Computers, Environment and Urban Systems 122 (2025) 102351

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

journal homepage: www.elsevier.com/locate/ceus

Urban visual uniqueness: A landmark-free framework to quantify city's identity and distinctiveness from everyday scenes

Song Guo [a,b,1], Kee Moon Jang [a,1], Fábio Duarte [a,*], Yuhao Kang [c], Carlo Ratti [a,d]

[a] Senseable City Lab, Department of Urban Studies and Planning, Massachusetts Institute of Technology, Cambridge, MA, United States
[b] School of Architecture, Tsinghua University, Beijing 100084, China
[c] GISense Lab, Department of Geography and the Environment, The University of Texas at Austin, Austin, TX, United States
[d] Politecnico di Milano, Milan, 20133, Italy

ARTICLE INFO

Keywords:
Urban uniqueness
Place identity
Place distinctiveness
Street view images
Landmark-free
Visual artificial intelligence

ABSTRACT

The visual appearance of a city is shaped by a complex interplay of factors, including cultural backgrounds, geographical features, historical developments, and policy decisions. But measuring cities' visual uniqueness remains a challenge. Previous studies often focused on iconic landmarks, neglecting everyday scenes that people are likely to encounter. By examining how and to what extent different visual patterns build up unique characteristics of cities, we propose a data-driven framework to measure visual uniqueness in terms of identity and distinctiveness. We performed bottom-up visual clustering on Google Street View (GSV) images in the six most visited Japanese cities. We found that 8 representative visual clusters explain each city's visual identity and relative distinctiveness. This research demonstrates how artificial intelligence applied to visual data can reveal subtle differences in urban environments. In the era of growing globalization, with frequent tourism and intercity visits, the cultivation of a city's unique visual characteristics can help avoid the homogenization of urban landscapes, and stimulate the development of urban tourism by shaping an imageable city.

1. Introduction

The visual appearance of a city is shaped by a complex interplay of factors, including cultural backgrounds, geographical features, historical developments, and policy decisions (Lynch, 1960; Salesses et al., 2013). It reflects not only the individuals who built and inhabit them (Rapoport, 1990) but also shapes the identities of the residents, thereby contributing to the concept of place (Duarte, 2017; Purves et al., 2019; Tuan, 1979). While serving as an external manifestation of local life, it also forms urban imageability which also allows strangers to better perceive the underlying features of a city through visual characteristics (Lynch, 1960). In the era of growing globalization, with frequent intercity visits, the cultivation of a city's unique visual characteristics can help avoid the homogenization of urban landscapes and stimulate the development of urban tourism. Therefore, analyzing and measuring the visual uniqueness of cities have been crucial in providing valuable references for urban planning, architectural design, and tourism industry strategies (Jarratt et al., 2019).

For the identification of city visual identities, previous research has utilized computer vision methods to recognize urban visual elements. For this purpose, classification tasks using Deep Convolutional Neural Networks have been mainly performed in order to map the relationship between architectural landmarks or detailed features with their corresponding cities (Doersch et al., 2015; Goel et al., 2012; Lee et al., 2015). Notably, Zhou et al. (2014) and Zhang et al. (2019) conducted multi-city comparisons using social media images to derive city-informative features and measure the visual similarity between different cities. However, earlier studies have often sought unique urban visual qualities from cases of global cities, with less attention given to comparison within the same national context (F. Zhang et al., 2019; Zhao et al., 2021). Given that geographical and cultural commonalities converge to greater visual similarities, and vice versa, representative cities of different countries tend to exhibit more pronounced distinctions in visual landscapes. In this regard, findings have gravitated around 'public symbols' that are easily identifiable in global cities (e.g, monumental buildings, religious spaces, public squares), while overshadowing places as 'fields of care' that are construed through repeated experiences in an everyday setting (Tuan, 1979; Wild, 1965).

* Corresponding author at: Senseable City Lab, MIT 9-216, 77 Massachusetts Avenue, Cambridge, MA 02139, USA.
E-mail address: fduarte@mit.edu (F. Duarte).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.compenvurbsys.2025.102351
Received 21 October 2024; Received in revised form 23 August 2025; Accepted 28 August 2025

Despite the low imageability, Tuan (1979) assumed that these day-to-day places (e.g., home, park, drugstore) project a visual identity primarily to local residents, whereas being inconspicuous and hardly discernible to outsiders in their physical form or appearances. Yet, it is these low-level visual features, or "many little things" as Jacobs puts it (Jacobs, 1958, p. 129), which help build a sense of place of a city (Schertz et al., 2018). However, this distinction is blurring in the context of globalization. Although local residents and visitors had traditionally been considered separate groups in urban tourism with needs of understanding urban identity from both perspectives (Bernardo et al., 2017; Palmer et al., 2013), the increased mobility of people has enabled tourists to repeatedly visit same locations and form an image of a place that are in common with the locals (Collins-Kreiner, 2020; Lim & Bouchon, 2017). This has led to the emergence of new visitor experiences to "live like a local", as advertised by home-sharing platforms such as Airbnb (Gurran, 2018). These new kinds of spaces can provide a novel setting to explore the visual uniqueness of cities through the subtle but distinguishable semantic cues in the urban landscapes (Mazzarello et al., 2024).

Among many works that established the theoretical grounds of place research, *The Image of the City* by Kevin Lynch (1960) provided a foundational taxonomy to explain how urban imageability is cognitively structured through five common elements, including city landmarks. While this framework identified urban elements that let people cognitively map cities, our study diverges from this perspective directed at legibility or wayfinding. Our motivation aligns more closely with and is guided by Tuan's (1979) and Jacobs' (1958) perspectives, which emphasized that sense of place arises not only from iconic features, but from the (often overlooked) 'little things' that are repeatedly encountered in daily life. To this end, this study aims to measure the uniqueness of a city at the compositional level of everyday scenes captured by street view images using urban visual intelligence (Zhang, Li, & Zhang, 2024). Specifically, this paper addresses two central research questions: First, can urban visual identity and distinctiveness be quantified without relying on city landmarks? Second, how do streetscape visual patterns, independent of landmarks, differentiate cities within the same cultural setting? In order to answer these research questions, six most visited cities of Japan were selected as a case study to explore their uniqueness whilst the homogeneity in cultural and historical backgrounds within Japan (Burgess, 2004). In particular, we identified the overlapping areas of two accommodation services, Airbnb and Booking.com, to capture street-level scenes perceived by both residents and visitors. Unsupervised clustering performed on feature representations of Google Street View (GSV) images extracted from a pre-trained deep learning model resulted in 8 clusters of visual characters for the studied Japanese cities. Then, we quantified the city's visual uniqueness in terms of two aspects, identity and distinctiveness, by comparing the proportion of different clusters possessed by each city and the deviations to their counterparts, respectively. Finally, a classifier trained on the clustering process was used to predict the cluster label for urban scenes of Japan and verify whether the visual uniqueness is consistent both within and across the most visited areas of each city. This study contributes a landmark-free framework for evaluating visual uniqueness in cities using unsupervised clustering of everyday street-level imagery. This provides empirical insights into how cities differentiate themselves through subtle scene-level features, rather than iconic landmarks, and proposes a method for assessing city-level identity and distinctiveness to capture nuanced forms of urban visual uniqueness.

2. Literature review

2.1. Urban visual uniqueness, identity, and distinctiveness

Urban visual uniqueness usually refers to the distinctive visual characteristics that make a city stand out from others. From existing studies, the concept of uniqueness is intrinsically linked to two related concepts, identity and distinctiveness (Ginting et al., 2017; Peng et al., 2020). Place identity has been an important topic in urban studies, on the one hand serving as an integral component of residents' self-identity that can be perceived through the structural elements of a place, whereas on the other, reflecting the internally coherent visual character of a city that makes it recognizable (Lynch, 1960; Proshansky et al., 2014; Twigger-Ross & Uzzell, 1996). In this sense, the identity of a place can be understood as recurring features that collectively distinguish itself from others (Relph, 1976). Although commonly considered as interconnected with place identity, distinctiveness can specifically refer to the degree to which the identity of a city differs from those of other cities (Phetsuriya & Heath, 2021). While identity illustrates a city's own coherence and internal composition, distinctiveness is a comparative notion which emerges based on the relational contrast between two or more different cities.

Traditional methods to measure place identity and distinctiveness typically involve qualitative approaches such as surveys, interviews, observations, and cognitive mapping (Lynch, 1960). However, they are often time-consuming and labor-intensive, which have limited the quantity and representativeness of samples (Y. Kang et al., 2023). Despite their effectiveness in offering researchers detailed insights into how individuals perceive and relate to specific places, it has been challenging to convincingly depict a city's collective image through conventional approaches (Jang & Kim, 2019). In addition, visual elements that contribute to the uniqueness of cities have been considered to a lesser extent.

The advent of crowdsourced digital data has revolutionized how researchers can gain a comprehensive understanding of places through visual means (Miranda et al., 2020; Zhou et al., 2014). Large-scale geotagged photos and social media texts have emerged to discern patterns and commonalities in people's experience and perception of urban spaces. For instance, Zhou et al. (2014) recognized city identity based on attributes detected from millions of geotagged images gathered from Panoramio, and Doersch et al. (2015) attempted to identify geographically informative elements and their particular styles of a given city from GSV images. In a broader sense, Zhang, Salazar-Miranda, et al. (2024) conceptualized an 'urban visual intelligence' framework to illustrate how artificial intelligence applied to visual data can help enhance the interpretation of cities. By using these data as input for generative artificial intelligence, it is also possible to visualize people's perceptions of a place (Jang et al., 2024).

To quantify the distinctiveness in the characteristics of cities, a common approach has been to measure the similarity between different cities through misclassification rates of a deep learning model (F. Wang et al., 2022; F. Zhang et al., 2019; Zhao et al., 2021). However, this approach only allows for a pairwise comparison, restricting the ability to directly measure multi-city differences. Furthermore, due to the bias in social media photos, this often highlights popular landmarks and tourist attractions, limiting the ability to capture everyday scenes that contribute to the city's identity (Schertz et al., 2018).

2.2. Feature extraction and clustering methods in urban studies

Since the number and nature of a city's characteristics are unknown, a bottom-up approach is more helpful in discovering patterns in everyday scenes compared to studying according to predetermined features. Unsupervised learning methods, such as clustering, can be used to uncover intrinsic structures within the data. Images can also serve as promising inputs for clustering, after feature extraction through deep learning models. Convolutional neural networks have been widely used to extract scene-level features from street view imagery. For example, ResNet18 has been employed to extract patterns of street view images (Fujiwara et al., 2024). When pretrained on place-related datasets such as Places365 (Zhou, Lapedriza, et al., 2017), it has demonstrated strong performance and high computational efficiency in scene-related tasks across multiple disciplines (M. Chen et al., 2024; Jiao & Wang, 2023).

More recently, contrastive learning methods and related models such as Contrastive Language-Image Pre-training (CLIP) (Radford et al., 2021) and OpenCLIP (Cherti et al., 2023) have been applied to infer the semantic information of urban imagery, including location, land use, and functions (Huang et al., 2024; Xu et al., 2024; Yan et al., 2024) for their capacity in extracting semantically aligned image features. In addition, self-supervised models such as DINOv2 have shown potential in learning general-purpose visual features when pretrained on a large quantity of curated data (Oquab et al., 2023), making it highly promising for extracting scene-related features in urban contexts.

The features of street view images derived, along with the spatial adjacency of image locations, can be jointly input into the K-means algorithm to discover potential patterns in urban landscapes (Liang et al., 2023). In addition to streetscapes, other representations of urban forms (N. Li & Quan, 2024), such as building figure-ground maps (J. Wang et al., 2024), can also be represented as vectors and then input into different clustering algorithms to discover urban morphological patterns.

These studies highlight the potential of unsupervised clustering to reveal patterns. However, current studies often lack comprehensive explanations of the clusters, with insufficient analysis of the resulting patterns. Merely naming clusters based on discovered patterns does not provide a deep understanding of the phenomena. Additionally, there is also a lack of exploration regarding the features that differentiate the clusters.

To transform clustering results into meaningful explanations and measurements, addressing differentiating features of the clusters is needed. In this paper, we integrate the random forest algorithm to identify the distinguishing features of clusters, which helps us better understand the essential differences between visual patterns. Furthermore, we utilize the clustering results for the computing of identity and distinctiveness, and further define the uniqueness index. This approach leverages the ability of clustering methods to discover and present potential patterns while also allowing for the quantitative measurement of the results.

3. Materials and methods

3.1. Overview

By examining how and to what extent different visual patterns build up unique characters of cities, we propose a data-driven framework to measure urban uniqueness in terms of identity and distinctiveness. As illustrated in Fig. 1, the overall framework consists of four steps: (a) research area identification and data preprocessing, (b) feature representation and clustering analysis, (c) cluster interpretation, and (d) urban visual uniqueness measurement. Part (a) determines the main study areas and performs necessary preprocessing on the collected GSV images. In Part (b), these images are represented as feature vectors that are used for clustering to discover visual patterns within the study areas. Part (c) and (d) identify visual attributes that characterize the cluster outputs and calculate the identity and distinctiveness of each city based on the frequency distribution of visual clusters. This framework provides a bottom-up representation of a city's unique visual profile using urban visual intelligence approach on street-level imagery.

3.2. Research area identification

As a geographically isolated environment, Japan has preserved a homogeneous cultural and historical background, letting its cities develop a similar sense of identity (Burgess, 2004). Meanwhile, their locational differences have resulted in varied urban environments, ranging from the bustling metropolis of Tokyo to the hometown of the Japanese heart, Kyoto, and the nation's kitchen, Osaka. In this regard, Japan offers a challenging yet meaningful context for examining a city's uniqueness, not merely from global cities and their landmark features but also from cities with shared but distinct visual qualities. In particular, this work takes six cities, Fukuoka, Kyoto, Nagoya, Osaka, Sapporo, and Tokyo, as the study sites. The six cases were selected based on the highest number of foreigner entries (by port of entry, updated in January 2024) as according to the latest data provided by the Japan National Tourism Organization.[2] Also, their locations in different parts of the country make it suitable for comparing city uniqueness across various regional settings.

We considered the administrative boundaries of cities instead of the entire prefectures in order to explore the urban visual profile at areas that make up the core of the city.[3] Subsequently, we delineate areas where local and non-local experiences mix by overlapping urban zones covered by hotels and short-term rental listings on Booking.com and Airbnb, respectively. With growing forms of repeated visitation and residential tourism, central areas defined by this overlap let us identify shared experiential spaces that are frequently encountered by both residents and tourists. While these are likely to be construed as tourism spaces located near city centers, Valentin and O'Neill (2019) and M. Li et al. (2015) found empirical evidence that accessibility to tourism attractions is not necessarily associated with hotel locations. Moreover, short-term rentals not only provide indoor living spaces to satisfy travelers' desire for authentic travel experiences (Mazzarello et al., 2024), but also stage local outdoor scenes in close proximity that fosters a stronger sense of attachment to the place they stay (Ding et al., 2023; González-Reverté et al., 2025; Guttentag, 2015). Street environments in zones frequented by visitors can still preserve local identity and distinctiveness by showcasing diverse aspects that are crucial in shaping unique characters of a city (Patandianan & Shibusawa, 2020).

In this sense, streetscapes near these accommodation services offer a unique setting to study the visual uniqueness of the ordinary public realm, beyond traditional landmarks (Madeira et al., 2023; Rodrigues et al., 2024). We identified agglomerations of both Booking.com and Airbnb listings by obtaining the convex hull of their DBSCAN clusters (Gao et al., 2017; Hu et al., 2015). 40 listings within 1 km radius were considered the threshold for determining the clusters based on the current number of accommodation facilities in most research cities[4] and manual experiments. Their overlapping areas (hereafter, inner city) in the six cities were selected as main study areas (see Fig. 2).

In the meantime, we distinguish these areas with the whole city areas to reflect coherent or differing patterns of visual experience across the urban fabric. While the selected inner city areas offer a good case to explore urban identity likely to be stronger and shared between residents and non-residents (Manahasa et al., 2025), there is also a need for an inter-scalar research that incorporates smaller neighborhoods or rural areas (Belanche et al., 2021; Sadeque et al., 2020). In this sense, we include the six cities' administrative boundaries (hereafter, whole city area; see Supplementary Materials Fig. S1) and verify whether localized uniqueness resonates across the city or remains confined to specific areas, addressing the challenge of spatial bias.

3.3. Data collection and preprocessing

In this study, we use Google Street View (GSV) images to obtain street-level scenes of the six cities. Using the Street View Download 360 software (https://svd360.istreetview.com/), we downloaded 360°

[2] https://statistics.jnto.go.jp/en/graph/#graph–foreigners–by–port–of–entry–by–country. The top-ranked airports in the table and their nearby cities are Narita (Tokyo), Kansai (Osaka), Haneda (Tokyo), Fukuoka (Fukuoka), Chubu (Nagoya), and Shinchitose (Sapporo). Kyoto is not on the list because it does not have an airport, but considering its high visitor rate, we included Kyoto as one of our research cities.

[3] For Tokyo, we considered the 23 special wards.

[4] https://www.fukushi.metro.tokyo.lg.jp/kiban/chosa_tokei/nenpou/2022.html

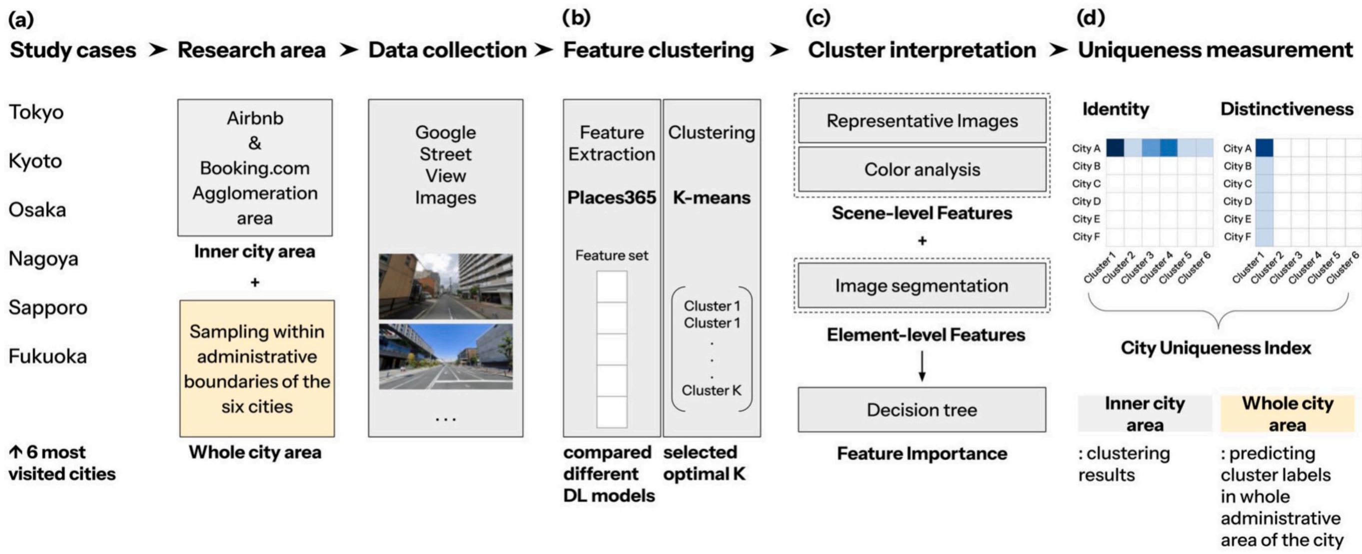

Fig. 1. The framework for measuring urban visual uniqueness in this study.

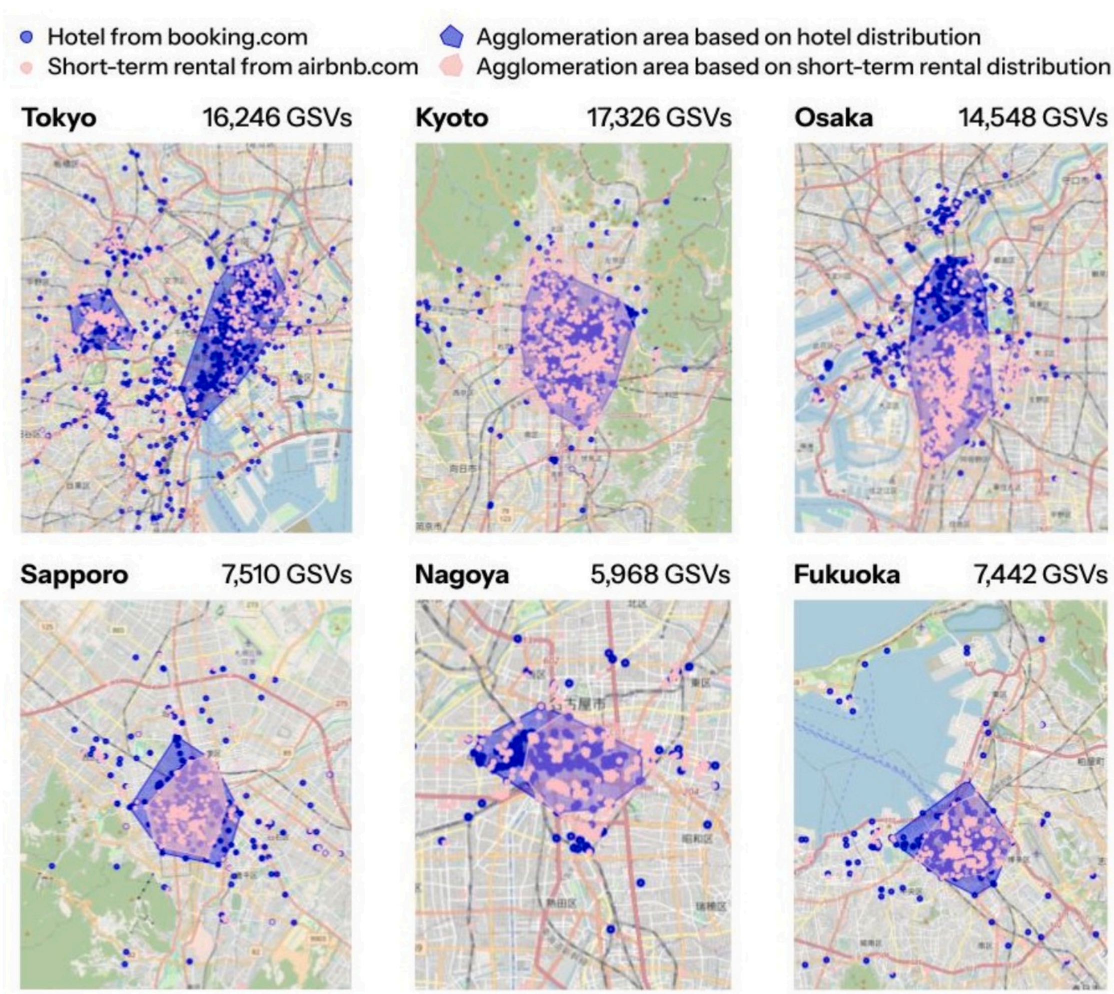

Fig. 2. The inner city areas in this study.

panorama GSV images along with their associated metadata at a 50-m interval and a resolution of 3328 × 1664 pixels. The collected images were then converted into four regular perspective images, from which we selected the front and rear views at 80° pitch angle that capture the street view as perceived by people. A total of nearly 70,000 GSVs were collected in the inner city area. Also, in order to verify the coherence of visual uniqueness across the whole city, we obtained another set of GSV images (N = 22,460) within the cities' administrative boundaries at a 500-m interval (see Supplementary Materials Table S1). It is noteworthy to mention that different sampling intervals were selected for the inner and whole city areas. In specific, we adopted a finer 50-m interval for central zones where visual elements are likely to vary over short distances, whereas a 500-m across the whole city areas to include peripheral regions that are generally more homogeneous and thus provide an overall perspective which extends the city's core visual uniqueness pattern to its broader urban context. This strategy balances data representativeness and redundancy, while accounting for variations in urban morphology at different areas of cities.

To ensure consistency in image quality and relevance of the dataset, a thorough manual filtering process was conducted. Images that were taken indoors (within train stations or under bridges), overexposed or blurred were considered erroneous and removed from analysis. In addition, the seasonal distribution of the collected GSV imagery (see Supplementary Materials Fig. S2) shows that the majority of images were captured between May and October, with a significantly lower number of images captured during the winter months (e.g., 788 in December and 664 in January). This concentration in warmer seasons can help reduce the impacts of seasonal outliers such as snow coverage or bare trees. Yet, considering this minimal representation of extreme winter conditions in the collected dataset, we do not remove images from these months as seasonal variations are to a lesser extent a factor of bias in urban visual uniqueness than street-level green view or exposure assessment. Table 1 summarizes the number of collected data in each city.

4. Discovering urban visual patterns of Japan

4.1. Feature representation and unsupervised clustering analysis

In order to identify visual patterns within the selected Japanese cities, we extracted place-related features from the collected GSV images. Deep learning (DL) models have recently improved visual place recognition tasks by enabling low-level representation of raw images (Wu et al., 2024). Furthermore, the introduction of large-scale datasets labeled with scene categories, such as Places365 (Zhou, Zhao, et al., 2017), have extended the applicability of DL-based scene representation of various urban settings.

For this study, we extracted visual features using ResNet18 pretrained on Places365, yielding a 512-dimensional embedding of a GSV image used for K-means clustering (Jain & Dubes, 1988) to categorize urban street scenes into distinct typologies. The feature extractor was selected for its strong task alignment with scene classification, compact feature dimensionality and stable performance (He et al., 2016).[5] The quality of clustering results was tested for several evaluation metrics.

Table 1
The number of Booking.com and Airbnb listings considered in this study.

City	Booking.com	Airbnb	GSV (inner city study area)	GSV (whole city area)
Tokyo	915	1517	16,246	4646
Kyoto	921	1154	17,326	5294
Osaka	876	1147	14,548	1616
Sapporo	630	530	7510	5838
Fukuoka	614	483	7442	2486
Nagoya	311	360	5968	2580

[5] We also tested DINOv2, OpenCLIP, and ResNet50 pretrained on Places365 as feature extractors and confirmed the qualitative results for clustering (see Supplementary Materials Figure S4).

First, we searched whether the within-cluster sum of squares (WCSS) form a significant bend when plotted against different numbers of clusters. This 'elbow' point indicates that an additional cluster will not significantly improve the clustering performance, thereby denoting the optimal number of clusters, K, within the dataset. In addition, we checked the Davies-Bouldin Index (DBI) that is measured as the ratio of within-cluster similarity to between-cluster similarity, with lower values indicating better cluster separation (Davies & Bouldin, 1979). Considering both evaluation metrics and diverse urban patterns in the six Japanese cities, ResNet18 trained on Places 365 as a feature extractor yields a distinct set of 8 clusters. It is worth noting that the Silhouette score (Rousseeuw, 1987) was also tested, although no significant optimal number of clusters was found (see Supplementary Materials Fig. S5(a)). In order to evaluate the clustering performance, silhouette score focuses on assessing how well individual points fit within a cluster, whereas DBI takes a more global approach by averaging intra-cluster similarity and inter-cluster separation (Halkidi et al., 2001; Tan et al., 2016). By doing so, DBI can help to inform the cluster number for K-means-based analysis with explained variance, less sensitivity to cluster shape, and demonstrated resilience in high-dimensional data (Liu et al., 2013; Singh et al., 2020; Thrun, 2021). Also, DBI identified the optimal number of clusters in accordance with WCSS, for which we choose DBI over Silhouette to be a more reliable metric for the interpretation of subtle differences of street scenes in this study.[6] Fig. 3 shows the selection criteria for determining the optimal number of clusters, K = 8, as well as a t-SNE visualization of the resulting clusters.

4.2. Interpretation of clustering results

As a result of the embedding based clustering on GSV images, we identified eight clusters that represent the typical streetscapes of the six Japanese cities considered in this study. First, we conducted qualitative evaluations to interpret distinct characteristics of each cluster. In order to gain an intuitive understanding of the clusters, we retrieved the nearest images to cluster centroids in the feature space. These are assumed to be indicative of the prominent visual pattern of each cluster. We also plotted the spatial distribution maps of the clusters to visually inspect where each cluster predominantly occurs and verify whether they align with the existing urban context.

Meanwhile, we incorporated quantitative assessments aimed at identifying both *element* and *scene* level observations (Y. Kang et al., 2020) that distinguish one cluster from another. For *element level* observation, semantic segmentation was performed using the Dense Prediction Transformer (DPT) model (Ranftl et al., 2021) trained on the ADE20K dataset (Zhou, Zhao, et al., 2017). Among the 150 object classes extracted from street view images, we aggregate some of the original classes to construct new categories used for analysis (Appendix Table A1). In addition, we constructed a measure of building height-to-street width ratio approximated by the proportion of pixels segmented as 'building' and 'road' classes as a proxy for the sense of enclosure.

Scene level observations, such as color, texture or shape, are another important factor that contain rich semantic information and contribute to visual perception (S. Chen & Biljecki, 2023; L. Ma et al., 2023; Rossetti et al., 2019). We obtained the ten most dominant hue, saturation and value (HSV) of pixels per GSV image, which were used to derive the ten most frequent colors from each cluster. These sets of colors and their variety are crucial for the quality of streetscape design (Fukahori & Kubota, 2003), and are expected to represent scene level visual patterns of the clusters. Note that colors with low saturation (<10) and value (<20) were excluded assuming that urban visual perception is less

[6] The evaluation of clustering results was also conducted based on 90-dimensional PCA-reduced-features, which retains 90 % of the explained variance and showed consistent results with the representative images of the clusters (see Supplementary Materials Figure S5(b)-(c)).

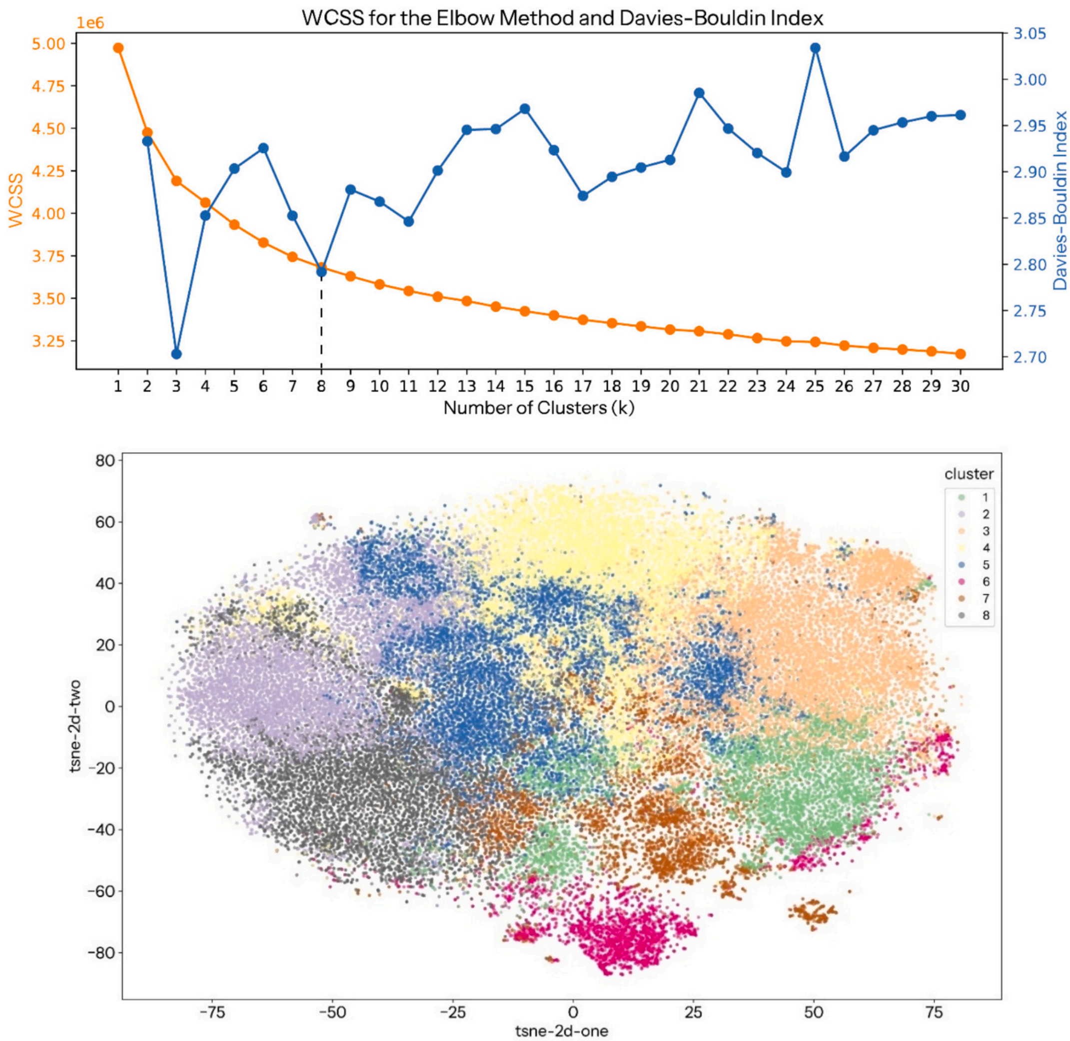

Fig. 3. Determining the optimal number of clusters and a t-SNE plot of the resulting clusters.

driven by a relatively monochrome color.

Therefore, urban visual cluster results are summarized in Fig. 4, and their spatial distributions are displayed in Appendix Fig. A1. Detailed descriptions of each cluster are the following, which illustrate distinct typologies of street environments that characterize the eight representative urban visual scenes in the major Japanese cities.

- **Cluster 1** exhibits a low building height-to-street width ratio along with a high proportion of sky as its main visual characteristics. These collectively contribute to a weak sense of enclosure and suggest an open urban environment. In addition, the color composition is mainly dominated by blue and gray tones, reflecting the strong presence of road pavements. This cluster represents large boulevards (or *Odori*) that carry heavy traffic, predominantly distributed along the major arterial roads or run parallel to Japan Rail (JR) lines in the study area.
- **Cluster 2** captures scenes of *Machiya* districts, where traditional wooden townhouses are organized on the sides of a narrow street. In contrast to Cluster 1, the built environment of *Machiya* neighborhoods is characterized by a higher building height-to-street width ratio and lower sky view factor, resulting in a stronger sense of enclosure. Despite the noticeable absence of greenery, streets are filled with warm tones of color attributing to natural materials used in the traditional houses.
- **Clusters 3, 4 and 5** represent various types of street environments in urban built-up areas. While they are similar with a balanced composition in their visual features, we observed differences that distinguish between the three clusters. For instance, Cluster 3 presents higher building heights and wider street widths than its counterparts, although their ratio falls in a comparable range, indicating a dense urban core. The distinction is far more subtle for Clusters 4 and 5. However, a qualitative evaluation of representative images demonstrated differences in building styles, where triplex apartments and flat-type apartments (Shin et al., 2011) were dominant in each cluster, respectively. In addition, the presence of signages on the ground floor implies a mixed-use function in Cluster 4, whereas a prevalence of mid-rise apartments in Cluster 5 indicates a medium-density residential area.
- **Cluster 6** is characterized by exceptionally high levels of green elements, along with the absence of buildings. The color variations also support this finding through various shades of green, reflecting abundant natural elements being captured in the example images. Accordingly, the geographic locations strongly verify that this cluster represents urban green spaces, where GSVs identified as Cluster 6 are observed in city parks or temple areas.
- **Clusters 7 and 8** are scenes of urban infrastructure built for specific functions. Cluster 7 typically appears along the riverfront of Japanese cities, noticeably in Tokyo, Kyoto and Sapporo. Notable visual features of this cluster include the highest level of open sky view and

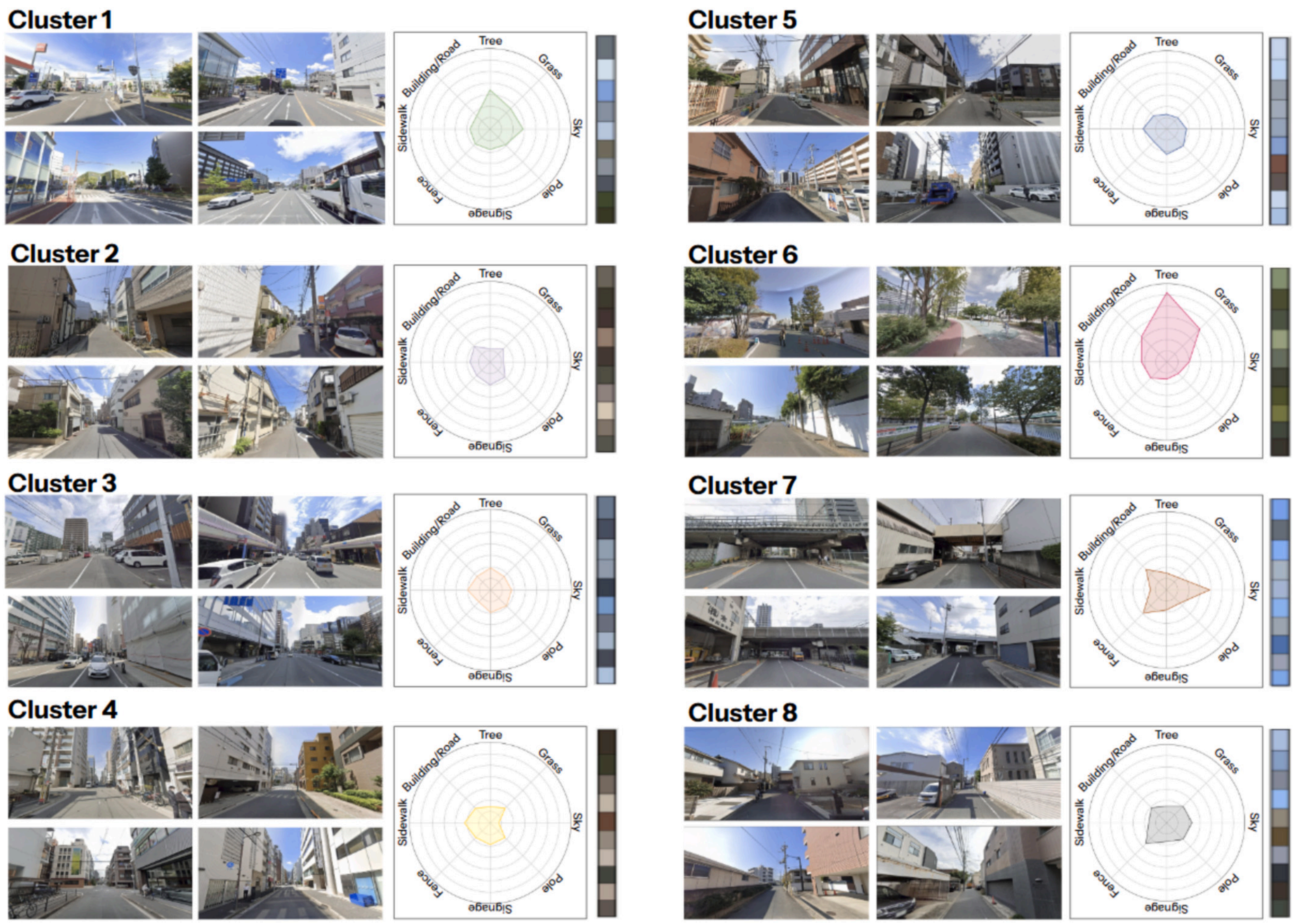

Fig. 4. Urban visual cluster results featuring representative examples, radar chart of mean proportions of selected segmented objects, and the ten most frequent colors of each cluster.

fences, while the least amount of sidewalks or green elements, which illustrate visual characteristics of urban infrastructure such as bridges or overpasses. Cluster 8 displays visual scenes of off-street parking lots located on the periphery of the study areas. This corroborates the history of parking in Japan, which shifted toward unmanned coin-parking facilities in response to the sharp increase in on-street parking (Axhausen et al., 2015; Seya et al., 2020).

These descriptions provide a detailed scientific characterization of each cluster, highlighting their spatial distributions, visual attributes, color compositions, and the implications of these features on the urban landscape. While some clusters contain familiar street-level elements (e.g., sky, grass, road) and resemble common visual scenes of urban environments, several offer meaningful domain insights by revealing localized patterns grounded in local planning history. For instance, Cluster 2 captures districts of *Machiya* houses, which represent a culturally embedded form of architectural practice and traditional urbanism. Identification of this cluster implies how such historical style of buildings remain visually distinct to construct the identity of Japan's streetscapes. Another is Cluster 8, which prominently features unmanned coin-parking lots created by Japan's policy-driven shift toward off-street parking. This particular finding underscores the strength of our landmark-free approach in revealing visually ordinary yet contextually significant patterns of urban visual uniqueness. These clusters embody the compositional qualities of street scenes without isolating iconic or other individual features, reinforcing our emphasis on the cumulative contribution of everyday urban elements to shaping a city's distinctive visual identity.

4.3. Random forest for differentiating clusters

In order to validate the urban visual cluster results and identify the specific features that may have contributed to the formation of clusters, we employed decision tree and random forest algorithms that have been effective in evaluating the feature importance for classification processes (Archer & Kimes, 2008; F. Zhang & Yang, 2020). Cluster labels were set as the target variable and the input dataset consisted of 25 features: 22 derived from semantic segmentation and 3 from color analysis (HSV). Based on model precision, the random forest classifier demonstrated higher accuracy and was selected as the final model for analysis.

We posited that the prediction accuracy serves as an indicator of how well GSV images were grouped into different clusters—higher precision suggests that a cluster was made easily identifiable and higher recall suggests that a cluster label was classified correctly based on the 25 selected features. As shown in Fig. 5(a), the model revealed that Cluster 6 had both the highest precision (0.80) and recall (0.73), indicating urban green spaces are the most distinguishing visual pattern compared to its counterparts. In contrast, Cluster 5 ("medium-density neighborhood") exhibited the lowest precision (0.47) and recall (0.45), suggesting that street-level scenes in medium-density neighborhoods are less likely to be predicted correctly.

Furthermore, feature importance analysis identified the building height-to-street width ratio and sky view factor as the most and third

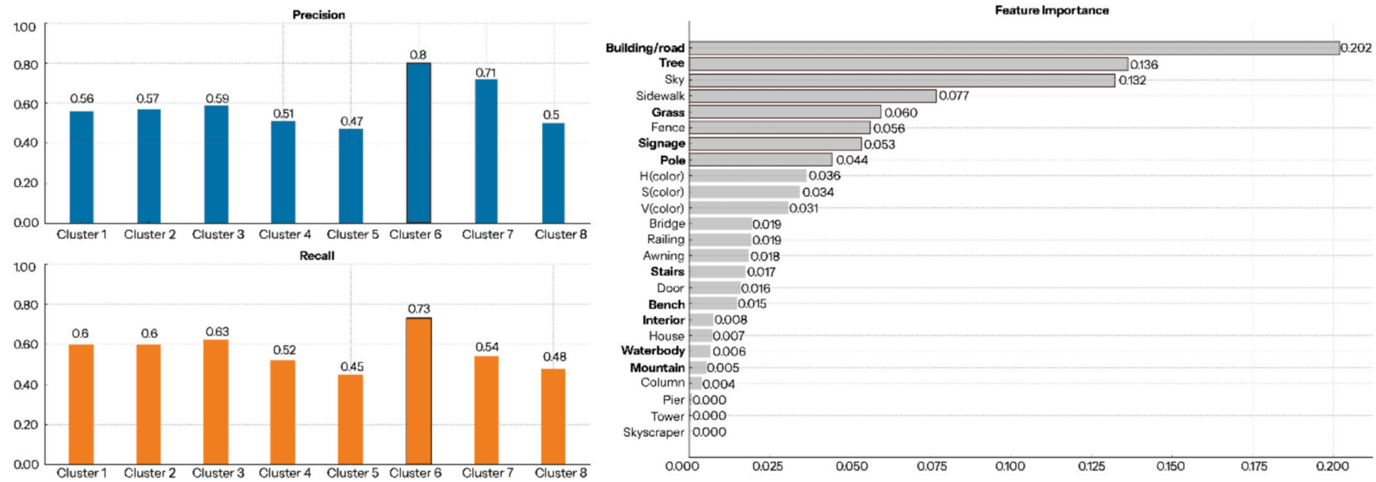

Fig. 5. Performance of the random forest model. Precision and recall (left) and feature importance (right).

most significant features, respectively, highlighting the role of spatial enclosure in distinguishing between open and densely built-up areas. The total amount of tree canopy was also found crucial for differentiating the clusters. Additionally, we also revealed that street design elements such as sidewalks, fences and poles, despite their small proportions, were among the significant contributors to visual cluster predictions in Japanese cities. These elements are often associated with a city's historical development or specific policies, underscoring their value as place-informative components in urban scene classification.

5. Urban visual uniqueness

5.1. Redefining city's visual uniqueness: Identity and distinctiveness

Eight visual characters that represent typical street scenes present in visitor-oriented areas of the studied Japanese cities were identified in the previous section. Different cities, however, may possess distinct patterns in the types and quantities of visual clusters, conveying a unique sense of place. In this section, we redefine a city's visual uniqueness in terms of two aspects: identity and distinctiveness.

City's visual identity refers to the unique composition of different visual clusters. We calculated the proportion of GSV images assigned to each cluster for each city, where the number of images of a particular cluster was divided by the total number of images collected for that city. A high proportion in a specific cluster suggests that the visual character represented by that cluster strongly defines the city's identity, and vice versa.

But a city's uniqueness should be understood not only within its own identity but also in comparison to other cities. City' visual distinctiveness refers to how much the composition of visual clusters differs from other counterparts. Not to be biased by the different number of samples collected from each city, we assessed distinctiveness by normalizing each city's cluster proportions (instead of an exact number of GSV images) by the sum of proportions in each cluster across all cities. This highlights whether a specific visual pattern is more distinctive to a certain city, indicated by a high normalized proportion.

According to Fig. 6, we can observe the overall results on the visual identity and distinctiveness for the main research areas in the six cities. First, as illustrated in Fig. 6(a), each row presents the identity of each city, revealing dominant visual patterns where higher proportions in specific clusters suggest what visual scene strongly characterizes the city's visual identity. Kyoto was mostly represented by Cluster 2 (25.65 %), which corresponds with the abundance of traditional architecture in its urban landscape. Also, Cluster 8 was found as the second most (21.96 %) visual scenes of Kyoto, indicating the city's effort to abide by the Parking Law and preserve the historical centers by providing sufficient off-street parking spaces. Tokyo's identity was predominantly represented by Clusters 2, 3, 4 and 5, likely reflecting various forms of urban

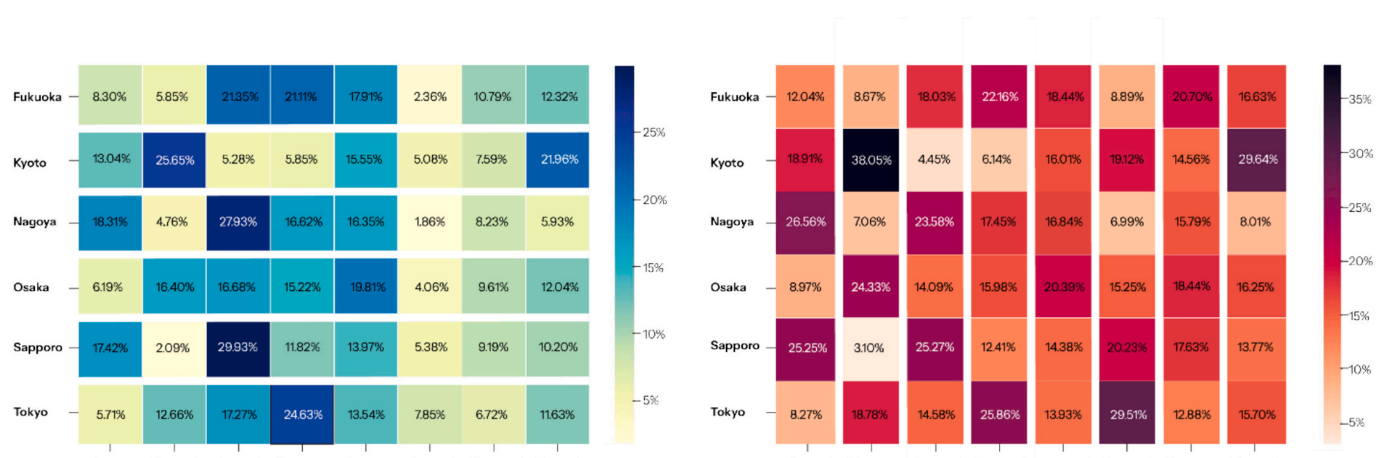

Fig. 6. The calculation and visualization of visual identity and distinctiveness of the inner city area in six Japanese cities. Identity (left) and Distinctiveness (right) heatmaps for main research areas.

built environments, among which mixed-use neighborhoods (Cluster 4) constituted the largest portion (24.63 %) of its modern dense cityscape. Meanwhile, we observed similarities in visual identities of Nagoya and Sapporo with strong representations in Cluster 3 (27.93 % and 29.93 %, respectively) along with Clusters 1, 4 and 5, whilst the absence of Cluster 2. This pattern shows that visual identities of both cities are shaped primarily by scenes with wide streets and a low sense of enclosure. Unlike other cases, Fukuoka and Osaka exhibited a relatively even distribution of visual components across multiple clusters, with only difference in their proportions of Cluster 2.

Fig. 6(b) displays the normalized proportions of images labeled as a certain cluster attributed to each city. This highlights whether a city has a higher share of specific visual patterns that make them distinctive relative to other cities for a certain visual scene. In general, we found that clusters that typified each city's identity (see Fig. 6(a)) were also prominent for their distinctiveness. A great portion of Cluster 2 (38.05 %) and Cluster 8 (29.64 %) were present in Kyoto. Also, Nagoya and Sapporo were highly distinctive in terms of Clusters 1 and 3, which were also among major components of their visual identities. On the other hand, we also observed discrepancies between identity and distinctiveness features. For instance, Tokyo was notably distinctive for Cluster 6 (29.51 %), indicating that green spaces significantly contribute to the city's visual uniqueness although they make up only a small fraction (7.85 %) of Tokyo's visual identity in the study area. Besides, nearly a quarter of (24.33 %) visual scenes represented by Cluster 2 were found in Osaka. Despite accounting for only 16.4 % of the city's visual identity, the traditional-style townhouses, or Machiyas, also serve as a distinctive feature for Osaka's uniqueness. We also noted that Clusters 5 and 7 are not distinctive scenes of any city, but rather evenly distributed across Japan.

5.2. Comparison with visual uniqueness in the whole city area

By considering both identity and distinctiveness, we identified the visual patterns that make each city unique based on ordinary street environments. This methodology advances the quantification of urban uniqueness by focusing on everyday scenes rather than prominent landmarks. It is worth noting, however, that the inner city area was derived based on agglomeration areas of accommodation services where non-local visitors are more likely to encounter. In this regard, we verified whether the identified visual uniqueness of cities is consistent across the whole city. As illustrated in Fig. 7(a), the urban visual clustering model from previous sections was used to assign one of the eight cluster labels to an independent set of GSV images collected within the cities' administrative boundaries, based on the distance in latent space between their feature representations and the centroids of the predefined clusters. This enables an exploration of urban visual uniqueness at intracity levels, offering insights into the generalizability of results.

As a result, Fig. 7(b) and (c) display the visual identity and distinctiveness in other parts of cities, respectively, from which we observed contrasting patterns to the inner city areas in Fig. 7. The most apparent is a weaker emphasis on Clusters 1 to 5, that represent various forms of urban built environments, but rather a stronger presence of Clusters 6 to 8. On the one hand, urban green spaces (Cluster 6) were found as the most prominent component for visual identities of Fukuoka (35.44 %), Kyoto (63.11 %) and Sapporo (34.46 %), with an exceptionally high proportion of Kyoto's urban landscape. On the other, Nagoya, Osaka, and Tokyo (23 districts) presented higher levels of functional features (e.g., overpass, off-street parking lot) over natural elements. Sapporo exhibits another unique finding in its richness of Cluster 1 (35.7 %). This corresponds with the particular context of Sapporo, whose post-war development has built a modernist city primarily with a grid-pattern layout in the city center along with urban corridors (e.g., highway, major road) in the outer edges for an effective transportation infrastructure (Hoshi, 2023).

Fig. 7(c) also revealed interesting findings regarding the distinctiveness of each city. Overall, more than half of the highly urbanized scenes (i.e., Clusters 2 to 5) were distributed in the entire city of Osaka and Tokyo, indicating that densely built environments distinguish them from the other four cases. Intriguingly, 43.62 % of Cluster 2, or Machiyas, were found in Osaka, followed by 30.34 % in Tokyo. This shows that while traditional-style detached houses were seemingly best-preserved and promoted in tourism hotspots of Kyoto (see Fig. 6), they are also likely to be scattered around cities of Osaka and Tokyo enabling

Fig. 7. (a) The approach for whole city area uniqueness analysis. (b) Identity (left) and Distinctiveness (right) heatmaps for whole city areas.

visitors to view and experience the cultural scenes of Japan. Meanwhile, Kyoto becomes more distinctive particularly in terms of Cluster 6 (40.11 %), suggesting its topographic condition in which the city is surrounded by mountains that provide rich green spaces.

5.3. City uniqueness index

Last, in order to quantify urban uniqueness through their visual scenes, we developed a *City Uniqueness Index* that combines both city's identity and distinctiveness. This is operationalized following the theoretical framing in Section 2.1, where identity refers to a city's unique character defined by its internal composition of visual patterns and distinctiveness indicates how dominantly each visual pattern appears in one city relative to others. Inspired by earlier studies (Pang et al., 2019; Taeuscher et al., 2022), we used *Z*-score as a measure of visual distinctiveness, which computes the deviation from the distribution mean of proportions of a cluster across the six cities. We referred to this as the 'adjusted distinctiveness' of city *i* in terms of cluster *j*, which is calculated as follows:

$$AdjustedDistinctiveness_{ij} = \left(P_{ij} - \mu_j\right) \Big/ \sigma_j$$

$$i \in \{Fukuoka, Kyoto, Nagoya, Osaka, Sapporo, Tokyo\}, j \in \{1, 2, \ldots, 8\}$$

where P_{ij} is the normalized proportion of city *i* among images of Cluster *j*, and μ_j and σ_j are the mean and standard deviation of P_{ij} across all cities, respectively.

However, the degree to which P_{ij} deviates from (either exceeds or subceeds) the average proportion only considers distinctiveness in terms of a single visual cluster. Moreover, it is plausible to assume that the stronger the visual pattern is within a city's identity, the more it contributes to the uniqueness of a city. Therefore, we weighted 'adjusted distinctiveness' values of city *i* for cluster *j* by their corresponding proportion in the visual identity composition. As a result, the *City Uniqueness Index* of city *i* is defined as the cumulative sum of deviations for each cluster, and finally calculated as follows:

$$CityUniqueness_i = \sum_{i=1}^{8} \left(\left| AdjustedDistinctiveness_{ij} \right| \times IdentityProportion_{ij} \right)$$

where $IdentityProportion_{ij}$ is the proportion of Cluster *j* in city *i*. Note that we took the absolute value of 'adjusted distinctiveness' assuming that containing scenes just as much as other cases (close to the distribution mean) does not satisfy as a unique characteristic. Thus, we considered both abundance and scarcity of a particular type of visual scene could contribute to city uniqueness.

In addition, a novel visualization was proposed as a means of presenting the two components of the *City Uniqueness Index* effectively. While $AdjustedDistinctiveness_{ij}$ and $IdentityProportion_{ij}$ convey separate meanings, having bivariate information in a single visualization can help portray unique visual characters of a city. In this regard, we used a variable width bar chart, whose bar height indicates $AdjustedDistinctiveness_{ij}$ and bar width indicates $IdentityProportion_{ij}$. Also, each bar was plotted both above and below the reference line, indicating the excess or deficit in P_{ij} compared to the distribution mean.

Fig. 8 shows variable width bar charts and Table 2 summarizes the *City Uniqueness Index* calculated for both the inner city areas and the whole city areas. As a result, Kyoto (1.141) was found with the highest uniqueness score in its visitor-oriented area, followed by Sapporo (0.846) and Tokyo (0.830); on the contrary, visitors are likely to

Table 2
City Uniqueness Index measured for both the inner city area and the whole city area.

City	Inner city area	Whole city area
Fukuoka	0.594	0.378
Kyoto	1.141	1.383
Nagoya	0.665	0.719
Osaka	0.602	0.905
Sapporo	0.846	1.079
Tokyo	0.829	0.885

Fig. 8. A variable width bar chart visualization for urban visual uniqueness. The widths and heights of the bars indicate *IdentityProportion* and *AdjustedDistinctiveness*, respectively.

experience the least unique urban scenes in Osaka (0.602) and Fukuoka (0.594). More notably, some visual patterns made cities distinctive regardless of their small portions, while others were not as contributive as expected from their high identity proportions. For instance, the lack of modern urban environments was one of the distinctive features of Kyoto (narrow but tall bars of Clusters 3 and 4), whereas their presence among Osaka's visual identity was barely effective for making the city distinctive (wide but short bars).

We also explored by taking the whole city area into account. While Kyoto (1.383) remained as the most unique city, we observed a sharp increase in the uniqueness of Osaka (0.905) presumably due to a myriad of urban built environment scenes (Clusters 2 to 5) distinguishing Osaka with the other cities. Sapporo's visual identity was greatly differentiated by the abundance of urban corridors throughout the city (Cluster 1), although its rich urban green spaces (Cluster 6) were indecisive. Similarly, Fukuoka was not very distinctive in terms of green spaces which constitute the highest proportion of its visual scenes, being the least unique city among its counterparts, which is the opposite of the case for Kyoto.

To assess the robustness of our sampling strategy, we tested the sensitivity of *City Uniqueness Index* within the inner city areas across different sampling intervals (every 50 m between 50 and 500 m). Supplementary Materials Fig. S6 shows that, while minor fluctuations exist (no statistical significance, $p > 0.05$), overall values of uniqueness scores and their relative rankings remained stable, supporting the validity of our approach.

6. Discussion

6.1. Rethinking urban visual uniqueness and its measurement

This study aimed to explore urban visual uniqueness from the cases of six major Japanese cities. As a result, we identified 8 distinct visual patterns that characterize urban landscapes in Japanese cities, including scenes of dense urban centers, various types of urban infrastructure (e.g., overpass, off-street parking lot), urban green spaces, and streets filled with traditional wooden townhouses. The novelty lies in the application of a deep feature-based clustering approach to street view imagery, which allows a bottom-up representation of the visual characters of cities. While previous research relied on iconic landmarks to examine urban uniqueness, our findings revealed that Japanese cities form clusters of visual similarity distinguished by visual semantic cues in everyday scenes, such as the building height-to-street width ratio or the presence of electric poles and signboards. Moreover, our attempt to redefine the city's visual uniqueness in terms of identity and distinctiveness explicates what is unique about a city and to what extent, reducing confusion among place-based concepts and terminologies that have been interchangeably used in earlier studies (Antonsich, 2010; Peng et al., 2020).

The theoretical significance of this finding is twofold. First, we add to the literature on how the subjective and ambiguous qualities of urban spaces can be captured. Indeed, it is these subtle differences that construct a unique sense of place as perceived by the people and their collective experiences (Longley & Duxbury, 2016). Cultural mapping has emerged as an effective tool to measure the non-measurable (or intangibles) in a participatory manner (Duxbury et al., 2015; Jeannotte, 2016; Radović, 2016), but still relies heavily on human input and interaction. In response, Liang et al. (2023) recently conducted an embedding-driven clustering to evaluate urban visual environments and assess their spatiotemporal change. However, only ten predefined physical or perceptual features were included for clustering analysis, and a challenge remains in considering other place-informative factors that point to the cultural or historical context of cities. By combining unsupervised clustering with post-hoc interpretation of results, the framework proposed in this study provides a comprehensive base for urban planning and design practitioners to better uncover underlying visual patterns and shape unique images of cities for placemaking strategies.

Second, this study expands the discourse on distinctiveness. While previous studies have tended to seek urban distinctiveness through the presence of unique characteristics, we also highlighted the relative deficiency of visual scenes as indicated by negative values of *AdjustedDistinctiveness* (plotted below the reference line in Fig. 8). This complies with the 'negative distinctiveness' from Social Identity Theory which refers to the unique differences of an individual (or a group) identified by the lack of certain characteristics (Mlicki & Ellemers, 1996; Tajfel, 1979). But does this necessarily imply a failure in urban branding? The notion of 'optimal distinctiveness' in social science literature posits that the highest performance of an organization can be achieved through the balance between differentiation and conformity (Brewer, 1991; Deephouse, 1999; Löhndorf & Diamantopoulos, 2014). This conforms to Lefebvre's (1996) idea of 'totality', where it is neither solely the presence nor absence but rather the sum of moments (or events) that constitute our everyday lives. Thus, we should view urban distinctiveness as a multi-faceted construct and not "privilege either the mundane or the exceptional" (Gardiner, 2004, p. 243).

Rethinking the definition and measurement of urban visual uniqueness also has practical implications. Residents' place identity as well as a sense of distinctiveness about their cities were found to have significant impacts on positive attitudes toward tourism (Wang & Xu, 2015). In this regard, it is suggested that city planners develop tourism strategies in a way that prioritizes preserving the unique character of cities, which can eventually improve place-based perceptions for both residents and visitors. This further provides insights into urban competitiveness in the recent smart city debate, by highlighting the importance of placemaking principles that help strengthen and create singularities of different cities (Abusaada & Elshater, 2021).

It is noteworthy to mention the perspective of street view images used in this study. The use of forward-facing GSV images differs from studies that evaluated relationships between people's perception and street environments focusing on active frontages (Heffernan et al., 2014) or architectural exteriors (Liang et al., 2024). Although façade-level details are integral to urban perception, we argue that urban visual uniqueness emerges from the cumulative composition of the full streetscape, not façades alone. In this sense, we adopted images with forward facing orientation that better captures the comprehensive scene of the outdoor space from one side of the street to the other, or in other words, the public realm "whose full-blown existence is what makes the city different" (Lofland, 2017, p. 9). Nonetheless, we recognize the advantages of side views that are perpendicular to the street in examining detailed architectural features or pedestrian-level interactions, and consider integrating such alternative perspectives for a more nuanced urban scene analysis tailored to specific elements of interest.

In sum, our novel landmark-free framework to identify distinct urban visual patterns can help expand our understanding of unique characteristics that define a city and distinguish it from others. By revealing the subtle differences among cities through inconspicuous yet place-informative urban elements, the output of the study is expected to inform future policy directions of urban design and tourism planning with both effective and feasible strategies. Moreover, coupled with the known effects of street morphology on urban activities (Zhang et al., 2023), this computational analysis of urban visual uniqueness can complement planning support systems (Hong, 2024) in promoting urban quality by hinting at the importance of "enormous collection of small elements" (Jacobs, 1958, p. 147).

6.2. Limitations and future work

Despite theoretical and practical contributions, we found several limitations of our research that can be improved in future work. First, we primarily focused on the six most visited cities of Japan according to the number of foreigner entries. Given the study's aim to examine subtle distinctions in urban visual uniqueness, the selected cities offer a controlled context characterized by shared sociocultural backgrounds. However, while the obtained inner city urban visual clusters were tested on a new set of street scenes in whole city areas (see Fig. 7), the geographic focus on Japanese cities limits the generalizability of our findings across different global contexts. To harness the full potential of the proposed landmark-free approach as a city-agnostic framework, future studies should investigate the transferability of unsupervised knowledge (i.e., urban visual clusters, uniqueness metrics) to other urban settings, which would enable comparative analyses and reveal whether international differences in urban uniqueness exceed or counterintuitively resemble those observed within a single country. In addition, the inner city areas were selected based on the aggregation of accommodation facilities, assuming a high likelihood of visitor experience in these sites, and the sampling area considered to collect GSV images may not fully cover the streets in the selected cities. These may limit the generalizability of results as the spatiotemporal coverage of street view imagery is not consistent across cities (Kim & Jang, 2023), and people's activity space may vary depending on trip purposes (Kang et al., 2017). In future studies, determining a more robust study area using travel surveys or high-resolution GPS records could be considered. Moreover, although we found the seasonal impact is minimal on the results, future research can consider removing winter images or explore seasonal variations in urban visual uniqueness by employing more balanced datasets.

Another drawback lies in the interpretability of the models used in this study. K-means clustering, which relies on Euclidean distance, tends to be less stable and less interpretable on high-dimensional data (Kappmeier et al., 2015). As dimensionality increases, Euclidean distances between data points become more similar, which can reduce the ability of k-means to clearly separate clusters. Future research could explore lower-dimensional embeddings or alternative clustering approaches that better capture the underlying data structure. For example, when considering spatial dependencies among street-view locations, multiple scenes can be modeled as a network. Graph-based clustering methods, such as community detection (Blondel et al., 2008), classify data according to their topological relationships rather than purely by distances. Such approaches may enhance the interpretability of results, particularly for understanding spatial structures and the interplay between visual similarity and geographic proximity. While the models effectively identified key visual features that distinguish different clusters (also agreeing with the common understanding of Japanese cities), understanding how specific features influence the results remains challenging. While we chose a scene-focused model for its interpretability and efficiency in understanding the streetscape, future work could explore contrastive learning or object-centric learning models for potentially more expressive street scene representations (Guo et al., 2024; Y. Li et al., 2025), and to evaluate the extent to which the models influence the results and investigate the underlying reasons. The explainability of the results can also be further enhanced by considering the spatial layouts. The spatial layouts of the urban scene can be represented through the construction of scene graphs (H. Ma et al., 2024; Zhang, Salazar-Miranda, et al., 2024). Scene graphs built based on key elements with semantic information, rather than all visible objects, can provide clearer insights about spatial configuration and human perception. Our future work will include developing an efficient approach for constructing scene graphs utilizing Vision Language Model and to further explain the relationship between scene graph, their prototypes and motifs, and urban visual patterns. Moreover, a comparative and interpretive urban research has been effective in examining distinctiveness of cities due to the subjective and explorative nature (Barbehön & Münch, 2017). Future research should provide more concrete and comprehensive explanations to offer deeper insights into the visual cues of urban scenes.

7. Conclusion

How to assess the collective identity of places has been questioned for decades by urban researchers, but both its conceptual meaning and quantification method remains to be established. In this research, we introduced a refined definition of urban uniqueness, encompassing visual identity and distinctiveness of cities. This offers a comprehensive framework for investigating a city's uniqueness, based on distinct clusters of visual scenes identified by leveraging deep feature representation and unsupervised clustering analysis on street view imagery. From a case study of six major Japanese cities, we revealed the existence of eight typical scenes that constitute the urban landscapes in Japan and contribute to their uniqueness. For city planners and policymakers, this provides valuable information in evaluating the visual qualities of urban environments and shaping visually distinctive cities. Additionally, for urban researchers, we offer a methodological contribution to *place* research by conceptualizing relevant concepts (i.e., identity, distinctiveness, uniqueness) and measuring them from everyday scenes, rather than geographic landmarks.

CRediT authorship contribution statement

Song Guo: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Kee Moon Jang:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Formal analysis, Data curation, Conceptualization. **Fábio Duarte:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Yuhao Kang:** Writing – review & editing, Validation, Investigation, Data curation. **Carlo Ratti:** Writing – review & editing, Validation, Supervision, Resources, Project administration.

Declaration of competing interest

The authors declare no conflict of interest.

Acknowledgments

Authors acknowledge the supportt of Toyota Woven City, and all members of the MIT Senseable City Lab Consortiumé.

Appendix A. Appendix

Table A1
Object categories considered for urban visual clustering interpretation.

Categories	Original classes of ADE20K
Building/road	'building;edifice'/'road;route'[a]
Tree	'tree'+'palm;palm;tree'
Sky	'sky'
Sidewalk	'sidewalk;pavement'
Grass	'grass' + 'flower' + 'plant;flora;plant;life' + 'rock;stone' + 'pot;flowerpot'
Fence	'fence;fencing'
Signage	'flag'+'bulletin;board;notice;board'+'poster;posting;placard;notice;bill;card' + 'signboard;sign'
Pole	'streetlight;street;lamp' + 'pole'+'traffic;light;traffic;signal;stoplight'
Bridge	'bridge;span'
Railing	'railing;rail'
Awning	'awning;sunshade;sunblind'
Stairs	'stairs;steps' + 'stairway;staircase' + 'step;stair' + 'escalator;moving;staircase;moving;stairway'
Door	'door;double;door'
Bench	'bench' + 'seat'
Interior	'windowpane;window' + 'ceiling'
House	'house'
Waterbody	'water' + 'sea' + 'river' + 'lake'
Mountain	'mountain;mount' + 'hill'
Column	'column;pillar''
Pier	'pier;wharf;wharfage;dock'
Tower	'tower'
Skyscraper	'skyscraper'

[a] For images where the value of 'road;route' is 0, fill the 'road;route' value with 10^{-6}. In the final results, remove any images where the ratio is greater than 1000 ($n = 1191$).

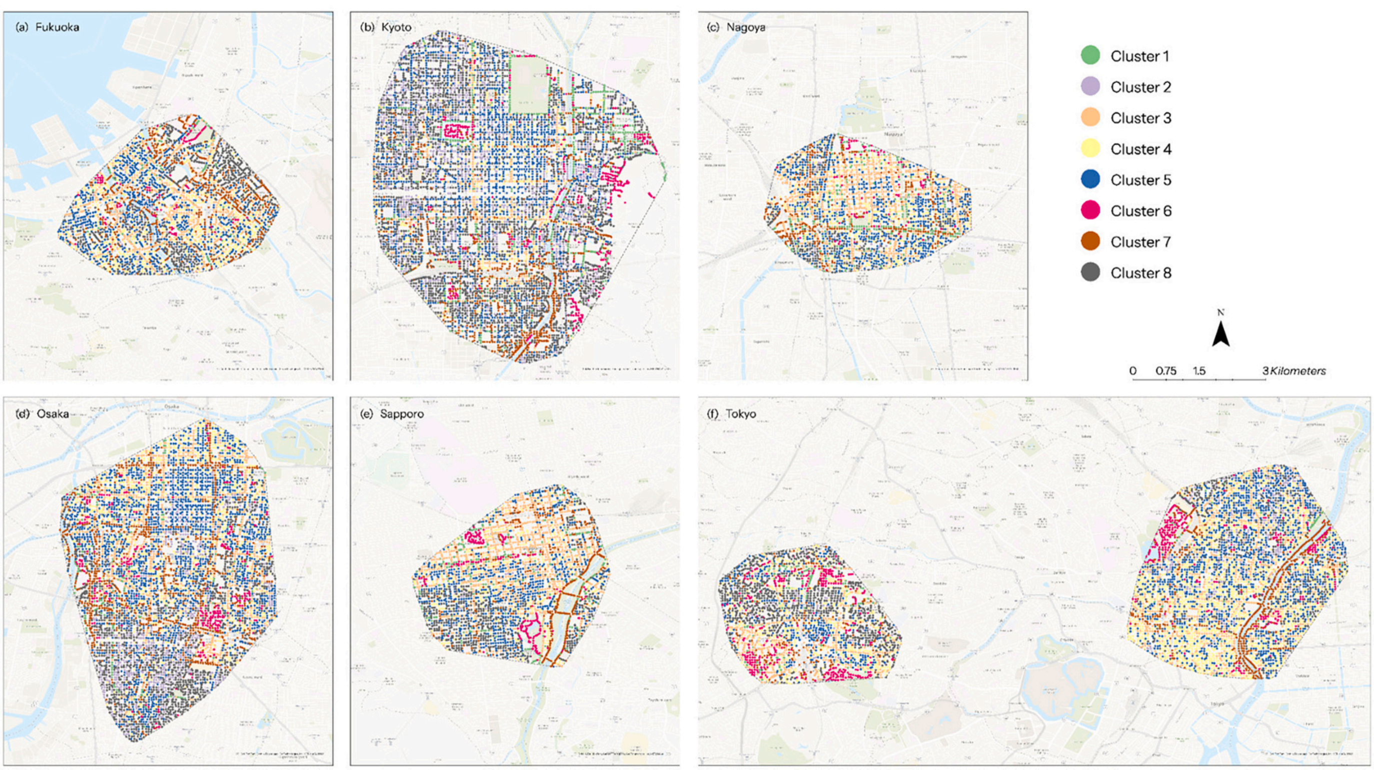

Fig. A1. Spatial distribution of clusters.

Appendix B. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compenvurbsys.2025.102351.

Data availability

All relevant data used and generated in the research are publicly available in the Figshare Repository at: https://figshare.com/s/3e296986411dec77312f.

References

Abusaada, H., & Elshater, A. (2021). Competitiveness, distinctiveness and singularity in urban design: A systematic review and framework for smart cities. *Sustainable Cities and Society, 68*, Article 102782.

Antonsich, M. (2010). Meanings of place and aspects of the self: An interdisciplinary and empirical account. *GeoJournal, 75*(1), 119–132.

Archer, K. J., & Kimes, R. V. (2008). Empirical characterization of random forest variable importance measures. *Computational Statistics & Data Analysis, 52*(4), 2249–2260.

Axhausen, K. W., Chikaraishi, M., & Seya, H. (2015). Parking: Learning from Japan. *Arbeitsberichte Verkehrs-Und Raumplanung, 1095*.

Barbehön, M., & Münch, S. (2017). Interrogating the city: Comparing locally distinct crisis discourses. *Urban Studies, 54*(9), 2072–2086.

Belanche, D., Casaló, L. V., & Rubio, M.Á. (2021). Local place identity: A comparison between residents of rural and urban communities. *Journal of Rural Studies, 82*, 242–252.

Bernardo, F., Almeida, J., & Martins, C. (2017). Urban identity and tourism: Different looks, one single place. *Proceedings of the Institution of Civil Engineers-Urban Design and Planning, 170*(5), 205–216.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment, 2008*(10), Article P10008.

Brewer, M. B. (1991). The social self: On being the same and different at the same time. *Personality and Social Psychology Bulletin, 17*(5), 475–482.

Burgess, C. (2004). Maintaining identities: Discourses of homogeneity in a rapidly globalizing Japan. *Electronic Journal of Contemporary Japanese Studies*. Retrieved from http://www.japanesestudies.org.uk/articles/Burgess.html.

Chen, M., Li, Z., Huang, W., Gong, Y., & Yin, Y. (2024). Profiling urban streets: A semi-supervised prediction model based on street view imagery and spatial topology. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (pp. 319–328).

Chen, S., & Biljecki, F. (2023). Automatic assessment of public open spaces using street view imagery. *Cities, 137*, Article 104329.

Cherti, M., Beaumont, R., Wightman, R., Wortsman, M., Ilharco, G., Gordon, C., Schuhmann, C., Schmidt, L., & Jitsev, J. (2023). Reproducible scaling laws for contrastive language-image learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 2818–2829).

Collins-Kreiner, N. (2020). Hiking, sense of place, and place attachment in the age of globalization and digitization: The Israeli case. *Sustainability, 12*(11), 4548.

Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 2*, 224–227.

Deephouse, D. L. (1999). To be different, or to be the same? It'sa question (and theory) of strategic balance. *Strategic Management Journal, 20*(2), 147–166.

Ding, K., Niu, Y., & Choo, W. C. (2023). The evolution of Airbnb research: A systematic literature review using structural topic modeling. *Heliyon, 9*(6).

Doersch, C., Singh, S., Gupta, A., Sivic, J., & Efros, A. A. (2015). What makes paris look like paris? *Communications of the ACM, 58*(12), 103–110.

Duarte, F. (2017). *Space, place and territory: A critical review on spatialities*. Routledge.

Duxbury, N., Garrett-Petts, W. F., & MacLennan, D. (2015). Cultural mapping as cultural inquiry: Introduction to an emerging field of practice. *Cultural Mapping as Cultural Inquiry*, 1–42.

Fujiwara, K., Khomiakov, M., Yap, W., Ignatius, M., & Biljecki, F. (2024). Microclimate vision: Multimodal prediction of climatic parameters using street-level and satellite imagery. *Sustainable Cities and Society, 114*, 105733.

Fukahori, K., & Kubota, Y. (2003). The role of design elements on the cost-effectiveness of streetscape improvement. *Landscape and Urban Planning, 63*(2), 75–91.

Gao, S., Janowicz, K., Montello, D. R., Hu, Y., Yang, J.-A., McKenzie, G., … Yan, B. (2017). A data-synthesis-driven method for detecting and extracting vague cognitive regions. *International Journal of Geographical Information Science, 31*(6), 1245–1271.

Gardiner, M. (2004). Everyday utopianism: Lefebvre and his critics. *Cultural Studies, 18*(2–3), 228–254.

Ginting, N., Nasution, A. D., & Rahman, N. V. (2017). More attractive more identified: Distinctiveness in embedding place identity. *Procedia Environmental Sciences, 37*, 408–419.

Goel, A., Juneja, M., & Jawahar, C. V. (2012). Are buildings only instances? Exploration in architectural style categories. In *Proceedings of the Eighth Indian Conference on Computer Vision, Graphics and Image Processing* (pp. 1–8).

González-Reverté, F., Corrons, A., & Gomis-López, J. M. (2025). Exploring the role of memorable tourism experiences and place attachment in the revisit intention of tourists staying in short-term rentals. *Journal of Hospitality and Tourism Insights, 8*(11), 79–99.

Guo, D., Yu, Y., Ge, S., Gao, S., Mai, G., & Chen, H. (2024). SpatialScene2Vec: A self-supervised contrastive representation learning method for spatial scene similarity evaluation. *International Journal of Applied Earth Observation and Geoinformation, 128*, Article 103743.

Gurran, N. (2018). Global home-sharing, local communities and the Airbnb debate: A planning research agenda. *Planning Theory & Practice, 19*(2), 298–304.

Guttentag, D. (2015). Airbnb: Disruptive innovation and the rise of an informal tourism accommodation sector. *Current Issues in Tourism, 18*(12), 1192–1217.

Halkidi, M., Batistakis, Y., & Vazirgiannis, M. (2001). On clustering validation techniques. *Journal of Intelligent Information Systems, 17*, 107–145.

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).

Heffernan, E., Heffernan, T., & Pan, W. (2014). The relationship between the quality of active frontages and public perceptions of public spaces. *Urban Design International, 19*, 92–102.

Hong, Q. (2024). Development of planning support systems: A systematic review of the Chinese language literature. *Transactions in Urban Data, Science, and Technology, 3*(1–2), 80–106, 27541231241251410.

Hoshi, T. (2023). Land use plan and changes in the land use planning system toward its realization: A case study of Sapporo City. *Japan Architectural Review, 6*(1), Article e12333.

Hu, Y., Gao, S., Janowicz, K., Yu, B., Li, W., & Prasad, S. (2015). Extracting and understanding urban areas of interest using geotagged photos. *Computers, Environment and Urban Systems, 54*, 240–254.

Huang, W., Wang, J., & Cong, G. (2024). Zero-shot urban function inference with street view images through prompting a pretrained vision-language model. *International Journal of Geographical Information Science, 38*(7), 1414–1442.

Jacobs, J. (1958). Downtown is for people. *The Exploding Metropolis, 168*, 124–131.

Jain, A. K., & Dubes, R. (1988). *Algorithms for clustering data*. Prentice Hall.

Jang, K. M., Chen, J., Kang, Y., Kim, J., Lee, J., Duarte, F., & Ratti, C. (2024). Place identity: A generative AI'S perspective. *Humanities and Social Sciences Communications, 11*(1), 1–16.

Jang, K. M., & Kim, Y. (2019). Crowd-sourced cognitive mapping: A new way of displaying people's cognitive perception of urban space. *PLoS One, 14*(6), Article e0218590.

Jarratt, D., Phelan, C., Wain, J., & Dale, S. (2019). Developing a sense of place toolkit: Identifying destination uniqueness. *Tourism and Hospitality Research, 19*(4), 408–421.

Jeannotte, M. S. (2016). Story-telling about place: Engaging citizens in cultural mapping. *City, Culture and Society, 7*(1), 35–41.

Jiao, J., & Wang, H. (2023). Forecasting traffic speed during daytime from Google street view images using deep learning. *Transportation Research Record, 2677*(12), 743–753.

Kang, B., Moudon, A. V., Hurvitz, P. M., & Saelens, B. E. (2017). Differences in behavior, time, location, and built environment between objectively measured utilitarian and recreational walking. *Transportation Research Part D: Transport and Environment, 57*, 185–194.

Kang, Y., Abraham, J., Ceccato, V., Duarte, F., Gao, S., Ljungqvist, L., Zhang, F., Näsman, P., & Ratti, C. (2023). Assessing differences in safety perceptions using GeoAI and survey across neighbourhoods in Stockholm, Sweden. *Landscape and Urban Planning, 236*, Article 104768.

Kang, Y., Zhang, F., Gao, S., Lin, H., & Liu, Y. (2020). A review of urban physical environment sensing using street view imagery in public health studies. *Annals of GIS, 26*(3), 261–275.

Kappmeier, J.-P. W., Schmidt, D. R., & Schmidt, M. (2015). Solving k-means on high-dimensional big data. *International Symposium on Experimental Algorithms*, 259–270.

Kim, J., & Jang, K. M. (2023). An examination of the spatial coverage and temporal variability of Google street view (GSV) images in small-and medium-sized cities: A people-based approach. *Computers, Environment and Urban Systems, 102*, Article 101956.

Lee, S., Maisonneuve, N., Crandall, D., Efros, A. A., & Sivic, J. (2015). Linking past to present: Discovering style in two centuries of architecture. In *IEEE International Conference on Computational Photography*.

Lefebvre, H. (1996). *The right to the city. Writings on cities*. Trans. Eleonore Kofman and Elizabeth Lebas.

Li, M., Fang, L., Huang, X., & Goh, C. (2015). A spatial–temporal analysis of hotels in urban tourism destination. *International Journal of Hospitality Management, 45*, 34–43.

Li, N., & Quan, S. J. (2024). Discovering urban block typologies in Seoul: Combining planning knowledge and unsupervised machine learning. *Cities, 150*, Article 104988.

Li, Y., Huang, Y., Mai, G., & Zhang, F. (2025). Learning street view representations with spatiotemporal contrast. *ArXiv Preprint* (ArXiv:2502.04638).

Liang, X., Chang, J. H., Gao, S., Zhao, T., & Biljecki, F. (2024). Evaluating human perception of building exteriors using street view imagery. *Building and Environment, 263*, Article 111875.

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

Lim, S. E. Y., & Bouchon, F. (2017). Blending in for a life less ordinary? Off the beaten track tourism experiences in the global city. *Geoforum, 86*, 13–15.

Liu, Y., Li, Z., Xiong, H., Gao, X., Wu, J., & Wu, S. (2013). Understanding and enhancement of internal clustering validation measures. *IEEE Transactions on Cybernetics, 43*(3), 982–994.

Lofland, L. H. (2017). *The public realm: Exploring the city's quintessential social territory*. Routledge.

Löhndorf, B., & Diamantopoulos, A. (2014). Internal branding: Social identity and social exchange perspectives on turning employees into brand champions. *Journal of Service Research, 17*(3), 310–325.

Longley, A., & Duxbury, N. (2016). Introduction: Mapping cultural intangibles. *City, Culture and Society, 7*(1), 1–7.

Lynch, K. (1960). *The image of the city*. MIT Press.

Ma, H., Zhang, Y., Liu, P., Zhang, F., & Zhu, P. (2024). How does spatial structure affect psychological restoration? A method based on graph neural networks and street view imagery. *Landscape and Urban Planning, 251*, Article 105171.
Ma, L., Guo, Z., Lu, M., He, S., & Wang, M. (2023). Developing an urban streetscape indexing based on visual complexity and self-organizing map. *Building and Environment, 242*, Article 110549.
Madeira, A. N., Rodrigues, R. I., Palrao, T., & Santos, V. R. (2023). The influence of web summit attendees' age and length of stay on leisure activity preferences and city image. *International Journal of Event and Festival Management, 14*(3), 326–343.
Manahasa, E., Zera, A., Manahasa, O., & Hysa, A. (2025). Neighbourhood identity versus city identity: An inter-scalar perspective via urban layers in a historical zone under transformation. *Frontiers of Architectural Research, 14*(5), 1147–1167.
Mazzarello, M., Klimenka, M., Sanatani, R., Eshtiyagh, J., Yao, Y., Santi, P., … Ratti, C. (2024). *A geography of indoors for analyzing global ways of living using computer vision*.
Miranda, F., Hosseini, M., Lage, M., Doraiswamy, H., Dove, G., & Silva, C. T. (2020). Urban mosaic: Visual exploration of streetscapes using large-scale image data. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–15).
Mlicki, P. P., & Ellemers, N. (1996). Being different or being better? National stereotypes and identifications of polish and Dutch students. *European Journal of Social Psychology, 26*(1), 97–114.
Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., … El-Nouby, A. (2023). Dinov2: Learning robust visual features without supervision. *ArXiv Preprint* (ArXiv:2304.07193).
Palmer, A., Koenig-Lewis, N., & Jones, L. E. M. (2013). The effects of residents' social identity and involvement on their advocacy of incoming tourism. *Tourism Management, 38*, 142–151.
Pang, G., Shen, C., & Van Den Hengel, A. (2019). Deep anomaly detection with deviation networks. In *Proceedings of the 25th ACM SIGKDD international conference on Knowledge Discovery & Data Mining* (pp. 353–362).
Patandianan, M. V., & Shibusawa, H. (2020). Importance and performance of streetscapes at a tourism destination in Indonesia: The residents' perspectives. *Frontiers of Architectural Research, 9*(3), 641–655.
Peng, J., Strijker, D., & Wu, Q. (2020). Place identity: How far have we come in exploring its meanings? *Frontiers in Psychology, 11*, 294.
Phetsuriya, N., & Heath, T. (2021). Defining the distinctiveness of urban heritage identity: Chiang Mai old city, Thailand. *Social Sciences, 10*(3), 101.
Proshansky, H. M., Fabian, A. K., & Kaminoff, R. (2014). Place-identity: Physical world socialization of the self (1983). In *The people, place, and space reader* (pp. 77–81). Routledge.
Purves, R. S., Winter, S., & Kuhn, W. (2019). Places in information science. *Journal of the Association for Information Science and Technology, 70*(11), 1173–1182.
Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., … Clark, J. (2021). Learning transferable visual models from natural language supervision. *International Conference on Machine Learning*, 8748–8763.
Radović, D. (2016). Measuring the non-measurable: On mapping subjectivities in urban research. *City, Culture and Society, 7*(1), 17–24.
Ranftl, R., Bochkovskiy, A., & Koltun, V. (2021). Vision transformers for dense prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 12179–12188).
Rapoport, A. (1990). *The meaning of the built environment: A nonverbal communication approach.* University of Arizona Press.
Relph, E. (1976). *Place and placelessness* (Vol. 67). Pion London.
Rodrigues, R., Madeira, A., & Palrão, T. (2024). Beyond a lasting impression: Exploring the impact of brand image on revisit intention, unveiling place attachment as a mediating mechanism. *Journal of Destination Marketing & Management, 33*, Article 100927.
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.
Sadeque, S., Roy, S. K., Swapan, M. S. H., Chen, C.-H., & Ashikuzzaman, M. (2020). An integrated model of city and neighborhood identities: A tale of two cities. *Journal of Business Research, 117*, 780–790.
Salesses, P., Schechtner, K., & Hidalgo, C. A. (2013). The collaborative image of the city: Mapping the inequality of urban perception. *PLoS One, 8*(7), Article e68400.
Schertz, K. E., Sachdeva, S., Kardan, O., Kotabe, H. P., Wolf, K. L., & Berman, M. G. (2018). A thought in the park: The influence of naturalness and low-level visual features on expressed thoughts. *Cognition, 174*, 82–93.
Seya, H., Axhausen, K. W., & Chikaraishi, M. (2020). Spatial unconditional quantile regression: Application to Japanese parking price data. *The Annals of Regional Science, 65*, 351–402.
Shin, S., Tae, S., Woo, J., & Roh, S. (2011). The development of environmental load evaluation system of a standard Korean apartment house. *Renewable and Sustainable Energy Reviews, 15*(2), 1239–1249.
Singh, A. K., Mittal, S., Malhotra, P., & Srivastava, Y. V. (2020). Clustering evaluation by davies-bouldin index (dbi) in cereal data using k-means. In *2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 306–310).
Taeuscher, K., Zhao, E. Y., & Lounsbury, M. (2022). Categories and narratives as sources of distinctiveness: Cultural entrepreneurship within and across categories. *Strategic Management Journal, 43*(10), 2101–2134.
Tajfel, H. (1979). An integrative theory of intergroup conflict. *The Social Psychology of Intergroup Relations/Brooks/Cole*.
Tan, P.-N., Steinbach, M., & Kumar, V. (2016). *Introduction to data mining.* Pearson Education India.
Thrun, M. C. (2021). Distance-based clustering challenges for unbiased benchmarking studies. *Scientific Reports, 11*(1), 18988.
Tuan, Y.-F. (1979). Space and place: Humanistic perspective. In *Philosophy in geography* (pp. 387–427). Springer.
Twigger-Ross, C. L., & Uzzell, D. L. (1996). Place and identity processes. *Journal of Environmental Psychology, 16*(3), 205–220.
Valentin, M., & O'Neill, J. W. (2019). The value of location for urban hotels. *Cornell Hospitality Quarterly, 60*(1), 5–24.
Wang, F., Yuan, C., Li, J., & Liu, B. (2022). What makes a place special? Research on the locality of cities in the Yellow River and Rhine River basins based on street view images. *Indoor and Built Environment, 31*(2), 435–451.
Wang, J., Huang, W., & Biljecki, F. (2024). Learning visual features from figure-ground maps for urban morphology discovery. *Computers, Environment and Urban Systems, 109*, Article 102076.
Wang, S., & Xu, H. (2015). Influence of place-based senses of distinctiveness, continuity, self-esteem and self-efficacy on residents' attitudes toward tourism. *Tourism Management, 47*, 241–250.
Wild, J. (1965). *Existence and the world of freedom*.
Wu, T. H., Amiruzzaman, M., Zhao, Y., Bhati, D., & Yang, J. (2024). Visualizing routes with AI-discovered street-view patterns. *IEEE Transactions on Computational Social Systems, 11*(5), 6380–6391.
Xu, S., Zhang, C., Fan, L., Meng, G., Xiang, S., & Ye, J. (2024). Addressclip: Empowering vision-language models for city-wide image address localization. In *European Conference on Computer Vision* (pp. 76–92).
Yan, Y., Wen, H., Zhong, S., Chen, W., Chen, H., Wen, Q., Zimmermann, R., & Liang, Y. (2024). Urbanclip: Learning text-enhanced urban region profiling with contrastive language-image pretraining from the web. In *, 2024. Proceedings of the ACM Web Conference* (pp. 4006–4017).
Zhang, E., Xie, H., & Long, Y. (2023). Decoding the association between urban streetscape skeletons and urban activities: Experiments in Beijing using Dazhong Dianping data. *Transactions in Urban Data, Science, and Technology, 2*(1), 3–18.
Zhang, F., Salazar-Miranda, A., Duarte, F., Vale, L., Hack, G., Chen, M., … Ratti, C. (2024). Urban visual intelligence: Studying cities with artificial intelligence and street-level imagery. *Annals of the American Association of Geographers, 114*(5), 876–897.
Zhang, F., & Yang, X. (2020). Improving land cover classification in an urbanized coastal area by random forests: The role of variable selection. *Remote Sensing of Environment, 251*, Article 112105.
Zhang, F., Zhou, B., Ratti, C., & Liu, Y. (2019). Discovering place-informative scenes and objects using social media photos. *Royal Society Open Science, 6*(3), Article 181375.
Zhang, Y., Li, Y., & Zhang, F. (2024). Multi-level urban street representation with street-view imagery and hybrid semantic graph. *ISPRS Journal of Photogrammetry and Remote Sensing, 218*, 19–32.
Zhao, L., Luo, L., Li, B., Xu, L., Zhu, J., He, S., & Li, H. (2021). Analysis of the uniqueness and similarity of City landscapes based on deep style learning. *ISPRS International Journal of Geo-Information, 10*(11), 734.
Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., & Torralba, A. (2017). Places: A 10 million image database for scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 40*(6), 1452–1464.
Zhou, B., Liu, L., Oliva, A., & Torralba, A. (2014). Recognizing city identity via attribute analysis of geo-tagged images. In *, 13. Computer vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part III* (pp. 519–534).
Zhou, B., Zhao, H., Puig, X., Fidler, S., Barriuso, A., & Torralba, A. (2017). Scene parsing through ade20k dataset. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 633–641).