This is the Accepted Manuscript version of an article published by Elsevier in the journal *Building and Environment* in 2024, which is available at:
`https://doi.org/10.1016/j.buildenv.2024.111875`
Cite as: Liang X, Chang JH, Gao S, Zhao T, Biljecki F (2024): Evaluating human perception of building exteriors using street view imagery. *Building and Environment*, 263: 111875

Evaluating Human Perception of Building Exteriors Using Street View Imagery

Xiucheng Liang[a,b], Jiat Hwee Chang[a], Song Gao[c], Tianhong Zhao[d], Filip Biljecki[a,e,*]

[a]*Department of Architecture, National University of Singapore, Singapore*
[b]*School of Life Sciences, Zhaoqing University, China*
[c]*Department of Geography, University of Wisconsin-Madison, United States*
[d]*College of Big Data and Internet, Shenzhen Technology University, Shenzhen, China*
[e]*Department of Real Estate, National University of Singapore, Singapore*

Abstract

Building appearances profoundly shape the urban visual landscape, influencing city images and the quality of urban life. Traditional methods for evaluating the perceptual and aesthetic qualities of building facades are often limited in scope. Despite recent studies that have sought to understand human perception of urban streetscapes, our grasp of how individuals perceive building exteriors on a broader scale and the subsequent impact on holistic street experiences, remains largely unexplored. In this study, we integrate a traditional survey-based evaluation framework with machine learning techniques to analyse human perception of over 250,000 building images from Singapore, San Francisco, and Amsterdam. Specifically, deep learning models trained on crowdsourced ratings of 1,200 building images across six perceptual attributes — complex, original, ordered, pleasing, boring, and style — achieve over 72% accuracy. This novel approach enables adaptive and comparative analyses of building appearances across regions, revealing spatial patterns in the perception of architectural exteriors and their relationships with functions, age, and location. Moreover, by applying propensity score matching to match images based on their features, we mark one of the first efforts to investigate the perceptual impacts of buildings on streetscape perceptions. The results show that streetscapes with higher levels of complex, pleasing, and historical ambience from buildings elicit more positive perceptions, whereas modern and monotonous exteriors often evoke holistic feelings of being "boring" and "depressing". These findings offer architects and city planners valuable insights into public sentiment towards city-level building exteriors and their influence on urban identity and perception.

Keywords: computer vision, built environment, building facade, architectural design, street-level imagery

Preprint submitted to Building and Environment

1. Introduction

Over the past decades, urban built environments have grown at an unprecedented rate, shaped by a confluence of socio-cultural, economic, and technological factors, resulting in diverse and complex urban landscapes. The visual appearance of a city, a crucial aspect of the environmental aesthetics, is believed to significantly influence people's perceptions of a place as well as their physical and mental well-being [1, 2, 3]. Studies have shown that an appropriate appearance can enhance the emotional appeal of the environment, thereby affecting human experiences related to outdoor thermal comfort [4, 5, 6], acoustic perception [7, 8] and perceived safety [9, 10, 11]. Exposure to aesthetically pleasing environments, such as natural scenes, has been shown to be a key factor in promoting residents' health [12, 13, 14, 15, 16], while visual disorder in urban settings may induce rule-breaking behaviours [17, 18]. Consequently, evaluating the visual appearance of built environments is pivotal in cultivating urban aesthetics and fostering a more liveable urban environment.

With the rapid development of urban informatics, the effective and quantitative measurement of visual landscapes has become a key area of interest in the built environment sector [19, 20, 21]. Recently, street-level imagery has emerged as a valuable data source, offering a unique perspective and detailed coverage [22]. This facilitates the scalable exploration of landscape intricacies from a pedestrian's viewpoint, including both the objective measurement of visual elements [23] and the subjective assessment of human perceptions [24, 25]. Further investigations have also illuminated the impact of visual components and image features on individuals' sentiments towards urban environments [11, 26, 27, 16, 28]. These studies not only evaluate human perceptions in broader extents but also underscore how landscapes interact with people's emotions, contributing to more human-centric urban environments [29].

However, the specific characteristics of visual elements in the built environment, such as building appearances, types of vegetation, and street designs, along with their spatial distributions and contributions to individuals' holistic sensory experiences, have not been fully explored. Among these elements, architectural exterior designs, which profoundly impact our daily visual experiences and cultural interactions, have received considerable attention in previous studies [30, 31, 32] due to their diverse appearances that range from historical to contemporary eras. Similar to urban street studies, architectural facade evaluation not only explores the impact of design on the environment, including thermal comfort and microclimates [33], but also incorporates psychological theories to investigate emotional

*Corresponding author

appraisals and elemental evaluations of buildings through aesthetic responses based on human cognition and perception [34]. Such perceptual assessments attempt to understand the interaction between human sentiments and architectural semantics [35, 36, 32], thereby aiding architects and urban designers in improving design principles for the public good [37]. However, these studies often rely heavily on traditional data collection methods such as interviews and questionnaires, typically focusing on specific building types or geographical areas. This approach confines the research to a narrow scope and limits scalability for cross-regional comparison, thereby posing challenges in evaluating architectural design at flexible spatial resolutions. Hence, this study aims to broaden our grasp of human perceptions of buildings within a large-scale urban environment, to compare their spatial distributions across cities, and to explore how building exteriors influence holistic human perception of streetscapes. Through this study, we seek to address the following research questions (RQ):

- RQ1: How well can machine learning be leveraged to describe and compare the building exteriors in a detailed and scalable manner?
- RQ2: How do the human perceptions vary across cities that are constituted by different types of building exteriors?
- RQ3: To what extent do the appearances of buildings impact the overall human perception of urban streetscapes?

In this research, we integrate a building exterior evaluation framework with deep learning techniques to measure human perception of building exteriors. First, a dataset of individual building images is compiled from street-level images taken in three cities: Singapore, San Francisco and Amsterdam. Second, a subset of these images is equally and manually assembled across different areas of the cities and various building types. This subset is then evaluated through a comprehensive online survey, which gathers 33,774 responses from 493 participants. The survey quantifies 1,200 images across six perceptual attributes: complex, original, ordered, pleasing, boring, and style, as identified from prior research. Third, deep learning models are trained on this dataset to analyse the entire set of about 250,000 building images, with the results applied to reveal spatial patterns of human perceptions on building facades. Moreover, this study conducts a correlation analysis between perceptions of buildings and their objective information (i.e., functions, age, and location), to pinpoint distinctive building identities across cities. Finally, through the lens of the six attributes, this research examines their influence on the holistic human perception of streetscapes. Our findings show that the proposed perceptual properties are effective in comparing architectural exteriors across large-scale built environments and elucidating the impacts of building designs on human perception of streetscapes.

The contributions of this research are threefold: (1) We introduce a scalable and efficient deep learning-based framework, featuring six perceptual attributes of building facades, for analysing and comparing human perceptions of architectural exteriors in different urban contexts. (2) The characteristics and distribution of building perceptions are revealed and compared across cities, to facilitate a broader understanding of architectural design strategies. (3) This work is among the first to investigate the influence of building appearances on human perception of streetscapes in larger regions. This investigation offers architects and city planners not only an effective framework for evaluating public sentiment towards city-level building exteriors, but also insights into their influence on urban identity and perceptions, underlining the significance of architectural design in urban planning and policy development.

2. Related work

2.1. Traditional approaches in architectural exterior evaluation

Over the past decades, the sensory connection between observers and objects, derived from psychological theory, has been widely applied in built environment studies [38]. Various studies have identified that an appropriate form or appearance may facilitate attractiveness and emotional appeal of environments, including the aesthetic quality of architecture, landscape, and streetscape [39, 37, 40, 41]. These studies highlight conceptual properties such as coherence, meaningfulness, enclosure, and mystery, which represent individual perceptions to measure cognitive and perceptual dimensions of the environment [37, 42, 43].

In architectural exterior assessment, conceptual properties offer designers a holistic view of an individual's aesthetic response and affective appraisal of buildings. Such assessments are typically conducted through surveys using interviews and questionnaires to collect people's ratings and opinions on a building's physical setting [31, 44, 45]. Different terminologies and combinations have been explored in various studies, playing a fundamental role in identifying specific and relevant aesthetic needs in building designs. Classic academic works have established correlations between specific perceptual attributes and preferences for certain buildings. For instance, perceived complexity and impressiveness have been shown to have a linear relationship, while a U-shaped relationship exists between complexity and preference criteria, indicating that people tend to favour buildings with a moderate level of complexity [30, 46]. Other properties, such as inclusion of historical elements and the exclusion of artificial elements, have been found to increase the pleasantness of buildings, while to optimise excitement, the use of more natural materials and higher levels of atypicality would be beneficial [37, 47].

Among the studies, *complexity* is regarded as a crucial factor in formal aesthetics and is frequently adopted as a key conceptual property in assessing building exteriors. Rapoport [48] describes complexity as tied to the number of noticeable differences in independent elements and the amount of usable information available to the viewer. A building with higher exterior complexity tends to deliver more intense visual information, increasing arousal experiences while typically minimising pleasantness at the extreme ends of high or low complexity [49]. *Order*, along with related variables like clarity, coherence, and fittingness, has been confirmed to influence preferences in streetscapes and housing appearances [50, 31, 32]. It refers to the degree to which a scene cohesive or sensible. Kaplan and Kaplan [51] highlight that an ordered appearance enhances legibility and identifiability, simplifying comprehension and recall for observers. Importantly, they also indicate that a scene can possess high complexity and high coherence simultaneously. Moreover, the contribution of *originality* to architecture design, similar to what Devlin and Nasar [31] and Canter [52] call "novelty", has also been extensively documented [42, 32, 36]. This term evaluates the uniqueness of a subject and is identified to significantly influence perceptual behaviour [53]. Similarly, architecture *style* represents an important symbolic variable, describing a general denotative meaning interpreted by individuals [37]. Such interpretations, as perceived by humans, may vary among different styles [54], illustrating how historical ambience and detailing can endow building facades with legibility, coherence, and harmony [55, 56], whereas certain modern cues may elicit pleasure and arousal in observers [35]. This interpretation significantly influences observers' emotional responses, to the extent that the presence of historical elements has been identified as beneficial to quality of life [57].

Besides, pleasure and arousal are key properties commonly investigated in academic discussions on building exteriors [31, 35, 58, 32]. These dimensions are closely linked to physiological responses and are essential for evaluating human reactions to architectural stimuli. Although these dimensions are interrelated, they represent independent aspects of experience. Russell et al. [59] clarify that pleasure and arousal maintain an orthogonal relationship, implying distinct influences on human reactions. Expanding on this, Gifford et al. [35] develop a graphic circumplex method to quantitatively assess pleasure and arousal levels elicited by images of modern buildings. Their findings indicate a strong correlation: buildings that are reflective, shiny, taller, more colourful, and ornate tend to provoke greater arousal, while these architectural qualities do not necessarily provide significant cues for pleasure. Hence, in this study, *pleasantness* is utilised to assess affective expression as the bipolar opposite of displeasure, while *excitement* (not boring) is employed to denote arousal, capturing the intensity of physiological activity.

In addition to these six aspects, various attributes such as friendliness, ruggedness, and ornateness have been employed to evaluate building facades. In this work, we aim to explore an adaptable framework to measure building exteriors with perceptual properties that are distinct and comprehensible. The framework can be progressively expanded to include other variables in future empirical and theoretical research.

2.2. Street-level imagery in human perception measurement

Street view imagery (SVI) has gained significant popularity in recent studies, providing a unique street-level vantage point of urban landscapes with broad coverage and precise spatial detail, and has thus been extensively employed across diverse scales to explore built environments [22]. These images are often customised to align with the specific objectives of various studies and are used to train machine learning models for predictive tasks.

In highlighted selection, we emphasise the urban perception studies that use perception-based labelling to quantitatively and extensively measure human responses to street environments. Typically, these studies employ SVI surveys to explore subjective feelings of participants, categorise images with perceptual labels, and convert these labels into quantifiable attributes [2]. The seminal work by Salesses et al. [24], introduces a methodology to obtain SVI ratings based on pairwise comparisons and responses to evaluative questions. This approach explores the perceptual inequalities of safety, class, and uniqueness in cities within the United States. Expanding upon this foundation, Dubey et al. [25] establish a crowdsourced dataset known as Place Pulse 2.0, which extends urban perception research to include six attributes — depressing, boring, beautiful, safe, lively, wealthy — and encompasses data from 56 cities. Due to their low-cost and high precision, these methodologies and datasets have seen extensive use in subsequent urban perception research (e.g., in China [11, 60], Singapore [61], the USA [62], Chile [26], and other global cities [63]) and serve as benchmarks for assessing the perceived quality of urban environments [64, 65, 66]. Moreover, several studies have identified specific image features associated with human perception in streetscapes. For example, Zhang et al. [11] highlight that objective elements, such as natural features, positively correlate with perceptions of beauty in streetscapes, whereas buildings often have a negative impact. This trend is also corroborated by Rossetti et al. [26], who found that images featuring buildings tend to be perceived as livelier but less beautiful and safe, and more boring and depressing. Additionally, clearer views of the sky negatively impact perceptions, while low-level features such as edges and blobs have a positive influence. Further expanding the scope, Zhao et al. [8] extend this framework to the measurement of soundscape perception, integrating perceptual attributes relevant to the soundscape domain, such as sound intensity, quality, and sources, along with other perceptual indicators. By incorporating street-level images, these frameworks hold significant potential to contribute to various urban research fields and deepen our understanding of how objective elements sensed by individuals correlate with their subjective perceptions.

To effectively model these complex and nuanced human perceptions, various advanced machine learning models have been employed. Among these, Convolutional Neural Networks (CNNs), have been predominant due to their proficiency in handling and analysing image data. These models are chosen for their ability to automatically extract and learn the most predictive features from images, which is crucial for accurately categorising and quantifying the attributes of urban environments. For instance, a common practice in urban perception research is employing pre-trained CNN models such as ResNet, fine-tuned on specific perception datasets like Place Pulse 2.0 [11, 61, 67]. As one of the primary visual components of the streetscape, buildings have attracted considerable attention in street-level research. CNNs, such as VGG, DenseNet, and ResNet, are also widely introduced to achieve, or serve as benchmarks for, the accurate classification of diverse functions [68], materials [69, 70], styles and ages [71, 72, 73] of buildings.

However, the ways in which individuals perceive and evaluate the exteriors of buildings from the street level, as well as the variation and distinction of such perceptions across different cities, and the overall impact of buildings on streetscape perceptions, remain underexplored. Therefore, aligning with previous studies for evaluating architectural exteriors and measuring street-level perception, this study introduces a synthetic and comparative framework. This framework is designed to assess the appearance of buildings by CNN models, uncover patterns of building perception in larger regions, and elucidate the influence of building design on streetscape perceptions.

3. Research framework

In this study, we introduce a comprehensive and comparative methodology for quantifying human perceptions of building exteriors utilising SVI. The research framework is structured into three main steps, as illustrated in Figure 1: (1) Section 3.1: Developing building perception models: We collect building exteriors by detecting and extracting images of individual buildings from SVI across cities. Then, a subset of images from this dataset is used to gather perceptual ratings through pairwise comparisons conducted by survey participants. These ratings span six dimensions: *complex, original, ordered, pleasing, boring, and style*. Subsequently, models to predict building perception scores are trained with these multidimensional labels. (2) Section 3.2: Analysing spatial characteristics: The trained building perception models are employed to estimate perception scores for the entire set of extracted SVIs. This process enables us to map these perceptual scores on broader scales and analyse the correlation between these scores and objective building attributes (i.e., functions, age, and location). (3) Section 3.3: Understanding streetscape perception:

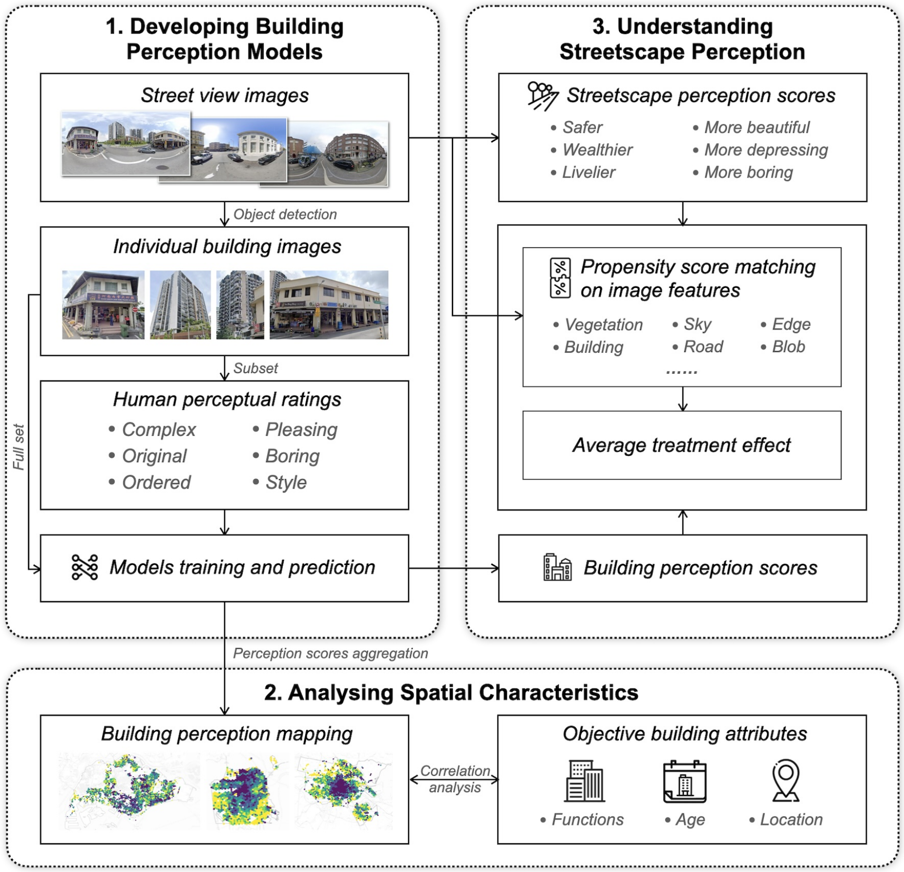


Figure 1: The framework to measure human perception of building exterior from street view images at the city-scale. Source of imagery: Google Street View.

the average perception scores of buildings and the streetscape are calculated separately for each panoramic images. Propensity score matching (PSM) is then applied to match and control SVIs with similar objective features, facilitating a detailed investigation into the effect of building exteriors on streetscape perceptions. Further details on these steps are provided below.

3.1. Developing building perception models

To gather human perceptions on individual buildings and assess them across broader regions, we develop a comprehensive workflow, as shown in Figure 2:

Extracting building images. Buildings, as primary visual elements of the streetscape, are often the focus of imagery datasets aimed at segmentation tasks. However, few datasets are tailored for the extraction of individual buildings from images. In this study, we endeavour to apply object detection methods to extract and isolate building images, thereby offering a more focused view of the buildings themselves. To achieve this, we employ the *GroundingDINO*, a state-of-the-art object detection model equipped with pre-trained weights capable of detecting arbitrary objects using human inputs such as category names or referring expressions [74]. Specifically, we utilise the "GroundingDINO-B" checkpoint, which has been trained on several widely-used object detection datasets, including COCO, O365, and Open Images, among others. By assigning the category name "building" to this open-set detector, we can acquire bounding boxes for buildings present in images. To ascertain the model's performance on this task, we evaluate it on the "building" category using the test set of the Open Images V7 dataset. The model demonstrates high accuracy and reliability in detecting buildings on the test set, achieving an Intersection over Union of 0.65, a precision of 0.86, and a recall of 0.75.

To collect diverse images of individual buildings, panoramic SVIs are retrieved from various urban locations. By applying the proposed building detection model to these SVIs, we obtain the coordinates of bounding boxes for buildings within the images. These bounding boxes are then processed with an algorithm specifically designed to correct lens distortion, resulting in normalised views of the buildings as seen from the street. Moreover, to assemble a targeted collection of building images for the survey, an equal number of images were collected from different regions of each city. This process adhered to two key principles: first, images were evenly sampled spatially to encompass a broad data distribution across urban areas; second, manual selection was conducted to ensure a wide array of building types and architectural styles were included, thereby creating a dataset that is both spatially and visually comprehensive for subsequent analysis.

Human perception labelling. To gather human subjective assessments of building exteriors, our study employs a perceptual survey designed based on prior urban perception studies [25, 75, 66, 8], structured around the six conceptual properties as discussed in Section 2.1, and further illustrated in Table 1. Specifically, the term "boring" is employed to denote an absence of excitement, thereby rendering the concept more accessible and comprehensible to participants. For *complex, original, ordered, pleasing*, and *boring*, participants are presented with pairs of building images during the survey. The central question guiding their evaluation is: "Given the pairs of building facades, which one do you think would be more [characteristic]?" The characteristic in question alternates among *complex, original, ordered, pleasing*, and *boring*. For each pairing, participants have the option to select either the left or right image or to denote that both images represent the quality to the same degree. "Style", considered an incomparable property, is designed to evaluate through a single-choice question with only one building image presented at a time. Participants are prompted to categorise the building's architectural style on a scale with five

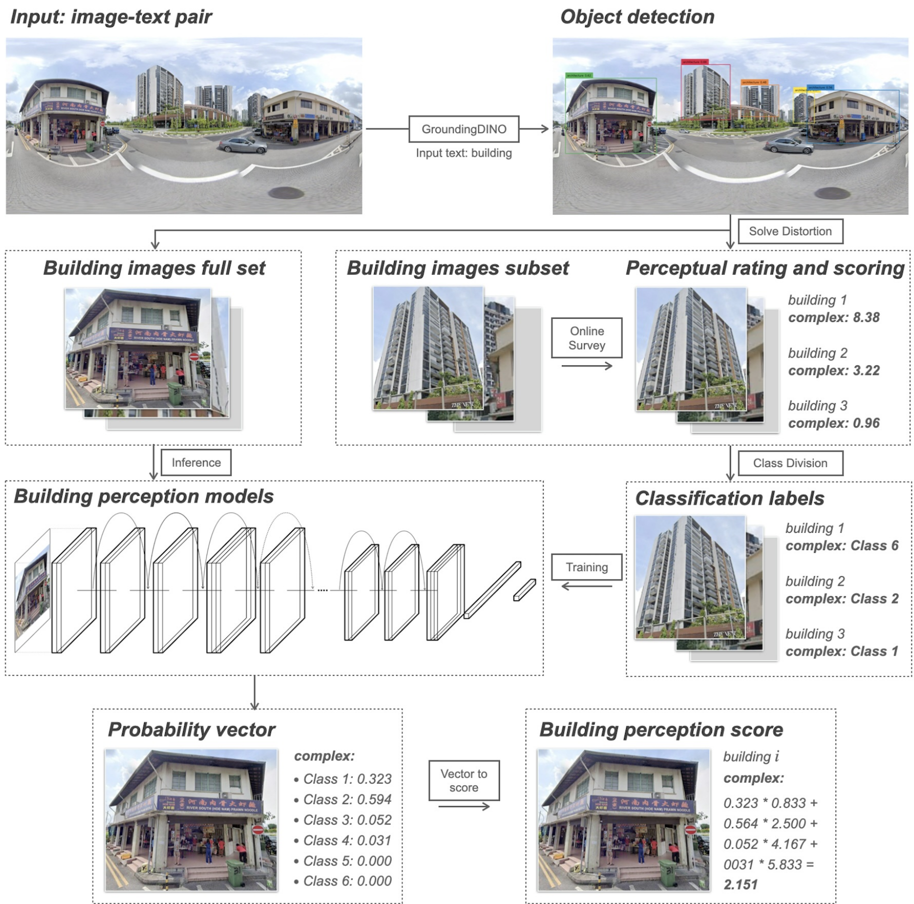


Figure 2: The detailed workflow for the development of building perception models. Source of imagery: Google Street View.

options: “modern”, “somewhat modern”, “no significant style”, “somewhat historical”, and “historical”. The selected images from different cities are randomly drawn and displayed to the participants, who are prompted to evaluate the building images in response to a series of questions that probe specific attributes. Figure 3 shows several examples from the survey platform.

The survey results are further applied for perception score calculation. In this study, Microsoft TrueSkill, a method applied in former visual perception studies [25, 76, 75], is used to ranked scores based on the pairwise comparison results. As a Bayesian ranking method, TrueSkill calculates a ranked score for each player (in our context, images of buildings) engaged in a two-entity comparison, by iteratively updating the ranked score of players after every comparison (in this case, the pairwise comparison results) [77]. For questions offering a single choice, we categorised architectural styles on a scale from 1 to 5, ranging from “modern” to “historical”, and computed the average scores of the architectural

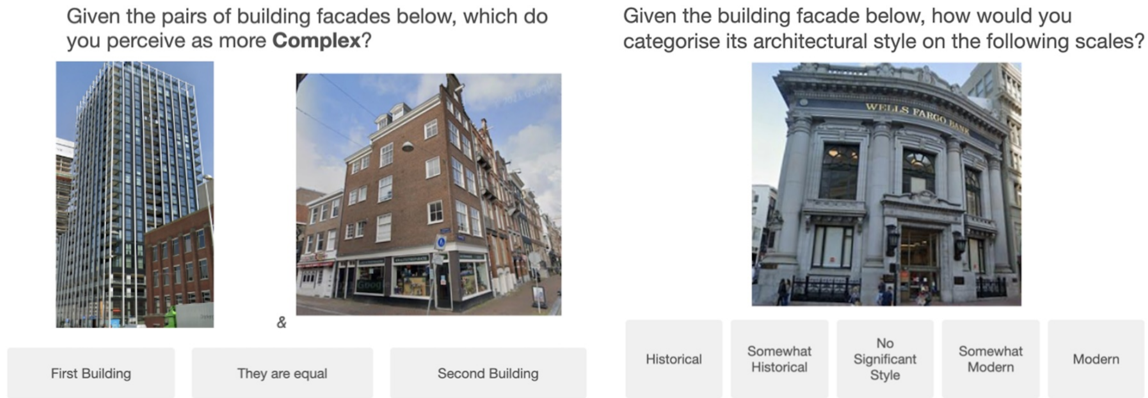


Figure 3: Examples of two types of questions on the survey platform used to collect human perceptions of building facades. Respondents are prompted to indicate their preferences through a series of questions targeting specific perceptual properties. Source of imagery: Google Street View.

style rankings received by each building. Thus, a building scoring higher in this style metric is perceived to be more historical, whereas a lower score suggests a modern perception. To enhance interpretability, these scores across six dimensions are normalised and adjusted to a continuous scale of 0 to 10. Through this methodology, we assign perceptual scores in six dimensions to each building image, establishing the basis of our training dataset for the study of building exterior perceptions.

Models training and prediction. Similar to Kang et al. [67], we then approach the prediction of perceptual attributes for building facades as supervised classification tasks, structured in three steps. Firstly, each perceptual score is assigned to one of six predefined ranges with equal score intervals: 0-1.67, 1.67-3.33, 3.33-5, 5-6.67, 6.67-8.33, and 8.33-10. This structuring, as shown in Figure 4, aims to enhance the machine's understanding of the perceptual characteristics of buildings across different ranges, and facilitate a more intuitive categorisation into low, medium, and high groups. This categorisation supports the computation of Top-2 accuracy metrics in subsequent model evaluations. For instance, class 1 includes building images with scores ranging from 0 to 1.67, while class 2 covers those with scores from 1.68 to 3.33; both are categorised as "low" within the Top-2 classes. Prior to model input, the data is divided into training and validation datasets with an 8:2 ratio across cities, ensuring a balanced representation of buildings from different cities in both sets of data.

To learn this compressed set of variables, we fine-tune CNN models, which were pre-trained on the ImageNet-1K dataset, using building images alongside their predefined classes of each perceptual property. Each model is individually tailored to predict one specific perceptual property in downstream prediction task. Lastly, the prediction task in-

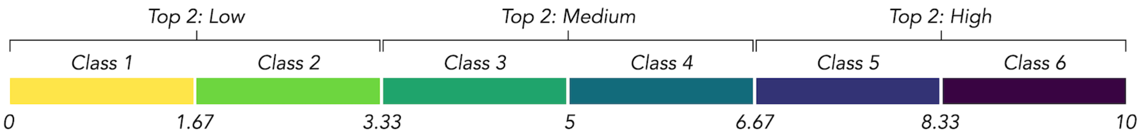


Figure 4: The scaling of predefined perceptual score ranges into six classes, and the subsequent categorisation into low, medium, and high groups.

Table 1: The conceptual properties and main survey questions applied in the study, along with their corresponding terminologies.

Conceptual Property	Corresponding Terminologies	
	Positive	**Negative**
Complex	complicated	simple
Original	unique, impressive, novel	uncreative, unoriginal
Ordered	ordered, coherent, unified	confusing, disorganised
Pleasing	enjoyable, pleasing, happy	unpleasant, depressing
Boring	exciting, arousal	dull, monotonous
Style	historical, somewhat historical, no significant style, somewhat modern, modern	

volves each model generating a six-dimensional vector for each building image, where each dimension represents the probability of the image belonging to one of the six predefined classes. These output probabilities P are used to compute scores for each building image, reflecting score of a specific perceptual property. These probabilities are multiplied by the median value of the ranges associated with their respective classes, transforming the probability distributions back into a numerical score S, for each image:

$$S_i = \sum_{n=1}^{6} \left(\frac{10}{6} \cdot \left(n - \frac{1}{2} \right) \cdot P_i^n \right) \tag{1}$$

where i denotes an individual building image, and P_i^n represents the probability of the image belonging to class n. The term $\frac{10}{6} \cdot \left(n - \frac{1}{2} \right)$ specifies the median value corresponding to class n, which scales the impact of each class's probability on the final score S_i.

3.2. Analysing spatial characteristics

For this phase of analysis, we integrate the H3 geospatial indexing system[1], specifically choosing a resolution at level 9, which corresponds to an average hexagon area of 0.105 km^2. The purpose of applying the spatial indexing system is twofold: Firstly, it allows for

[1]Hexagonal hierarchical geospatial indexing system: https://h3geo.org/

the aggregation and visualisation of perceptual information within discrete units, standardising the scale for comparative cross-city analysis. In practice, the perceptual attribute values for each cell are computed by averaging the scores of all buildings within that cell, thus reflecting the area's collective perceptual trend. Secondly, we compile additional building attributes within these same units to examine their correlation with the perceptual ratings generated by the model. This analysis of correlation offers further insights into the interplay between subjective perceptions of buildings and their objective characteristics, as well as variations in facade features across different cities.

3.3. Understanding streetscape perception

To uncover the influence of buildings within a streetscape, we apply the Place Pulse 2.0 dataset [25] as the benchmark for streetscape perception analysis. This dataset contains 110,988 images from 56 cities across 28 countries and quantifies human perception of the streetscape across six dimensions (depressing, boring, beautiful, safe, lively, and wealthy) based on volunteer labelling. Following a similar approach to our building perception model training, we first train deep learning models using Place Pulse 2.0 dataset across these six-dimensional features. Second, we segment our panoramic SVIs into perspective views to match the format used in the dataset. These perspective images are then assessed using the streetscape perception models, and the results are averaged to produce comprehensive streetscape perception scores for each geographical location.

Furthermore, as identified by previous studies, certain visual elements and low-level features within the images, such as buildings, vegetation, edges, and blobs, potentially influence people's perceptions [11, 26, 27]. As a result, perceptions may vary due to multiple factors present in images. For instance, a place with a higher visual factor of vegetation tends to be perceived as more beautiful, even if it contains unpleasing building exteriors compare to other images that have pleasing building designs. To minimise the impact and facilitate a robust discussion on the influence of building appearance, we introduce the propensity score matching (PSM) to control and match the SVIs. This method involves dividing SVIs based on high and low building perceptual properties, then matching them into strata according to their characteristics (e.g., having a similar amount of visual elements), allowing us to calculate the average treatment effect of that perceptual property on street perception.

To accomplish this, we first employ semantic segmentation using a deep learning model trained on the Cityscapes dataset [78] to extract physical features from each perspective SVI. This process quantifies the proportions of various visual elements, such as buildings, vegetation, roads, and sky in images. Second, we retrieve pixel-level features, including the number of edges and blobs as well as the mean values of hue, saturation, and lightness, using algorithms from the OpenCV library. For each SVI shooting point, we aggregate the average streetscape perception scores and visual features from perspective views, together with its average six building perception scores derived from the individual buildings within the image. Thirdly, the visual features are used in binary logistic models to calculate propensity scores for the matching process. Each respondent in the treatment group (e.g., having a high complexity level of building exteriors) is matched to one in the control group (e.g., having a low complexity level of building exteriors) according to the propensity scores. Finally, we evaluate the effects of building perceptual properties on overall streetscape perception by calculating the average treatment effect on the treated group (ATT) [79]:

$$ATT = E(Y_1 - Y_0 | D = 1) \tag{2}$$

where $D = 1$ indicates the treatment group, characterised by buildings received high perceptual values. The variable Y_1 refers to the streetscape perception score of the treatment group, and Y_0 refers to the streetscape perception score of the control group.

4. Research areas and data

4.1. Research areas

This study aims to explore and assess the reliability and scalability of a building perception dataset by examining buildings in three distinct cities: Singapore, San Francisco, and Amsterdam. These cities were chosen due to their diverse architectural landscapes, which provide a comprehensive base for analysis. Singapore is a highly urbanised metropolis that has rapidly developed over recent decades, containing buildings of various types and styles. A significant aspect of the city's urban character is its abundance of high-rise buildings, which are not merely typical but also functional, housing the majority of the resident population [80]. In contrast, San Francisco presents a mix of architecture that combines historical landmarks with contemporary structures, characterised by a wide range of building styles from Victorian to modern structures. Amsterdam differs from the aforementioned cities with its well-preserved historic mid- and low-rise buildings, some of which are complemented by modern additions, creating a diverse and contrasting architectural experience. By the examination on these cities, the study aims to evaluate the dataset's adaptability and to gather insights into how building appearances and styles are perceived across different urban environments.

4.2. Data description

4.2.1. Street-level building images

This study evaluated perceptions of buildings using individual building images derived from panoramic SVIs sourced from Google Street View (GSV)[2]. To standardise the gathering of SVIs across the three cities, we employed the H3 geospatial indexing system at a resolution of level 11, corresponding to hexagons of 0.002 km^2 each. For each hexagon, the central coordinate was used to search for and retrieve the nearest panoramic SVI within the unit. The collection process yielded 30,520, 24,292, and 25,062 images from Singapore, San Francisco, and Amsterdam, respectively. As illustrated in Figure 2, these images were then utilised for an object detection task to identify buildings, followed by addressing the image distortion. To ensure accurate predictions, images smaller than 100x100 dpi were excluded from the analysis. Finally, 94,632, 63,884, and 92,501 building images were generated for Singapore, San Francisco and Amsterdam, respectively. For this study, 400 images from each city were selected according to the principles outlined in Section 3.1, for inclusion in the dataset to gather human perceptual ratings. As a result, a total of 1,200 images were used for the online participant survey, and there were total 249,817 images remaining for the prediction task.

In this work, we conducted an online survey among students at the National University of Singapore. To ensure a representative sample for the building perception survey, a pre-survey was initially conducted to gather general information from potential participants. A balanced group was then selected based on gender, age, and academic programs for the building perception survey. Ultimately, the survey garnered responses from 493 participants, collecting a total of 33,774 responses related to 1,200 building images. Each building image was rated more than five times to minimise individual bias, ensuring a more balanced evaluation. While the survey participants were primarily students, we acknowledge that this group may not fully represent the broader population. The primary aim of this study was to demonstrate the feasibility of our method rather than to provide definitive perceptual scores. For future studies or applications requiring precise, demographically tailored insights, this methodology can be replicated with a broader and more diverse participant pool to enhance the representativeness and reliability of the results.

Utilising the scoring method mentioned in Section 3.1, the scores for six different perceptual attributes were calculated for each building image based on the results of the survey. The samples of building images with their corresponding perceptual scores, as shown in Figure 5, indicate that different building exteriors may induce different human

[2]https://www.google.com/maps/

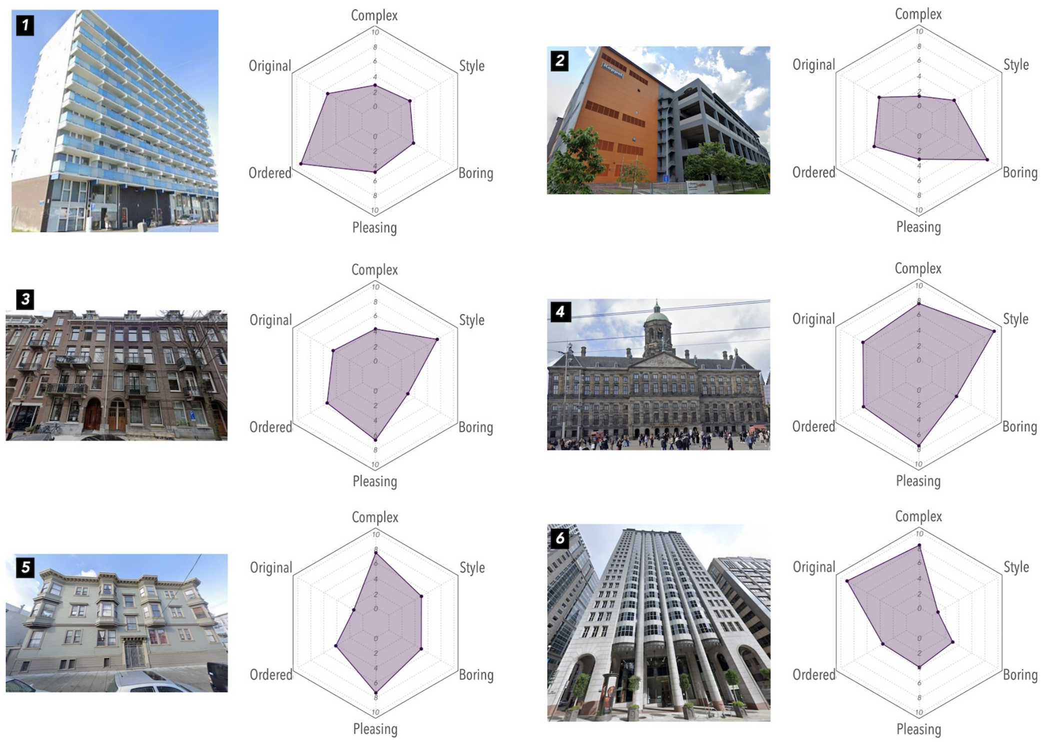


Figure 5: Building image samples of different cities with their perceptual score of the 6 dimensions calculated based on survey responses. Source of imagery: Google Street View.

perceptions. For instance, the first building image, featuring blue balconies and prominent horizontal lines, tends to be perceived as a modern structure and ranks higher in "more ordered" and relatively lower in terms of being perceived as "more boring". The fourth image, which depicts a historical building with rich visual elements, has relatively higher scores for complexity, originality, and pleasantness compared to the others. Moreover, Image 6, with its unique shape and abundant architectural features, is perceived as the most original and complex among the others. Furthermore, such complexity induces a lesser sense of boredom but does not necessarily contribute to a feeling of pleasantness. After categorising the scores into six classes as described in Section 3.1, the extreme classes exhibit fewer labels compared to others. To mitigate this imbalance, we implemented data augmentation techniques, including horizontal flipping and centre cropping, to enhance the representation of underrepresented classes in our training set.

[3] `https://data.gov.sg`

[4] `https://data.sfgov.org`

[5] `https://maps.amsterdam.nl`

Table 2: Objective building indices, derived from various data sources, are calculated for the H3 cells across three cities. The mean and standard deviation for each index value are indicated.

	Variables	Description	Mean	Std.	Data source
City: Singapore					
Function	Residential density	Residential building area (*ha*)	2.369	2.421	Master Plan 2019[3]
	Commercial density	Commercial building area (*ha*)	0.443	1.439	
	Industrial density	Industrial building area (*ha*)	0.215	0.777	
	Public housing density	Public housing area (*ha*)	1.389	0.946	HDB Existing Building[3]
Location	Distance to city centre	Straight-line distance (*km*)	11.081	4.680	GIS analysis
City: San Francisco					
Function	Residential density	Residential building area (*ha*)	2.144	1.420	Land Use 2020[4]
	Commercial density	Commercial building area (*ha*)	0.215	0.427	
	Office density	Office building area (*ha*)	0.256	0.606	
Age	Building age	The average of building age since completion	68.826	35.357	
Location	Distance to city centre	Straight-line distance (*km*)	4.474	2.127	GIS analysis
City: Amsterdam					
Function	Residential density	Residential building area (*ha*)	1.041	0.957	Amsterdam Government Maps Data[5]
	Commercial density	Commercial building area (*ha*)	0.266	0.519	
	Office density	Office building area (*ha*)	0.328	0.631	
Age	Building age	The average of building age since completion	56.151	37.544	
Location	Distance to city centre	Straight-line distance (*km*)	4.773	2.379	GIS analysis

4.2.2. *Objective building attributes*

Aiming to provide a comprehensive profile of building perceptions and their spatial patterns, objective building attributes are collected from three aspects: functions, ages, and locations. These attributes are believed to have potential relationships with urban visual appearances [65, 11, 81]. Due to the variation in format and openness of the data released by different cities, we utilised diverse datasets to represent these three aspects, as shown in Table 2. On the one hand, the building functions and distance to city centre are common attributes for the three cities. On the other hand, in Singapore, public housing developed by the Housing and Development Board (HDB) represents a typical building type with available data for analysis. Conversely, in San Francisco, the age of buildings is the specific building-level data that warrants further exploration in this study. The attributes of each research unit are aggregated as mean values to serve as contextual representations of the area.

5. Results and analysis

5.1. *Building perception models*

5.1.1. *Models evaluation*

To identify the suitable model for this study, we fine-tune commonly used CNN architectures in urban research, as discussed in Section 2.2, using our building perception dataset. We compared the Top-1 and Top-2 accuracy metrics of these models to select the best one for predicting perception scores across the entire image set of individual buildings. Top-1 accuracy measures the percentage of the validation set which the model accurately predicts the exact class, while Top-2 accuracy assesses whether the model can correctly predict the category closest to the original one. As shown in Figure 4, the Top-2 categories cover perception scores ranging from 0 to 3.33, 3.33 to 6.67, and 6.67 to 10, reflecting a low, medium and high ranking trend.

Table 3 presents the average values of performance metrics for various CNN models across the six attributes, with ResNet50 generally outperforming the others. Consequently, the ResNet50 architecture has been adopted for building perception models in this study. Table 4 provides further details on the performance of ResNet50 models across perceptual properties, including accuracy, recall, precision, and F1-score for both Top-1 and Top-2 classes. The models demonstrate strong performance in classifying building images based on human perception trends, with Top-2 accuracies exceeding 72%, and attributes such as “complex”, “boring”, and “style” achieving higher predictive accuracies. This suggests that the complexity and excitement of building exteriors, as well as their architectural styles, tend to display more consistent visual features across cities, facilitating more accurate

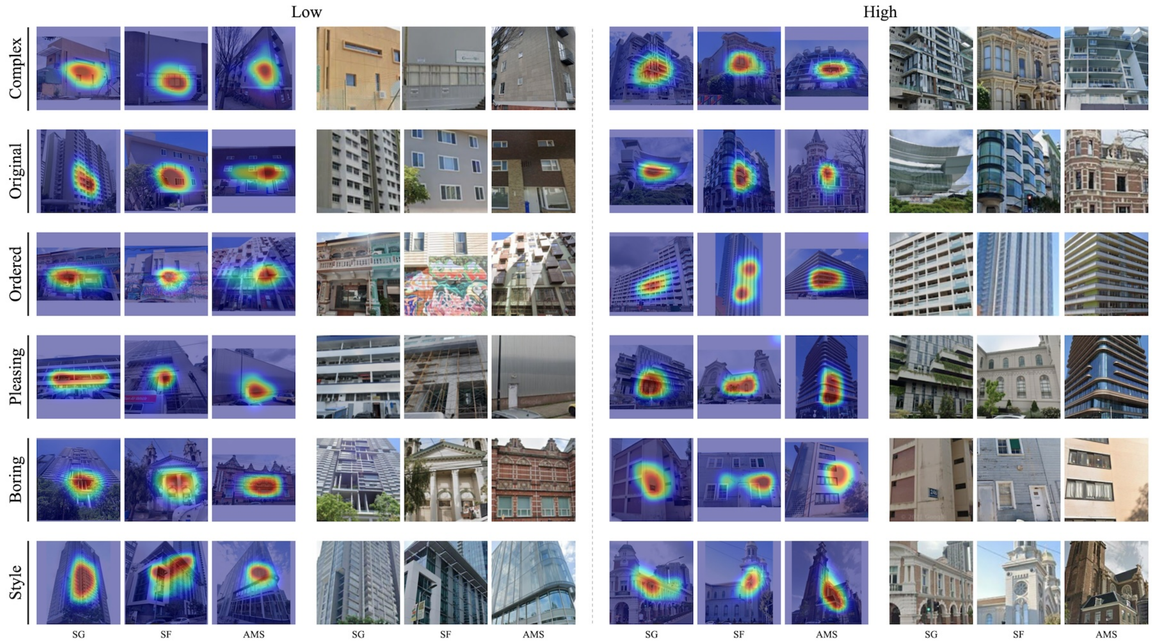


Figure 6: The results of Class Activation Maps (CAM) overlaid on original images and the building images cropped according to CAM showing the specific discriminative regions. Source of imagery: Google Street View.

predictions by the models. Nonetheless, the relatively modest performance on the Top-1 benchmark should be acknowledged. This discrepancy may be attributed to the inherently subjective nature of human perception, influenced by cultural, historical, and personal factors that are not easily captured by visual features alone.

Building on architecture research by Lee et al. [82] and Sun et al. [72], Class Activation Maps (CAM) are further employed to identify discriminative regions that CNN models use for predictions. This explainable AI technique allows us to understand which parts of an image the model focuses on, influencing the final classification across the six perceptual properties. Figure 6 illustrates the CAM results and the discriminative regions extracted from the original images. The CAM results demonstrate clear yet varied patterns on the semantic information of the buildings based on different perceptual dimensions, indicating models' capacity to discern and learn pertinent architectural features from the imagery. For example, regions containing rich visual information may prompt predictions of high complexity, whereas distracting details might result in disordered rankings. Furthermore, chaotic and uninviting elements in images often lead to low "pleasing" classifications, while blank and unadorned designs are typically defined as "boring" buildings. Modern buildings are frequently categorised based on their sleek structures and cold colours, whereas warmer tones and ornamental architectural features such as columns, arches, and belfries are generally associated with "historical" identifications. These patterns not only illustrate the models' ability to classify buildings based on various perceptual attributes but also their adaptability across diverse building types in the cities. The models also prioritise building features over unrelated elements like vegetation, cars, and pedestrians, enhancing their focus on relevant architectural details. Overall, these evaluations affirm the robustness and generalisability of our approach in large-scale urban environments, demonstrating the models' capability to align building designs with human perceptual judgements.

Table 3: Average validation performance of CNN models across all perceptual properties.

Model	**Benchmark**	**Accuracy**	**Precision**	**Recall**	**F1-score**
VGG16	Top-1	0.424	0.356	0.326	0.320
	Top-2	0.693	0.725	0.676	0.684
DenseNet201	Top-1	0.452	0.356	0.347	0.339
	Top-2	0.709	0.725	0.662	0.679
ResNet34	Top-1	0.440	0.356	0.329	0.327
	Top-2	0.731	0.747	0.684	0.697
ResNet50	Top-1	0.480	0.386	0.353	0.358
	Top-2	**0.770**	**0.809**	**0.699**	**0.732**
ResNet101	Top-1	0.471	0.426	0.364	0.370
	Top-2	0.732	0.758	0.675	0.697

5.1.2. Prediction results

All building images from the three cities are processed by the trained models, with the output probabilities being converted into scores across six perceptual dimensions, as defined in Equation 1. Figure 7 displays a selection of buildings from Singapore, San Francisco, and Amsterdam, categorised by their "original" scores. Generally, the model assigns higher originality scores to buildings with innovative and distinctive architectural features that carry historical or contemporary significance. Buildings have the lowest originality scores tend to lack distinctive visual elements and are characterised by flat mundane appearances, often representing residential or industrial structures. In Singapore, the most original buildings are typically contemporary in design, featuring unique shapes and modern elements. In San Francisco, buildings that are considered original often possess historical significance and are adorned with rich ornamentation. Amsterdam displays a harmonious blend of historical and modern buildings among those rated as most original, reflecting a region that integrates tradition with modernity. Furthermore, Figure 8 show-cases a collection of buildings among cities, categorised by a predictive model based on

Table 4: Detailed validation metrics for ResNet50: accuracy, precision, recall, and F1-score in building perception classification tasks.

Properties	**Benchmark**	**Accuracy**	**Precision**	**Recall**	**F1-score**
Complex	Top-1	0.512	0.356	0.355	0.350
	Top-2	0.777	0.782	0.666	0.704
Original	Top-1	0.475	0.398	0.332	0.347
	Top-2	0.752	0.842	0.627	0.683
Ordered	Top-1	0.450	0.367	0.335	0.344
	Top-2	0.719	0.747	0.653	0.681
Pleasing	Top-1	0.438	0.337	0.304	0.311
	Top-2	0.764	0.828	0.669	0.721
Boring	Top-1	0.508	0.338	0.312	0.313
	Top-2	0.777	0.810	0.744	0.761
Style	Top-1	0.498	0.520	0.479	0.484
	Top-2	0.828	0.846	0.833	0.839

their “ordered” score. In this case, buildings with lower-order scores often exhibit unique or unconventional features, while those with higher scores display more standardised, symmetrical, and orderly designs. Contrary to the high original buildings, it is common across the three cities that the highly ordered buildings tend to be modern structures with a pragmatic design. Despite the variety in building designs across different urban regions, these results highlight the models’ adeptness at recognising and distinguishing architectural uniqueness in varied urban settings.

Figure 9 compares the quartiles of predicted scores for buildings in the three cities. Similar to the insights provided earlier, buildings in Amsterdam indicate a wide range of architectural style indices, reflecting the city’s rich historical legacy and its harmonious integration of modern designs. Singapore’s architecture also stands out in the “style” dimension, with more buildings receiving lower scores compared to the other cities. This, incorporated the prevalence of ordered and complex buildings, suggests a sense of regularity and urbanisation in its cityscape. In contrast, buildings in San Francisco are generally perceived as less ordered. Moreover, while all cities display similar levels of “boring” facades, San Francisco’s buildings show a wider range in this dimension, which may reflect the city’s diverse and distinctive architectural types and styles. This analyses provide global insights into the unique architectural identities of each city and how they are perceived by people.

Figure 7: Image samples from Singapore, San Francisco and Amsterdam that were predicted with low original scores (left) and high original scores (right). Source of imagery: Google Street View.

Figure 8: Image samples from Singapore, San Francisco and Amsterdam that were predicted with low ordered scores (left) and high ordered scores (right). Source of imagery: Google Street View.

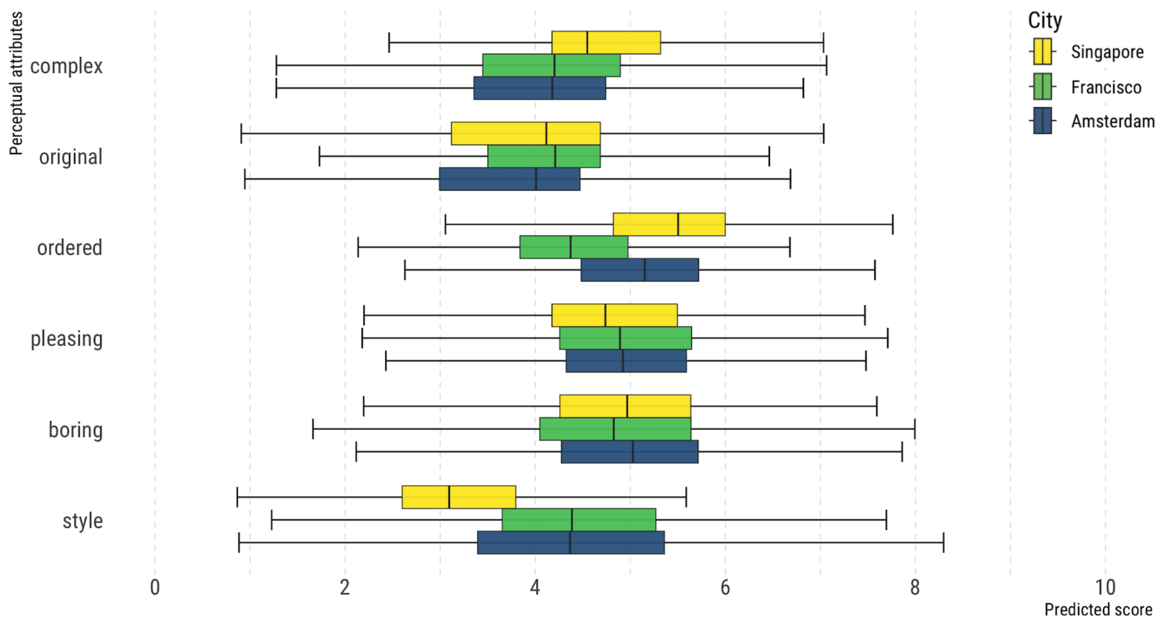


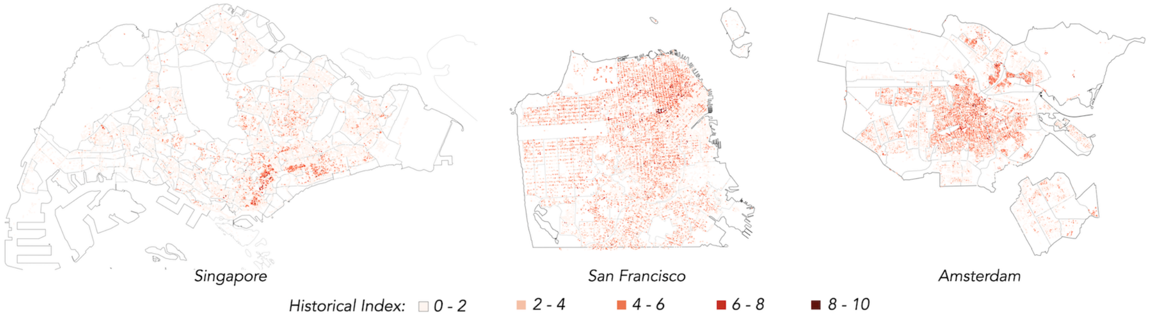
Figure 9: The comparative distribution of predicted scores for six perceptual dimensions of buildings in Singapore, San Francisco, and Amsterdam.

To enhance the understanding of building style, the “style” category is divided into two continuous values: “historical” and “modern” for subsequent analysis. This division aims to clarify the extent to which buildings are recognised within these styles based on equations:

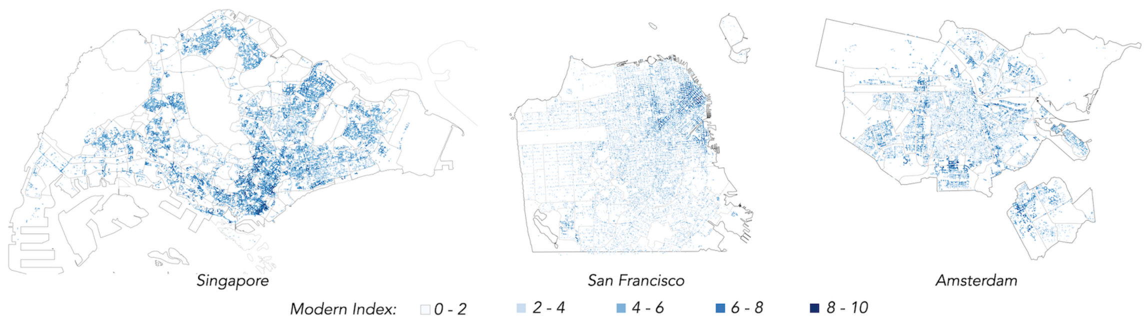
$$H = \begin{cases} S - 5 & \text{if } S > 5 \\ 0 & \text{if } S \leq 5 \end{cases} \tag{3}$$

$$M = \begin{cases} 5 - S & \text{if } S < 5 \\ 0 & \text{if } S \geq 5 \end{cases} \tag{4}$$

where H and M represent the transformed score representing the building’s historical and modern significance, respectively, and S is the original “style” score assigned to the buildings. If “style” scores above 5, indicating a certain historical significance, the score of H will be set to $S - 5$, and M will be set to 0. For scores lower that 5, indicating modern significance, H will be set to 0, while M will be calculated as $5 - S$. These attributes are then normalised and adjusted to a scale from 0 to 10. Following the score adjustment, Figure 10a and Figure 10b illustrate the spatial distributions of the trends in perceiving historical and modern building exteriors, respectively.



(a) Historical index of each view-point



(b) Modern index of each view-point

Figure 10: The geographical distribution of locations identified to have different perceptual levels of historical and modern buildings in Singapore, San Francisco, and Amsterdam.

5.1.3. Correlation analyses

As indicated by Zhang et al. [11], perceptual properties have potential relationship with each others, which may vary among cities. Thus, to better elucidate the relationship of building perceptual properties, a Pearson cross-correlation analysis of the perceptual indicators is conducted. Due to the insufficient number of building images identified as having a “historical” sense, this investigation focuses on the relationship between complex, original, ordered, pleasing, boring, and modern.

As illustrated in Figure 11, the relationship between perceptual scores shares similar patterns among the three cities. The attribute “boring” has a significant negative relationship with the attributes of complexity, pleasantness, and originality. Among these, higher levels of complexity exert the strong positive associations with inciting observer excitement, aligning with with previous studies [49, 30, 46]. Likewise, buildings that are original and complex in design also have a positive relationship with perceptions of pleasantness. There is also a strong positive association between “original” and “complex”, suggesting that architecturally impressive designs are often seen as visually rich from a human perspective. Specifically, San Francisco stands out with the strongest relationship

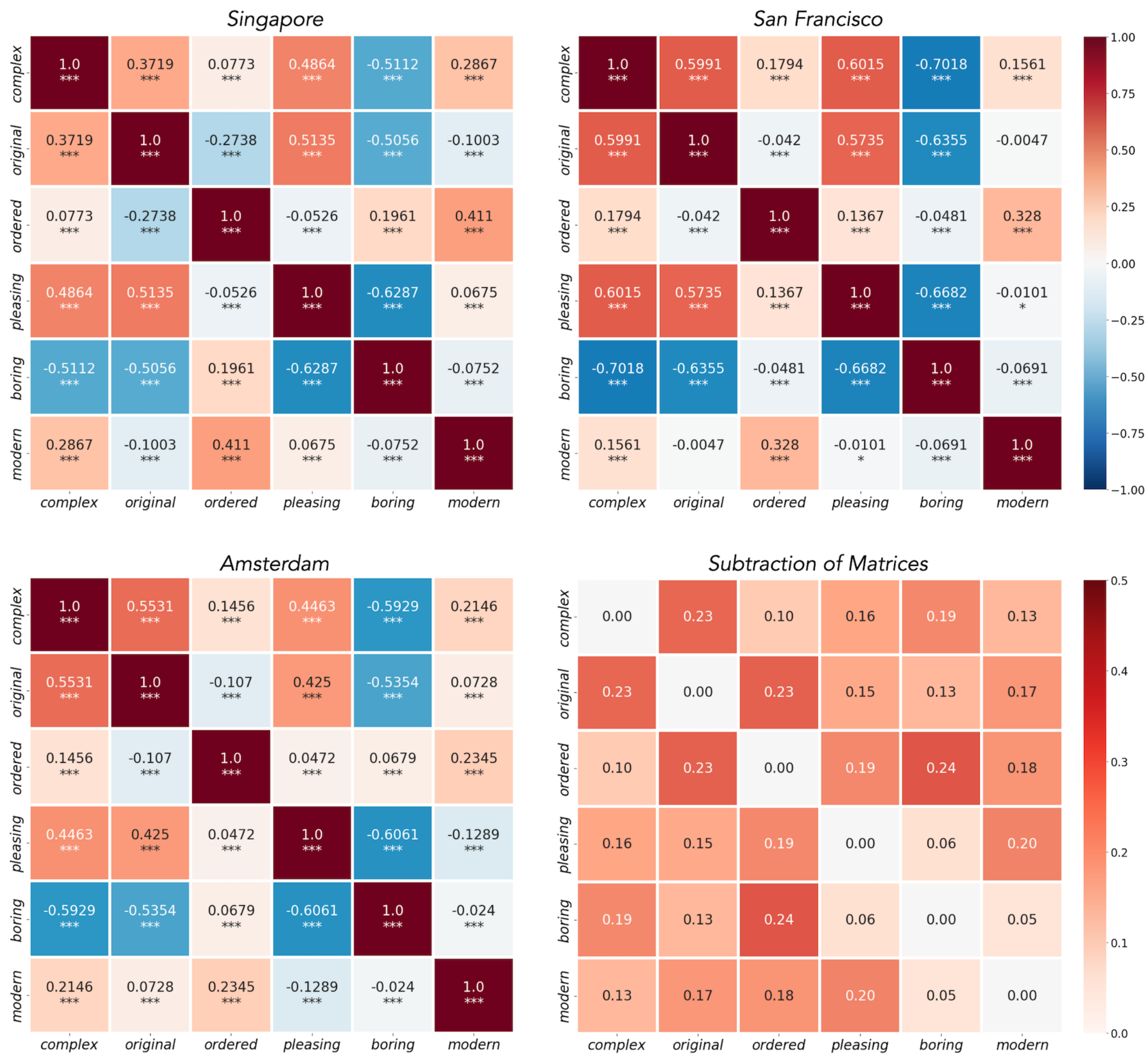


Figure 11: Cross-correlation matrices among the six perceptual properties using data from Singapore, San Francisco and Amsterdam, with subsequent subtraction to highlight the significant differences among the cities.

between these two attributes, suggesting that the unique architectural design in the city resonates strongly with observers' sense of complexity. In Singapore, "complex" and "ordered" are significantly positively associated with modernity, reflecting that buildings with more modern characteristics are often visually complex and organised. The subtraction of the matrices further highlights the variances in the "boring—ordered", "original—ordered" and "original—complex" relationships among the three cities. These variations are mainly attributable to the perceptions of Singapore's building exteriors. Unlike in San Francisco and Amsterdam, the orderly nature of Singapore's architecture negatively correlates with perceptions of originality and excitement. This trend may be associated with the design of Singapore's residential housing, which, while aligned and balanced, is also highly uniform and predominantly high-rise. This induces a perception of high visual complexity but lower originality, appearing more monotonous to observers. A more comprehensive study would be beneficial to delve deeper into the specific architectural characteristics in these cities and to validate these hypotheses.

5.2. Spatial analysis of building perceptions

To reveal building perception patterns, the scores from six perceptual dimensions for each building image are aggregated to represent the overall perception of building exteriors (within the H3 tessellation units). Figure 12 provides a comparative spatial analysis of building perception across the cities. Generally, areas with highly complex and original facade designs are clustered in city centres, where development is dense and the architecture often features rich innovation and visual appeal. However, the distribution of order, pleasantness, and perceived monotony of building appearances varies across the three cities, reflecting their distinct planning strategies, cultures, and historical contexts.

In Singapore and Amsterdam, the spread of high complexity and originality extends beyond their central areas, indicating a diffusion of innovative architecture into other districts. In contrast, San Francisco's complex and orderly facades are primarily concentrated in and near the city centre, suggesting a modularly developed urban core. In Amsterdam, buildings perceived as well-ordered are distributed in a pattern that inversely correlates with complexity and originality, highlighting a unique architectural stratification within the city. These outcomes are likely due to the city's historically small building blocks, which contribute to a visually diverse and rich environment, potentially clashing with perceptions of symmetry and uniformity. In San Francisco, extensive areas characterised by varied types of small residential blocks are rated low in terms of "complex" and "ordered", while areas known for their high-quality Victorian and Edwardian homes are associated with the highest levels of originality and pleasure. In Singapore, buildings generally score higher in terms of orderliness compared to the other cities, yet these areas are perceived as less pleasant. This perception may be attributed to Singapore's extensive urbanisation and the prevalence of high-rise residential buildings, which can evoke feelings of congestion and stress, thereby impacting the perception of pleasantness. Amsterdam displays the opposite trend, with its central areas offering a variety of low-density and historical districts, which contribute to a higher perception of pleasantness.

To deepen our understanding of the interaction between human perceptions and the built environment, we undertook a cross-correlation analysis, comparing perceptual scores against the objective characteristics of buildings as discussed in Section 4.2.2. Table 5 shows the relationship between various urban land uses — residential, commercial, and office or industrial — and their correlation with perceptual qualities, alongside other attributes such as the prevalence of HDB flats in Singapore, the age of buildings in San Francisco and Amsterdam, as well as the locations of areas (distance to city centre).

In general, the three cities indicate unique fabrics and architectural features in the relationship between perceptions and building functions. In San Francisco and Amsterdam, residential areas with high coverage are significantly positively correlated with pleasing architectural designs and strongly negatively correlated with perceptions of being boring and modern. This suggests that these areas resonate with appealing design and visual excitement while offering a non-modern aesthetic. Conversely, in Singapore, area covered by HDB buildings correlates positively with orderliness but conveys a sense of monotony. This observation, as discussed in Section 5.1.3, reflects a standardised architectural approach in public housing that priorities functionality and uniformity. While this approach contributes to coherent urban order, it may lack the variability needed to stimulate excitement and interest. In San Francisco, districts with commercial and office buildings are associated with complexity and organisation, but in all three cities, these areas do not show a strong relationship with building perceptions. Future studies could incorporate additional factors such as building height, colours, and materials to deepen the analysis.

Additionally, in both San Francisco and Amsterdam, areas with older buildings exhibit a strong negative correlation with perceptions of being "boring" and a positive correlation with originality and pleasantness. In San Francisco, older districts feature significantly more complex and original designs compared to those in Amsterdam, highlighting distinct variations in design across different building age groups in the former. Regarding distance to the city centre, both San Francisco and Amsterdam display similar trends: complexity, originality, and pleasantness decrease towards the periphery, while perceptions of buildings as boring increase. Interestingly, the two cities show opposing trends in the correlation between "modern" and distance to the city centre. As illustrated in Figure 12, modern buildings in Amsterdam are typically located outside the central area, while in San Francisco, they are concentrated in the city centre. Singapore exhibits a weak correlation in

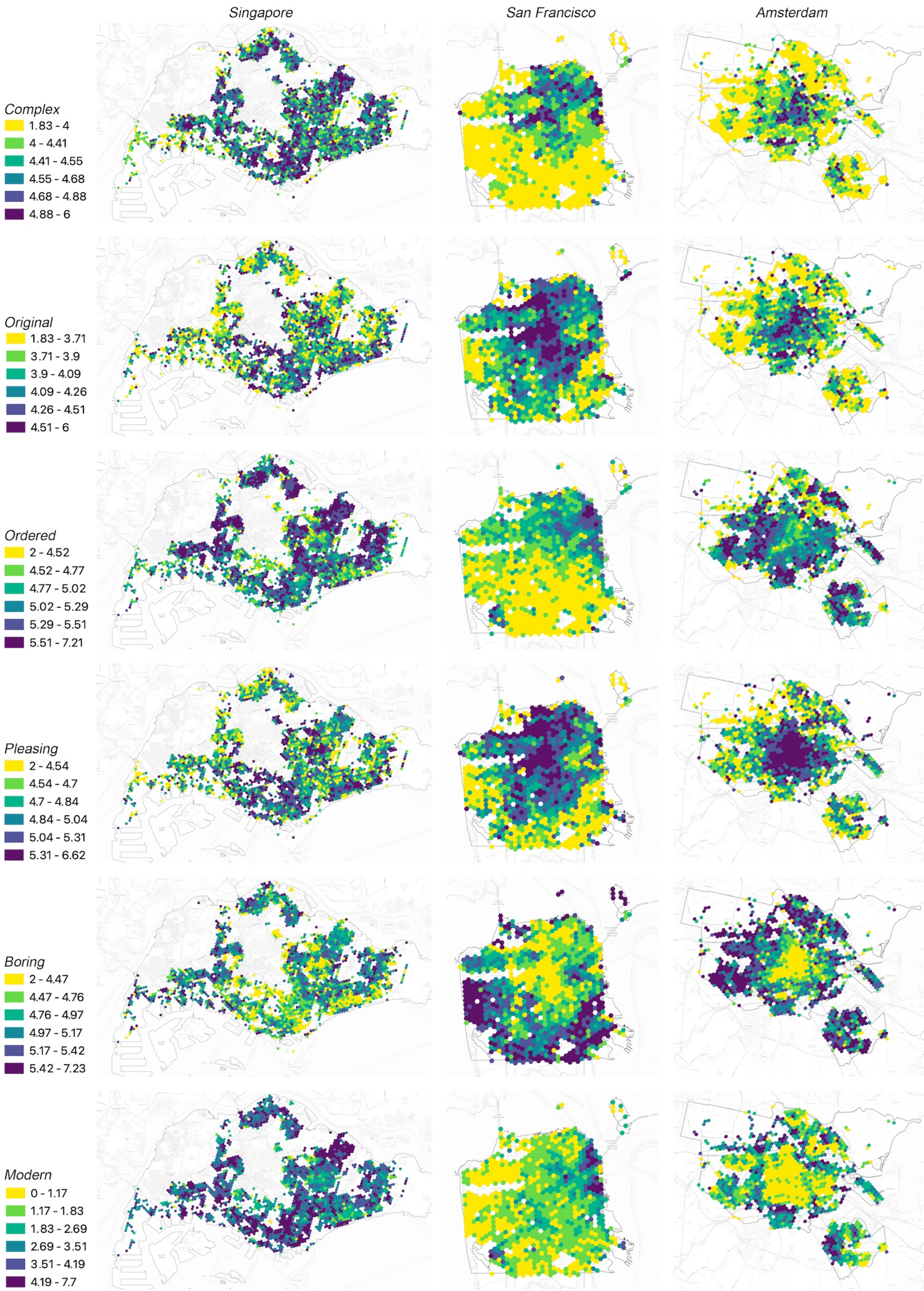


Figure 12: The spatial distribution of six perceptual properties of Singapore, San Francisco and Amsterdam. Basemap: (c) OpenStreetMap contributors.

Table 5: Spearman correlations between perceptual properties and objective building attributes. The combinations with moderate or strong correlations are highlighted.

	Complex	Original	Ordered	Pleasing	Boring	Modern
City: Singapore						
Residential density	0.1232 ***	0.020	-0.1023 **	0.0911 **	-0.0308	-0.0775 *
Commercial density	-0.0763 *	0.1088 ***	-0.257 ***	-0.0407	-0.0239	-0.1613 ***
Industrial density	-0.0618	0.1081 ***	-0.1746 ***	0.0382	-0.0579	0.0198
Public housing density	0.0605	-0.1678 ***	0.3785 ***	-0.1467 ***	0.2594 ***	0.2626 ***
Distance to city centre	0.1131 ***	-0.1682 ***	0.312 ***	-0.0969 **	0.1879 ***	0.1102 ***
City: San Francisco						
Residential density	0.2138 ***	0.366 ***	-0.1082 **	0.3159 ***	-0.3045 ***	-0.3722 ***
Commercial density	0.255 ***	0.0542	0.2388 ***	0.0299	-0.1361 ***	0.0008
Office density	0.2633 ***	0.0227	0.3103 ***	-0.0046	-0.1216 ***	0.2366 ***
Building age	0.452 ***	0.5436 ***	0.0433	0.4833 ***	-0.519 ***	-0.4201 ***
Distance to city centre	-0.6848 ***	-0.515 ***	-0.4977 ***	-0.5296 ***	0.609 ***	-0.2963 ***
City: Amsterdam						
Residential density	0.1597 ***	0.2335 ***	0.0958 **	0.5034 ***	-0.3609 ***	-0.542 ***
Commercial density	0.2538 ***	0.2251 ***	0.1165 ***	0.1109 ***	-0.1791 ***	-0.0497
Office density	0.2197 ***	0.1241 ***	-0.0248	-0.1284 ***	-0.031	0.3034 ***
Building age	0.1861 ***	0.2784 ***	-0.2093 ***	0.5417 ***	-0.4467 ***	-0.6356 ***
Distance to city centre	-0.4438 ***	-0.4811 ***	0.0288	-0.5102 ***	0.5579 ***	0.3325 ***

*Note: All correlations presented are Spearman correlation coefficients. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

these relationships, indicating a varied distribution of perceptual types of buildings across the city. These findings underscore the unique dynamics between urban form and perceived attributes, showcasing distinct trends in how architectural and urban features are perceived across different cities.

5.3. Building exteriors in streetscape perceptions

In this section, we discuss the influence of buildings' perceptual properties on holistic streetscape perceptions. Using SVIs with comparable features, as described in Section 3.3, we categorise them into control groups (with lower perceptual property values) and treatment groups (with higher perceptual property values) by employing PSM. Following this, we evaluate the standardised difference (denoted as δ) for each covariate to ensure there are no significant statistical disparities between the treatment and control groups after the matching procedure. As detailed in Appendix A, the δ values post-matching are below

Figure 13: Examples of matched pairs from the control group (low building complexity) and the treatment group (high building complexity), based on propensity score matching conducted using the proposed image features. Source of imagery: Google Street View.

10%, indicating an acceptable level of balance consistent with previous studies [83, 84]. This effective pairing, exemplified in Figure 13, sets the stage for further analysis of how perceptions of buildings affect streetscape perceptions.

Figure 14 reports the average treatment effect of treated (ATT) of each building perceptual properties. Here, the scale of streetscape perception score is set to 0 to 10, and the coefficients of ATT mean that the average values of streetscape perception score in the treatment group are higher than that of the matched respondents in control group. This indicates that to what extent the level of building perceptions positively (yellow bar) or negatively (blue bar) contributed to each indicator of streetscape perception.

Generally, certain perceptions of buildings–complex, original, pleasing, and historical—are significantly positively correlated with higher scores of positive streetscape perception, including safer, wealthier, livelier, and more beautiful environments. Among these attributes, pleasing architectural design is noted for its particularly strong influence on streetscape perception, underscoring the significance of aesthetically appealing appearances in urban settings [9, 10, 5]. Furthermore, buildings that convey a sense of higher complexity and historical ambiance are also identified as making substantial positive contributions to streetscape perceptions. While ordered buildings may impart a monotonous aspect to streetscapes, they contribute significantly to perceptions of safety and wealth. This

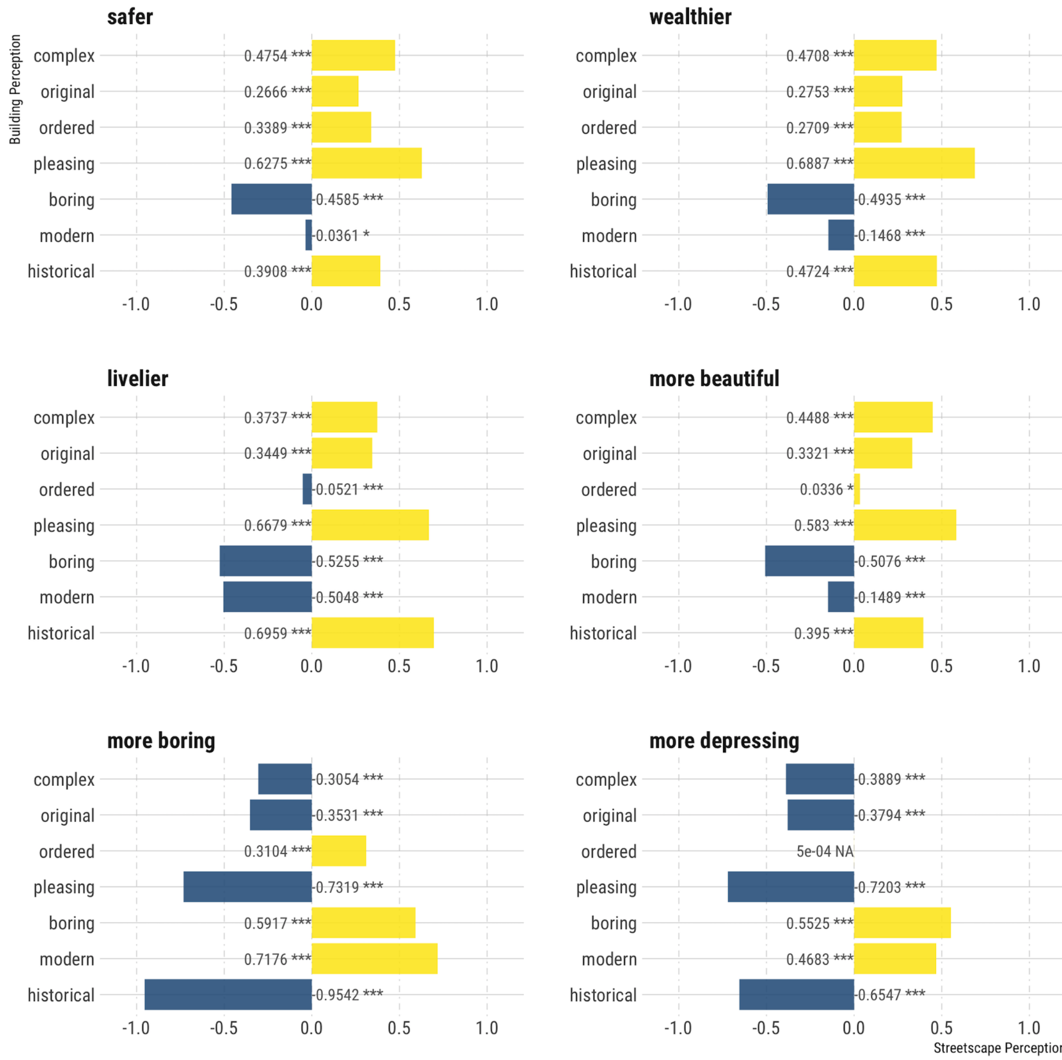


Figure 14: Average effect of treatment on the treated regarding building perception properties across various streetscape perception attributes. (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. NA means "not applicable".)

aligns with prior research suggesting that disordered physical environments can diminish a sense of safety [85]. Moreover, a higher level of modernity in building design is found to have significant negative effects on positive streetscape perception, particularly decreasing the sense of liveliness in streetscapes, with a coefficient of -0.504. In contrast, original buildings can enhance the perception of liveliness, wealth, and beauty in a place.

Regarding the perceptions of "boring" and "depressing" in streetscapes, a higher degree of modern and monotonous building appearances is found to be positively associated with these negative sentiments. Additionally, the findings highlight the importance of historical elements within streetscapes and the quantity of visual elements on building exteriors, which can mitigate negative feelings towards streetscapes. This investigation confirms that the perceptual properties proposed in this study are significantly related to human perception in the urban environment. Although the effect of building exteriors may not be as strong as the visual ratio of vegetation, sky, and buildings identified in previous studies, this study real the trends and extent to which such properties affect human perception in the built environment, which is critical for future urban renovation and building design.

6. Discussion

6.1. Application of building exterior evaluation

Buildings, as the dominant elements in cityscape, have attracted attention from a wide variety of fields that explore urban physical appearance and the visual factors influencing human perceptions. Various insightful studies have conducted architectural design evaluations on groups of buildings, but rarely on a large scale. Therefore, understanding their roles and impacts is vital in shaping urban environments that are both aesthetically pleasing and functionally efficient. The contributions of this work are significant in three aspects: First, it integrates building exterior evaluation criteria with the current urban perception framework, offering a novel approach to assess architectural design perceptions across large-scale urban regions. Second, the study delves into the relationship between human perceptions and building attributes, identifying unique characteristics and spatial patterns for different cities, thereby enhancing our understanding of urban architectural diversity and its perception by residents and visitors. Finally, we deepen the exploration of current urban perception research by assessing the impact of building perceptions, uncovering the trends and extent of their influence. By integrating perceptual properties into an existing urban perception framework, this study offers a comprehensive and comparative understanding of human perceptions of architectural designs across different cities. This approach aids in summarising and balancing the architectural imagery within different city regions, promoting a more human-centric urban design by understanding how constructions and designs contribute to the overall image of these areas. Looking ahead, incorporating other robust and multifaceted building perceptual attributes into this framework can further enhance our ability to create more liveable cities.

Benefiting from the geo-referencing capabilities of street-level imagery, this study represents a preliminary effort to identify, investigate, and understand the spatial patterns of human perception regarding building exteriors. Utilising geo-tagged images allows for the mapping and uncovering of the typical perceptual identity of architectural designs in various urban environments. In our analysis, the concentrated modern areas in the northeast of San Francisco suggest a strong, centralised architectural identity, while the city centre of Amsterdam showcases the significance of historical and exciting constructions. By revealing these building perception patterns, urban planners and architects can gain a macroscopic perspective of entire cities and urban regions. This method can optimise the cityscape in various ways, including monitoring changes in urban appearance, organising visually appealing districts for tourism, and identifying buildings that require exterior renewal. Furthermore, as highlighted by Qiu et al. [75], subjective streetscape perceptions exhibit strong strength in explaining housing prices. Incorporating the perspective of building perception into street-level studies of other urban factors—such as socioeconomic [86, 87], urban activities [88, 89], and residents' mental well-being [90, 91, 16]—this approach can enhance the dimensions of urban informatics and promote more engaging cities.

In this study, the geographic information of buildings is connected with the perception levels of different areas for the first time, providing novel insights into urban design. For instance, Amsterdam's residential areas, which display a less modernised ambience compared to other cities, have a strong positive correlation with the complexity and excitement levels of buildings. Conversely, Singapore's HDB flats, while correlating with orderliness, are perceived as dull, reflecting a functional but monotonous design. This underscores the importance of considering both physical characteristics and human perceptions in urban planning and architectural design to create environments that are both aesthetically pleasing and functionally relevant. Aligning with the existing practices of generative design [92, 93], the potential benefit of these results lies in the creation of comprehensive building profiles, which can be enriched with additional labels and keywords related to visual features, perceptions, geographical locations, and cultural backgrounds. This enhances the ability of machines to better understand and generate exterior designs for architects and city planners.

Urban perception, as a focal topic in urban studies, has received unprecedented attention in recent times. While numerous studies have established connections between visual elements of street views and human perceptions [11, 26], the influence of building exteriors in this process has remained largely unexplored. This research seeks to bridge this gap by employing PSM to assess the impact of architecture designs on the perception of urban streetscapes. Our findings highlight the importance of a building's historical ambience in enhancing the quality of streetscapes, closely followed by its pleasantness and complexity. Extending the works of Zhang et al. [11] and Rossetti et al. [26], which identify cars, sidewalks, and vegetation as influential factors in safety perception, our study reveals that streetscapes featuring more complex and aesthetically pleasing building designs are also deemed to be safer. Besides, to foster a livelier, more exciting, and impressive urban environment, minimising the monotonous and modern sense of architectural design would be advantageous. This approach reveals the significant contributions of building exteriors to individuals' perceptions of their urban surroundings, providing direct insights for future urban development and design strategies.

6.2. Limitations and future work

There are some limitations in our study that should be addressed in future research. First, participant backgrounds play a crucial role in evaluating building exteriors, as preferences for facade designs may vary among different groups [31, 30, 36, 46, 32]. For instance, Imamoglu [30] found that non-architecture students tend to rate both traditional and modern house facades as more complex compared to architecture students. Additionally, the human perception of urban environment is influenced not only by visual factors but also by non-visual elements such as human activities, cultural familiarity, and historical context [94, 95, 96]. Kang et al. [67] identifies disparities in safety perception between deep learning-based measurements and survey-based measurements within neighbourhoods. Future studies could delve deeper into these aspects, seeking to provide a global understanding of the perception biases regarding building exteriors that arise from different socio-demographic backgrounds.

Second, although our current dataset serves as a general representation of urban buildings, balancing thoroughness and efficiency, its scope remains relatively limited, comprising only 400 labelled building images per city. This constraint may affect the precision of our predictions across all building types. When applying our models to new urban areas, the potential biases introduced by visual variances in building designs among different regions should be considered. We believe that the current dataset is suitable for cities that share similar architectural features with those selected for the study, while further investigation into its reliability and sensitivity in culturally distinct regions would be beneficial. Building on our framework, future endeavours can focus on gathering a more comprehensive and diverse collection of building images, thereby covering a broader spectrum of architectural appearances. Moreover, while our study offers a method for capturing building exteriors in urban settings, it focuses mainly on the most visible aspects of facades from roads due to the nature of street-level imagery. To expand the scope of our research, exploring other data sources like Mapillary, which accesses areas beyond drivable roads and the typical scope of GSV [97], could offer a more extensive study of building appearances.

7. Conclusion

Architectural appearance, shaped by buildings that dominate cityscapes, is integral to city planning and human well-being [57, 5, 3]. Traditional studies on architectural evaluation, however, have often concentrated on small, specific groups of buildings, focusing on their visual aspects without considering their broader distribution within urban spaces. To address this gap, we leverage street-level imagery to develop a comprehensive urban perception framework that combines traditional survey-based building evaluation methods with advanced deep learning methods, thereby broadening the scope of building exterior evaluations to encompass larger urban areas. In line with the escalating interest in urban perception, this research dives into an unexplored domain: the perception of building exteriors, and does so within the broader streetscape context and their interplay with objective building attributes and streetscape perceptions. Our methodology includes conducting surveys to gather human perception scores for building images, employing machine learning techniques to assess a wider array of buildings across different urban regions, and conducting the propensity score matching to evaluate how buildings affect streetscape perception. This comprehensive and innovative approach allows us to uncover and compare the distinctive characteristics of architectural designs in various cities, highlighting their potential connections to the built environment and human perceptual behaviours.

To address the research questions posed in the Section 1, our experiments demonstrate:

- RQ1: The proposed framework is capable of precisely capturing the degrees of various perceptual attributes associated with building exteriors in Singapore, San Francisco, and Amsterdam, achieving over 72% accuracy in perception classification tasks.

- RQ2: Cities exhibit distinct yet homogeneous patterns in the perception of building exteriors, linked to their unique urban functions, cultural attributes, and planning strategies. For instance, Amsterdam's residential areas show pleasing, non-modernised development; San Francisco's office and commercial zones feature complex, orderly structures; and Singapore's public housing designs suggest order but a sense of monotony.

- RQ3: Building perceptual properties demonstrate varying degrees of influence on holistic streetscape perceptions. Pleasing appearances and complex, historical features of buildings generally elicit positive responses on streetscapes, while modern and monotonous facades are tend to evoke holistic streetscape perceptions of being “boring” and “depressing”.

This work supports urban design theories and practices by decoding individuals’ opinions on architectural designs of cities. It provides an effective method for city governments or planners to capture the overall image of city’s building design and pinpoint key areas primed for future developmental enhancements and targeted interventions. We believe that this method also holds a great opportunity to be integrated with broader urban informatics studies, fostering a deeper understanding of the interrelations between human experiences and the built environment, which will contribute to future building architectural design and urban planning.

Acknowledgements

We gratefully acknowledge the participants of the survey and the input data. We thank the members of the NUS Urban Analytics Lab for the discussions, especially Koichi Ito, Winston Yap and Zicheng Fan for their helpful suggestions. The Ethical Review Committee of College of Design and Engineering, at the National University of Singapore has reviewed and approved the ethical aspects of this research (reference code 2023/DOA/009). The first author acknowledges the NUS Graduate Research Scholarship granted by the National University of Singapore (NUS) - GRSUR0600042. This research is part of the project Large-scale 3D Geospatial Data for Urban Analytics, which is supported by the National University of Singapore under the Start Up Grant R-295-000-171-133.

Appendix A. Balance check before and after propensity score matching for each building perceptual properties

Table A.1: Balance check before and after propensity score matching for perception scores "complex".

Covariate	Before matching			After matching		
	Treatment (mean)	**Control (mean)**	δ (%)	**Treatment (mean)**	**Control (mean)**	δ (%)
Building	0.144	0.161	-8.874	0.144	0.151	-3.794
Sky	0.306	0.332	-20.820	0.306	0.302	3.449
Road	0.417	0.386	21.109	0.417	0.413	3.039
Vegetation	0.109	0.102	6.468	0.109	0.110	-0.831
Edge Count	0.168	0.152	24.018	0.168	0.169	-0.475
Blob Count	0.114	0.095	22.756	0.114	0.113	0.772
Hue Mean	0.440	0.421	16.351	0.440	0.441	-1.052
Saturation Mean	0.614	0.616	-1.486	0.614	0.612	2.937
Lightness Mean	0.268	0.268	-0.301	0.268	0.268	-0.294

Table A.2: Balance check before and after propensity score matching for perception scores "original".

Covariate	Before matching			After matching		
	Treatment (mean)	**Control (mean)**	δ (%)	**Treatment (mean)**	**Control (mean)**	δ (%)
Building	0.163	0.142	10.554	0.163	0.151	5.759
Sky	0.308	0.331	-18.350	0.308	0.311	-2.357
Road	0.394	0.408	-9.631	0.394	0.396	-1.213
Vegetation	0.115	0.096	16.914	0.115	0.121	-5.295
Edge Count	0.164	0.156	11.926	0.164	0.163	1.849
Blob Count	0.094	0.115	-26.077	0.094	0.093	1.039
Hue Mean	0.440	0.421	16.436	0.440	0.441	-0.762
Saturation Mean	0.609	0.621	-12.123	0.609	0.607	2.393
Lightness Mean	0.278	0.258	18.208	0.278	0.275	2.752

Table A.3: Balance check before and after propensity score matching for perception scores "ordered".

Covariate	Before matching			After matching		
	Treatment (mean)	**Control (mean)**	δ (%)	**Treatment (mean)**	**Control (mean)**	δ (%)
Building	0.137	0.168	-16.148	0.137	0.143	-3.538
Sky	0.302	0.336	-27.416	0.302	0.292	8.009
Road	0.417	0.385	22.256	0.417	0.409	6.746
Vegetation	0.116	0.095	18.863	0.116	0.129	-10.082
Edge Count	0.176	0.144	47.041	0.176	0.183	-9.734
Blob Count	0.130	0.080	63.404	0.130	0.125	4.541
Hue Mean	0.429	0.431	-2.005	0.429	0.430	-1.103
Saturation Mean	0.612	0.618	-5.929	0.612	0.603	9.367
Lightness Mean	0.262	0.274	-11.380	0.262	0.260	2.238

Table A.4: Balance check before and after propensity score matching for perception scores "pleasing".

Covariate	Before matching			After matching		
	Treatment (mean)	**Control (mean)**	δ (%)	**Treatment (mean)**	**Control (mean)**	δ (%)
Building	0.140	0.165	-12.490	0.140	0.136	2.265
Sky	0.317	0.321	-3.049	0.317	0.324	-5.610
Road	0.399	0.403	-2.260	0.399	0.401	-0.911
Vegetation	0.120	0.090	27.169	0.120	0.116	3.708
Edge Count	0.168	0.152	23.071	0.168	0.164	5.364
Blob Count	0.096	0.114	-22.286	0.096	0.092	5.301
Hue Mean	0.438	0.422	13.859	0.438	0.437	0.993
Saturation Mean	0.608	0.622	-14.265	0.608	0.610	-1.316
Lightness Mean	0.274	0.263	10.325	0.274	0.275	-1.336

Table A.5: Balance check before and after propensity score matching for perception scores “boring”.

Covariate	Before matching			After matching		
	Treatment (mean)	**Control (mean)**	δ (%)	**Treatment (mean)**	**Control (mean)**	δ (%)
Building	0.146	0.159	-6.618	0.146	0.152	-3.323
Sky	0.332	0.307	19.688	0.332	0.332	0.007
Road	0.404	0.398	4.466	0.404	0.401	2.065
Vegetation	0.096	0.114	-16.351	0.096	0.093	3.399
Edge Count	0.150	0.170	-28.767	0.150	0.150	-0.205
Blob Count	0.110	0.099	13.823	0.110	0.109	1.706
Hue Mean	0.418	0.443	-21.502	0.418	0.417	0.831
Saturation Mean	0.619	0.611	9.011	0.619	0.618	1.465
Lightness Mean	0.259	0.278	-17.014	0.259	0.258	1.057

Table A.6: Balance check before and after propensity score matching for perception scores “historical”.

Covariate	Before matching			After matching		
	Treatment (mean)	**Control (mean)**	δ (%)	**Treatment (mean)**	**Control (mean)**	δ (%)
Building	0.153	0.152	0.652	0.153	0.153	0.096
Sky	0.344	0.313	24.914	0.344	0.341	1.912
Road	0.386	0.405	-13.178	0.386	0.386	-0.097
Vegetation	0.093	0.109	-14.588	0.093	0.095	-2.607
Edge Count	0.160	0.160	0.022	0.160	0.162	-2.838
Blob Count	0.099	0.106	-10.038	0.099	0.098	1.809
Hue Mean	0.417	0.434	-14.835	0.417	0.414	2.434
Saturation Mean	0.610	0.616	-6.932	0.610	0.610	0.583
Lightness Mean	0.267	0.269	-1.577	0.267	0.269	-2.338

Table A.7: Balance check before and after propensity score matching for perception scores “modern”.

Covariate	Before matching			After matching		
	Treatment (mean)	**Control (mean)**	δ (%)	**Treatment (mean)**	**Control (mean)**	δ (%)
Building	0.158	0.147	5.260	0.158	0.163	-2.932
Sky	0.295	0.343	-38.906	0.295	0.284	8.123
Road	0.414	0.389	17.206	0.414	0.402	8.175
Vegetation	0.111	0.100	10.001	0.111	0.126	-8.585
Edge Count	0.166	0.154	17.092	0.166	0.171	-6.945
Blob Count	0.115	0.094	25.535	0.115	0.109	6.446
Hue Mean	0.436	0.424	10.335	0.436	0.433	3.124
Saturation Mean	0.617	0.613	3.930	0.617	0.612	5.571
Lightness Mean	0.261	0.275	-12.652	0.261	0.259	1.948

References

[1] Y. Kang, F. Zhang, S. Gao, H. Lin, Y. Liu, A review of urban physical environment sensing using street view imagery in public health studies, Annals of GIS 26 (2020) 261–275.

[2] K. Ito, Y. Kang, Y. Zhang, F. Zhang, F. Biljecki, Understanding urban perception with visual data: A systematic review, Cities 152 (2024) 105169.

[3] P. St-Jean, O. G. Clark, M. Jemtrud, A review of the effects of architectural stimuli on human psychology and physiology, Building and Environment 219 (2022) 109182.

[4] M. Nikolopoulou, K. Steemers, Thermal comfort and psychological adaptation as a guide for designing urban spaces, Energy and buildings 35 (2003) 95–101.

[5] K. K.-L. Lau, C. Y. Choi, The influence of perceived aesthetic and acoustic quality on outdoor thermal comfort in urban environment, Building and Environment 206 (2021) 108333.

[6] M. H. Elnabawi, E. Jamei, The thermal perception of outdoor urban spaces in a hot arid climate: A structural equation modelling (sem) approach, Urban Climate 55 (2024) 101969.

[7] J. Liu, J. Kang, H. Behm, T. Luo, Effects of landscape on soundscape perception: Soundwalks in city parks, Landscape and urban planning 123 (2014) 30–40.

[8] T. Zhao, X. Liang, W. Tu, Z. Huang, F. Biljecki, Sensing urban soundscapes from street view imagery, Computers, Environment and Urban Systems 99 (2023) 101915.

[9] H. W. Schroeder, L. M. Anderson, Perception of personal safety in urban recreation sites, Journal of leisure research 16 (1984) 178–194.

[10] D. D. Perkins, J. W. Meeks, R. B. Taylor, The physical environment of street blocks and resident perceptions of crime and disorder: Implications for theory and measurement, Journal of environmental psychology 12 (1992) 21–34.

[11] F. Zhang, B. Zhou, L. Liu, Y. Liu, H. H. Fung, H. Lin, C. Ratti, Measuring human perceptions of a large-scale urban region using machine learning, Landscape and Urban Planning 180 (2018) 148–160.

[12] R. S. Ulrich, Visual landscapes and psychological well-being, Landscape research 4 (1979) 17–23.

[13] R. C. Smardon, Perception and aesthetics of the urban environment: Review of the role of vegetation, Landscape and Urban planning 15 (1988) 85–106.

[14] L. E. Jackson, The relationship of urban design to human health and condition, Landscape and urban planning 64 (2003) 191–200.

[15] B. Jiang, D. Li, L. Larsen, W. C. Sullivan, A dose-response curve describing the relationship between urban tree cover density and self-reported stress recovery, Environment and behavior 48 (2016) 607–629.

[16] L. Xiang, M. Cai, C. Ren, E. Ng, 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 (2021) 108273.

[17] G. L. Kelling, C. M. Coles, Fixing broken windows: Restoring order and reducing crime in our communities, Simon and Schuster, 1997.

[18] H. P. Kotabe, O. Kardan, M. G. Berman, The order of disorder: Deconstructing visual disorder and its effect on rule-breaking., Journal of Experimental Psychology: General 145 (2016) 1713.

[19] R. Ewing, S. Handy, Measuring the unmeasurable: Urban design qualities related to walkability, Journal of Urban design 14 (2009) 65–84.

[20] M. Purciel, K. M. Neckerman, G. S. Lovasi, J. W. Quinn, C. Weiss, M. D. Bader, R. Ewing, A. Rundle, Creating and validating gis measures of urban design for health research, Journal of environmental psychology 29 (2009) 457–466.

[21] N. C. Inglis, J. Vukomanovic, J. Costanza, K. K. Singh, From viewsheds to viewscapes: Trends in landscape visibility and visual quality research, Landscape and Urban Planning 224 (2022) 104424.

[22] F. Biljecki, K. Ito, Street view imagery in urban analytics and GIS: A review, Landscape and Urban Planning 215 (2021) 104217.

[23] F.-Y. Gong, Z.-C. Zeng, F. Zhang, X. Li, E. Ng, L. K. Norford, Mapping sky, tree, and building view factors of street canyons in a high-density urban environment, Building and Environment 134 (2018) 155–167.

[24] P. Salesses, K. Schechtner, C. A. Hidalgo, The collaborative image of the city: mapping the inequality of urban perception, PloS one 8 (2013) e68400.

[25] A. Dubey, N. Naik, D. Parikh, R. Raskar, C. A. Hidalgo, Deep learning the city: Quantifying urban perception at a global scale, in: European conference on computer vision, Springer, 2016, pp. 196–212.

[26] T. Rossetti, H. Lobel, V. Rocco, R. Hurtubia, Explaining subjective perceptions of public spaces as a function of the built environment: A massive data approach, Landscape and urban planning 181 (2019) 169–178.

[27] L. Dai, C. Zheng, Z. Dong, Y. Yao, R. Wang, X. Zhang, S. Ren, J. Zhang, X. Song, Q. Guan, Analyzing the correlation between visual space and residents' psychology in wuhan, china using street-view images and deep-learning technique, City and Environment Interactions 11 (2021) 100069.

[28] B. Wu, B. Yu, S. Shu, H. Liang, Y. Zhao, J. Wu, Mapping fine-scale visual quality distribution inside urban streets using mobile lidar data, Building and Environment 206 (2021) 108323.

[29] D. Verma, A. Jana, K. Ramamritham, Predicting human perception of the urban environment in a spatiotemporal urban setting using locally acquired street view images and audio clips, Building and Environment 186 (2020) 107340.

[30] Ç. Imamoglu, Complexity, liking and familiarity: Architecture and non-architecture turkish students' assessments of traditional and modern house facades., Journal of Environmental Psychology (2000).

[31] K. Devlin, J. L. Nasar, The beauty and the beast: Some preliminary comparisons of 'high'versus 'popular'residential architecture and public versus architect judgments of same, Journal of environmental psychology 9 (1989) 333–344.

[32] M. Ghomeishi, Aesthetic preferences of laypersons and its relationship with the conceptual properties on building façade design, Journal of Asian Architecture and Building Engineering 20 (2021) 12–28.

[33] S. M. Hosseini, M. Mohammadi, A. Rosemann, T. Schröder, J. Lichtenberg, A morphological approach for kinetic façade design process to improve visual and thermal comfort, Building and environment 153 (2019) 186–204.

[34] R. S. Ulrich, Aesthetic and affective response to natural environment, in: Behavior and the natural environment, Springer, 1983, pp. 85–125.

[35] R. Gifford, D. W. Hine, W. Muller-Clemm, D. J. Reynolds JR, K. T. Shaw, Decoding modern architecture: A lens model approach for understanding the aesthetic differences of architects and laypersons, Environment and behavior 32 (2000) 163–187.

[36] G. Brown, R. Gifford, Architects predict lay evaluations of large contemporary buildings: whose conceptual properties?, Journal of environmental psychology 21 (2001) 93–99.

[37] J. L. Nasar, Urban design aesthetics: The evaluative qualities of building exteriors, Environment and behavior 26 (1994) 377–401.

[38] R. Parsons, The potential influences of environmental perception on human health, Journal of environmental psychology 11 (1991) 1–23.

[39] S. Kaplan, The restorative benefits of nature: Toward an integrative framework, Journal of environmental psychology 15 (1995) 169–182.

[40] D. Appleyard, M. Lintell, The environmental quality of city streets: the residents' viewpoint, Journal of the American institute of planners 38 (1972) 84–101.

[41] T. C. Daniel, Whither scenic beauty? visual landscape quality assessment in the 21st century, Landscape and urban planning 54 (2001) 267–281.

[42] R. Gifford, D. W. Hine, W. Muller-Clemm, K. T. Shaw, Why architects and laypersons judge buildings differently: Cognitive properties and physical bases, Journal of architectural and Planning Research (2002) 131–148.

[43] T. R. Herzog, A cognitive analysis of preference for urban spaces, Journal of environmental psychology 12 (1992) 237–248.

[44] H. D. Arslan, K. Yıldırım, Perceptual evaluation of stadium façades, Alexandria Engineering Journal 66 (2023) 391–404.

[45] F. He, Y. He, L. Sun, Gender differences in color perceptions and preferences of urban façades based on a virtual comparison, Building and Environment 245 (2023) 110907.

[46] A. Akalin, K. Yildirim, C. Wilson, O. Kilicoglu, Architecture and engineering students' evaluations of house façades: Preference, complexity and impressiveness, Journal of environmental psychology 29 (2009) 124–132.

[47] R. Weber, J. Schnier, T. Jacobsen, Aesthetics of streetscapes: Influence of fundamental properties on aesthetic judgments of urban space, Perceptual and motor skills 106 (2008) 128–146.

[48] A. Rapoport, History and precedent in environmental design, (No Title) (1990).

[49] D. E. Berlyne, Studies in the new experimental aesthetics: Steps toward an objective psychology of aesthetic appreciation., Hemisphere, 1974.

[50] J. L. Nasar, Visual preferences in urban street scenes: a cross-cultural comparison between japan and the united states, Journal of cross-cultural psychology 15 (1984) 79–93.

[51] R. Kaplan, S. Kaplan, The experience of nature: A psychological perspective, Cambridge university press, 1989.

[52] D. Canter, An intergroup comparison of connotative dimensions in architecture, Environment and behavior 1 (1969) 37.

[53] H. M. Parsons, Work environments, in: Human Behavior and Environment: Advances in Theory and Research. Volume 1, Springer, 1976, pp. 163–209.

[54] C. F. Ng, Perception and evaluation of buildings: The effects of style and frequency of exposure, Collabra: Psychology 6 (2020) 44.

[55] D. J. Levi, Does history matter? perceptions and attitudes toward fake historic architecture and historic preservation, Journal of Architectural and Planning Research (2005) 148–159.

[56] A. Hossein Askari, K. B. Dola, S. Soltani, An evaluation of the elements and characteristics of historical building façades in the context of malaysia, Urban Design International 19 (2014) 113–124.

[57] T. G. Yahner, D. J. Nadenicek, Community by design: contemporary problems—historic resolve, Landscape and Urban Planning 39 (1997) 137–151.

[58] T. Heath, S. G. Smith, B. Lim, Tall buildings and the urban skyline: The effect of visual complexity on preferences, Environment and behavior 32 (2000) 541–556.

[59] J. A. Russell, A. Weiss, G. A. Mendelsohn, Affect grid: a single-item scale of pleasure and arousal., Journal of personality and social psychology 57 (1989) 493.

[60] Y. Liu, M. Chen, M. Wang, J. Huang, F. Thomas, K. Rahimi, M. Mamouei, An interpretable machine learning framework for measuring urban perceptions from panoramic street view images, Iscience 26 (2023) 106132.

[61] X. Liang, T. Zhao, F. Biljecki, Revealing spatio-temporal evolution of urban visual environments with street view imagery, Landscape and Urban Planning 237 (2023) 104802.

[62] Z. Wang, K. Ito, F. Biljecki, Assessing the equity and evolution of urban visual perceptual quality with time series street view imagery, Cities 145 (2024) 104704.

[63] Y. Hou, M. Quintana, M. Khomiakov, W. Yap, J. Ouyang, K. Ito, Z. Wang, T. Zhao, F. Biljecki, Global streetscapes—a comprehensive dataset of 10 million street-level images across 688 cities for urban science and analytics, ISPRS Journal of Photogrammetry and Remote Sensing 215 (2024) 216–238.

[64] C. Harvey, L. Aultman-Hall, S. E. Hurley, A. Troy, Effects of skeletal streetscape design on perceived safety, Landscape and Urban Planning 142 (2015) 18–28.

[65] N. Naik, S. D. Kominers, R. Raskar, E. L. Glaeser, C. A. Hidalgo, Computer vision uncovers predictors of physical urban change, Proceedings of the National Academy of Sciences 114 (2017) 7571–7576.

[66] J. Luo, T. Zhao, L. Cao, F. Biljecki, Water view imagery: Perception and evaluation of urban waterscapes worldwide, Ecological Indicators 145 (2022) 109615.

[67] Y. Kang, J. Abraham, V. Ceccato, F. Duarte, S. Gao, L. Ljungqvist, F. Zhang, P. Näsman, C. Ratti, Assessing differences in safety perceptions using geoai and survey across neighbourhoods in stockholm, sweden, Landscape and Urban Planning 236 (2023) 104768.

[68] J. Kang, M. Körner, Y. Wang, H. Taubenböck, X. X. Zhu, Building instance classification using street view images, ISPRS journal of photogrammetry and remote sensing 145 (2018) 44–59.

[69] F. Ghione, S. Mæland, A. Meslem, V. Oye, Building stock classification using machine learning: A case study for oslo, norway, Frontiers in Earth Science 10 (2022) 886145.

[70] D. Raghu, M. J. J. Bucher, C. De Wolf, Towards a 'resource cadastre'for a circular economy–urban-scale building material detection using street view imagery and computer vision, Resources, Conservation and Recycling 198 (2023) 107140.

[71] T. Lindenthal, E. B. Johnson, Machine learning, architectural styles and property values, The Journal of Real Estate Finance and Economics (2021) 1–32.

[72] M. Sun, F. Zhang, F. Duarte, C. Ratti, Understanding architecture age and style through deep learning, Cities 128 (2022) 103787.

[73] Y. Ogawa, C. Zhao, T. Oki, S. Chen, Y. Sekimoto, Deep learning approach for classifying the built year and structure of individual buildings by automatically linking street view images and gis building data, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 16 (2023) 1740–1755.

[74] S. Liu, Z. Zeng, T. Ren, F. Li, H. Zhang, J. Yang, C. Li, J. Yang, H. Su, J. Zhu, et al., Grounding dino: Marrying dino with grounded pre-training for open-set object detection, arXiv e-prints (2023) arXiv–2303.

[75] W. Qiu, Z. Zhang, X. Liu, W. Li, X. Li, X. Xu, X. Huang, Subjective or objective measures of street environment, which are more effective in explaining housing prices?, Landscape and Urban Planning 221 (2022) 104358.

[76] A. Larkin, A. Krishna, L. Chen, O. Amram, A. R. Avery, G. E. Duncan, P. Hystad, Measuring and modelling perceptions of the built environment for epidemiological research using crowd-sourcing and image-based deep learning models, Journal of Exposure Science & Environmental Epidemiology 32 (2022) 892–899.

[77] R. Herbrich, T. Minka, T. Graepel, Trueskill™: a bayesian skill rating system, Advances in neural information processing systems 19 (2006).

[78] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, B. Schiele, The cityscapes dataset for semantic urban scene understanding, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 3213–3223.

[79] C. Heinrich, A. Maffioli, G. Vazquez, A primer for applying propensity-score matching (2010).

[80] B. Yuen, A. Yeh, S. J. Appold, G. Earl, J. Ting, L. Kurnianingrum Kwee, High-rise living in singapore public housing, Urban Studies 43 (2006) 583–600.

[81] Y. Yao, J. Wang, Y. Hong, C. Qian, Q. Guan, X. Liang, L. Dai, J. Zhang, Discovering the homogeneous geographic domain of human perceptions from street view images, Landscape and Urban Planning 212 (2021) 104125.

[82] S. Lee, N. Maisonneuve, D. Crandall, A. A. Efros, J. Sivic, Linking past to present: Discovering style in two centuries of architecture, in: IEEE International Conference on Computational Photography, 2015.

[83] R. Shao, B. Derudder, Y. Yang, Metro accessibility and space-time flexibility of shopping travel: A propensity score matching analysis, Sustainable Cities and Society 87 (2022) 104204.

[84] L. Cheng, J. De Vos, K. Shi, M. Yang, X. Chen, F. Witlox, Do residential location effects on travel behavior differ between the elderly and younger adults?, Transportation research part D: transport and environment 73 (2019) 367–380.

[85] G. L. Kelling, J. Q. Wilson, et al., Broken windows, Atlantic monthly 249 (1982) 29–38.

[86] S. Yang, K. Krenz, W. Qiu, W. Li, The role of subjective perceptions and objective measurements of the urban environment in explaining house prices in greater london: A multi-scale urban morphology analysis, ISPRS International Journal of Geo-Information 12 (2023) 249.

[87] Z. Fan, T. Su, M. Sun, A. Noyman, F. Zhang, A. entland, E. Moro, Diversity beyond density: Experienced social mixing of urban streets, PNAS nexus 2 (2023) pgad077.

[88] J. Kruse, Y. Kang, Y.-N. Liu, F. Zhang, S. Gao, Places for play: Understanding human perception of playability in cities using street view images and deep learning, Computers, Environment and Urban Systems 90 (2021) 101693.

[89] X. Li, Y. Li, T. Jia, L. Zhou, I. H. Hijazi, The six dimensions of built environment on urban vitality: Fusion evidence from multi-source data, Cities 121 (2022) 103482.

[90] R. Wang, Y. Yuan, Y. Liu, J. Zhang, P. Liu, Y. Lu, Y. Yao, Using street view data and machine learning to assess how perception of neighborhood safety influences urban residents' mental health, Health & place 59 (2019) 102186.

[91] M. Helbich, Y. Yao, Y. Liu, J. Zhang, P. Liu, R. Wang, Using deep learning to examine street view green and blue spaces and their associations with geriatric depression in beijing, china, Environment international 126 (2019) 107–117.

[92] J. Chen, R. Stouffs, From exploration to interpretation: Adopting deep representation learning models to latent space lnterpretation of architectural design alternatives, Proceedings of the 26th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2021 1 (2021) 131–140.

[93] C. Sun, Y. Zhou, Y. Han, Automatic generation of architecture facade for historical urban renovation using generative adversarial network, Building and Environment 212 (2022) 108781.

[94] R. Kaplan, E. J. Herbert, Cultural and sub-cultural comparisons in preferences for natural settings, Landscape and urban planning 14 (1987) 281–293.

[95] S. Bell, Landscape pattern, perception and visualisation in the visual management of forests, Landscape and Urban planning 54 (2001) 201–211.

[96] D. Quercia, N. K. O'Hare, H. Cramer, Aesthetic capital: what makes london look beautiful, quiet, and happy?, in: Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing, 2014, pp. 945–955.

[97] F. Biljecki, T. Zhao, X. Liang, Y. Hou, Sensitivity of measuring the urban form and greenery using street-level imagery: A comparative study of approaches and visual perspectives, International Journal of Applied Earth Observation and Geoinformation 122 (2023) 103385.