMDetection: Open MMLab Detection Toolbox and Benchmark. doi:10.48550/arXiv.1906.07155, arXiv:1906.07155.

Cliff, A.D., Ord, J.K., 1981. Spatial Processes: Models & Applications. Pion.

De Nadai, M., Vieriu, R.L., Zen, G., Dragicevic, S., Naik, N., Caraviello, M., Hidalgo, C.A., Sebe, N., Lepri, B., 2016. Are Safer Looking Neighborhoods More Lively? A Multimodal Investigation into Urban Life. arXiv:1608.00462.

Ewing, R., Clemente, O., Neckerman, K.M., Purciel-Hill, M., Quinn, J.W., Rundle, A., 2013. Measuring Urban Design. Island Press/Center for Resource Economics, Washington, DC. doi:10.5822/978-1-61091-209-9.

Ewing, R., Handy, S., 2009. Measuring the Unmeasurable: Urban Design Qualities Related to Walkability. Journal of Urban Design 14, 65–84. doi:10.1080/13574800802451155.

Filomena, G., Verstegen, J.A., Manley, E., 2019. A computational approach to 'The Image of the City'. Cities 89, 14–25. doi:10.1016/j.cities.2019.01.006.

Florio, P., Leduc, T., Sutter, Y., Brémond, R., 2023. Visual complexity of urban streetscapes: Human vs computer vision. Machine Vision and Applications 35, 7. doi:10.1007/s00138-023-01484-1.

Gehl, J., 1987. Life between Buildings. volume 23. New York: Van Nostrand Reinhold.

Haupt, C., Huber, A.B., 2008. How axons see their way–axonal guidance in the visual system. Frontiers in Bioscience: A Journal and Virtual Library 13, 3136–3149. doi:10.2741/2915.

He, K., Gkioxari, G., Dollár, P., Girshick, R., 2018. Mask R-CNN. doi:10.48550/arXiv.1703.06870, arXiv:1703.06870.

Ibrahim, M.R., Haworth, J., Cheng, T., 2020. Understanding cities with machine eyes: A review of deep computer vision in urban analytics. Cities 96, 102481. doi:10.1016/j.cities.2019.102481.

India, C., 2023. Pixellib: A Python Library for Easy Image Segmentation.

Jacobs, J., 1961. The Death and Life of Great American Cities. 1st ed., New York : Random House.

Lee, J., Kim, D., Park, J., 2022. A Machine Learning and Computer Vision Study of the Environmental Characteristics of Streetscapes That Affect Pedestrian Satisfaction. Sustainability 14, 5730. doi:10.3390/su14095730.

Li, S., Ma, S., Tong, D., Jia, Z., Li, P., Long, Y., 2022. Associations between the quality of street space and the attributes of the built environment using large volumes of street view pictures. Environment and Planning B: Urban Analytics and City Science 49, 1197–1211. doi:10.1177/23998083211056341.

Li, X., Ratti, C., Seiferling, I., 2017. Mapping Urban Landscapes Along Streets Using Google Street View, in: Peterson, M.P. (Ed.), Advances in Cartography and GIScience, Springer International Publishing, Cham. pp. 341–356. doi:10.1007/978-3-319-57336-6_24.

Liang, Q., 2020. Machine Mediated Human Perception. Thesis. Massachusetts Institute of Technology.

Liu, L., Sevtsuk, A., 2024. Clarity or confusion: A review of computer vision street attributes in urban studies and planning. Cities 150, 105022. doi:10.1016/j.cities.2024.105022.

Long, Y., Ye, Y., 2019. Measuring human-scale urban form and its performance. Landscape and Urban Planning 191, 103612. doi:10.1016/j.landurbplan.2019.103612.

Lou, S., Stancato, G., Piga, B.E.A., 2024. Assessing In-Motion Urban Visual Perception: Analyzing Urban Features, Design Qualities, and People's Perception, in: Giordano, A., Russo, M., Spallone, R. (Eds.), Advances in Representation: New AI- and XR-Driven Transdisciplinarity. Springer Nature Switzerland, Cham, pp. 691–706. doi:10.1007/978-3-031-62963-1_42.

Nagata, S., Nakaya, T., Hanibuchi, T., Amagasa, S., Kikuchi, H., Inoue, S., 2020. Objective scoring of streetscape walkability related to leisure walking: Statistical modeling approach with semantic segmentation of Google Street View images. Health & Place 66, 102428. doi:10.1016/j.healthplace.2020.102428.

Naik, N., Philipoom, J., Raskar, R., Hidalgo, C., 2014. Streetscore – Predicting the Perceived Safety of One Million Streetscapes, in: 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 793–799. doi:10.1109/CVPRW.2014.121.

PaperswithCode, . Cityscapes test Benchmark (Semantic Segmentation). https://paperswithcode.com/sota/semantic-segmentation-on-cityscapes.

Paydar, M., Kamani Fard, A., Gárate Navarrete, V., 2023. Design Characteristics, Visual Qualities, and Walking Behavior in an Urban Park Setting. Land 12, 1838. doi:10.3390/land12101838.

Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You Only Look Once: Unified, Real-Time Object Detection. doi:10.48550/arXiv.1506.02640, arXiv:1506.02640.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. doi:10.48550/arXiv.1505.04597, arXiv:1505.04597.

Salesses, M.P., 2012. Place Pulse Measuring the Collaborative Image of the City. Ph.D. thesis. Massachusetts Institute of Technology.

Simonyan, K., Zisserman, A., 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. doi:10.48550/arXiv.1409.1556, arXiv:1409.1556.

Tang, J., Long, Y., 2019. Measuring visual quality of street space and its temporal variation: Methodology and its application in the Hutong area in Beijing. Landscape and Urban Planning 191, 103436. doi:10.1016/j.landurbplan.2018.09.015.

van der Walt, S., Schönberger, J.L., Nunez-Iglesias, J., Boulogne, F., Warner, J.D., Yager, N., Gouillart, E., Yu, T., 2014. Scikit-image: Image processing in Python. PeerJ 2, e453. doi:10.7717/peerj.453.

Wang, R., Feng, Z., Pearce, J., Yao, Y., Li, X., Liu, Y., 2021. The distribution of greenspace quantity and quality and their association with neighbourhood socioeconomic conditions in Guangzhou, China: A new approach using deep learning method and street view images. Sustainable Cities and Society 66, 102664. doi:10.1016/j.scs.2020.102664.

Whyte, W.H., 2001. The Social Life of Small Urban Spaces. 8th ed. edition ed., Project for Public Spaces, New York, NY.

Xiang, L., Cai, M., Ren, C., Ng, E., 2021. Modeling pedestrian emotion in high-density cities using visual exposure and machine learning: Tracking real-time physiology and psychology in Hong Kong. Building and Environment 205, 108273. doi:10.1016/j.buildenv.2021.108273.

Zachary, D., Dobson, S., 2021. Urban Development and Complexity: Shannon Entropy as a Measure of Diversity. Planning Practice & Research 36, 157–173. doi:10.1080/02697459.2020.1852664.

Zare, Z., Yeganeh, M., Dehghan, N., 2022. Environmental and social sustainability automated evaluation of plazas based on 3D visibility measurements. Energy Reports 8, 6280–6300. doi:10.1016/j.egyr.2022.04.064.

Zhong, T., Lü, G., Zhong, X., Tang, H., Ye, Y., 2020. Measuring Human-Scale Living Convenience through Multi-Sourced Urban Data and a Geodesign Approach: Buildings as Analytical Units. Sustainability 12, 4712. doi:10.3390/su12114712.