International Journal of Geo-Information

MDPI

Article

Crowd-Sourced City Images: Decoding Multidimensional Interaction between Imagery Elements with Volunteered Photos

Yao Shen [1,2,3,*], Yiyi Xu [1] and Lefeng Liu [1]

[1] College of Architecture and Urban Planning, Tongji University, Shanghai 200092, China; xuyiyi@tongj.edu.cn (Y.X.); 2130100@tongji.edu.cn (L.L.)
[2] Key Laboratory of Ecology and Energy-saving Study of Dense Habitat, Ministry of Education, Shanghai 200092, China
[3] The Bartlett Centre for Advanced Spatial Analysis, University College London, London W1T 4TJ, UK
* Correspondence: eshenyao@tongji.edu.cn; Tel.: +86-6598-2345

Abstract: The built environment reshapes various scenes that can be perceived, experienced, and interpreted, which are known as city images. City images emerge as the complex composite of various imagery elements. Previous studies demonstrated the coincide between the city images produced by experts with prior knowledge and that are extracted from the high-frequency photo contents generated by citizens. The realistic city images hidden behind the volunteered geo-tagged photos, however, are more complex than assumed. The dominating elements are only one side of the city image; more importantly, the interactions between elements are also crucial for understanding how city images are structured in people's minds. This paper focuses on the composition of city image–the various interactions between imagery elements and areas of a city. These interactions are identified as four aspects: co-presence, hierarchy, heterogeneity, and differentiation, which are quantified and visualized respectively as correlation network, dendrogram, spatial clusters, and scattergrams in a framework using scene recognition with volunteered and georeferenced photos. The outputs are interdependent elements, typologies of elements, imagery areas, and preferences for groups, which are essential for urban design processes. In the application in Central Beijing, the significant interdependency between two elements is complex and is not necessarily an interaction between the elements with higher frequency only. The main typologies and the principal imagery elements are different from what were prefixed in the image recognition model. The detected imagery areas with adaptive thresholds suggest the spatially varying spill over effects of named areas and their typologies can be well annotated by the detected principal imagery elements. The aggregation of the data from different social media platforms is proven as a necessity of calibrating the unbiased scope of the city image. Any specific data can hardly capture the whole sample. The differentiation across the local and non-local is found to be related to their preference and activity space. The results provide more comprehensive insights on the complex composition of city images and its effects on placemaking.

Keywords: city image; imagery area; geotagged photos; crowdsourcing; image recognition

Citation: Shen, Y.; Xu, Y.; Liu, L. Crowd-Sourced City Images: Decoding Multidimensional Interaction between Imagery Elements with Volunteered Photos. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 740. https://doi.org/10.3390/ijgi10110740

Academic Editor: Wolfgang Kainz

Received: 16 August 2021
Accepted: 27 October 2021
Published: 1 November 2021

1. Introduction

An image of a city bridges the built environment and people's perceptions [1]. Such an image is a key to understanding and interpreting the urban form with the elements that can be adopted for organizing the design plan. Lynch's work has extensive implications for urban design. Analysing city images and their qualities is a tool for urban designers to assess the deliverability of their spatial intervention from blueprints to daily experience. However, the actual city images perceived by humans might vary greatly across demographic groups and even across individuals from time to time. Even city images for a certain group of people are hardly the same as supposed in the model of city images by experts from the top down. Therefore, decoding the complexity of the collective city image

is a requirement for an in-depth understanding of the mental translation from cities to city images.

Emerging volunteered data are now becoming a new normal approach to sense cities. As an image-based type of volunteered geographical information (VGI), user-generated photos from various social media provide a novel way of recording people's perceptions of cities and their daily lives embedded in them, complementing the conventional ways of surveying people's spatial experience, e.g., recording mental maps and conducting questionnaires. In the established studies, the spatial and semantic information in the volunteered photos has been validated as the new, useful substitutes of city images or tourist destinations and their spatial distributions [2–4]. These efforts mainly tried to measure the dominant imagery elements or to detect the spatial clustering of photography behaviours but neglected the interactions between elements forming the whole images. The city image is informatively rich, formed and reformed from place to place by various imagery elements, rather than the dominant elements only. Decoding these interactions between elements, therefore, is a necessity of understanding the complexity of the formation of city images and of designing a more meaningful built environment for citizens.

City image is complex due to the varying relationships between elements. Given this, the aim of this research is to uncover the composition of city imagery elements in people's perceptions more comprehensively. It is achieved by using a data-driven framework equipped with image recognition models and geotagged photos reflecting urban scenes. The interactions between city imagery elements in this work are four folded: co-presence (among elements), hierarchy (of elements), heterogeneity (across space), and differentiation (between groups). The co-presence dimension reflects the interdependency between any two elements in photos. The hierarchy dimension records the emergence of the principal components of city images with different combinations of individual or groups of elements. The heterogeneity dimension denotes the borders of imagery areas with adaptive density thresholds reflecting varying local levels of spatial discrepancy. The differentiation dimension replies to the changing properties of city image across various social groups. These four dimensions are essential for addressing key issues in urban design research, e.g., for a given city or area, what are the imagery elements and their interrelationships in people's real perceptions? How are they configured as the main groups of imagery components in different scenes that people are willing to record? Where are the boundaries of imagery areas with proper annotations? Who are the groups preferring some city images with specific features? By answering these What-How-Where-Who questions, this research produces an evidence-based action plan for city image research and relevant urban design practice, which demonstrates a data-driven, comprehensive, and reproducible solution for the employment of image-based open-source [5].

This research introduces a fully data-driven framework to decode the collective city images that are documented in volunteered photos. The first part is image element detection. It identifies the various types of elements in the recorded images by using image recognition models, classifies these elements in accordance with their functions in the models, and maps the interrelationship among those elements as a network pedigree. The second part is the delineation of the imagery areas. By using the density gradient principle, it captures the imagery areas where photo-shooting behaviours agglomerate, validates them with named areas, and discriminates them with the typologies of the combination of different elements. The third part concerns data sensitivity. It tests the extent to which different social groups can be distinguished by their preference for city images and discusses the necessity and the risk of data aggregation.

The structure of this research is organized as follows. The next section reviews the established efforts and background before, and the methodology section delivers the research design, data, and study area in Beijing. The fourth section reports the results of empirical studies, and the last section delivers the concluding remarks and discusses the future steps to take.

2. Background

2.1. Previous Studies on City Image

People's perception of the built environment has been the main topic for urban designers, planners, and decision-makers. From an architecture research perspective, human experience in spatial form is assumed to fit together with morphological properties. The built form, therefore, is interpreted as a composition of structural elements in different types that shape the 'image' in people's minds. With a cognitive map survey of residents in three American cities (Boston, Jersey City and Los Angeles), Kevin Lynch summarized the "five-elements" schema of urban perception in his benchmark book, The Image of the City: paths, edges, districts, nodes, and landmarks [1]. Subsequently, Appleyard redefined cognitive maps into two main types: sequential and spatial. Sequential elements contain four subtypes, namely, the fragment, chain, branch/loop, and network, while spatial factors include the scatter/cluster, mosaic, link and pattern [6]. Morello and Ratti proposed a more precise framework of Lynch's visual elements with 2D and 3D isovists [7]. The framework of city image was applied to many cities across Europe, America and Asia [6,8] and to the city images perceived by various groups of populations, e.g., residents, tourists, and commuters [9,10]. Therefore, the structural perspective is a key for understanding and designing the image of cities by (re)locating the city elements in different types.

Studies on city images from a structure or typology perspective have provided possibilities for urban designers and relevant decision-makers to estimate people's perceptions of cities. However, mapping the elements of city images from a bird's-eye view does not always conform with the real perceptions of non-professionals or the ways they express their observations, thereby leading to inaccurate estimations [11]. It is confirmed that city images vary across different social classes, ages, genders, educational backgrounds, ethnic groups, etc. [12–14]. Due to the existence of this inter-group difference, it is difficult to calibrate a common, objective image for all [15]. Moreover, as a representation of personal perceptions, cognitive maps can hardly avoid omissions and distortions of spatial information in complex cognition processes [16]. This demonstrates the difficulty of validating city images and the complexity regarding how they are formed.

With the development of complexity theory since the 1980s, structural element-based studies have been criticized for overemphasizing the physical aspects, but neglecting other essential dimensions of city images, leading to the popularization of simple geometries in design rather than sophisticated spatial solutions [11]. It is noted that the image of a city is multifaceted, involving spatial dimensions (e.g., structure, function, quality, changes, sense of place) and socioeconomic, ecological, and cultural dimensions (e.g., history, prosperity, civic life, local customs, and natural environment) [11]. Moreover, as a key concept mediating the sense of place and the sense of occasion, the city image was argued to be related to temporary activities showing people's engagement in various events. In the last decade, the non-spatial dimensions of city images have been a new focus for destination image studies in the fields of environmental psychology, tourism management and planning, and city marketing. In these works, city images are normally decomposed as a series of factors that impact their emergence, namely, the image components or dimensions [17–20]. Gartner postulated that destination images are formed by cognitive, affective, and conative components [18]. Stern & Krakover introduced a conceptual model of image formation based on literature analysis comprising seven determinants: urban aesthetics, distance, level of activities, population size, population trait, climate, and residential appeal [21]. Nasar delivered the evaluation factors of city images and highlighted the importance of two measures: the imageability and likability of city images [8]. The former is three-fold with distinctiveness, visibility and use/symbolic significance, while the latter embraces five aspects: naturalness, upkeep/civilities, openness, historical significance and order. Anholt identified a 6-P model of city image: presence, place, potential, pulse, people, and prerequisites [17]. Luque-Martinez et al. constructed an overall model of the city image formation and evaluation with 12 dimensions: architectural and urbanistic attractiveness, transport and communication infrastructure and traffic, historical heritage, environment,

social problems, culture, innovation and business culture, economy and commerce, range of services, education-university, international projection, and citizen self-perception [22]. These efforts uncover the hidden aspects of city images with statistical tools, e.g., component analysis and regression, demonstrating the complexity of city images. However, these researchers usually neglected that the concept of the city image is inherently visual [10]. This factor-based perspective, therefore, is parallel to the structural element-based scope to measure the aspects of city image, showing the necessity of interconnection.

Enabled by information technology, an 'image society' is now emerging with a new fashion of digital photography shared via various social media platforms. Within such a society, people dynamically read the city by many formats of images. Given this, a city is not only a built image as Lynch has argued but also a graphic image. Urry introduced the tourist gaze theory and reclaimed the importance of the nature of seeking new visual experience as the 'gaze' at different 'signs' [23]. The analysis of 35 million Flickr images indicated that many of the photos taken by visitors were what had already been frequently photographed [24]. Instead of looking for surprises, visitors prefer to reproduce pictures confirming the expected images of the city [9]. Photos are the visual products of selecting, shaping and structuring elements of the physical environment to reflect the photographer's mental images [25], and they are the condensation of destination images [26]. In recent years, as the crowd-sourced photos have offered a wealth of information about people's perception of the city, a new framework based on imagery elements and areas has gained popularity in city image research (Table 1).

Table 1. Three Types of Research Frameworks of City Image.

Aspects	Structural-Element-Based Study	Factor-Based Study	Imagery-Element-and-Area-Based Study
Representative research	City image [1,6,7].	City image and destination image [8,17–20].	City image and tourist gaze, etc. [1,23].
Conceptual nature	The generalization of the "built image", i.e., basic types of spatial carriers of city image.	The analysis of complex factors related to city perception, i.e., influential factors and evaluation factors.	The analysis of the "graphic image" and its spatial distribution, i.e., people's "gaze" to the city and its concentrated areas.
Cognitive subject	Mainly residents.	Mainly tourists and investors.	Multi groups including residents, commuters, and tourists.
Cognitive object	Built environment.	All aspects including physical environment, human activities, economic development, etc.	Built environment, natural landscape, and human activities, etc.
Cognitive form	Mental maps.	Mental images and descriptions.	Mental images and mental maps.
Cognitive View	Bird's-eye view (urban planners' perspective).	Abstract view (psychologists' perspective).	Human's-eye view (publics' perspective).
Data	Sketched maps, photo identification survey, etc.	Likert-scale-based questionnaire or interview results.	User-generated geo-tagged photos.
Methodology	Cognitive mapping.	Literature analysis or statistical analysis.	Data-driven methods, e.g., image recognition, spatial clustering, etc.
Image Reconstruction	Mainly through blueprints of spatial structure planning.	Mainly through city marketing and tourist publicity strategies	Virtually through images in the media or physically through planning strategies.

2.2. Related Works on Geotagged Photos

On the current social media platforms, massive amounts of photos are shared voluntarily by smart device users with geo-references [27]. As a type of socially sensed data, these volunteered images are basically individual-based with large spatial coverage, a fine resolution, and a large subset of them is city-related. This facilitates studies on collective, fine-resolution, and multi-dimensional city images across the population and their distributional effects and meanings.

The geo-references of crowd-sourced photos can scale cognitive mapping to a large coverage without forfeiting resolution [28,29]. Many tourism studies worked on hotspot and landmark detection through the spatial concentration of geotagged photos or tweets [4,14,30–32]. For instance, Kennedy et al. firstly employed a spatial clustering method to identify landmarks and events through geo-tagged photos, and then used a location-driven approach to generate representative tags for these landmarks [2]. Crandall et al. applied a non-parametric clustering method named mean shift to recognize landmarks within several cities at a global scale [24]. Jankowski et al. presented a spatio-temporal analytic approach to discover landmarks and movement patterns from Flickr photos [33]. Their results helped to distinguish between sites that are occasionally popular, and sites known as city landmarks. Ji et al. used a spectral clustering approach to identify landmark regions and then mined representative photos [34]. Moreover, Liu et al. proved that Lynch's "five elements" can now be partially identified from the spatial distribution of geotagged photos [3].

The textual and visual contents of crowd-sourced photos can reflect what people prefer to perceive in cities and thus contain abundant information about city images in their minds. Regarding textual contents, Rattenbury et al. proposed two novel approaches, TagMaps and scale-structure identification, to extract place or event semantics from Flickr tags automatically [35,36]. Dunkel proposed a visualization method to map tag features of Flickr photos and evaluate city perception [37]. As to visual contents, Kennedy et al. demonstrated a location-tag-vision-based approach to extract representative images of landmarks [2]. Papadopoulos et al. performed clustering on image visual and tag similarity graphs by means of community detection to automate the detection of landmarks and events [38]. Using Flickr data, Miah et al. proposed a conceptual framework for tourist behaviour analysis comprising four parts: textual meta-data processing, geographical data clustering, representative photo identification and time-series data modelling [39]. Recently, computer vision models have been used to successfully localize multiple urban objects or classify urban scenes based on a group of themes [40–42]. For example, scene recognition techniques can reveal the specific type of shooting place to reveal the spatial distribution of different imagery elements [3,43–45] and investigate the characteristics, similarities and differences among cities [3,14,43]; discriminant clustering and image object detection are used to extract and map visual elements of local characteristics from photos [46,47]. The applications of computer vision techniques with volunteered, geotagged photos in social media are providing new perspectives about people's images of specific places or areas with the advantages of wide coverage, instant updates, vector-based resolution, and an individual focus.

How to use the user-generated photos for interpreting the properties of city images properly and comprehensively is still an incomplete mission with ambiguity. Most of the established research captured the dominant city images and their spatial distributions by extracting the semantic information of geotagged photos but failed to address the interrelationships between elements. This constrains relevant applications in real urban design processes with far more elements than normally defined by conventional urban design studies. The image of city is not only formed by the elements with the highest frequency in scenes and density across places, but the configuration in which all elements are interrelated and organized in people's perception [48]. Understanding cities as systems that are composed of elemental interactions is a valuable way to measure, present and interpret urban complexity [49]. Therefore, the key to explaining the complexity of city

image is interactions between elements, and which is a tool that can be used to explore how we as planners engage in science of design on city images.

3. Methodology and Data

3.1. Framework

Cities are complex systems with elements and their interactions, which are 'the hidden hands' help these systems retain their own integrity from the bottom up. The interactions between elements then are complex as well in multiple dimensions. By conceptualizing a geographical system at time T as an Element-Space-People (ESP) structure, we define crucial dimensions of the interactions according to the ways how they are projected and mapped for proper analysis and interpretation. For a given city system in a time cross-section, its ESP structure can be mapped and analysed by elements, space, and people. In other words, a cross-sectional ESPT structure can be analysed and interpreted from the element-based, space-based, and people-based scopes. From the element-based scopes, there are two basic dimensions between and of elements, namely, co-presence and hierarchy, showing structural semantics of the interactions between various elements, which are pairwise and groupwise, respectively. The space-based and people-based perspectives demonstrate the heterogeneity and differentiation dimensions, recording the spatial and social semantics of these interactions, respectively. The interactions between elements, therefore, are structurally, spatially, and socially distinct from city to city. All the four basic dimensions are complementary to one another and none of which can be arbitrarily neglected or substituted so that a comprehensive understanding of the complex interactions between elements can be obtained.

Addressing the dimensions of co-presence, hierarchy, heterogeneity, and differentiation is imperative for understanding, applying, and interpreting cities with volunteered geographic information. Volunteered geographic information, as a typical, geo-referenced open data, is new for its structural information, providing more metadata into the conventional geographic information. This is very true for the image-based and text-based VGI in which geographic information is restructured as visual and literal contexts beyond the spatial [50]. VGI is also constrained by urban space due to the fact that people's behaviours are naturally heterogeneous from every place to one another. VGI suffers the data quality issues that are very common for many social media data, as various portals might filter some user groups out. So, considering those four dimensions is a necessity of using VGI data appropriately for the city image research and for other studies on relevant topics in a similar data environment. In so doing, we can suit our interventions for better city images with more effectively bounded elements to distinctive conditions in terms of locality and persons involved. In short, co-presence, hierarchy, heterogeneity, and differentiation are four dimensions of VGI that we should study, and that we could study due to the emergence of new data and open-source geospatial science [5].

The workflow in this research is stepwise (Figure 1). The raw datasets of volunteered photos were gathered from various social media platforms and then aggregated to reduce potential bias as much as possible. In the step of data pre-processing, the raw photos were all recognized as scenes, and then the invalid images were removed. For scene recognition, a well-established trained dataset is required for the convolutional neural network (CNN) model. This research then employed the trained data compiled by MIT with over 1.8 million images labelled with 365 scene semantic categories and 10 main types, comprising a large and diverse type list of environments to be used for scene classification and other visual recognition tasks. Among the different open CNN models trained on Places 365, the 152-layer residual neural network (ResNet152) outperformed the others with more than 85% accuracy in detecting the top five scenes for every image that was labelled [51]. ResNet152 was used in this study to produce the outputs of the top five most likely scenes and their corresponding likelihoods. Since some geotagged photos on social media are irrelevant to city images or it is difficult to identify their scene types (e.g., close-ups of faces, objects, and sky), data cleaning and categorization refinement are necessary [3]. The scene here is

the photo with only one type of imagery element; accordingly, the scene probability is then the likelihood of one imagery element's presence. In this work, we plotted the recognition accuracy against the recall rate to calibrate a combination of thresholds to filter invalid images out. The outputs of data pre-processing are packaged in a geodatabase where the point-based records are stored with the features documenting the recognized probabilities for 365 imagery elements as defined in Place 365, the coordinates where the photos were taken and the social media platforms from which the data were obtained.

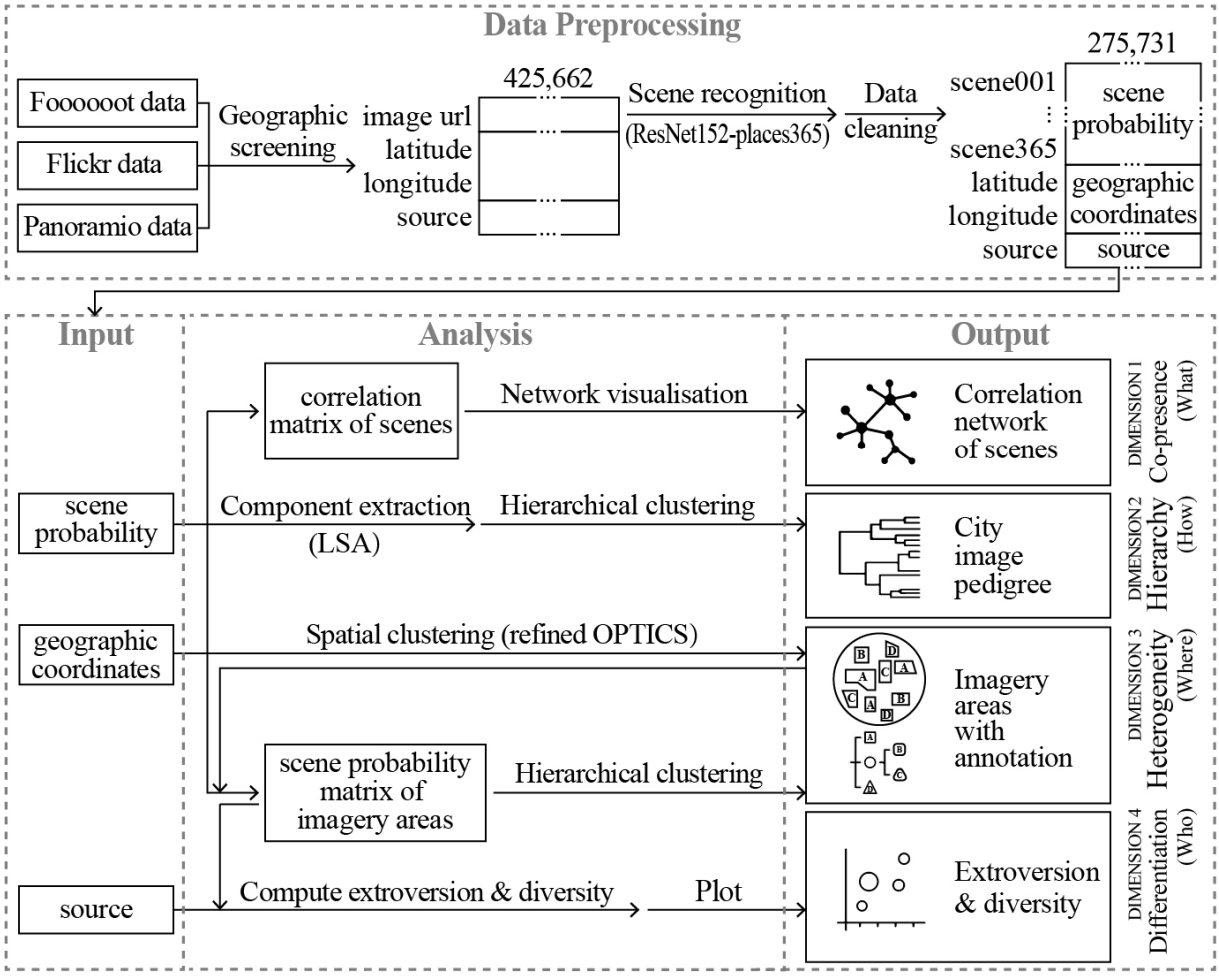


Figure 1. Research workflow to unfold multi-dimensional interactions between the elements of city images.

In the second step, the point-based geodatabase is used to compute and present various dimensions of interactions between imagery elements. In this research, we demarcate four essential dimensions: co-presence (among elements), hierarchy (of elements), heterogeneity (across space), and differentiation (between groups). Specifically, the co-presence dimension reflects the interdependency between any two imagery elements in the given set(s) of photos, measured by the co-existence likelihood between, and represented as links in a correlation network. It is produced by generating a correlation matrix of all imagery elements. The result can be helpful to answer the 'what' question in urban design–what are the twin-like imagery elements that should be put together? The hierarchy dimension captures the emergence of the typical, unrelated, main groups of elements, namely the principal components of city images, combining the imagery elements that are possibly correlated. It is generated by using a truncated SVD model with hierarchical clustering and represented as a dendrogram showing the pedigree of city image from the basic elements to the hybrid components. The output reflects the 'how' question–how are multiple imagery elements grouped as themes and then distinguished from others?

The heterogeneity dimension sketches the natural boundaries of the imagery areas -the hotspots of geotagged photos with adaptive density thresholds. It is achieved by the refined OPTICS model that has been proven to be more effective for detecting spatial agglomeration with varying densities than the DBSCAN model with an optimized number of clusters [52]. The clusters are then annotated by the principal components that are detected. Relevant outcomes in this dimension provide references for responding to the 'where' question–where people are more likely to record the city images? The differentiation dimension investigates the preference on imagery elements varying across social subgroups as reflected in different data sources. We introduce the extroversion index as a measure of differentiation by estimating the share of non-local subset data in the whole dataset. Diversity index in an entropy form is used to measure the informational richness or non-randomness of elements within each photo. They are plotted as two axes in a scattergram to represent who are the groups preferring the city images with more information than others–the 'who' question. This dimension, in turn, is a validation of the necessity of data aggregation for reducing potential bias.

3.2. Study Area

The proposed methodology was applied in an empirical study on the Central Beijing, the Metropolitan Areas of Beijing (MAB), the capital of PR China (Figure 2). The MAB covers Beijing with six districts, Xicheng, Dongcheng, Haidian, Chaoyang, Fengtai and Shijingshan, where the social media data are agglomerated. One-twelfth of Greater Beijing accounted for nearly half of the data, which helped to avoid data sparsity in other areas where fewer people visit. Moreover, the MAB covers various typical city landscapes in Beijing, from the city to natural geographies and from gardens to hills. The representativeness of this study area ensures that the outputs can sufficiently reflect Beijing's city image.

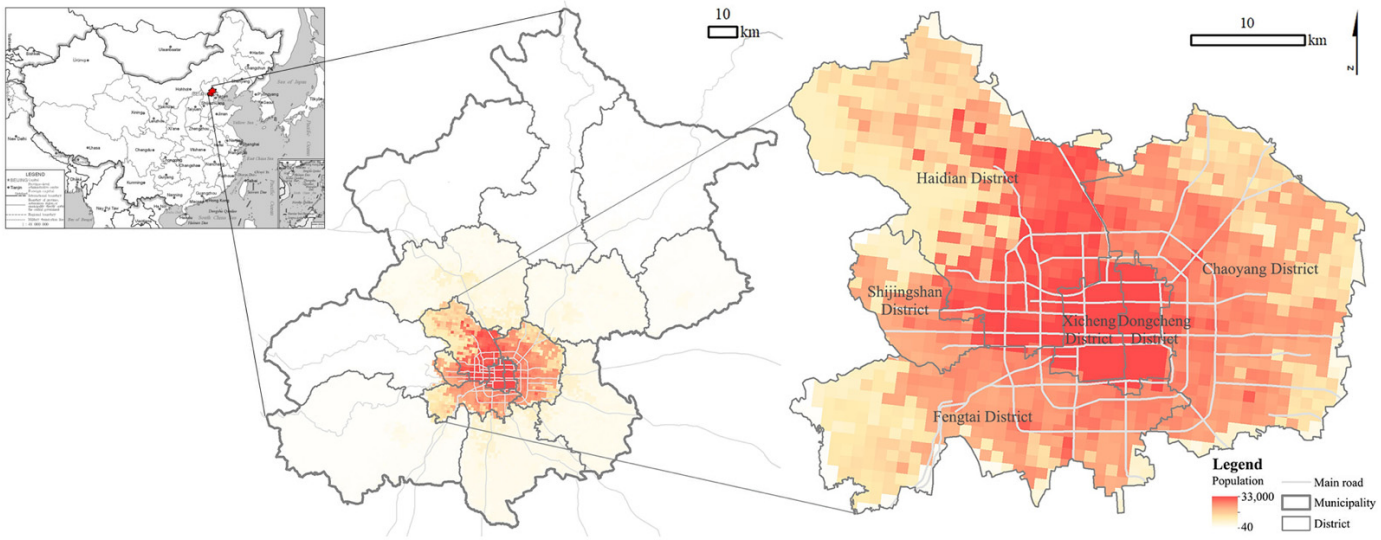


Figure 2. Location and population distribution of Central Beijing.

3.3. Data

The data used in this research are volunteered and georeferenced photos from three different social media platforms. The reason why we used the data from different panels is to avoid sampling problems, e.g., data spatial scarcity and potential bias. In fact, any specific social media might focus on certain type(s) of users that are normally subsets of the total sample. Therefore, one safe strategy is to aggregate the data from different channels to approach the common reality. Here, the data were gathered from three social media platforms: Fooooooot, Flickr and Panoramio. Fooooooot is a GPS travel community and trip sharing website designed to facilitate travellers displaying and sharing their travel

footprints and stories. There were 248,826 valid photos with coordinates that were located within the study area, which were obtained in September 2018. Flickr is a Yahoo property that is among the largest photo-sharing websites. Panoramio is a geotagged photo-sharing website owned by Google that contains only landscape photos. The Flickr data in this study came from the open-source "YFCC-100M" dataset, and the Panoramio data came from the MIT "City Perception" dataset [52], with a total of 123,933 Flickr images (as of 27 April 2014) and 63,594 Panoramio images (as of 2 November 2016). Among all three datasets, the Fooooooot data are relatively domestic, whereas the users of Flickr and Panoramio data are more likely to be non-local.

The size of the raw, aggregated, volunteered data in the study areas is 425,662 images. There are 275,731 valid images, accounting for 64.78% of the original data after data pre-processing. We used a stepwise procedure to filter out the invalid data. (1) Initial recognition. We conducted an initial scene recognition to assign metadata to the geographic information for describing the visual content. The metadata includes two categorical layers: 365 subtypes and 13 main types with names and associating probabilities as defined by the Place 365 model. (2) Non-urban photos removal. We removed the non-urban records from the raw data, e.g., selfies, sky-only, animals, etc., and those classified as working or housing types with very small percentages. So only 11 urban-related urban main types were kept. (3) Non-representative photos filter. There were a few records wrongly geo-tagged. So, we deleted those that were nearly completely different (99%) from their neighbouring records at a 50-m radius. (4) Cross-validation. For validating the data and securing the scene recognition is close enough to our knowledge, we built a testing dataset with 1000 photos that were randomly selected from 11 main types and used it to obtain the recall rate to validate the training results.

Users using different social media platforms have an obviously differentiated preference for urban scenes (Figure 3). The local citizens, as defined as the Fooooooot users, took more photos of natural scenes, but experienced fewer activities than the users of Flickr and Panoramio did. However, some consistency still exists. The shares of historic heritage sites, modern buildings, transportation, water, etc., are high across all subsets of datasets. The complementation between these datasets suggests the necessity of data aggregation for a comprehensive understanding of the city image for most social groups.

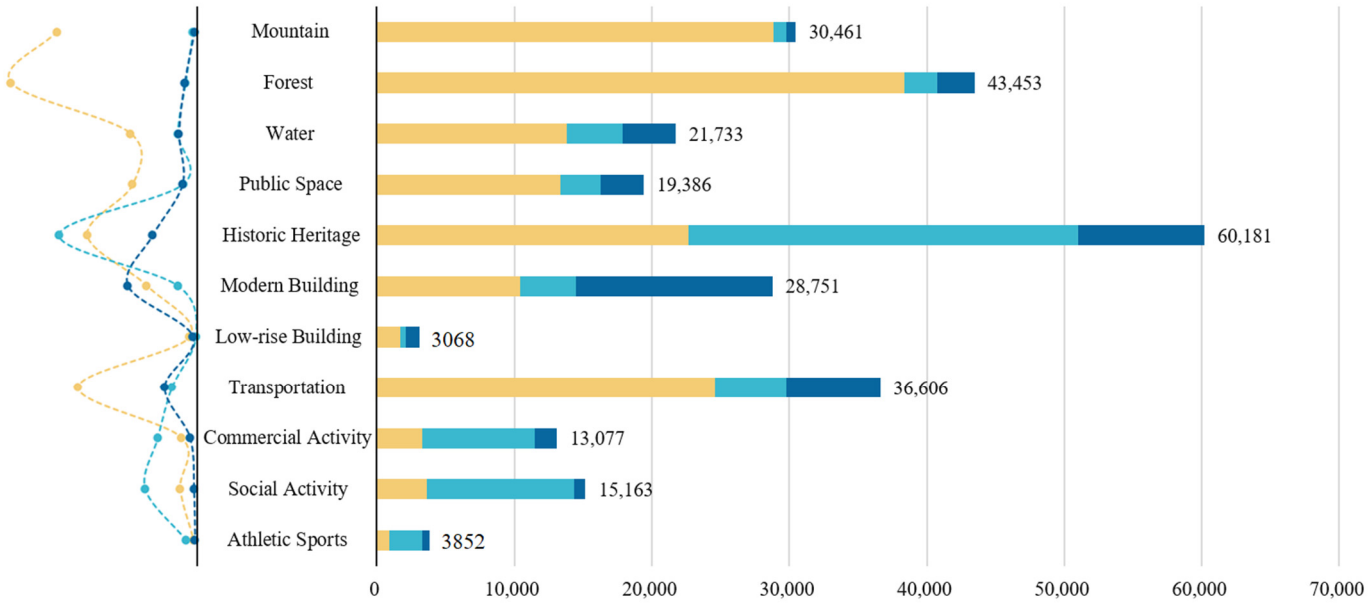


Figure 3. Composition of the image datasets in Central Beijing across eleven main scenes.

3.4. Setting and Evaluation

Validating results is vital for proving the effectiveness of the methods and the correctness of the results. Nevertheless, it is not always easy for the exploratory data mining tasks in urban studies since the results are indicators of urban realities rather than the facts. This is very true for some useful indices, e.g., geographic accessibility, which cannot be proven to be correct due to the absence of the associating grand truth. City image is something alike. So, our strategy for validation is two-folded: one is based on statistical optimisation to define the thresholds that are required so that the exploratory outputs are more statically significant and the other is based on a comparison with urban reality patterns.

The whole process in this research is data-driven. However, we still need to define thresholds that are used. The first is the threshold of the scene probability for filtering out invalid data. By plotting the accuracy and recall rate as percentages, we selected one pair of turning points along the curves, maximizing these two aspects of model performance. Therefore, the probability threshold was calibrated as 0.44 with a recall rate of 81.81% for the testing dataset (Figure 4). Another parameter that must be calibrated is the reachability distance parameter in the OPTICS algorithm. This parameter is used to control the size of the clusters when the density threshold is relaxed. One challenge for adaptive density-based clustering techniques is to balance the density and size effects, avoiding the underestimation of the size of the clusters with higher density and the overestimation of the size of those with lower density. By comparison with several well-known named areas in Beijing, we selected a threshold reachability distance of 0.01 to define the cluster boundaries. These thresholds were proven to be optimized than others in various criteria and the results were validated preliminarily.

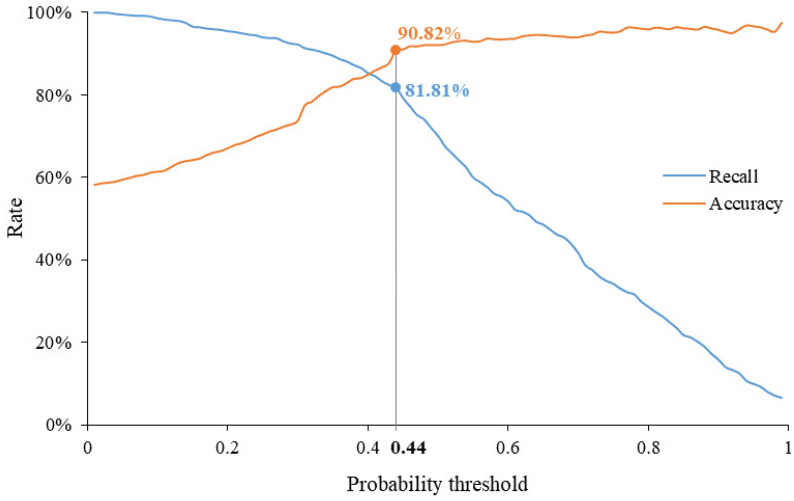


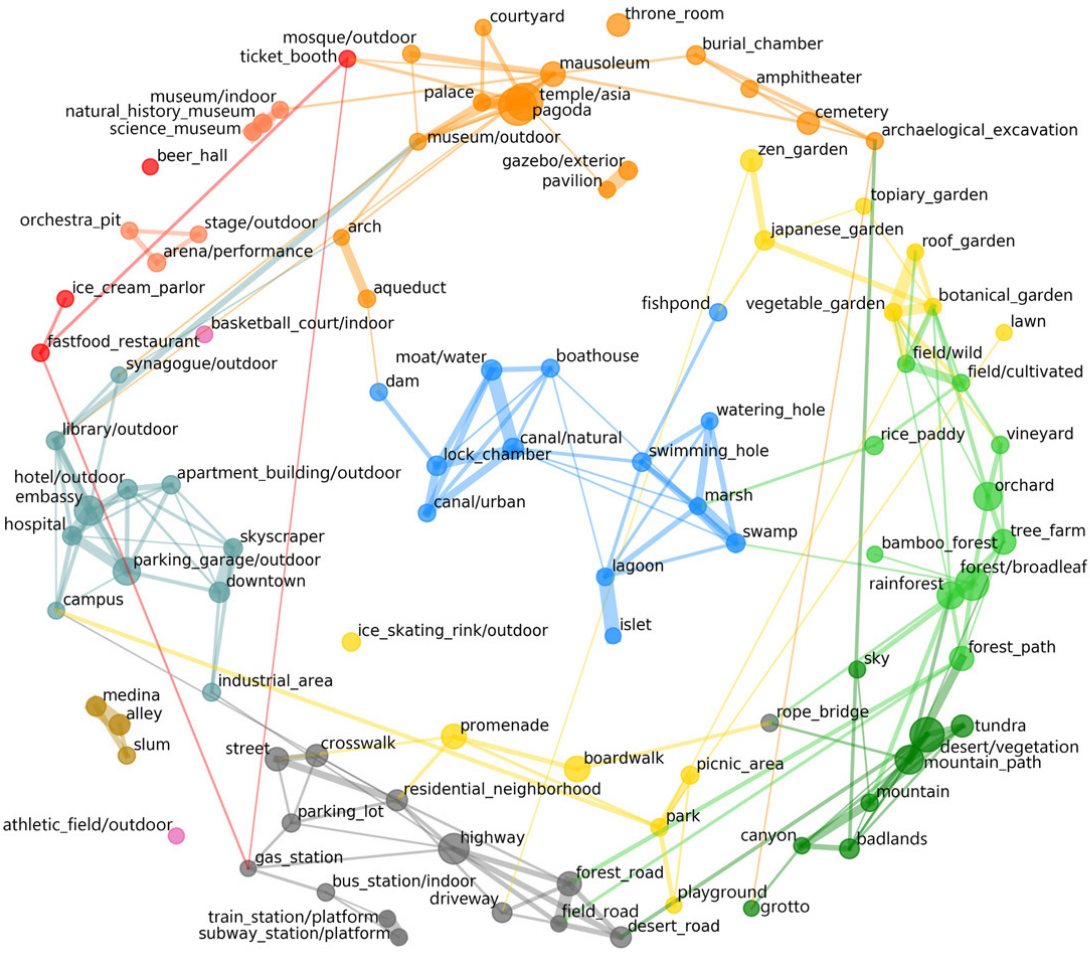
Figure 4. Curves of the recognition accuracy and recall rate of valid data for the testing data.

4. Results

4.1. Co-Presence between Imagery Elements: Interdependent Elements

Imagery elements are not independent layouts in urban reality but are interdependent in shaping city scenes from place to place. Understanding of the interdependent elements in any given area can be manual for designers to reshape more meaningful images than those using those unrelated elements. The association among the detected imagery elements of Beijing is mapped as a correlation network. In the detected imagery elements, the top three elements of the images of Beijing that are most often captured in people's photos are "pagoda" (4.46%), "temple/Asia" (4.14%), and "desert/vegetation" (3.43%). These elements portray the historical characteristics of this ancient capital and landscape features due to the dry climate in North China. Figure 5 illustrates the similarity network encompassing the top 100 imagery elements with stronger linkages that were recognized. The links between imagery elements either reflect their co-presence in the photos that

were taken by the citizens, e.g., "library" and "museum", "park" and "campus", and "vegetable garden" and "field/wild", or reflect their co-occurrence, e.g., "industrial area" and "highway", "rope bridge" and "forest/broadleaf", "fishpond" and "Japanese garden", and "archaeological excavation" and "badlands". These interdependent elements document the important pairs of imagery elements, which provides additional information on the element ranking in terms of frequency. It is noticed that the elements ranked higher are also more likely to be co-present with other elements that are also interconnected. These co-presence interrelationships, however, exist between the elements with various levels of frequency, showing the necessity of considering the interdependency as a characteristic of city images.



Rank	Imagery Element	Frequency
1	pagoda	4.46%
2	temple/asia	4.14%
3	desert/vegetation	3.43%
4	forest/broadleaf	3.23%
5	highway	2.60%
6	embassy	2.31%
7	mountain_path	2.21%
8	orchard	2.11%
9	parking_garage/o…	1.95%
10	rainforest	1.90%
11	boardwalk	1.65%
12	promenade	1.52%
13	tree_farm	1.51%
14	mausoleum	1.44%
15	forest_path	1.42%
16	forest_road	1.39%
17	street	1.27%
18	throne_room	1.14%
19	cemetery	1.08%
20	crosswalk	1.08%
21	tundra	1.06%
22	zen_garden	1.03%
23	residential_neigh…	0.90%
24	desert_road	0.87%
25	alley	0.87%
…		

Figure 5. Correlation network of the top 100 imagery elements of Central Beijing (Node size represents the frequency of recognized elements. Link thickness represents the correlation between two interdependent elements).

We colour the nodes and links according to 11 main types pre-fixed in Place 365 model showing the communities that each element might belong to. The correlation networks are normally very dense inside than outside. The intergroup connections, then suggest the composition of Beijing's unique city images. For instance, water-related elements are tightly connected to historic heritage sites, gardens, and natural landscapes. Commercial activities are more likely to be present with historical heritage sites and transportation services. It is worth noting that some elements drift away from the main network, resulting in their isolation. Slum-like areas with alleys intertwining among low-rise buildings are a relatively unique type of city image. Some predefined main types of elements are split into various clusters that are proximal to different elements. The elements in the type of public space

are mainly divided into two groups: one is more tightly connected to the forest group, and the other is closer to the transportation and modern buildings.

The co-presence network between imagery elements might vary across places. For a clear comparison between different areas, Figure 6 records the correlation networks of 11 main types of imagery elements of Beijing disaggregated by administrative districts. These maps show a series of schematic diagrams of the raw graphs. It is apparent that the tendency of the inter-element interactions varies from area to area. These sub-graphs further concretize the patterns of the overall graph, illustrating the spatial difference of the city images and their compositions. More specifically, Central Beijing, Xicheng and Dongcheng are three districts with a strong city image of historic buildings and its co-presence with commercial and social activities. Haidian and Shijingshan are those two areas where urban greenness relevant elements are dominant with very few connections to other elements, despite that the interaction between the historical buildings and commercial or social activities is still significant. Chaoyang and Fengtai are perceived as images with more modern buildings and transportation elements than others. These districts can also be distinguished from each other by looking at the varying co-presence between two given elements. For instance, public-space-related elements are more likely to be together with water elements in citizens' photos in Chaoyang only than others. And in Central Beijing, modern buildings and social activities are more often in the same frames within people's perceptions than that in other districts. Disaggregating the correlation network of imagery elements is useful for uncovering the sub-structures of the city images of Beijing, which is featured by the nodes–the high-frequency elements, and the edges–the interactions between elements. The city image of Beijing, and its areas, therefore, is not the element list only, but the ways how the elements are constructed, designed, perceived, and recorded.

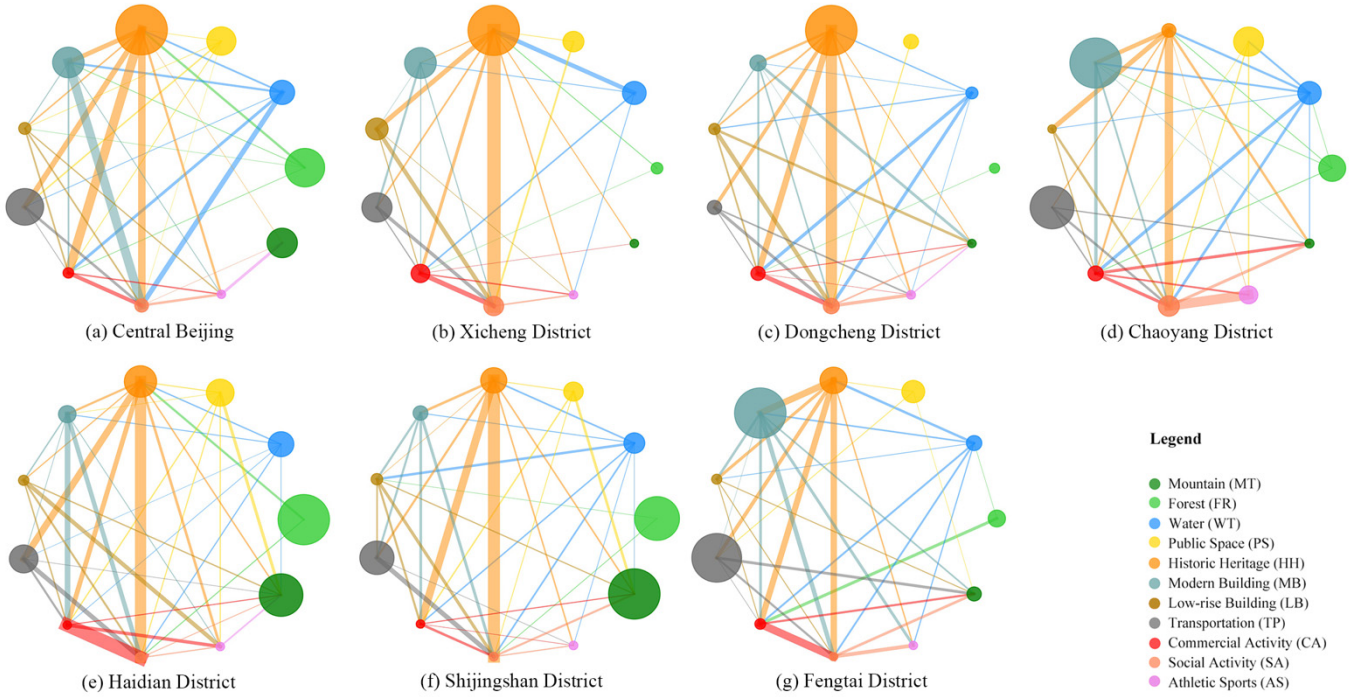


Figure 6. Correlation networks between 11 main predefined main types of imagery elements for different areas in Central Beijing.

4.2. Hierarchy of Imagery Elements: Principal Components and Main Typologies

Components are a summary of multiple elements and a record of the emergence of city image typologies. It represents the interaction among multiple elements rather than two, as shown in the co-presence dimension, showing the main types of imagery elements from the bottom up. By mapping the dendrogram, we unfolded the pedigree of the component extracted from every detected element (Figure 7). In our model, 80 principal components (PCs) were detected through the truncated SVD model, in which the absolute values of probability are input, and these principal components can explain 83.64% of the raw variation in the detected elements. For annotating each PC, we use the dominating elements with the top 2 loading as the first and second elements. These PCs can be grouped into 7 typologies according to their rank in terms of similarity: urban space (24.61%), natural landscapes (22.29%), historical architectures (14.82%), water (8.05%), sports & social activities (7.50%), catering (2.07%) and others (4.30%). This is different from what has been pre-defined in the Place 365 model regarding with the number and inclusion relations. The first and second elements for each PC demonstrate that the PC detected is a new organization of 365 elements.

For Beijing, urban space is the main city image, a composite of public buildings, streets, alleys, pedestrian paths, gardens, and high-raised buildings. The images of historic heritage sites highlight scenes with pagodas, temples, and palaces. The natural landscape is characterized by the presence of desert vegetation, broadleaf forests, tree farms, boardwalks, etc. Human social activities are also captured as some type of image suggesting the ways in which people interact with the built environment are the images in people's minds. This family of images highlights sports amenities, conference sites and catering locations. It demonstrates that city images are shaped not only by the urban form but also by people's engagements that are also scenes for others.

It is noted that there are some elements in the city image pedigree that are in pairs with different signs, showing their functions of discriminating scenes that are very similar in all other dimensions. These elements are, therefore, named as discriminate elements. Eighteen pairs of discriminate components were detected, as listed in Figure 7 (also highlighted in red in Figure 8). We selected the most representative photos for every element. It is obvious that most of those elements are very similar to each other but are also highly likely to be spatially proximal. For instance, in cities, crosswalks and streets are normally spatially interconnected, but they might be fully different scenes for pedestrians. These components, therefore, are the basic elements to distinguish one image from another at the very bottom level in the hierarchical structure of complex images, as shown in the pedigree tree. It is worth noting that these components might be case-wise, varying from city to city, suggesting alternative information to the formation of city images as the similarity between elements has shown.

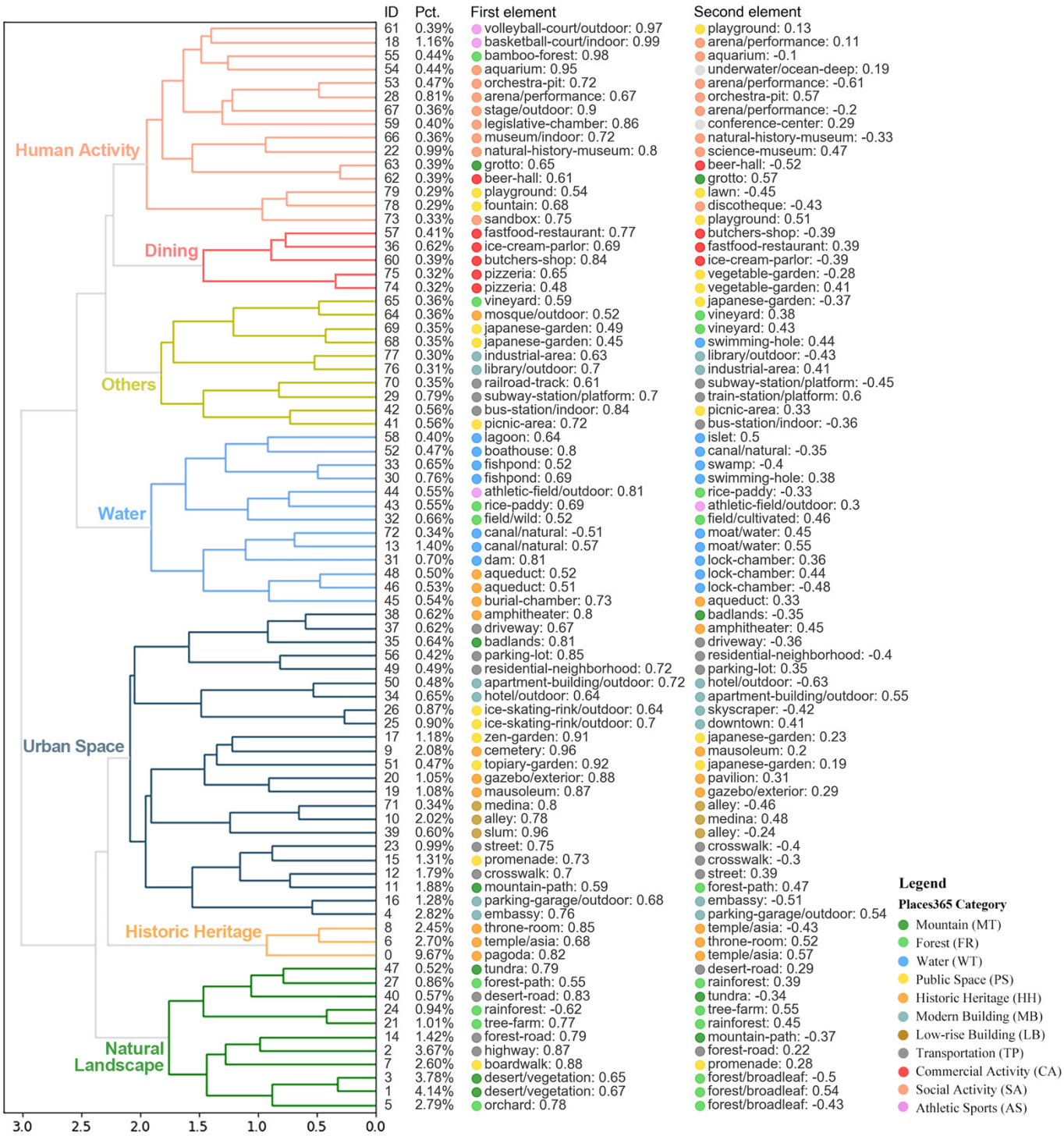


Figure 7. City image pedigree of Central Beijing: 7 main types extracted from 80 principal components with first and second elements having the largest loadings.

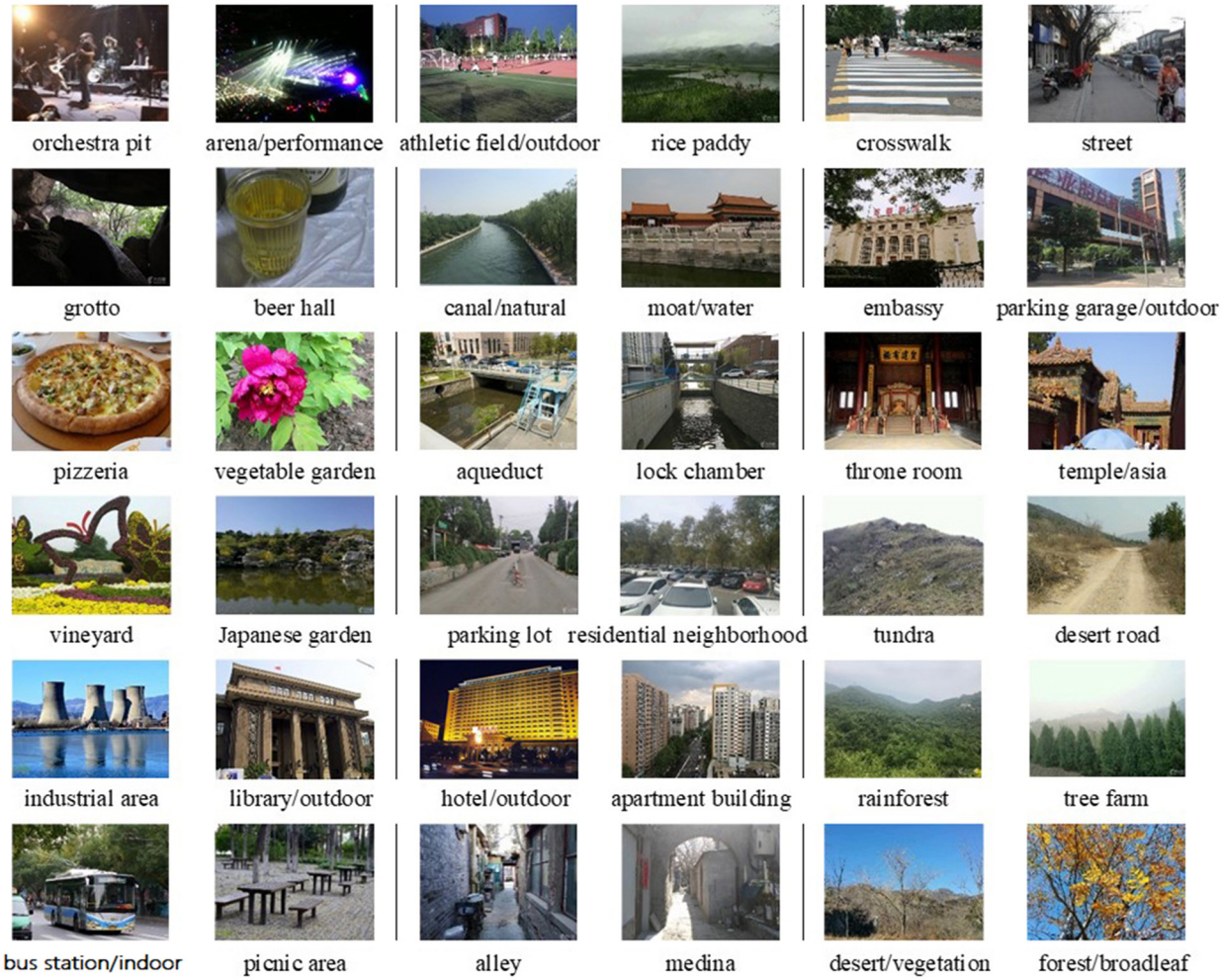


Figure 8. Discriminative components in Central Beijing.

4.3. Heterogeneity of Imagery Elements across Space: Imagery Areas

The morphologies of imagery elements are geographically heterogeneous, whereby shaping the imagery areas where citizens are more likely to record the city images they see than other places. There are 29 imagery areas detected in Central Beijing, accounting for 76.32% of the volunteered photos, as shown in Figure 9. The map shows a heterogeneous pattern where some imagery areas are agglomerated as larger, spatially continuous clusters. The dendrogram of the hierarchical clustering yields that these individual imagery areas can be grouped into seven main clusters according to the density variation of photo-taken behaviours (Figure 9). The first cluster is the core region (the central zones centred in the Forbidden City and surrounded by the Third Ring Road), where ten detected imagery areas are spatially agglomerated. Four clusters are moderately diverse, featuring several individual imagery areas, including the areas around the Olympic Village, Pan Summer Palace, Small West Hills, and Big West Hills. The Wangji area and Fengtai Science Park, however, are two clusters that are spatially independent from the context in which they are embedded. These results demonstrate that the spatial interaction between imagery areas varies significantly from place to place, and it might be at a higher level for long-lasting, historic, or developed areas but at a lower level for newly built destinations.

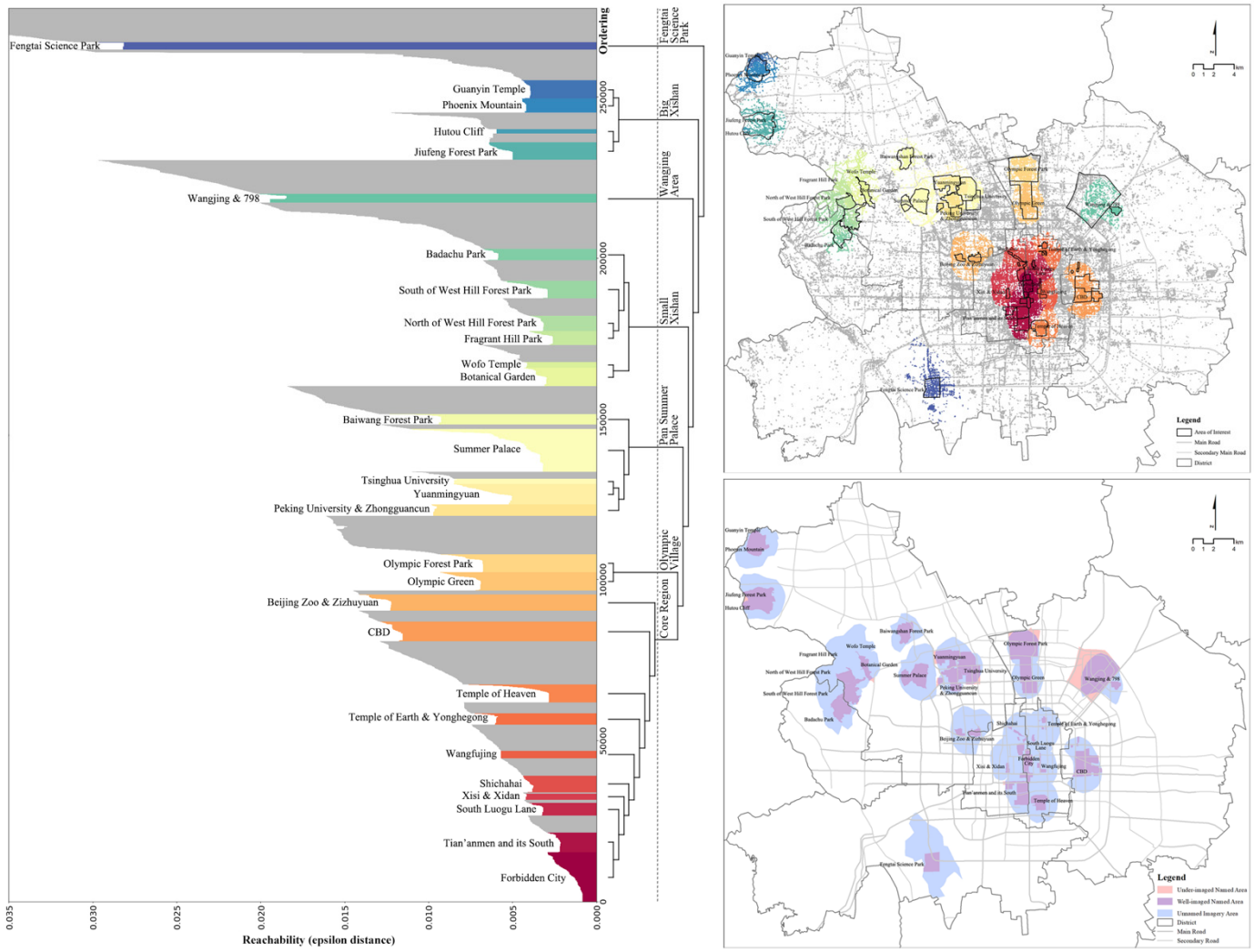


Figure 9. Imagery areas detected by the adaptive density thresholds produced by the refined OPTICS algorithm with reachability plot and dendrogram profiles.

The imagery areas are detected with clues to some well-known named areas in Beijing. The named areas are easily obtained from any navigation service provider by searching for the so call area of interest (AOI). It is a polygon-based identification of an area that someone may find useful or interesting. The effectiveness of these imagery areas is validated preliminarily. However, spatial inconsistency still exists. For most of the imagery areas, the spill-over effect from the imagery area cores is obvious: their boundaries are larger than the named areas. The directions in which the imagery areas from the named areas are not random. The imagery area around the CBD, for instance, extends towards the north and south rather than the east and west, and the imagery area around Wangjing is largely 'stretched' by 798 art villages. This reminds us that imagery areas delineated by the density gradient of geotagged photos might be the functional aspect of the named areas with a clear spatial definition. We can further define the areas as three categories of areas, according to the way the imagery areas and named areas are overlapped. The well-imaged areas are those AOIs fully recognized as imagery areas, which means that those areas are well-design and perceived. The named areas that are not detected as imagery places are annotated as the under-imaged areas where city images are not well delivered. The third areas are called unnamed imagery areas, denoting the places where people often visit and record the images, but they are not named as a space identity. This type of areas maps the spatial extents to which city images spill-over across places, can be very informative for urban planners who might require references for a new spatial plan. Within the unnamed imagery

areas, they can decide to assign similar, or complementary areas to enhance or change the imagery areas. In turn, we can monitor the transformation of these imagery areas and their relations to named areas, or social communities, and then capture where, when, and how imagery areas change across time due to an individual, or a series of planning projects.

For a better understanding of the composition of imagery elements, we clustered the imagery areas into 6 basic types that are featured with the main categories of elements detected in Figure 7 (Figure 10). These 6 basic types can be further grouped into two overarching categories, namely the urban and natural environment. The imagery areas of the urban environment are more diverse with more dining and human activities than that of the natural that are highlighted with natural landscapes and urban spaces only. To present more information about the element composition for each basic type, we plot the clouds with the first and second elements in each PC and normalize their presence probabilities to secure the comparability. The imagery areas of historic architecture are the ones with the most distinctive composition of elements. Pagoda, Asian temple, throne room, mausoleum are the leading elements with larger weights than any other elements in other groups. The areas of historic conservation areas are also featured with some similar elements which, however, are less weighted than those in the areas of historic architecture. The city images in the modern cityscape areas are more likely to be shot outdoors with city squares/streets, high-raised or public buildings, and social activities. These images are also informatively richer than most of the other areas due to the largest number of elements with relatively even weights in the cloud. For the natural environment, the areas of forests differ from areas of mountains as they maintain more significant elements. This annotation information enhances the interpretability of the pattern imagery areas, and the detected PCs are the features enabling this step.

4.4. Differentiation of Imagery Elements across Groups: Diversity and Extroversion

Different types of imagery areas also maintain distinctive characteristics in terms of image diversity and extroversion. The imagery areas of the urban built environment can be easily extracted since they normally have higher levels of image diversity (>0.70) than that of the natural landscape due to the presence of modern spaces, e.g., modern buildings, transport hubs, commercial and social activities, and sports amenities. The only exception is the historic heritage sites that are strictly preserved with less likelihood of being exposed to other imagery elements, including the Forbidden City and Temple of Heaven. In regard to the extroversion of city images, the differentiation of natural and urban imagery areas repeats. The photos taken in natural imagery areas are from local people who are more likely to use domestic social media platforms, whereas the users of foreign applications more often visit urban imagery areas. This trend was successfully recorded by the scattergram plotting the diversity and extroversion, in which the goodness-of-fit reaches 0.519 (Figure 11). This finding further demonstrates that non-local users prefer more diverse urban scenes than local people. Moreover, the detected imagery clusters can also be denoted by the quadrants they fall into. The modern cityscape and historic conservation areas are city images with higher levels of diversity and extroversion, whereas the natural landscapes are images with lower levels. Historic heritage sites and suburban parks feature higher extroversion and diversity. This suggests that the diversity of city images might be an indicator that could be helpful to reflect social performance. The imagery areas near the centre of the scattergram might be the essential places for social inclusion between the local and non-local communities. This impact is still recognized to be statistically significant in the regression model where other factors, e.g., distance to CBD, are controlled.

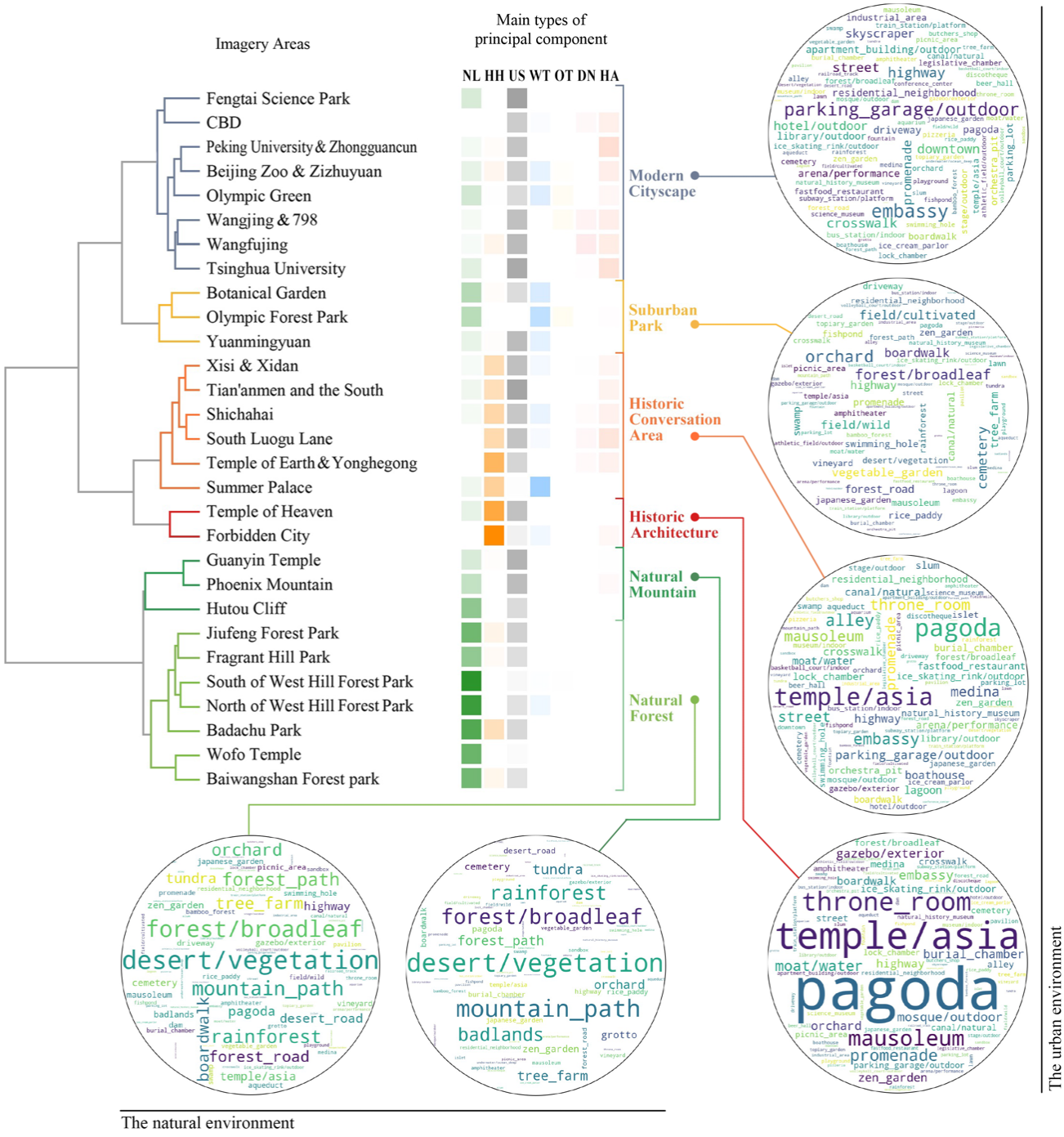


Figure 10. Dendrogram of 29 imagery areas in Central Beijing with annotations by the detected main types of principal components and by the word clouds of imagery elements with normalized weights (NL: natural landscape; HH: historical heritage; US: urban space; WT: water; OT: others; DN: dining; HA: human activities).

Figure 11. Image extroversion and diversity of 6 types of imagery areas of Central Beijing (Note: Bubble size represents the data size of the imaging area).

5. Conclusions and Discussion

Due to the widespread use of mobile devices with ITC Infotech, digital photography and social media, both facilitate an unprecedented 'image society'. Volunteered images are then an essential way for citizens to record their perception of the built environment, and city images emerge in these records as they share the imagery experience. The detection of city images with these socially sensed photos, therefore, complements existing models of calibrating the city image with prior knowledge, providing a bottom-up perspective to detect city images and areas and their typologies. This perspective, in comparison with others (e.g., structural-element-based studies and factor-based studies), is distinguishable in many dimensions. The contribution of this work, therefore, is three-fold: (1) This work echoes the importance of the complex interactions between imagery elements and provides an approach to address various aspects of these interactions comprehensively. There are four aspects defined as co-presence, hierarchy, heterogeneity, and differentiation, referring to the interaction between elements, among multiple elements, across space, and across groups. (2) The framework introduced to uncover all these four dimensions is data-driven, and it can be easily redone and applied in other contexts or in a comparison between contexts. Relevant results have also been validated preliminarily. (3) The outputs of the framework can contribute to giving in-depth feedbacks to inherently interrelated issues in city image research and design. These issues are summarized as the What-How-Where-Who questions–what are the interdependent elements? How are they combined as different types of city images? Where are the elements clustered as imagery areas? And who might prefer a certain subset of the city image? By answering these questions based on the method we proposed, a design action plan might be generated more easily than before. This introduced working flow can be employed in toolboxes for urban design, historic conservation, and tourism development as an upgrade.

The specific results of the application in Beijing also showcase several crucial findings of the interactions between imagery elements in the four dimensions that we proposed. First, the interdependency between two elements with high frequency is not necessarily stronger than that between those with lower frequency. The pair-wise understanding of the co-presence between elements is spatially varying providing structural information about city images. Furthermore, the detected typologies of city elements differ dramatically from

those main types that are predefined. Some pairs of so-called discriminant elements are detected to be statistically opposite, but spatially proximal. They are the key factors for solving the hierarchy structure of the emergence of the main types. Moreover, the imagery areas are detected by adoptable density thresholds, showing the spatial boundaries of city images. Imagery areas are more spatially continuous in urbanized areas or historical sites than in developing areas. The typologies of imagery areas can be well-annotated by the detected types of elements and principal components. The overlapping relationship between imagery areas and named areas can be a scope of portraying the spill-over effects of city images. Finally, the differentiation from social groups is proven to be related to the diversity of elements and the distance to the city centre. Local people might prefer pure and natural images as an escape from the city, whereas non-local people prefer diverse, ample scenes. These results, in turn, validate the necessity of data aggregation for reducing bias.

Future steps can be taken to reduce the potential limitations of the current work. First, two parameters must be further debugged in more cases to validate their universality. Second, invisible elements regarding the senses of city images are currently absent from this work, but they could be considered in subsequent work with new training data containing relevant labels. Third, the variation in city images among different social groups can be further researched by feeding different datasets into our models. Fourth, the time dimension of city images could be further studied with time records assigned to the photos to illustrate the temporal (diurnal, seasonal, etc.) variation of city images. Accordingly, we can detect the imagery elements and areas that are sensitive to the time change.

Author Contributions: Conceptualization, Yao Shen and Yiyi Xu; methodology, Yao Shen; software, Yiyi Xu and Lefeng Liu; validation, Yao Shen and Yiyi Xu; formal analysis, Yao Shen and Lefeng Liu; investigation, Yao Shen and Yiyi Xu; resources, Yiyi Xu and Lefeng Liu; data curation, Yiyi Xu; writing—original draft preparation, Yao Shen and Yiyi Xu; writing—review and editing, Yao Shen; visualization, Yiyi Xu; supervision, Yao Shen; project administration, Yao Shen; funding acquisition, Yao Shen. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China (NSFC), grant number 51908413 and by Pujiang Talent Project, grant number 19PJC106.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

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

References

1. Lynch, K. *The Image of the City*; The MIT Press: Cambridge, MA, USA, 1960.
2. Kennedy, L.; Naaman, M.; Ahern, S.; Nair, R.; Rattenbury, T. How flickr helps us make sense of the world: Context and content in community-contributed media collections. In Proceedings of the 15th international conference on Multimedia (MULTIMEDIA'07), Augsburg, Germany, 24–27 September 2007; Lienhart, R., Prasad, A.R., Hanjalic, A., Choi, S., Bailey, B.P., Sebe, N., Eds.; ACM: New York, NY, USA, 2007; pp. 631–640.
3. Liu, L.; Zhou, B.; Zhao, J.; Ryan, B. C-IMAGE: City cognitive mapping through geo-tagged photos. *GeoJournal* **2016**, *81*, 817–861. [CrossRef]
4. I Agustí, D.P. Characterizing the location of tourist images in cities. Differences in user-generated images (Instagram), official tourist brochures and travel guides. *Ann. Tour. Res.* **2018**, *73*, 103–115. [CrossRef]
5. Mobasheri, A. (Ed.) *Open Source Geospatial Science for Urban Studies: The Value of Open Geospatial Data*; Springer Nature: Berlin/Heidelberg, Germany, 2020.
6. Appleyard, D. Styles and Methods of Structuring a City. *Environ. Behav.* **1970**, *2*, 100–117. [CrossRef]
7. Morello, E.; Ratti, C. A digital image of the city: 3D isovists in Lynch's urban analysis. *Environ. Plan. B Plan. Des.* **2009**, *36*, 837–853. [CrossRef]
8. Nasar, J.L. The Evaluative Image of the City. *J. Am. Plan. Assoc.* **1990**, *56*, 41–53. [CrossRef]
9. Hospers, G.-J. Lynch's The Image of the City after 50 Years: City Marketing Lessons from an Urban Planning Classic. *Eur. Plan. Stud.* **2010**, *18*, 2073–2081. [CrossRef]
10. Pearce, P.L.; Fagence, M. The legacy of Kevin Lynch: Research implications. *Ann. Tour. Res.* **1996**, *23*, 576–598. [CrossRef]
11. Lynch, K. Reconsidering the image of the city. In *City Sense and City Design: Writings and Projects of Kevin Lynch*; Banerjee, T., Southworth, M., Eds.; The MIT Press: Cambridge, MA, USA, 1985; pp. 247–256.

12. Francescato, D.; Mebane, W. How citizens view two great cities: Milan and Rome. In *Image and Environment: Cognitive Mapping and Spatial Behavior*; Downs, R., Stea, D., Eds.; Edward Arnold: London, UK, 1973; pp. 182–220.
13. Deng, N.; Liu, J.; Dai, Y.; Li, H. Different cultures, different photos: A comparison of Shanghai's pictorial destination image between East and West. *Tour. Manag. Perspect.* **2019**, *30*, 182–192. [CrossRef]
14. Peng, X.; Bao, Y.; Huang, Z. Perceiving Beijing's "City Image" Across Different Groups Based on Geotagged Social Media Data. *IEEE Access* **2020**, *8*, 93868–93881. [CrossRef]
15. Carmona, M.; Heath, T.; Oc, T.; Tiesdell, S. *Public Places, Urban Spaces: The Dimensions of Urban Design*; Architectural Press: London, UK, 2003.
16. Lloyd, R.; Heivly, C. Systematic distortions in urban cognitive maps. *Ann. Assoc. Am. Geogr.* **1987**, *77*, 191–207. [CrossRef]
17. Anholt, S. The Anholt-GMI City Brands Index: How the world sees the world's cities. *Place Brand.* **2006**, *2*, 18–31. [CrossRef]
18. Gartner, W.C. Image Formation Process. *J. Travel Tour. Mark.* **1994**, *2*, 191–216. [CrossRef]
19. Gilboa, S.; Jaffe, E.D.; Vianelli, D.; Pastore, A.; Herstein, R. A summated rating scale for measuring city image. *Cities* **2015**, *44*, 50–59. [CrossRef]
20. Stylos, N.; Vassiliadis, C.A.; Bellou, V.; Andronikidis, A. Destination images, holistic images and personal normative beliefs: Predictors of intention to revisit a destination. *Tour. Manag.* **2016**, *53*, 40–60. [CrossRef]
21. Stern, E.; Krakover, S. The Formation of a Composite Urban Image. *Geogr. Anal.* **1993**, *25*, 130–146. [CrossRef]
22. Luque-Martínez, T.; Del Barrio-García, S.; Ibáñez-Zapata, J.; Molina, M.R. Modeling a city's image: The case of Granada. *Cities* **2007**, *24*, 335–352. [CrossRef]
23. Urry, J. *The Tourist Gaze*; Sage: London, UK, 1990.
24. Crandall, D.J.; Backstrom, L.; Huttenlocher, D.; Kleinberg, J. Mapping the world's photos. In Proceedings of the 18th International Conference on World Wide Web, Madrid, Spain, 20–24 April 2009; pp. 632–637.
25. Crawshaw, C.; Urry, J. Tourism and the photographic eye. In *Touring Cultures: Transformations of Travel and Theory*; Ojek, C.R., Urry, J., Eds.; Routledge: London, UK, 1997; pp. 176–195.
26. Pan, S.; Lee, J.; Tsai, H. Travel photos: Motivations, image dimensions, and affective qualities of places. *Tour. Manag.* **2014**, *40*, 59–69. [CrossRef]
27. Xu, Z.; Liu, Y.; Yen, N.Y.; Mei, L.; Luo, X.; Wei, X.; Hu, C. Crowdsourcing Based Description of Urban Emergency Events Using Social Media Big Data. *IEEE Trans. Cloud Comput.* **2020**, *8*, 387–397. [CrossRef]
28. Naaman, M.; Becker, H.; Gravano, L. Hip and trendy: Characterizing emerging trends on Twitter. *J. Am. Soc. Inf. Sci. Technol.* **2011**, *62*, 902–918. [CrossRef]
29. Crooks, A.; Pfoser, D.; Jenkins, A.; Croitoru, A.; Stefanidis, A.; Smith, D.; Karagiorgou, S.; Efentakis, A.; Lamprianidis, G. Crowdsourcing urban form and function. *Int. J. Geogr. Inf. Sci.* **2015**, *29*, 720–741. [CrossRef]
30. Lu, X.; Wang, C.; Yang, J.; Pang, Y.; Zhang, L. Photo2Trip: Generating travel routes from geo-tagged photos for trip planning. In Proceedings of the 18th International Conference on Multimedia, Firenze, Italy, 25–29 October 2010; pp. 143–152.
31. García-Palomares, J.C.; Gutiérrez, J.; Mínguez, C. Identification of tourist hot spots based on social networks: A comparative analysis of European metropolises using photo-sharing services and GIS. *Appl. Geogr.* **2015**, *63*, 408–417. [CrossRef]
32. Samany, N.N. Automatic landmark extraction from geo-tagged social media photos using deep neural network. *Cities* **2019**, *93*, 1–12. [CrossRef]
33. Jankowski, P.; Andrienko, N.; Andrienko, G.; Kisilevich, S. Discovering Landmark Preferences and Movement Patterns from Photo Postings. *Trans. GIS* **2010**, *14*, 833–852. [CrossRef]
34. Ji, R.; Gao, Y.; Zhong, B.; Yao, H.; Tian, Q. Mining flickr landmarks by modeling reconstruction sparsity. *ACM Trans. Multimed. Comput. Commun. Appl.* **2011**, *7S*, 31. [CrossRef]
35. Rattenbury, T.; Good, N.; Maaman, M. Towards automatic extraction of event and place semantics from flickr tags. In Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Amsterdam, The Netherlands, 23–27 July 2007; ACM: New York, NY, USA, 2007; pp. 103–110.
36. Rattenbury, T.; Naaman, M. Methods for extracting place semantics from Flickr tags. *ACM Trans. Web* **2009**, *3*, 1–30. [CrossRef]
37. Dunkel, A. Visualizing the perceived environment using crowdsourced photo geodata. *Landsc. Urban. Plan.* **2015**, *142*, 173–186. [CrossRef]
38. Papadopoulos, S.; Zigkolis, C.; Kompatsiaris, I.; Vakali, A. Cluster-Based Landmark and Event Detection for Tagged Photo Collections. *IEEE Multimed. Mag.* **2011**, *18*, 52–63. [CrossRef]
39. Miah, S.J.; Vu, H.Q.; Gammack, J.; McGrath, M. A Big Data Analytics Method for Tourist Behaviour Analysis. *Inf. Manag.* **2017**, *54*, 771–785. [CrossRef]
40. Ibrahim, M.R.; Haworth, J.; Cheng, T. Understanding cities with machine eyes: A review of deep computer vision in urban analytics. *Cities* **2020**, *96*, 102481. [CrossRef]
41. Wang, R.; Luo, J.; Huang, S. Developing an artificial intelligence framework for online destination image photos identification. *J. Destin. Mark. Manag.* **2020**, *18*, 100512. [CrossRef]
42. Sheng, F.; Zhang, Y.; Shi, C.; Qiu, M.; Yao, S. Xi'an tourism destination image analysis via deep learning. *J. Ambient. Intell. Humaniz. Comput.* **2020**, in press, corrected proof. [CrossRef]
43. Zhou, B.; Liu, L.; Oliva, A.; Torralba, A. Recognizing City Identity via Attribute Analysis of Geo-tagged Images. In Proceedings of the 13th European Conference on Computer Vision (ECCV), Zurich, Switzerland, 6–12 September 2014; pp. 519–534. [CrossRef]

44. Zhang, K.; Chen, Y.; Li, C. Discovering the tourists' behaviors and perceptions in a tourism destination by analyzing photos' visual content with a computer deep learning model: The case of Beijing. *Tour. Manag.* **2019**, *75*, 595–608. [CrossRef]
45. Chen, M.; Arribas-Bel, D.; Singleton, A. Quantifying the Characteristics of the Local Urban Environment through Geotagged Flickr Photographs and Image Recognition. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 264. [CrossRef]
46. Doersch, C.; Singh, S.; Gupta, A.; Sivic, J.; Efros, A.A. What makes Paris look like Paris? *ACM Trans. Graph.* **2012**, *31*, 101. [CrossRef]
47. Zhang, F.; Zhou, B.; Ratti, C.; Liu, Y. Discovering place-informative scenes and objects using social media photos. *R. Soc. Open Sci.* **2019**, *6*, 181375. [CrossRef]
48. Batty, M. Cities as Complex Systems: Scaling, Interaction, Networks, Dynamics and Urban Morphologies. *UCL CASA Work. Pap. Ser.* **2009**, *131*, 1041–1071. [CrossRef]
49. Batty, M. *The New Science of Cities*; MIT Press: Cambridge, MA, USA, 2013. [CrossRef]
50. Senaratne, H.; Mobasheri, A.; Ali, A.L.; Capineri, C.; Haklay, M. A review of volunteered geographic information quality assessment methods. *Int. J. Geogr. Inf. Sci.* **2017**, *31*, 139–167. [CrossRef]
51. Zhou, B.; Lapedriza, A.; Khosla, A.; Oliva, A.; Torralba, A. Places: A 10 Million Image Database for Scene Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 1452–1464. [CrossRef]
52. Ankerst, M.; Breunig, M.M.; Kriegel, H.P.; Sander, J. OPTICS: Ordering points to identify the clustering structure. In Proceedings of the ACM SIGMOD International Conference on Management of Data, Philadelphia, PA, USA, 31 May–3 June 1999; pp. 49–60.