This is the Accepted Manuscript version of an article published by Elsevier in the journal *Landscape and Urban Planning* in 2023, which is available at:
https://doi.org/10.1016/j.landurbplan.2023.104802
Cite as: Liang X, Zhao T, Biljecki F (2023): Revealing spatio-temporal evolution of urban visual environments with street view imagery. *Landscape and Urban Planning*, 237: 104802.

Revealing spatio-temporal evolution of urban visual environments with street view imagery

Xiucheng Liang[a,b], Tianhong Zhao[b,c], Filip Biljecki[b,d,*]

[a] *School of Life Sciences, Zhaoqing University, China*
[b] *Department of Architecture, National University of Singapore, Singapore*
[c] *School of Architecture and Urban Planning, Shenzhen University, China*
[d] *Department of Real Estate, National University of Singapore, Singapore*

Abstract

The visual landscape plays a pivotal role in urban planning and healthy cities. Recent studies of visual evaluation focus on either objective or subjective approach, while describing the visual character holistically and monitor its evolution remains challenging. This study introduces an embedding-driven clustering approach that integrates both physical and perceptual attributes to infer the spatial structure of the visual environment, and investigates its spatio-temporal evolution. Singapore, a highly urbanised yet green city, is selected as a case study. Firstly, a visual feature matrix is derived from street view imagery (SVI). Then, a graph neural network is constructed based on road connections to encode visual features and spatial dependency leading to a clustering algorithm that is used to discover the underlying characteristics of the visual environment. The implementation characterises streetscapes of the city-state into six types of clusters. Finally, taking advantage of historical SVI, a longitudinal analysis reveals how visual clusters have evolved in the past decade. Among them, one of the clusters represents high-density visual experience, affirming the work as such streetscape dominates the central business district and it is evolving elsewhere, mirroring the expansion of new towns. In turn, another identified cluster, indicating sparse landscapes, decreased, while areas that are considered to be in the most visually pleasant cluster, increased. For the first time, this study demonstrates a novel method to understand the urban visual structure and analyse its spatio-temporal evolution, which could support future planning decision-making and urban landscape betterment.

Keywords: computer vision, urban perception, street-level, urban pattern, change detection, GraphSAGE

*Corresponding author
Email addresses: xiucheng@u.nus.edu (Xiucheng Liang), zhaotianhong2016@email.szu.edu.cn (Tianhong Zhao), filip@nus.edu.sg (Filip Biljecki)

1. Introduction

Urban landscape — the appearance of the city — provides an abstract depiction of the urban function, identity, socio-demographics, and history (Krause, 2001). As defined in the European Landscape Convention (Council of Europe, 2000), the landscape of an area is generated from the action and interaction of natural and/or human factors that are further perceived by humans, highlighting the sensory connection between the observer and the landscape (Nijhuis et al., 2011). This description also emphasises that the visual environment is not only an enumeration of objective elements, such as physical cues and natural environment (Lynch, 1960; Dramstad et al., 2006), but also the aggregation of subjective perception from human beings (Kaplan and Herbert, 1987; Kaplan and Kaplan, 1989; Bell, 2001). In particular, this environment would not only influence how people perceive and experience their surroundings, but also have potential relationships with physical health (Jackson, 2003), psychological well-being (Smardon, 1988; Quercia et al., 2014; Jiang et al., 2021), and the social environment (Kelling and Coles, 1997; Kotabe et al., 2016). Therefore, assessing the visual quality of urban environment plays a critical role in environmental improvement for urban planning and better quality of life for residents (Kaplan, 1995).

Considering that multifaceted factors are interwoven in the urban environment, a systematic approach is required to quantify visual quality. Traditional studies of the visual environment commit to establishing representative indicators to monitor and evaluate the visual quality (Kaplan and Herbert, 1987), which can be summarised into two typical perspectives: objective and subjective (Lothian, 1999; Daniel, 2001). These approaches predominantly rely on manual assessment of field images by experts or respondents, which is cost-intensive and has a low coverage, inhibiting the implementation at the urban scale. Street view imagery (SVI), with its wide coverage, fine spatial sampling, as well as pedestrian perspective, has attracted widespread attention in the assessment of urban visual quality (Biljecki and Ito, 2021). Typically, various computer vision models and training datasets are applied, such as object detection, semantic segmentation, and image classification, to extract physical objects from images. At the same time, large-scale perception measurement is also conducted by labelling and pre-training models (Zhang et al., 2018; Yao et al., 2019; Zhao et al., 2023). Altogether, SVI-driven studies offer unparalleled opportunities to bridge the gaps of previous visual assessment approaches.

However, due to the complex mechanisms between physical and perceptual attributes (Zhang et al., 2018; Rossetti et al., 2019; Dai et al., 2021), there are significant research gaps remain: first, previous evaluation of urban visual landscape focus either on objective or subjective approach, while few of them established a framework to comprehensively describe the environment from both aspects. Second, the approach to describe and observe visual environment change is still vague. According to the urban visual similarity research of Doersch et al. (2012) and Zhou et al. (2014), as well as homogeneous geographic domain proposed by Yao et al. (2021), the distribution of visual environment is believed to have potential similarity and spatial aggregation in urban environment. Furthermore, SVI across years also provides valuable opportunities to detect changes of streetscapes (Naik et al., 2017; Tang and

Long, 2019; Byun and Kim, 2022; Li et al., 2022; Hou and Biljecki, 2022; Han et al., 2023), providing insight into how the visual environment evolves with urban (re)development. Inspired by them, we determine to investigate the evolution of urban visual environment through portraying its holistic patterns over time, raising three research questions:

1. Do visual features reveal certain spatial patterns of the visual environment? And what visual characteristics they represent?
2. How can we utilise SVI to indicate the dynamics of the built environment?
3. How does the visual pattern evolve within the urban fabric?

To answer these questions, discovering the underlying structure and evaluating the visual environment, this study introduces a comprehensive and comparative framework based on multi-year SVI. First, we start by selecting features from two most important aspects: physical and perceptual, relying on previous visual landscape assessment frameworks. Then, deep learning methods are employed to extract the features and embed the visual properties with spatial structure. Once the compressed representation is generated, clustering analysis is engaged to unpack the features of multiple research periods into distinct partitions. Finally, following the interpretation of the characteristic of the clustering result, the spatial distribution and temporal evolution are conducted to analyse the variation patterns of visual environment in case study area. The key contributions of this study are a new approach to characterise cities and their street spaces at a high spatial resolution, a method to monitor changes in the built environment, a new use case of SVI, and one that relies on data from multiple periods, a rarity.

2. Background and Related Work

2.1. Traditional approaches in visual landscape studies

Preserving and promoting visual environment is one of the crucial missions of urban planning. Previous studies have identified urban visual landscape as a comprehensive environment condition (Tveit et al., 2006; Silver and Clark, 2016), which consists of not only objective components, such as physical characteristics, natural conditions, and spatial forms (Alexander, 1977; Dramstad et al., 2006; Quercia et al., 2014; Harvey et al., 2015), but also contextual factors that could affect subjective perception, such as weather, human activities, cultural familiarity, as well as social and historical conditions (Kaplan and Herbert, 1987; Bell, 2001). Furthermore, the integration of such visual and non-visual cues would not only invisibly but profoundly affect an individual's perception of a place and their psychological well-being (Lynch, 1960; Smardon, 1988; Quercia et al., 2014). Specifically, exposure to greenery has been evidenced in many aspects as principal keys to human health (Jackson, 2003; Jiang et al., 2016), while visual disorder in urban appearance might induce rule-breaking behaviours (Kelling and Coles, 1997; Kotabe et al., 2016).

Taking into account the intertwined factors in visual environment, various evaluative frameworks have been established to investigate the impact of multiple indicators on visual landscape performance. As summarised by Lothian (1999) and

Daniel (2001), previous visual landscape assessment can be generally divided into two different paradigms: objectivist and subjectivist approaches, which focus on physical components and psychological impression, respectively. The objectivist approach assumes that visual quality can be measured by experts through the physical characteristics inherent in environment, while subjectivist evaluates the quality from beholders' visual perception. For example, a comprehensive visual quality assessment method, introduced by Daniel and Vining (1983), contains five different models ranging from expert-based to perception-based approach. In addition, building on the proposal by Daniel and Vining (1983), Kaplan et al. (1989) examine the relative merit of judgemental and perceptual indicators in environmental preference and further revealed the importance of combining multiple predictor domains in visual environment studies.

Recently, advances in geographical techniques triggered a data-driven stream of large-scale assessment in visual environment studies, of which the previous judgemental components have been translated into computer measurable indicators, such as density, naturalness, transparency, complexity, and enclosure (Ewing and Handy, 2009; Purciel et al., 2009; Tveit and Sang, 2014). For example, by measuring the skeletal streetscape of New York City, Harvey (2014) indicates the physical conditions, especially building enclosure and tree canopy geometry, have a significant relationship with visual appeal. Besides, enclosure, the indicator significantly related to a place's sense of liveability and security (Porta and Renne, 2005; Jorgensen et al., 2002), has been measured by various new approaches, such as sky visibility, proportion of street wall, and section height to width ratio (Ewing and Handy, 2009; Carmona and Tiesdell, 2007). Visibility modelling, another advanced application in this field, is also widely used to simulate human point of view for assessing visual quality (Štefunková and Cebecauer, 2006; Inglis et al., 2022), while these valuations are mainly based on geographic information rather than authentic visual condition observed by pedestrian. In short, traditional studies conclude the emphasis in the terms of integrated perspective in visual environment, as well as the limitation on large-scale and street-level evaluation.

2.2. Street view imagery in urban studies

Providing a unique ground-level perspective of cityscape with wide coverage and fine spatial sampling, SVI has been widely used in urban built environment studies across multiple scales (Biljecki and Ito, 2021). The images are typically labelled according to different research purposes and are further applied in the training of semantic segmentation models. In the field of urban studies, the most commonly labeled elements are inherent ones such as green, sky, and building. These elements are not only quantified as enclosure, openness, and greenery features of streetscape (Li et al., 2018; Zhou et al., 2019; Tang and Long, 2019; Gong et al., 2019b), but also utilised to investigate correlations with other urban information that has underlying geographical characteristics, including urban form (Gong et al., 2019a; Li and Ratti, 2019; Ito and Biljecki, 2021), socio-economic factors (Glaeser et al., 2018; Meng et al., 2020), and land use (Srivastava et al., 2020; Hu et al., 2020; Li et al., 2021). For example, classifying visual components into six categories: sky, trees & plants, buildings, impervious surfaces, pervious surfaces and non-permanent objects, Middel et al. (2019) analyse their spatial distribution and correlation in

Philadelphia, and further identify the visual characteristics within three different types of neighbourhoods. Additionally, by detecting visual elements, computer vision can also help us to understand physical environment and its spatial structure in the city: Doersch et al. (2012) investigate visual geographies from different geospatial scales based on SVI, and further explore the visual identities of three central districts of Paris based on their representative elements. On that note, geotagged images are also used by Zhou et al. (2014) to conduct identity recognition experiments in global cities, indicating visual differences and similarities geographically.

Besides sensing the appearance of the built environment, expert-based and perception-based labelling are utilised for scalable visual landscape assessment based on previous theoretical frameworks. To automatically measure the physical street quality, Liu et al. (2017) establish criteria for expert rating to label images based on the maintenance quality of building façade. From the non-physical aspects, Place Pulse, a crowdsourced dataset containing multiple perceptual attributes (depressing, boring, beautiful, safe, lively, wealthy), is created based on human-derived rankings (Salesses et al., 2013). Building on this, Naik et al. (2014) introduce a model called Streetscore-CNN, to predict the safety index of streetscape for multiple cities. Thanks to the advantage of low-cost and high accuracy, both the dataset and model have been widely used not only in later urban perception studies (Li et al., 2015; Zhang et al., 2018; Rossetti et al., 2019; Liu et al., 2023), but also as the indicators representing perceived quality of the urban environment (Harvey et al., 2015; Naik et al., 2017; Luo et al., 2022). Similarly, Yao et al. (2019) developed a human–machine adversarial scoring framework, with high throughput and accuracy for perception assessment. By integrating the framework with a directed graph abstracted from the road network, further exploration is carried out to discover the homogeneous geographic domains of different perception attributes in Beijing, indicating that perceptual communities exhibit specific characteristics (Yao et al., 2021).

In general, deep learning methods enable the scalability of visual assessment, and recent studies explore the possibilities of predicting both physical or perceptual attributes for SVI with high accuracy, bridging the gap of its traditional approach. However, as the cityscape is a complex of subjective and objective factors, how these visual factors affect the image of cities as a whole and how they converge into different visual domains offer ample room for further research. Therefore, in line with the traditional frameworks for the evaluation of visual environment, this study introduces a synthesis method based on material–space view and space–perception view, including the most representative attributes of urban visual environment identified by previous studies. Further, it takes advantage of the availability of SVI at the same locations from different time periods (i.e. over a decade) to assert a temporal component to this important aspect of urban planning.

2.3. Clustering analysis based on urban informatics

Although supervised learning enables highly accurate urban information detection on the basis of known semantic relationships, cities, as sophisticated products of diverse human activities, house a myriad of interrelated information that are difficult to penetrate (Bettencourt and West, 2010). Therefore, unleashing the implication of urban complexity beyond human's a prior knowledge is another essential task in urban studies. Unsupervised learning (UL), with its ability to automatically explore high-dimensional data, has been routinely used for various purposes, such as urbanisation evaluation (Deng et al., 2009; Ye and Chen, 2015), environmental sound classification (Salamon and Bello, 2015; Oldoni et al., 2015), city substantiality assessment (Akande et al., 2019), geographical data translation (Wu and Biljecki, 2022) and so on. Under this category, clustering analysis is the most commonly used application for disentangling hidden patterns among heterogeneous information based on feature similarity, with k-means as the most efficient and implementable algorithm (Jain, 2010).

To identify geographical units and further portray the potential urban structure that is distinct from administrative divisions, various studies perform a cluster analysis involving multivariate geotagged data, including visual components features (Li et al., 2017), socioeconomic indexes (Li and Xie, 2018), social media activities (Steiger et al., 2016; Huang et al., 2021), built environment metrics (Bobkova et al., 2021; Niu et al., 2021; Jochem et al., 2021), etc. Among them, Li and Xie (2018) summarise the features by principal component analysis into four categories: race-poverty, education-employment, housing and age-mobility, then further engage the k-means algorithm to establish diverse neighbourhood typologies, which indicate the underlying socioeconomic pattern of Metro Detroit. Such hidden patterns, on the one hand, are usually interpreted by representative features, since label referencing is not capable for UL (Tessler et al., 2016; Comber et al., 2020). On the other, the contextual information is also typically used as supplement data for cluster definition and investigation. For example, after generating multiple clusters based on landscape metrics by k-means, Schmiedel et al. (2015) and Ferrara et al. (2017) profile the relationship between environmental characteristics and human activities. Besides, automatic urban pattern recognition is also conducted based on the integration of data from various aspects (Gao et al., 2017; Tu et al., 2018; Zhang et al., 2019; Cai et al., 2019). To be specific, Tu et al. (2018) apply hierarchical clustering to identify six types of urban functional zones by incorporating landscape (remote sensing images) and human activity metrics (mobile phone positioning data), with further exploration on gradient analysis of urban patterns in Shenzhen. Furthermore, having revealed the underlying structure, the clustering approach also facilitates the detection of large-scale urban change based on dynamic data sources (Li and Xie, 2018; Qi et al., 2019; Tao et al., 2019).

However, traditional clustering approaches, such as k-means, mainly focus on thematic characteristics of input variables without spatial constraints (MacQueen et al., 1967; Gao et al., 2017), while the spatial adjacency relationship is a crucial factor in urban structure analysis (Tobler, 1970). Therefore, some studies consider spatial constraint as the connection between different venues in the city and further compile nodes (geotagged information) and edges (spatial relationship) into a graph to detect spatial aggregation (Yu et al., 2014; Steiger et al., 2016; Sun et al., 2016; Gao et al., 2017; Yao et al., 2021; You, 2022). In this list, we feature the work of Yao et al. (2021), who examine the homogeneous domain of human perception in Beijing. Integrating the road network and image-based perception scores to establish a weighted directed graph, they detect the irregular partition of each perceptual attribute based on the Infomap algorithm, demonstrating that it is possible to explore the spatial aggregation of mental attributes at community scale and their relationship with urban functions.

Moreover, other methods, such as Delaunay-triangulation-spatial-constraints clustering (Gao et al., 2017), geographical self-organising map (Steiger et al., 2016), and graph sample and aggregate network (GraphSAGE) (Fan et al., 2021) are also employed for network-based pattern detection. Taken together, these investigations highlight the importance and possibility of considering spatial constraint to capture hidden patterns in cities.

In summary, previous visual landscape research emphasises the importance of both physical environment and human perception, which are further measured separately by different SVI-based studies in terms of low-cost and large-scale. The related efforts of streetscape clustering are introduced by Li et al. (2017) and Gong et al. (2019b), focusing on the characteristics of physical visual features. We advance the evaluation frameworks taken by them and aim to consider spatial dependency to gather visual environment structure comprehensively. Moreover, it is important to underline that our work is not confined to SVI in single period, as our clustering approach not only examines the possibility of partitioning visual features from multiple years, but also provides first exploration of visual character change.

3. Methodology

A research framework is established with three main steps (Figure 1):

1. Physical and perceptual features extraction: measurable physical component indices and perceptual attributes of multi-year SVI are automatically assessed by deep learning models, and the average value of representative features is then calculated for each research unit.
2. Features embedding: to consider multiple features as well as spatial dependency, the convolutional neural networks GraphSAGE is engaged to integrate feature matrix (established from representative features in Step 1) and undirected graph (generated by adjacency matrix based on road connections), resulting in a vector representation of visual condition for each unit.
3. Visual features clustering: the agglomerative clustering method K-means is employed to discover distinct visual clusters based on embedding features, with the cluster number determined by silhouette method. Further, the most representative variables and complementary built environment indicators of clusters are utilised to generate substantive interpretations for their characteristics. The detailed description of each part of the research framework is as follows.

3.1. Physical and perceptual features extraction

In the computer vision community, traditional street visual quality indicators (e.g. enclosure, greenery, openness, visual pavement) are commonly evaluated by physical components extracted from SVI (Tang and Long, 2019; Zhou et al., 2019), with the view index of vegetation, building, sky, and road as the most representative indicators among them (Gong et al., 2018). Therefore, Cityscapes, a large-scale dataset designed to capture urban scenes (Cordts et al., 2016), is selected as the training dataset due to its

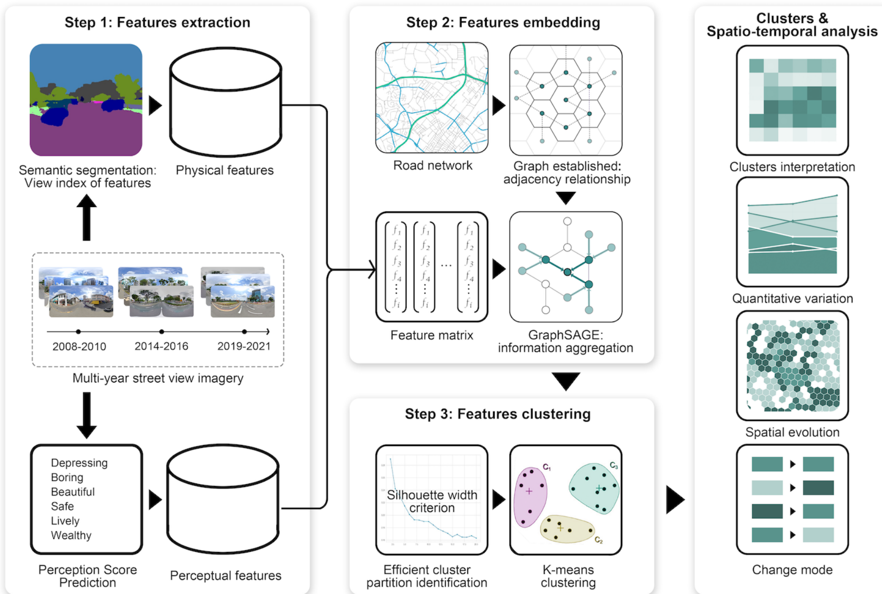


Figure 1: A comprehensive and comparative framework for revealing visual environment patterns based on multi-year street view imagery.

collection of 5000 high-quality pixel-level images with more than 19 urban semantic annotations. The deep learning model we employ in this study is DeepLabV3+, which attains a high accuracy in semantic segmentation task when trained on Cityscapes (Chen et al., 2018). To capture the inherent visual environment, we further exclude the temporary objects (e.g. person, cars, and bicycles) and aggregate other physical components into four indicators: sky, vegetation, building and road, following by six classes proposed by Middel et al. (2019). Therefore, the images in study area are segmented into different ratios of street composition, and further summarised into four inherent environment indices based on the categories shown in Table 1.

For perceptual scores prediction, we emphasise Place Pulse 2.0 (PP 2.0), the seminal work by Dubey et al. (2016), which introduces a crowdsourced dataset to quantify human perception of urban appearance from six aspects (depressing, boring, beautiful, safe, lively and wealthy) based on volunteer labelling. It contains 110,988 images from 56 cities in 28 countries. Based on examination of its initial version, no significant cultural or individual preference biases were found, indicating its feasibility for global studies (Salesses et al., 2013). To learn this compressed set of variables, we employ the ResNet model (He et al., 2016), which introduces residual blocks to tackle the degradation problem of deep convolutional neural network, and it is widely used because of its practicality and high accuracy in solving complicated tasks. Thus, the images are valued by ten dimensional features from physical and perceptual attributes extraction based on the deep learning models and dataset we implemented in this study.

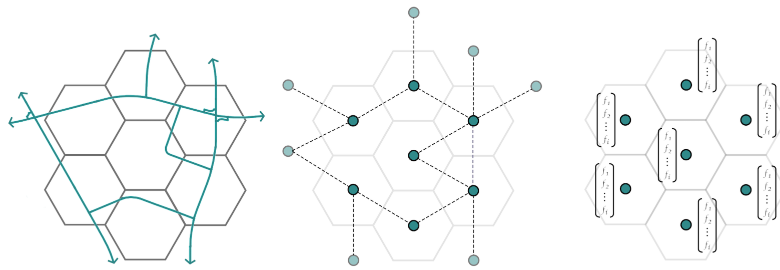
Table 1: Summarising 19 features from the Cityscapes dataset into five categories, with the first four as indicators representing inherent physical environment.

1. Sky	**2. Vegetation**	**3. Building**	**4. Road**	**5. Non-inherent features**
sky (11)	vegetation (9)	building (3)	road (1)	person (12)
	terrain (10)	wall (4)	sidewalk (2)	rider (13)
		fence (5)	pole (6)	car (14)
			traffic light (7)	truck (15)
			traffic sign (8)	bus (16)
				train (17)
				motorcycle (18)
				bicycle (19)

3.2. Feature embedding

Potential dynamic factors (e.g. location, weather, traffic conditions) are common biases of SVI research that could affect the inherent visual features. In this study, we draw inspiration from Tobler's first law of geography as well as other urban studies (Tobler, 1970; Hong and Yao, 2019; Yao et al., 2021) — urban areas have higher spatial variable similarity and connectivity tend to share similar characteristics. We assume that visual features are correlated in space and spatial constraint is a crucial factor in shaping urban visual clusters. Therefore, network analysis is introduced to capture spatial similarity and smooth the visual features according to geographical relationships, in order to provide a better representation for performing clustering and reveal the holistic patterns of visual environment. To consider spatial dependency, an undirected graph ($G = V, E$), consisting of nodes (V) and edges (E), is established. In this case, we apply the hexagonal hierarchical geospatial indexing system (H3) provided by Uber Technologies at the scale of resolutions-10 (average hexagon area 0.017 km^2), as our research units, which are considered as nodes (V) in graph network. Following previous studies on urban patterns, road networks, as conspicuous patterns in cities, are believed to have a shaping role in urban communities (Hong and Yao, 2019), such as urban functional areas and transport analysis zones. Therefore, the OpenStreetMap (OSM) road map is employed to detect the physical connection of adjacent nodes: first, the research units are overlaid with road network (Figure 2a). Second, once each two units are linked by the same road, an undirected edge (e, $\forall e \in E$) is generated according to adjacent matrix, representing their spatial interaction. Accordingly, the undirected graph G, which captures spatial similarity and topological connectivity, is established as shown in Figure 2b.

Furthermore, the physical and perceptual features (ten dimensional features mentioned in Section 3.1) of SVI are combined to represent the visual environment of each research unit, belonging to which the features of images are averaged as initial characteristic indices (a_i), and are further normalised to values f_i in the range [0, 1] with Min-Max normalisation, as estimated by Eq.(1). Thus, a feature matrix is established as $F^v = (f_1^v, f_2^v, \ldots, f_i^v)$, where v equal to the total number of research units, and each vector represents a set of feature values to represent nodes (V) (Figure 2c).



(a) Research units & road network (b) Undirected graph G (c) Feature matrix F^v

Figure 2: Generating graph and feature matrix based on research units at resolution-10.

$$f_i = \frac{a_i - \min(a_i)}{\max(a_i) - \min(a_i)} \tag{1}$$

i = sky, vegetation, buildings, surface, depressing, boring, beautiful, safe, lively, and wealthy

To further incorporate spatial adjacency matrix (G) and features matrix (F^v), an appropriate feature embedding model is required to abstract the information. Among different methods of network analysis, graph embedding methods are commonly employed to encode structure and property of network model, while rather than identifying global topological structure, such methods focus more on the local topology of the network and without considering propagation of features attributes (Kim and Yoon, 2022). To address this dilemma, spatial-based graph neural networks (GNNs) are introduced to capture spatial dependency and aggregate node information based on graph convolution operation. Among them, in order to portray visual environment as an overall image of a place, GraphSAGE, a spatial-based GNNs introduced by Hamilton et al. (2017), is adopted in this study. As an inductive framework designed to generate node embedding, GraphSAGE treats features sampling and aggregating from the node neighbourhoods while considering network structure, resulting in a low-dimensional vector representation of nodes that can be further applied to downstream tasks, such as classification and clustering.

For brevity, the detailed technical developments of this model are moved to Appendix A. In general, GraphSAGE is utilised because of two reasons: first, it is an efficient unsupervised technique that considers node's feature and graph information and fits naturally to solve clustering type of graph problem. Second, this model can propagate feature information based on graph network, which, since scenes in SVI are often connected by road network, fits our purpose that bring the features of each research unit closer to its neighbours, reducing the influence of dynamic factors such as offset, time, etc. of SVIs. Through this case, we aim to validate the feasibility of this method in urban studies, and to extend the application of network-based pattern detection, alongside other approaches mentioned in Section 2.3.

3.3. *Visual features clustering*

Having encoded all features and spatial structures of research units based on neural network embedding, the vector representation can be further clustered to portray homogeneous visual partitions between physical and perceptual features. In this task, we employed K-means, a partitional algorithm commonly used in pattern recognition due to its simplicity, effectiveness and veritable results (Jain, 2010). The clustering is accomplished by relocating statistical centroids and grouping points (the feature vectors of units) that are closest to particular centroid that identify the minimum squared error between points and centroid. In this approach, data points inside each cluster have relatively similar attributes, while clusters exhibit distinct characteristics based on the cluster number specified by users. Following studies on urban pattern recognition (Schmiedel et al., 2015; Spielman and Singleton, 2015; Comber et al., 2020), the average silhouette width criterion (Rousseeuw, 1987) is applied to identify the most efficient partition. To enable cluster comparison and interpretation, the indicators of different clusters are further normalised from the average value of feature attributes, combining with other urban functional data, such as point-of-interest (POI), land use distribution, and urban morphology. Among them, indicators are quantified into density based on research units, which are further averaged and normalised to reveal the similarities and characteristics of certain clusters.

4. Case study and data collection

4.1. *Study area*

The main island of Singapore is selected as the study area, which houses a diverse landscape and built environment from historical to modern times, due to its multiple local ethnic cultures, architecture, and urbanisation of the city (Yuen, 2006). Singapore is a highly urbanised metropolis that includes 23 towns and 3 estates that contribute to its urban morphological features. It is also considered as a representative city of green urbanism that balances the built and ecological environment (Newman, 2010; Palliwal et al., 2021; Wu and Biljecki, 2021). Among planning zones, Singapore is characterised by the Central Area, which includes the most densely developed places like the Central Business District; the subzones in Western Region, which accommodate the vast majority of industrial estates; the Central and Western Water Catchments, which are nature reserves with rich natural resources and less urban development; and other areas, which largely made up of residential towns (Figure 3). In line with the vision of sustainable urban development of cities, it features the necessity of capturing the distribution of visual patterns that have potential effects on physical and mental well-being and monitoring the transformation and determinants of visual landscape for future planning decision making.

4.2. *Street view imagery data collection and processing*

In this study, we divide the research span into three periods: 2008-2010, 2014-2016, and 2019-2021 (with 2009, 2015, and 2021 being the main years of interest), to obtain Google Street View (GSV) images covering all over Singapore for each period (the periods are based on the imaging campaigns of the service). Then, datasets

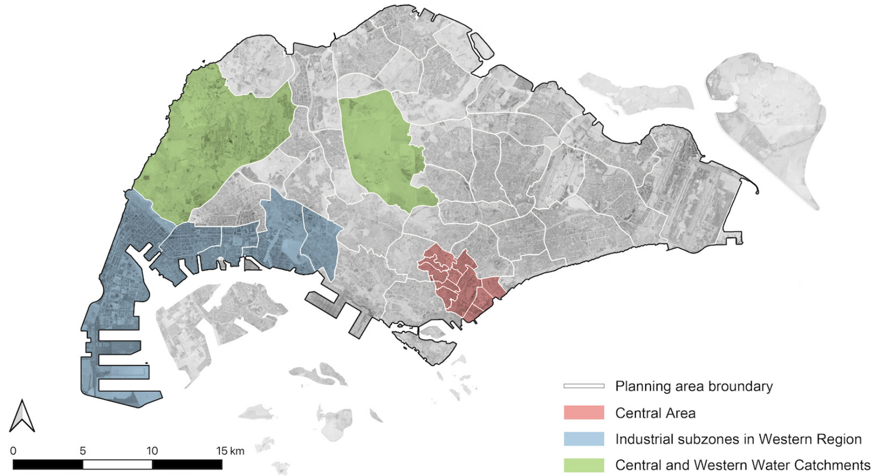


Figure 3: Map of the study area (main island of Singapore). The white lines represent the boundary of planning areas, with the Central Area, industrial zones in the Western Region, the Central Water Catchment, and the Western Water Catchment highlighted. These areas have distinct environment identities according to their function (commercial, industrial, and nature), which will be automatically identified with our novel method.

from the main years (2009, 2015, 2021) are overlayed with the H3 geospatial indexing system at resolution-12 (average hexagon area 347 m^2) to identify locations that have been covered in all three years, and one image from each year is retained for each identified hexagon. Subsequently, images from adjacent years (2008, 2010, 2014, 2016, 2019, and 2020) are included as supplementary data for the remaining resolution-12 hexagons, where no information is available in the three main years. Hence, this results in three images (from three different periods) collected in each of the 111,397 hexagons, which are identified to contain data from all the periods. Figure 4 indicates the number of images collected per month within the periods under consideration. In total, 334,191 images are used for further processing (111,397 images each period) within 14,938 research units at resolution-10, as mentioned in Section 3.1, and the volume of images contained in each research unit is shown in Figure 5.

4.3. Contextual data

With the objective of providing a comprehensive profile of visual patterns and a foundation for spatial variation analysis, contextual urban indicators are calculated from three dimensions: urban function, morphology, and human activity, which are believed to have distinct influences on visual environment (Naik et al., 2017; Zhang et al., 2018; Gong et al., 2019b; Yao et al., 2021).

The indicators of urban function originate from the Singapore Master Plan 2019 (Urban Redevelopment Authority, 2019), which is established as a guide to urban development over the next 10 to 15 years. It is summarised into four categories of indices based on the areas of land use types: industrial, living, transportation

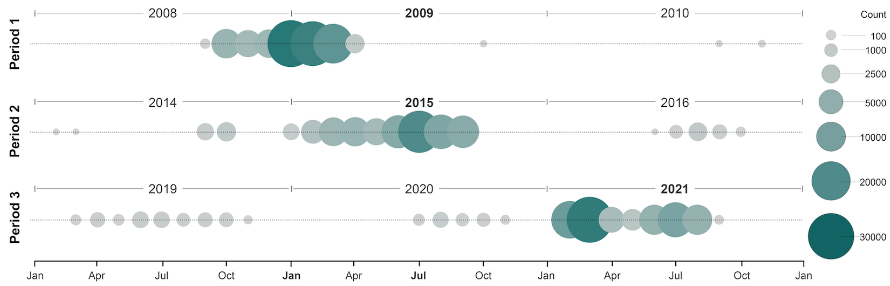


Figure 4: The number and temporal distribution of street view images utilised in this study (breakdown by the considered periods).

Figure 5: The 14,938 research units (average hexagon area 0.017 km^2 at resolution-10 according to the H3 system) and volume of imagery in each.

Table 2: Contextual indices derived from different data sources are calculated for 14,938 units. The Mean represents the normalisation of the average values of all hexagon units.

Category	**Variables**	**Description**	**Mean**	**Data Source**
Land use	Road index	Proportion of road area	0.217	Master Plan 2019
	Residential index	Proportion of residential area	0.527	
	industrial index	Proportion of industrial area	0.070	
Urban morphology	Building index	Proportion of footprint area of buildings	0.302	OpenStreetMap
	Building distance index	Average building distance within 25 meter buffer	0.492	GBMI
Human activity	POI density index	Average POI density	0.096	OpenStreetMap

and open spaces, revealing urban function distributions of the city. In addition, POI data is also adopted to investigate the index of services in clusters. The urban morphological indicators include building density and average building distance. The first index originates from building footprints obtained from OSM, while the second one is calculated based on the average building distance within 25 m of each building according to the Global Building Morphology Indicators (GBMI) open dataset (Biljecki and Chow, 2022).

Combining these datasets, as shown in the Table 2, the values of each study unit are averaged and further aggregated to form contextual representations of the different types of clusters. We acknowledge that the original datasets only reflect the urban context at certain time periods, and therefore the study period closest to the release date of dataset is chosen to further interpret the characteristics of visual pattern.

5. Results and analysis

Having the vector representation of visual features in case study area, silhouette method is employed to determine the numbers of clusters following by (Spielman and Singleton, 2015; Comber et al., 2020), and the average silhouette width of 100-class k-means solution is calculated (Figure 6). Thus, considering an interpretable number of clusters with higher silhouette width, a six-cluster solution is adopted in this study, generating from the integration of feature embedding result of three periods. In this section, we will describe our empirical findings from two perspectives. First, significant features and contextual data in certain period are used to unravel the hidden characteristic of six visual domains. Then, the spatial evolution and temporal variation analysis are further conducted to reveal the dynamic changes of visual environment in case study area.

5.1. Clustering results and interpretation

Non-hierarchical clustering identified six distinct visual clusters in three periods. To generate substantive interpretations of our empirical findings, the third research period (2019-2021) is selected, with the average value of visual features shown in Table 3. The distributions of each value are further visualised in density plots (Figure 7),

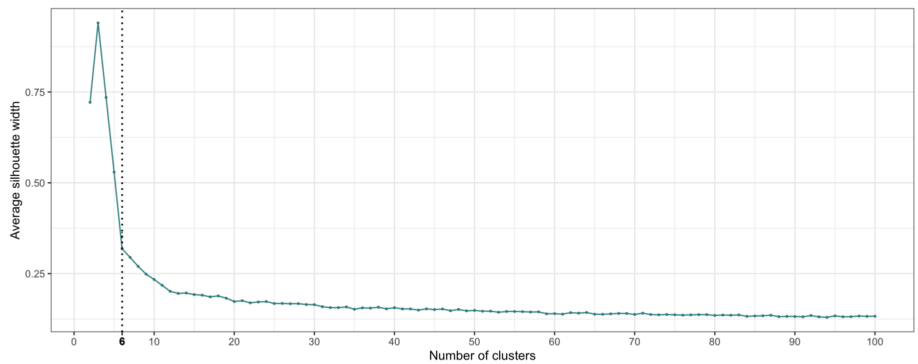


Figure 6: The average silhouette width of 100-class k-means solution, with the applied six-class solution indicated by a dotted line.

Table 3: Cluster characteristics represented by average value of visual features (maximum value in bold).

Indicators	road	building	vegetation	sky	wealthy	beautiful	lively	depressing	safe	boring
Cluster 1	0.414	**0.265**	0.170	0.577	0.608	0.475	0.581	0.493	0.520	0.476
Cluster 2	0.439	0.027	**0.586**	0.343	0.713	**0.708**	0.617	0.312	0.657	0.400
Cluster 3	0.418	0.099	0.192	**0.704**	0.562	0.468	0.487	**0.530**	0.478	**0.546**
Cluster 4	0.423	0.063	0.391	0.518	**0.747**	0.682	**0.669**	0.293	**0.700**	0.383
Cluster 5	0.406	0.118	0.269	0.607	0.690	0.591	0.634	0.375	0.630	0.426
Cluster 6	0.419	0.070	0.376	0.531	0.643	0.591	0.581	0.422	0.579	0.458

with images to present examples and with descriptor names we defined for each cluster. Additionally, incorporating contextual data as multifaceted profile (mentioned in Section 4.3), the statistical relationships between clusters and built environment are displayed in Figure 8. The identities of visual environment reflected by six clusters are interpreted as follows.

Cluster 1: Urban Jungle. As the second smallest group that comprises 1,448 units (about 10%), it is distinguished by the highest view factor of urban construction and the lowest proportion of vegetation. This is also verified by the fairly dense and compact built environment indicated in contextual indicators (Figure 8), implying the dominance of man-made structures over the visual landscape. Although it enjoys relatively high scores in terms of liveliness, it has the lowest beauty index and tends to be perceived as depressing. In general, Cluster 1 is a highly urbanised or vibrant area with relatively crowded visual senses created by frigid cityscape, resulting in a negative impact on human perception.

Cluster 2: Flanked by Nature. Cluster 2 characterises only 1,306 units and is differentiated by the highest proportion of green view factor and extremely low urbanised features (e.g. roads and buildings), conjuring the image of a place rich in natural recourse and low in development. On the note that vegetation (e.g. tree or grass) serves as a significant contributor of landscape preference (Jiang et al., 2015;

Zhang et al., 2018), such as beauty and safety, these units also stand out for their extremely high rate of beauty and safety, with relatively low rates of depressing and boring. Overall, Cluster 2 projects the most nature-dominated landscape in the case study area, providing a delightful and pleasant visual experience for observers.

Cluster 3: Expansive Horizons. Identified as Cluster 3, 2,099 research units share visual similarity in their large proportion of sky view factors and relatively low in other visual components (e.g. vegetation and buildings), which are also reinforced by urban functional characteristic that these visual landscapes are mostly made up of urban-industrial and traffic areas, with the least residential development. The second identity of Cluster 3 is the significantly lower level in terms of wealthy, safety, and lively compared to other clusters. These identities depict the scene of underdeveloped neighbourhoods that lack proper consideration of the aesthetics of the visual landscape and have a potential opportunity for future improvement.

Cluster 4: Balanced Living. There are 3,264 units contained by Cluster 4, sharing the similarity of the highest index of safety, wealthy, lively, and beauty for its visual environment. Compared to Cluster 2 that is extremely well endowed with greenery which might reduce the sense of safety (Dai et al., 2021), the visual characteristic of this cluster reflects a good balance between urban development and natural landscape, resulting in a homogeneous domain that has a high visual quality in urbanised areas.

Clusters 5 & 6: Concrete Heights and Sparse Neighbourhoods. Consisting of a large amount of research units, Cluster 5 and Cluster 6 are highly similar to each other, albeit a few features can set apart the two clusters. These clusters represent the intermediate state that both physical and perceptual attributes occur at medium magnitude among other groups above, without particular visual feature dominating. In fact, Cluster 5 is characterised by the second highest building view factor, reflecting a high density developed urban area with lower greenery coverage compared to Cluster 6. This identity suggests a cross-balanced environment of visual features, revealing relatively high-density but low-greenery areas. Additionally, although the physical features of Cluster 6 are highly similar to Cluster 4, the values of perceptual factors are slightly opposite, representing a lower visual quality of landscape. The overall image of this cluster is sparse, lacking a dominant feature, and the overall visual perception of the scene is relatively negative.

In general, the first three clusters exhibit different distinctive visual characteristics, while others have a high similarity on physical features. In a synthesis of characteristics and external circumstances in Figure 8, we can further reveal that firstly, as mentioned above, the most representative visual features are consistent with contextual information. For example, Cluster 1, a group with a high visual intensity of buildings tends to allocate in the area with high building coverage and road density. Also, such visual environment tends to distribute in a non-residential area but contains vibrant services, indicated by highest POI density and lowest residential usage. On the contrary, the lowest development and POI density are associated with Cluster 2, which is the most well-vegetated cluster in the case study area. Moreover, we feature the strong spatial relationship between the last three clusters and residential function,

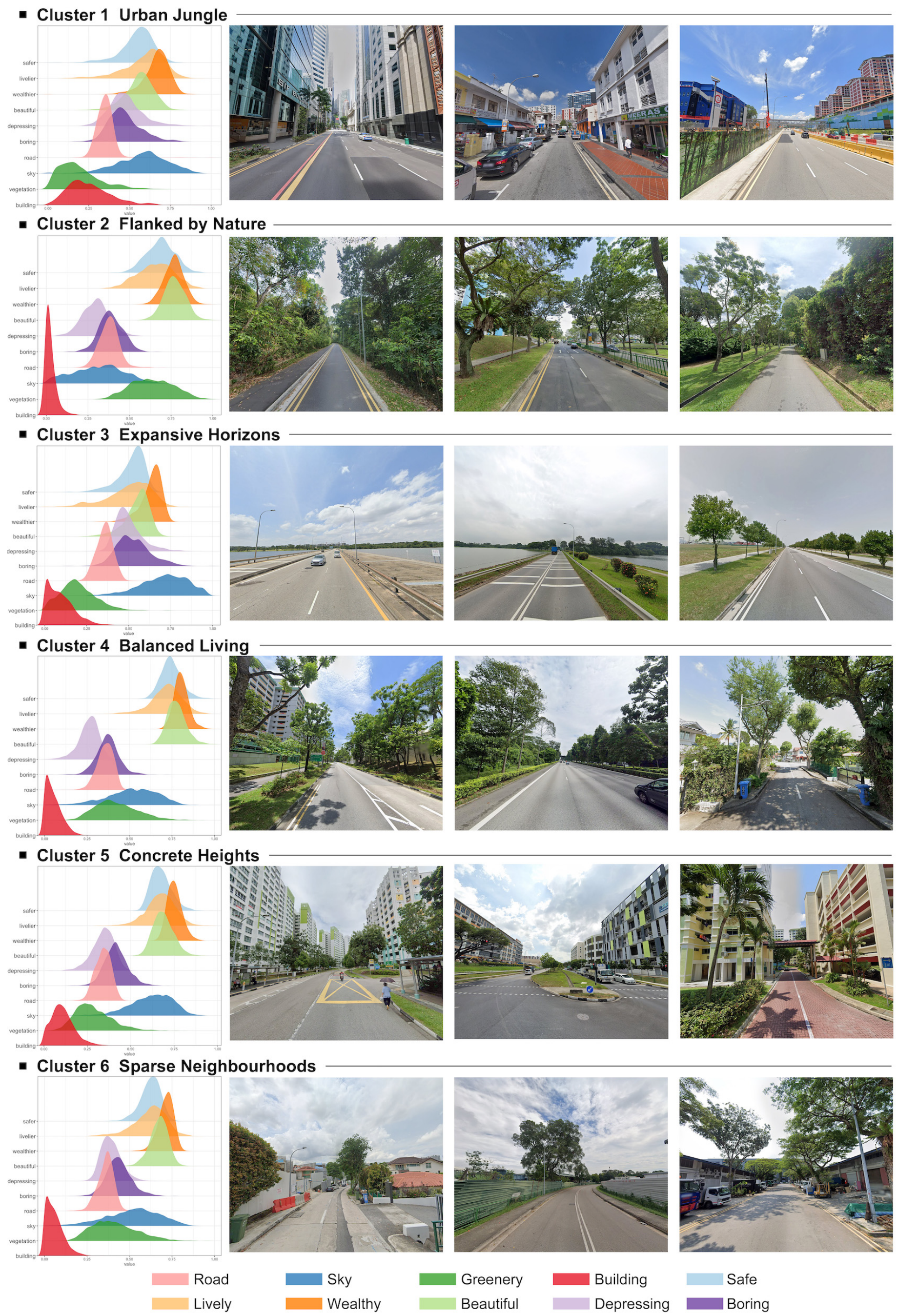


Figure 7: The six identified clusters of the visual environment in Singapore with epithets we assigned to each. The average value distribution of 10 different visual features of each cluster is included, together with their image samples.

Figure 8: The average value of contextual information (land use, urban morphology, and human activity) among the six clusters.

implying that such environments mainly represent a visual experience within living neighbourhoods according to different perceptions. Such correlation between visual landscape and land use is also present in Cluster 3, which tends to locate around industrial areas in contrast to the other groups. Aligning with Yao et al. (2021), who suggest that urban functions have a relationship with perceptual attributes, we also identify the relationship among urban function, built environment, human activity, and visual landscape, further enriching confidence in the validity of cluster representation.

5.2. Spatio-temporal evolution analysis

According to clustering analysis among three periods, sequential investigations on the evolution of visual environment are conducted from three perspectives: 1) Quantitative variation: detecting the overall changes of visual clusters longitudinally. 2) Spatial evolution: observing the location and changes of the spatial visual patterns. 3) Course and frequency of change: summarising the changing patterns and further exploring the prominent forms of evolution.

5.2.1. Quantitative variation

Generally speaking, Singapore has witnessed an urbanisation and landscape improvement according to changes in visual clusters (Figure 9). For example, the total number of visually urbanised clusters (Clusters 1, 4, 5 and 6) has altogether increased, representing 73.1%, 74.6% and 77.2% of the total units in each period (2008-2010, 2014-2016 and 2019-2021), respectively. Furthermore, the visual environment in Singapore is largely occupied by the three living-focused clusters (4, 5, 6), which comprise the common cityscapes of Singapore. Among them, the areas of Cluster 4

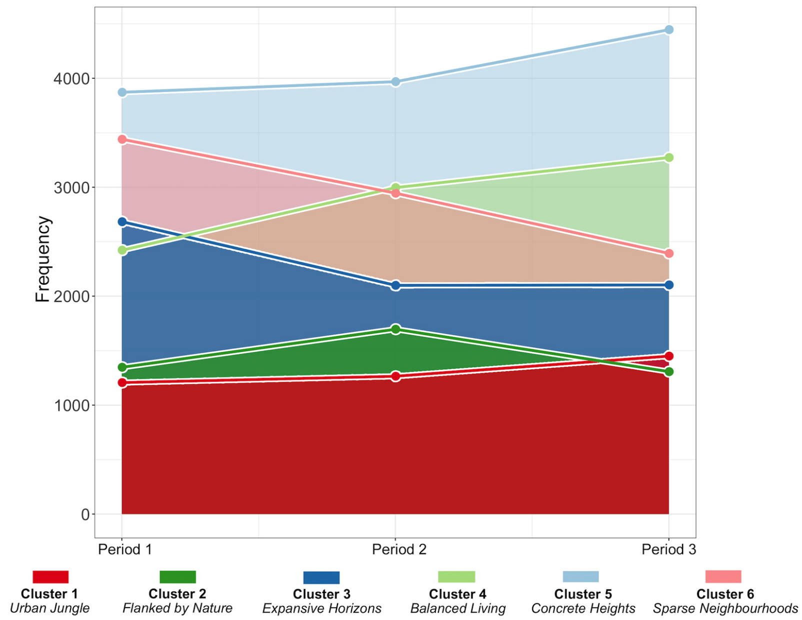


Figure 9: The variation in the total number of each cluster from Period 1 to Period 3.

and 5 with relatively good visual quality have expanded, while the number of Cluster 6 has decreased from 3,431 to 2,387, indicating the improvement of visual environment.

The specific variation between cluster types can be observed in Table 4, with almost half of the previous clusters remaining the same between periods. The global cluster evolution is further visualised in Figure 10, indicating that Period 1 to 2 and Period 2 to 3 share similar trends. For example, highly developed clusters (1, 5) and well-perceived clusters (2, 4) tend to remain, with their conversion rates ranging from 41.5% to 72.9% during these periods. Among them, the visual environment in Cluster 2 (best vegetated domain) and Cluster 1 (highest urbanised domain) are the most stable. In contrast, there is a higher probability of conversion from low visual quality or underdeveloped clusters (Clusters 3 and 6) to other clusters, ranging from 34.7% to 41.8%. This can be explained by the fact that the clusters contain more dominant visual components that are difficult to be overridden, while fragmented visual environment tends to evolve concomitantly with the urbanisation of Singapore. Moreover, the transformation among living-orientated clusters (4, 5 and 6) have also account for large proportions of their changes. The overall growth in Cluster 4 is largely due to conversions from other two clusters (5 and 6), with 1,666 units evolving into it from Period 1 to 2. Similarly, large numbers of research units labelled by Cluster 6, have transferred to Cluster 4 and 5, indicating general visual enhancement. However, in Cluster 2, a reduction from Period 1 to 2 can be observed, suggesting that the cityscape occupied by adequate greenery has been compressed. Therefore, the overall variation of clusters demonstrates the visual development of Singapore, reflecting the

Table 4: The proportion of cluster variation between different periods.

Periods		Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Period 1 (2008-2010) to Period 2 (2014-2016)	Cluster 1	**55.39%**	0%	8.79%	2.74%	32.26%	0.83%
	Cluster 2	0.22%	**72.88%**	1.93%	9.66%	1.86%	13.45%
	Cluster 3	10.80%	0.49%	**35.79%**	5.83%	33.43%	13.67%
	Cluster 4	1.16%	11.96%	6.29%	**41.54%**	14.15%	24.91%
	Cluster 5	5.85%	0.41%	13.80%	18.23%	**48.12%**	13.57%
	Cluster 6	1.46%	11.51%	9.33%	28.04%	13.09%	**36.58%**
Period 2 (2014-2016) to Period 3 (2019-2021)	Cluster 1	**64.95%**	0%	9.97%	1.98%	22.47%	0.63%
	Cluster 2	0.53%	**59.03%**	1.83%	22.43%	1.83%	14.34%
	Cluster 3	9.31%	0.19%	**41.81%**	6.44%	32.41%	9.83%
	Cluster 4	1.81%	4.68%	4.78%	**46.60%**	22.32%	19.81%
	Cluster 5	8.11%	0.13%	13.87%	13.39%	**56.47%**	8.03%
	Cluster 6	1.63%	5.34%	12.73%	27.26%	18.31%	**34.72%**

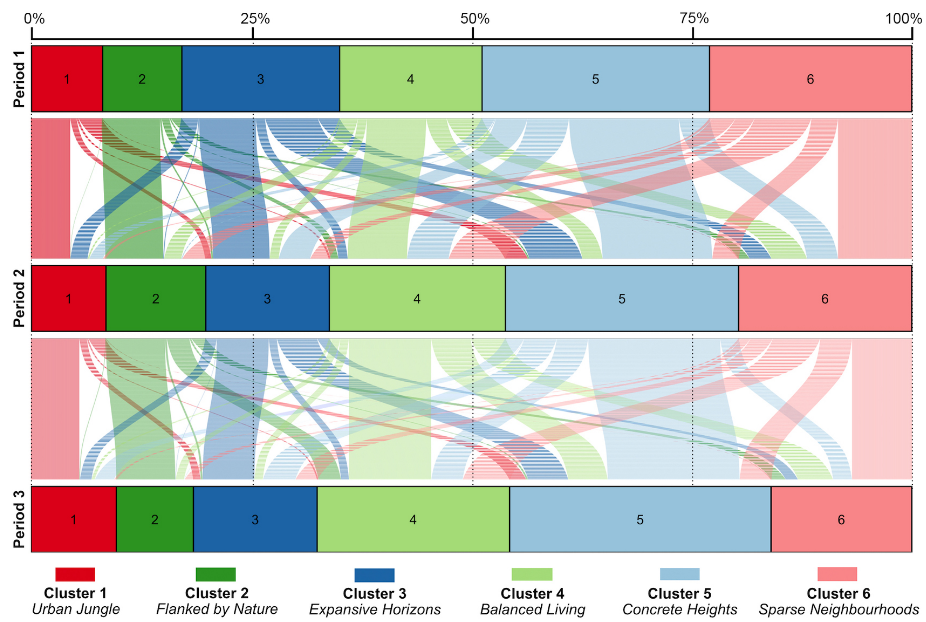


Figure 10: Temporal evolution of different visual clusters from Period 1 to Period 3.

urbanisation and landscape upgrading being the main themes.

5.2.2. Spatial evolution

Besides quantitative analysis, further investigation is conducted on the dynamic spatial distribution of urban visual patterns, with the first and last periods chosen as examples to indicate the evolution of the visual environment over the course of a decade (Figure 11). As a whole, clusters with significant characteristics indicate strong spatial aggregation, while the three living-orientated clusters indicate a relatively equal distribution. Cluster 1 is the main visual domain aggregated in the Central Area, indicating an expansion associated with urban development. This visual condition is also reflected in new built-up areas, with high-density developments that have not yet been landscaped, whereas such areas are often transformed into other clusters in the next period, as a result of post-construction greening improvements. Cluster 3, the sky-dominated cluster, tends to be observed along major roads, reflecting a driving visual experience with poor vegetation. It is spatially distant from urban areas, and tends to aggregate in suburban or industrial areas.

As the dominated clusters in Singapore, Cluster 4, 5, and 6 are distributed more equally in both urban and suburban areas where visual components are abundant, but at the same time, their distributions exhibit their spatial characteristics. For example, the old residential towns in the centre of Singapore are mainly occupied by clusters with well-vegetated environment, while Cluster 5 is dominated in the East Region, where most of the new mature towns are located. These spatial distributions are related to the period of urban development in Singapore, suggesting that earlier built communities provide a more pleasant and inclusive visual experience. This possibly not only due to their relatively low development density, but also because vegetation grows over a longer period of time. In contrast, the new towns contain a higher proportion of Cluster 5 units, followed by Cluster 6. Such groups are especially associated with public housing areas, representing the typical interior visual environment in non-gated communities.

To be specific, spatial evolution also reflects changes in physical environment in certain areas. Tuas, a newly developed heavy industrial district located in the West Region (Figure 11, Area A), has witnessed a visual landscape upgrade. The first period of this area is mainly covered by sky-dominated cluster, while the clusters change can be detected because of the further construction and greenery improvement. Similar changes also appear in Punggol District (Figure 11, Area C), where a new town started to develop since 2007. The dominated clusters have transferred from high dense domain to neighbourhood domain, representing a visual environment evolution from underdeveloped area to fully constructed new town. Furthermore, the visual environment in central residential area has also experienced an evolution (Figure 11, Area B), varying mainly between Cluster 4 and 6. This represents that the changes in physical compositions of these areas are negligible, while their evolution can be differentiated from the variation of perception level, which strongly influenced by other visual factors, such as vehicles, grass, sidewalks, enclosure and openness (Zhang et al., 2018; Dai et al., 2021). These perceptual indicators facilitate the consideration of the visual influence from non-dominant visual elements and thus avoid examining

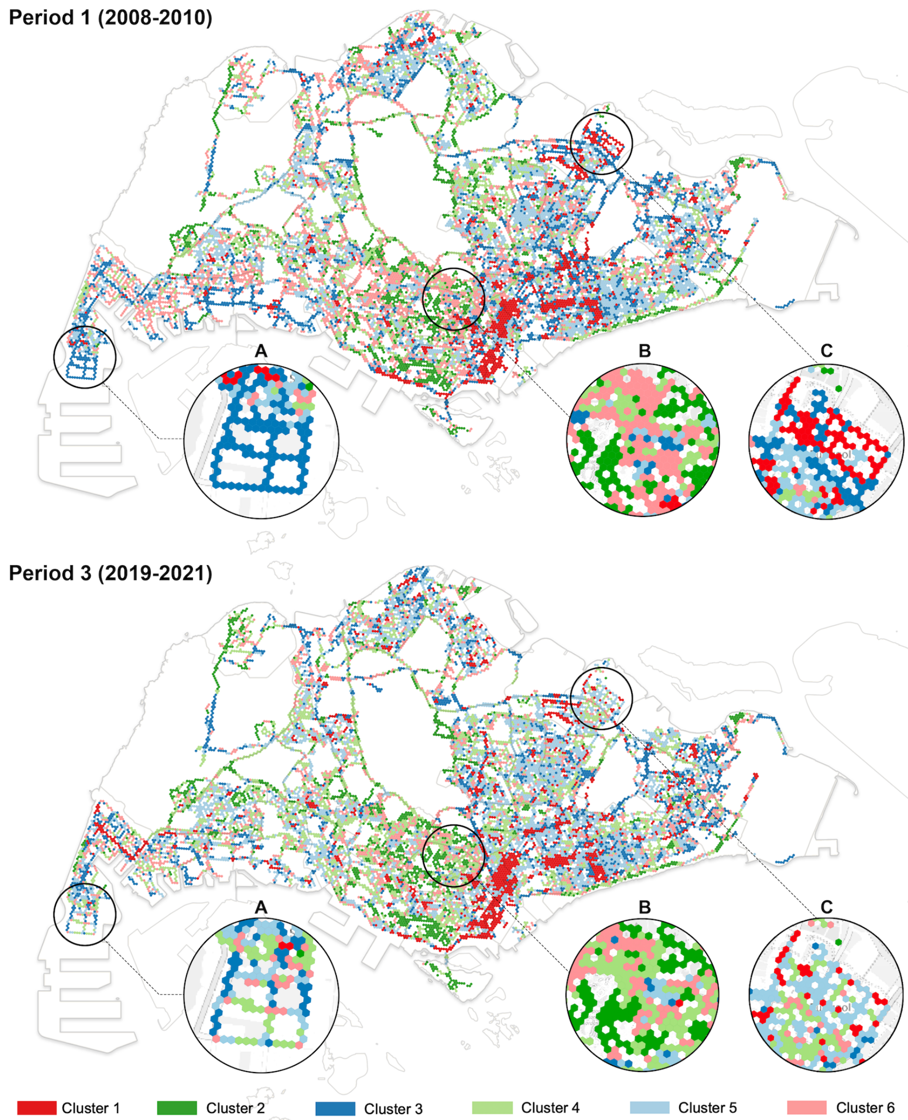


Figure 11: Maps visualising the spatial patterns of visual environment within Period 1 & 3 in Singapore, with representative areas highlighted. Maps visualising the spatial patterns of visual environment within different periods in Singapore, with some areas highlighted as examples.

the complex mechanisms of these attributes, highlighting the importance of integrating visual and non-visual factors in the study of urban visual environment.

5.2.3. Course and frequency of change

To unpack the changing patterns of visual environment, the most 10 frequently appeared change courses are extracted from Period 1 to Period 2 and Period 2 to Period 3 (Figure 12a), excluding the units remained as the same characteristic. As we identified in Section 5.1, we further summarise the main change courses of visual clusters into two directions: perception improved and perception declined, according to Figure 7. In general, the changes appeared mainly between clusters having similar characteristics, reflecting that the overall visual environment in Singapore is stable, with other specific areas having significant changes. For example, the change from Cluster 6 to Cluster 4 (c6-c4) is the most common direction of change and does not indicate an apparent change in the physical environment, but rather an increase in the general level of perception. The change from Cluster 3 to Cluster 5 (c3-c5) is the second common direction of change across periods, which usually occurs in areas of development and construction in the first period, while greening is improved in the second. Correspondingly, changes in the opposite direction of these clusters (c4-c6 and c5-c3) also account for high proportions of perception level decline. Moreover, change courses that reflect significant transformations in the physical environment are also observed across different periods. For instance, there are two significant change courses from Period 1 to Period 2: Cluster 6 to Cluster 2 (c6-c2) and Cluster 1 to Cluster 5 (c1-c5). They all indicate improvements of certain area from urbanised to better vegetated (Figure 13). In Period 2 to 3, 321 units underwent significant visual changes from Group 5 to Group 1 (c5-c1), mainly as a result of road maintenance, which strongly affect the visual environment because it always occurs with the removal of greenery and the increase of construction equipment, resulting in a typical change course from good perceived environment to urbanised visual experience. Taken together, these directions of change provide an understanding of dominant environmental transformation in different periods and reveal potential changing trends between visual clusters.

Furthermore, the distributions of most 10 frequently appeared courses are mapped based on the change directions of perception level Figure 12b. For Period 1 to Period 2, the improvement of visual environment is concentrated in most of the residential areas of Singapore, while the industrial estates, such as the Tuas and northern region, experienced a decrease in visual quality. This trend reversed in the second period: although visual improvement was still the main theme, the perceptual level decline mostly taken place in eastern Singapore. In accordance with this, spatial shifts in the evolution of the visual environment are revealed, and potential areas for visual remediation and enhancement are identified. This also indicates that the overall visual environment tends to remain stable, while areas underwent visual quality degradation tend to improve in the next period.

All in all, the results reveal that the characteristics of visual environment can be differentiated by multivariate extracted from geotagged images at pedestrian level. The evolution of visual clusters also reflects the change in the overall image of a place behind physical changes. This finding validates the possibility to evaluate visual environment from multifaceted view, as well as to detect the dynamic changes

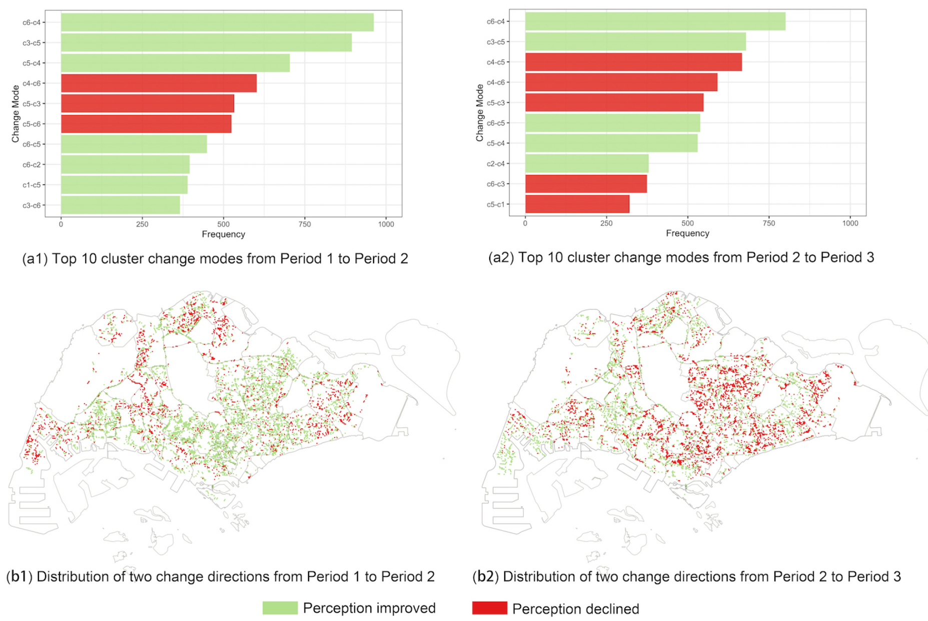


Figure 12: The most frequent courses of change between adjacent periods and the distribution of units that indicate improvement/decline in the level of perception.

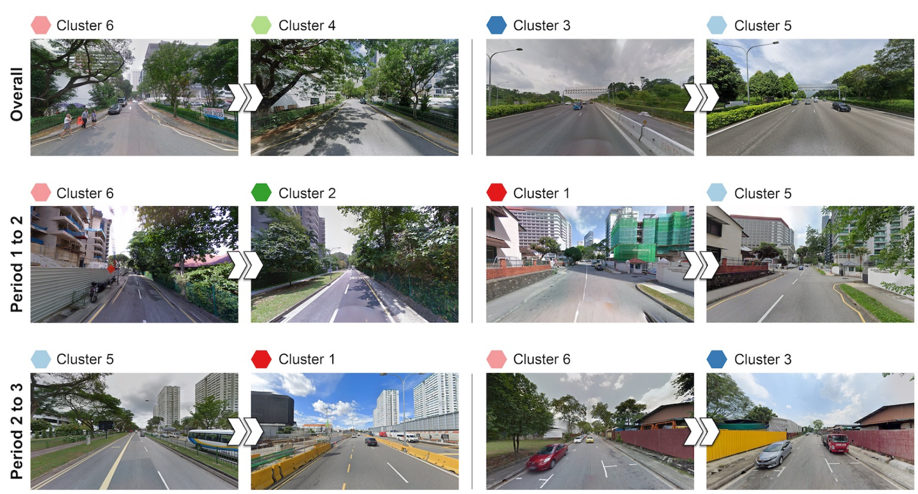


Figure 13: Examples of change courses in different periods.

in a long period based on SVI. These findings give valuable insights for future environment planning and place-making, with the aim of promoting urban liveability and contributing to health and well-being.

6. Discussion

Visual landscape of the urban environment impacts both image of city and health of its dwellers. Various assessment frameworks of visual landscapes have been discussed in the past decades. This study introduced a new comparative approach to discover urban visual pattern and its evolution. The contributions of this work are as follows: first, we summarise and integrate representative visual features in large-scale urban region for evaluation. Second, spatial-based graph neural networks is employed to encode spatial dependency with the features, from which the sensible aggregation of distinct visual environment is revealed. Third, for the first time, multi-year data are applied to explore spatio-temporal evolution of the visual character in a city.

Various studies assess visual landscapes in terms of their physical components or human perception, while a few of them integrate physical and perceptual factors because of the uncertainty of the weighting among them. Through selecting the typical visual components, incorporating them with perceptual attributes that might be affected by other underlying features (Zhang et al., 2018; Rossetti et al., 2019), we are able to investigate cityscapes from a systematic and quantitative perspective. For example, Cluster 4 and Cluster 6 share similar physical elements visually, while perception indicators distinguish their characteristics based on perception levels. This can balance the contribution of each visual element in evaluating visual landscape, providing more robust and stable experimental results. We believe that this exploration shows great potential and establishes a benchmark for visual environment clustering and sensing. In the future, robust and multifaceted variables can be gradually added to this framework, helping us to better understand the underlying visual patterns and process our surrounding environment. Although the distribution of certain visual clusters may come as no surprise, e.g. Cluster 1 is mostly concentrated in the Central Area of Singapore and Cluster 3 corresponds to areas that are less urbanised, the result indicates the structure and aggregation of visual environment beyond individuals' images of certain city areas. For instance, instead of Cluster 1, there are certain units within Central Area that reflect pleasant visual experience, and visual environment in different urban functional areas can share similar visual characteristic. This approach highly relies on transparent and consistent datasets for both model training and prediction, which meet the visual character assessment framework proposed by Tveit et al. (2006), offering a possibility to investigate the impacts of landscape change for visual character and promote a better understanding of the visual environment in urban design.

Rather than evaluating the environment in certain geographical locations (Naik et al., 2017; Tang and Long, 2019; Zhou et al., 2019; Nagata et al., 2020), this study introduces a deep learning approach to consider urban environment from a holistic perspective by learning compressed representation based on multiple features and geographical relationship. We applied GraphSAGE to reveal visual homogeneous areas at community scale, following the human perception homogeneity domain proposed by Yao et al. (2021). This can help us to aggregate urban informatics as a whole based on their spatial structure and minimise the bias created by street view images that capture the scene from different angle, weather, timing, and traffic conditions. The same methodology can also be applied to other urban studies that involve multivariate with strong spatial correlation.

Even though SVI is becoming a prominent data source in temporal studies (Naik et al., 2017; Tang and Long, 2019; Kagan et al., 2021; Byun and Kim, 2022), few studies have used it to investigate the overall evolution of the cityscape. This study applied vast quantities of SVI from various years to conduct time series analysis on visual environment, demonstrating the possibility of utilising it to reveal visual similarity among periods. We summarise the characteristics of visual patterns changes from three aspects: quantitative variation, spatial evolution, as well as course and frequency of change. The results suggest that different types of visual environments are linked to their contextual settings, raising a noteworthy opportunity that visual environment has a potential evolutionary pattern. This could be further explored alongside other temporal urban studies, such as urban morphology evolution and socioeconomic variation, to investigate urban transformation, detect factors of change, and monitor the effects of policies and urban planning.

Furthermore, there are also limitations that should be addressed in future research. First, in addition to the representative factors identified in this study, visual quality of an environment is fostered not only by dynamic visual factors (e.g., human activities, urban function, traffic conditions, etc.), but also the non-visual factors (e.g., urban soundscape, cultural familiarity, historical context, etc.), which simultaneously affect human perception of a place (Kaplan and Herbert, 1987; Bell, 2001; Quercia et al., 2014). Future evaluation of visual landscape could be extended to wider dimensions by formulating a model or digital environment to investigate how such factors can be integrated or transferred to visual sense. Second, SVI is not available everywhere for all the periods. Thus, for newly urbanised areas for which SVI became available only in the last period, the evolution of the visual environment from emergence to stability can be observed only in the future. Finally, in our implementation, we focus on SVI data, which has common limitation that it mainly represents the environment from driving perspective (Zhang et al., 2018; Yao et al., 2021). Besides, despite combining imagery from adjacent years to maximise research coverage, imagery-uncovered area, e.g., parks, open spaces and communities, also constitute as one part of the urban visual environment. Future research could account for off-road imagery (e.g., sidewalk, bicycle lane), evaluating visual landscape from different viewshed and perspective. Other data sources, such as Mapillary and KartaView, which may penetrate areas beyond driveable roads and the typical reach of GSV (Yap et al., 2023), warrant further exploration in future visual environment studies.

7. Conclusions

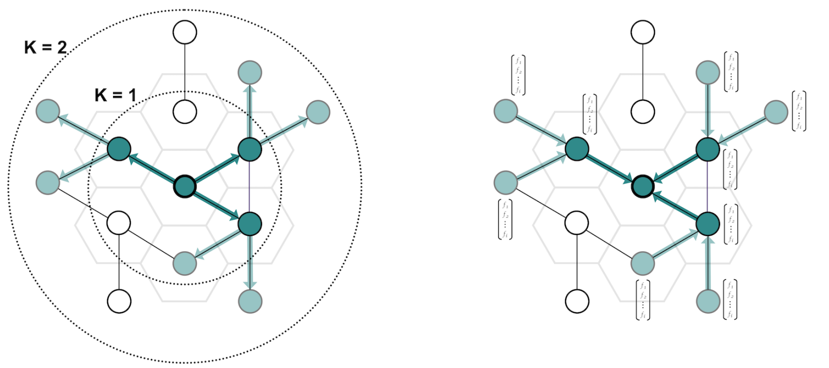
The urban visual environment, as the appearance of cities and streetscapes, plays an essential role in urban planning and human well-being. However, visual evaluations performed by previous studies focus mainly on a specific geographical location from either physical or perceptual aspects, and how the environment evolves remains unknown. As SVI is now an established urban data source, the same streets are being revisited, providing imagery from more than one period, a tremendous benefit that has not yet been taken much advantage of in urban studies. Our work also follows the recent trends in urban informatics, which is increasingly regarding multi-period instances of emerging urban datasets for longitudinal analyses (Morlighem et al., 2022; Li et al., 2022; Chen and Biljecki, 2022; Yu et al., 2022). In this research, we established a comparative framework based on multi-year SVI and investigated the change courses of visual domains. Graph neural network is introduced to associate the visual properties and spatial dependency generated from features extraction and road network structure, respectively. Specifically, this enables us to reveal the characteristics of visual environment at city scale, which indicate potential association with built environment and human activities. We have identified that the global quality of visual landscape in the city has improved, with urbanisation and landscape upgrading as main themes, and same visual cluster always indicate similar course and frequency of change.

This work examined the possibility to utilise vast street level-imagery for urban changes observation at city scale, providing a freely available method for governments or planners to capture the overall image of a city and identify places for future improvement. We believe that this method also has a great opportunity to be expanded to other urban informatics studies, helping us to obtain urban patterns beyond prior knowledge, which will contribute to future urban design and landscape planning.

Appendix A. GraphSAGE

The lack of consideration on spatial constraints is an inherent limitation of traditional clustering approaches, while the spatial adjacency relationship is a crucial factor in urban structure analysis. With this in mind, we employed a GNN model, GraphSAGE algorithm, which uses both feature and graph information and fits naturally to solve clustering type of graph problem, thereby taking spatial constraint into account. As an inductive framework designed to generate node embedding, GraphSAGE treats features sampling and aggregating from the node's neighbourhood while considering network structure (Figure A.14), resulting in a low-dimensional vector representation of nodes that can be further applied to downstream tasks, such as classification and clustering.

In this model, the inputs are the spatial adjacency matrix (G) and features matrix (F^v) obtained in section 3.2, with the parameters of K ('search depths') representing the number of hops for information collection and aggregation from neighbourhoods nodes. Then, GraphSAGE is adopted to encode the information and relationships among urban areas, generating representative embeddings as output. To be specific, the model first combines the two matrices and result in one hidden matrix with multiple dimensions (ten in this case). Second, sampling is operated to identify target node's neighbours that needed for subsequent computation. Third, concatenation layer propagate the information from local neighbours of a node according to the adjacency matrix, with different aggregators that can be operated to aggregate the information. For example, in the kth layer, the embedding representation h_i^k of a given node v_i can be calculated by:



(a) Sample node's neighbourhood

(b) Aggregate feature information from neighbours

Figure A.14: Visual illustration of the GraphSAGE sample and aggregate approach (Hamilton et al., 2017)

$$h_v^k = \sigma\left(W^k \cdot \ Concat \ \left(h_v^{k-1}, \ Aggregate \ \left(h_u^{k-1}, \forall u \in N\left(v\right)\right)\right)\right) \tag{A.1}$$

where all the weight of neighbours of node v are defined as $N(v)$, W is the learnable weight parameter, *Concat* is the operation of concatenation, $\sigma(\cdot)$ is the activation function ReLU and *Aggregate* is the aggregation functions that can be predefined, including three candidate aggregator: mean, Long Short Term Memory, and pooling based different aggregating strategies shown in Hamilton et al. (2017). In practice, we choose Pooling aggregator, which assumes that all the input feature vectors are sufficiently distinct, at search depth K = 2. Through this inductive approach, the vector representation of each research unit is obtained based on multivariate feature embedding and spatial interaction leveraging, thereby smoothing data based on spatial structure of the units and achieving our goal of detecting the holistic representation of the visual environment.

Appendix B. Perception differences between clusters

As a supplementary presentation, this appendix provides detailed visualisation of experimental results to the main text. Addition to Figure 7, the differences of perception values between each pair of clusters are further shown in Figure B.15. The red colour represents certain perception values increase when a cluster transfer from y-axis to x-axis, while blue indicates a decrease. For example, Cluster 3 (C3), identified as Expansive Horizons in previous introduction, increases in value as it evolves to other clusters in terms of safety, lively beautiful and wealthy, especially a significant surge when it changes to Cluster 4 (C4), but a slight improvement in its change to Cluster 1 (C1). According to the figure, we can further evaluate the visual environment change when the cluster evolves.

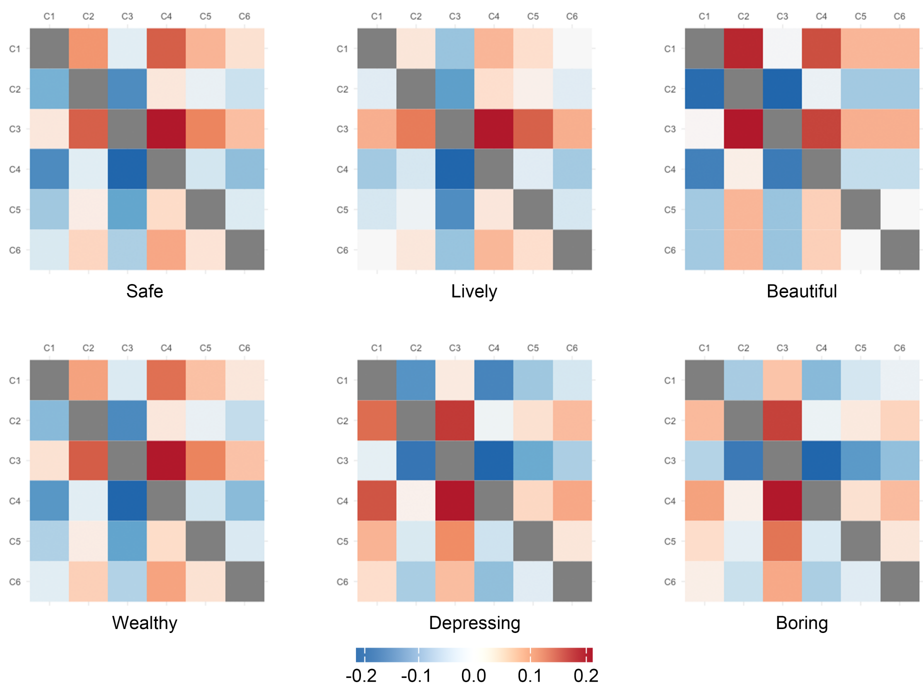


Figure B.15: The comparison of perception values between each pair of clusters (red - x-axis cluster higher than y-axis cluster, blue - x-axis cluster lower than y-axis cluster).

Acknowledgements

We thank the members of the NUS Urban Analytics Lab for the discussions, and the editor and the reviewers for their constructive comments and suggestions. The second author has been supported by Department of Education of Guangdong Province. This research is part of the projects (i) 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; and (ii) The Technical Key Project of Shenzhen Science and Technology Innovation Commission Grant JSGG20201103093401004.

References

Akande, A., Cabral, P., Gomes, P., Casteleyn, S., 2019. The lisbon ranking for smart sustainable cities in europe. Sustainable Cities and Society 44, 475–487.

Alexander, C., 1977. A pattern language: towns, buildings, construction. Oxford university press.

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

Bettencourt, L., West, G., 2010. A unified theory of urban living. Nature 467, 912–913.

Biljecki, F., Chow, Y.S., 2022. Global Building Morphology Indicators. Computers, Environment and Urban Systems 95, 101809.

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

Bobkova, E., Berghauser Pont, M., Marcus, L., 2021. Towards analytical typologies of plot systems: Quantitative profile of five european cities. Environment and Planning B: Urban Analytics and City Science 48, 604–620.

Byun, G., Kim, Y., 2022. A street-view-based method to detect urban growth and decline: A case study of midtown in detroit, michigan, usa. PloS one 17, e0263775.

Cai, L., Xu, J., Liu, J., Ma, T., Pei, T., Zhou, C., 2019. Sensing multiple semantics of urban space from crowdsourcing positioning data. Cities 93, 31–42.

Carmona, M., Tiesdell, S., 2007. Urban design reader. Routledge.

Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation, in: Proceedings of the European conference on computer vision (ECCV), pp. 801–818.

Chen, X., Biljecki, F., 2022. Mining real estate ads and property transactions for building and amenity data acquisition. Urban Informatics 1, 12.

Comber, S., Arribas-Bel, D., Singleton, A., Dolega, L., 2020. Using convolutional autoencoders to extract visual features of leisure and retail environments. Landscape and Urban Planning 202, 103887.

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

Council of Europe, 2000. European landscape convention, in: Report and convention.

Dai, L., Zheng, C., Dong, Z., Yao, Y., Wang, R., Zhang, X., Ren, S., Zhang, J., Song, X., Guan, Q., 2021. 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, 100069.

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

Daniel, T.C., Vining, J., 1983. Methodological issues in the assessment of landscape quality, in: Behavior and the natural environment. Springer, pp. 39–84.

Deng, J.S., Wang, K., Hong, Y., Qi, J.G., 2009. Spatio-temporal dynamics and evolution of land use change and landscape pattern in response to rapid urbanization. Landscape and urban planning 92, 187–198.

Doersch, C., Singh, S., Gupta, A., Sivic, J., Efros, A., 2012. What makes paris look like paris? ACM Transactions on Graphics 31.

Dramstad, W.E., Tveit, M.S., Fjellstad, W., Fry, G.L., 2006. Relationships between visual landscape preferences and map-based indicators of landscape structure. Landscape and urban planning 78, 465–474.

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

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

Fan, C., Yang, Y., Mostafavi, A., 2021. Neural embeddings of urban big data reveal emergent structures in cities. arXiv preprint arXiv:2110.12371 .

Ferrara, C., Carlucci, M., Grigoriadis, E., Corona, P., Salvati, L., 2017. A comprehensive insight into the geography of forest cover in italy: Exploring the importance of socioeconomic local contexts. Forest Policy and Economics 75, 12–22.

Gao, S., Janowicz, K., Couclelis, H., 2017. Extracting urban functional regions from points of interest and human activities on location-based social networks. Transactions in GIS 21, 446–467.

Glaeser, E.L., Kominers, S.D., Luca, M., Naik, N., 2018. Big data and big cities: The promises and limitations of improved measures of urban life. Economic Inquiry 56, 114–137.

Gong, F.Y., Zeng, Z.C., Ng, E., Norford, L.K., 2019a. Spatiotemporal patterns of street-level solar radiation estimated using google street view in a high-density urban environment. Building and Environment 148, 547–566.

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

Gong, Z., Ma, Q., Kan, C., Qi, Q., 2019b. Classifying street spaces with street view images for a spatial indicator of urban functions. Sustainability 11, 6424.

Hamilton, W., Ying, Z., Leskovec, J., 2017. Inductive representation learning on large graphs. Advances in neural information processing systems 30.

Han, Y., Zhong, T., Yeh, A.G., Zhong, X., Chen, M., Lü, G., 2023. Mapping seasonal changes of street greenery using multi-temporal street-view images. Sustainable Cities and Society 92, 104498.

Harvey, C., 2014. Measuring streetscape design for livability using spatial data and methods. The University of Vermont and State Agricultural College.

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

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778.

Hong, Y., Yao, Y., 2019. Hierarchical community detection and functional area identification with osm roads and complex graph theory. International Journal of Geographical Information Science 33, 1569–1587.

Hou, Y., Biljecki, F., 2022. A comprehensive framework for evaluating the quality of street view imagery. International Journal of Applied Earth Observation and Geoinformation 115, 103094.

Hu, F., Liu, W., Lu, J., Song, C., Meng, Y., Wang, J., Xing, H., 2020. Urban function as a new perspective for adaptive street quality assessment. Sustainability 12, 1296.

Huang, J., Obracht-Prondzynska, H., Kamrowska-Zaluska, D., Sun, Y., Li, L., 2021. The image of the city on social media: A comparative study using "big data" and "small data" methods in the tri-city region in poland. Landscape and Urban Planning 206, 103977.

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

Ito, K., Biljecki, F., 2021. Assessing bikeability with street view imagery and computer vision. Transportation Research Part C: Emerging Technologies 132, 103371.

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

Jain, A.K., 2010. Data clustering: 50 years beyond k-means. Pattern recognition letters 31, 651–666.

Jiang, B., He, J., Chen, J., Larsen, L., Wang, H., 2021. Perceived green at speed: a simulated driving experiment raises new questions for attention restoration theory and stress reduction theory. Environment and Behavior 53, 296–335.

Jiang, B., Larsen, L., Deal, B., Sullivan, W.C., 2015. A dose–response curve describing the relationship between tree cover density and landscape preference. Landscape and Urban Planning 139, 16–25.

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

Jochem, W.C., Leasure, D.R., Pannell, O., Chamberlain, H.R., Jones, P., Tatem, A.J., 2021. Classifying settlement types from multi-scale spatial patterns of building footprints. Environment and Planning B: Urban Analytics and City Science 48, 1161–1179.

Jorgensen, A., Hitchmough, J., Calvert, T., 2002. Woodland spaces and edges: their impact on perception of safety and preference. Landscape and urban planning 60, 135–150.

Kagan, D., Alpert, G.F., Fire, M., 2021. Automatic large scale detection of red palm weevil infestation using street view images. ISPRS Journal of Photogrammetry and Remote Sensing 182, 122–133.

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

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

Kaplan, R., Kaplan, S., Brown, T., 1989. Environmental preference: A comparison of four domains of predictors. Environment and behavior 21, 509–530.

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

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

Kim, N., Yoon, Y., 2022. Effective urban region representation learning using heterogeneous urban graph attention network (hugat). arXiv preprint arXiv:2202.09021 .

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

Krause, C.L., 2001. Our visual landscape: Managing the landscape under special consideration of visual aspects. Landscape and Urban planning 54, 239–254.

Li, L., Tompkin, J., Michalatos, P., Pfister, H., 2017. Hierarchical visual feature analysis for city street view datasets, in: Workshop on Visual Analytics for Deep Learning.

Li, M., Liu, J., Lin, Y., Xiao, L., Zhou, J., 2021. Revitalizing historic districts: Identifying built environment predictors for street vibrancy based on urban sensor data. Cities 117, 103305.

Li, M., Sheng, H., Irvin, J., Chung, H., Ying, A., Sun, T., Ng, A.Y., Rodriguez, D.A., 2022. Marked crosswalks in us transit-oriented station areas, 2007–2020: A computer vision approach using street view imagery. Environment and Planning B: Urban Analytics and City Science , 23998083221112157.

Li, X., Ratti, C., 2019. Mapping the spatio-temporal distribution of solar radiation within street canyons of boston using google street view panoramas and building height model. Landscape and urban planning 191, 103387.

Li, X., Santi, P., Courtney, T.K., Verma, S.K., Ratti, C., 2018. Investigating the association between streetscapes and human walking activities using google street view and human trajectory data. Transactions in GIS 22, 1029–1044.

Li, X., Zhang, C., Li, W., 2015. Does the visibility of greenery increase perceived safety in urban areas? evidence from the place pulse 1.0 dataset. ISPRS International Journal of Geo-Information 4, 1166–1183.

Li, Y., Xie, Y., 2018. A new urban typology model adapting data mining analytics to examine dominant trajectories of neighborhood change: a case of metro detroit. Annals of the American Association of Geographers 108, 1313–1337.

Liu, L., Silva, E.A., Wu, C., Wang, H., 2017. A machine learning-based method for the large-scale evaluation of the qualities of the urban environment. Computers, environment and urban systems 65, 113–125.

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

Lothian, A., 1999. Landscape and the philosophy of aesthetics: is landscape quality inherent in the landscape or in the eye of the beholder? Landscape and urban planning 44, 177–198.

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

Lynch, K., 1960. The image of the city. MIT press.

MacQueen, J., et al., 1967. Some methods for classification and analysis of multivariate observations, in: Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, Oakland, CA, USA. pp. 281–297.

Meng, Y., Xing, H., Yuan, Y., Wong, M.S., Fan, K., 2020. Sensing urban poverty: From the perspective of human perception-based greenery and open-space landscapes. Computers, Environment and Urban Systems 84, 101544.

Middel, A., Lukasczyk, J., Zakrzewski, S., Arnold, M., Maciejewski, R., 2019. Urban form and composition of street canyons: A human-centric big data and deep learning approach. Landscape and Urban Planning 183, 122–132.

Morlighem, C., Labetski, A., Ledoux, H., 2022. Reconstructing historical 3D city models. Urban Informatics 1, 11. doi:10.1007/s44212-022-00011-3.

Nagata, S., Nakaya, T., Hanibuchi, T., Amagasa, S., Kikuchi, H., Inoue, S., 2020. Objective scoring of streetscape walkability related to leisure walking: Statistical modeling approach with semantic segmentation of google street view images. Health & Place 66, 102428.

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

Naik, N., Philipoom, J., Raskar, R., Hidalgo, C., 2014. Streetscore-predicting the perceived safety of one million streetscapes, in: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp. 779–785.

Newman, P., 2010. Green urbanism and its application to singapore. Environment and urbanization Asia 1, 149–170.

Nijhuis, S., Van Lammeren, R., van der Hoeven, F., 2011. Exploring the visual landscape: advances in physiognomic landscape research in the Netherlands. volume 2. TU Delft.

Niu, S., Hu, A., Shen, Z., Huang, Y., Mou, Y., 2021. Measuring the built environment of green transit-oriented development: A factor-cluster analysis of rail station areas in singapore. Frontiers of Architectural Research 10, 652–668.

Oldoni, D., De Coensel, B., Bockstael, A., Boes, M., De Baets, B., Botteldooren, D., 2015. The acoustic summary as a tool for representing urban sound environments. Landscape and Urban Planning 144, 34–48.

Palliwal, A., Song, S., Tan, H.T.W., Biljecki, F., 2021. 3d city models for urban farming site identification in buildings. Computers, Environment and Urban Systems 86, 101584.

Porta, S., Renne, J.L., 2005. Linking urban design to sustainability: formal indicators of social urban sustainability field research in perth, western australia. Urban Design International 10, 51–64.

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

Qi, J., Liu, H., Liu, X., Zhang, Y., 2019. Spatiotemporal evolution analysis of time-series land use change using self-organizing map to examine the zoning and scale effects. Computers, Environment and Urban Systems 76, 11–23.

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

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

Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of computational and applied mathematics 20, 53–65.

Salamon, J., Bello, J.P., 2015. Unsupervised feature learning for urban sound classification, in: 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE. pp. 171–175.

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

Schmiedel, I., Bergmeier, E., Culmsee, H., 2015. Plant species richness patterns along a gradient of landscape modification intensity in lower saxony, germany. Landscape and urban Planning 141, 41–51.

Silver, D.A., Clark, T.N., 2016. Scenescapes. University of Chicago Press.

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

Spielman, S.E., Singleton, A., 2015. Studying neighborhoods using uncertain data from the american community survey: a contextual approach. Annals of the Association of American Geographers 105, 1003–1025.

Srivastava, S., Vargas Munoz, J.E., Lobry, S., Tuia, D., 2020. Fine-grained landuse characterization using ground-based pictures: a deep learning solution based on globally available data. International Journal of Geographical Information Science 34, 1117–1136.

Štefunková, D., Cebecauer, T., 2006. Visibility analysis as a part of landscape visual quality assessment. Ekológia (Bratislava) 25, 229–239.

Steiger, E., Resch, B., Zipf, A., 2016. Exploration of spatiotemporal and semantic clusters of twitter data using unsupervised neural networks. International Journal of Geographical Information Science 30, 1694–1716.

Sun, Y., Fan, H., Li, M., Zipf, A., 2016. Identifying the city center using human travel flows generated from location-based social networking data. Environment and Planning B: Planning and Design 43, 480–498.

Tang, J., Long, Y., 2019. Measuring visual quality of street space and its temporal variation: Methodology and its application in the hutong area in beijing. Landscape and Urban Planning 191, 103436.

Tao, H., Wang, K., Zhuo, L., Li, X., 2019. Re-examining urban region and inferring regional function based on spatial–temporal interaction. International Journal of Digital Earth 12, 293–310.

Tessler, Z.D., Vörösmarty, C.J., Grossberg, M., Gladkova, I., Aizenman, H., 2016. A global empirical typology of anthropogenic drivers of environmental change in deltas. Sustainability science 11, 525–537.

Tobler, W.R., 1970. A computer movie simulating urban growth in the detroit region. Economic geography 46, 234–240.

Tu, W., Hu, Z., Li, L., Cao, J., Jiang, J., Li, Q., Li, Q., 2018. Portraying urban functional zones by coupling remote sensing imagery and human sensing data. Remote sensing 10, 141.

Tveit, M., Ode, Å., Fry, G., 2006. Key concepts in a framework for analysing visual landscape character. Landscape research 31, 229–255.

Tveit, M.S., Sang, A.O., 2014. Landscape assessment in metropolitan areas–developing a visual indicator-based approach. SPOOL 1, 301–316.

Urban Redevelopment Authority, 2019. Singapore master plan 2019.

Wu, A.N., Biljecki, F., 2021. Roofpedia: Automatic mapping of green and solar roofs for an open roofscape registry and evaluation of urban sustainability. Landscape and Urban Planning 214, 104167.

Wu, A.N., Biljecki, F., 2022. GANmapper: geographical data translation. International Journal of Geographical Information Science 36, 1394–1422.

Yao, Y., Liang, Z., Yuan, Z., Liu, P., Bie, Y., Zhang, J., Wang, R., Wang, J., Guan, Q., 2019. A human-machine adversarial scoring framework for urban perception assessment using street-view images. International Journal of Geographical Information Science 33, 2363–2384.

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

Yap, W., Chang, J.H., Biljecki, F., 2023. Incorporating networks in semantic understanding of streetscapes: Contextualising active mobility decisions. Environment and Planning B: Urban Analytics and City Science , 239980832211388.

Ye, S., Chen, D., 2015. An unsupervised urban change detection procedure by using luminance and saturation for multispectral remotely sensed images. Photogrammetric Engineering & Remote Sensing 81, 637–645.

You, G., 2022. Spatiotemporal data-adaptive clustering algorithm: An intelligent computational technique for city big data. Annals of the American Association of Geographers 112, 602–619.

Yu, B., Shu, S., Liu, H., Song, W., Wu, J., Wang, L., Chen, Z., 2014. Object-based spatial cluster analysis of urban landscape pattern using nighttime light satellite images: A case study of china. International Journal of Geographical Information Science 28, 2328–2355.

Yu, X., Her, Y., Huo, W., Chen, G., Qi, W., 2022. Spatio-temporal monitoring of urban street-side vegetation greenery using Baidu Street View images. Urban Forestry & Urban Greening 73, 127617. doi:10.1016/j.ufug.2022.127617.

Yuen, B., 2006. Reclaiming cultural heritage in singapore. Urban Affairs Review 41, 830–854.

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

Zhang, L., Ye, Y., Zeng, W., Chiaradia, A., 2019. A systematic measurement of street quality through multi-sourced urban data: A human-oriented analysis. International journal of environmental research and public health 16, 1782.

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

Zhou, B., Liu, L., Oliva, A., Torralba, A., 2014. Recognizing city identity via attribute analysis of geo-tagged images, in: European conference on computer vision, Springer. pp. 519–534.

Zhou, H., He, S., Cai, Y., Wang, M., Su, S., 2019. Social inequalities in neighborhood visual walkability: Using street view imagery and deep learning technologies to facilitate healthy city planning. Sustainable cities and society 50, 101605.