Landscape and Urban Planning 212 (2021) 104125

Contents lists available at ScienceDirect

Landscape and Urban Planning

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

Research Paper

Discovering the homogeneous geographic domain of human perceptions from street view images

Yao Yao [a], Jiale Wang [b], Ye Hong [c], Chen Qian [d], Qingfeng Guan [a,*], Xun Liang [a], Liangyang Dai [a], Jinbao Zhang [e]

[a] School of Geography and Information Engineering, China University of Geosciences. Wuhan 430078, Hubei Province, China
[b] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430075, Hubei Province, China
[c] Institute of Cartography and Geoinformation, ETH Zurich, Stefano-Franscini-Platz 5, Zürich 8093, Switzerland
[d] Department of Engineering Systems and Environment, University of Virginia, 22904 VA, United States
[e] School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, Guangdong Province, China

HIGHLIGHTS

- The distribution of human perceptions in urban area was obtained.
- This study first focuses on the spatial homogeneity of human perceptions.
- A method is proposed to discover the homogeneous geographic domain of human perceptions.
- This study explored the role of urban function in shaping human perceptions.

ARTICLE INFO

Keywords:
Street view
Human perception
Urban function
Geographic domain
Infomap

ABSTRACT

Human perception of place refers to residents' psychological feelings about urban areas. Many studies of human perceptions have focused on a specific geographic location. Whether the distribution of human perceptions in continuous city space shows specific characteristics and how to disclose these phenomena remains a direction worth exploring. Due to cities' heterogeneity, quantitatively identifying the homogeneous perception regions at a fine scale within large urban regions is challenging. This study proposed a novel method to discover the homogeneous geographic domain of human perception using massive street view images. First, human perceptions of the urban visual environment were evaluated using street view images. Next, perception network models were constructed based on the road network and perception assessment results. Then, the Infomap community detection algorithm was used to identify homogeneous human perception communities. The qualitative and quantitative results verified our approach's effectiveness for capturing human perceptions' homogeneous geographic domain. Moreover, driving factor analysis was conducted to determine the urban function that may cause a community to be perceived differently based on point-of-interest (POI) data. In general, our method for combining human perceptions and the topology of urban roads could identify the homogeneous perception domain, which is valuable for urban structure studies and human perception assessment.

1. Introduction

Human perception of place refers to residents' psychological feelings about an urban locale (Ordonez and Berg 2014, Tuan 2013). A city has specific functions and carries the psychological and emotional attachment of urban residents to their living environment (Dubey et al. 2016, Goodchild 2011, Zhang et al. 2018a). Different places characterize different visual information, built environments, and urban function, affecting people's sense of the urban environment (Goodchild 2011) and leads to varying psychological perception levels (Yao et al. 2019, Zhang et al. 2018a). Measuring the human perceptions of place can help researchers understand the interaction between the built environment and

* Corresponding author.
E-mail addresses: yaoy@cug.edu.cn (Y. Yao), wangjiale@whu.edu.cn (J. Wang), hongy@ethz.ch (Y. Hong), cq4sk@virginia.edu (C. Qian), guanqf@cug.edu.cn (Q. Guan), liangxun@cug.edu.cn (X. Liang), daily@cug.edu.cn (L. Dai), zhangjb27@mail2.sysu.edu.cn (J. Zhang).

https://doi.org/10.1016/j.landurbplan.2021.104125
Received 23 July 2020; Received in revised form 16 April 2021; Accepted 19 April 2021

residents' mental health (Wang et al. 2019b, Wang et al. 2019c, Wolch et al. 2014). Human perceptions can enrich the potential semantic information of urban places (Zhang et al. 2018a). Accurately gathering the distribution of human perceptions in urban space has essential significance for urban planning and public health studies (Li et al., 2015b, Zhang et al. 2018a, Yao et al. 2019, Wang et al., 2019b).

Human perception assessment of urban regions has been intensively studied during recent years. Traditional human perception research mainly uses low-throughput and data collection methods, such as time-consuming and labor-intensive questionnaires and interviews (Cresswell 1992, Dadvand et al. 2016, Kabisch et al. 2015), which cannot be expanded to a large scale. With the rapid development of multi-source geo-spatial data (Liu et al. 2015), researchers could obtain a massive amount of publicly available geo-tagged images that could effectively reflect every corner of a city. As typical geo-tagged data, street view images have enabled us to observe, perceive, and understand our built environment. Moreover, what has been proven is that these data could accurately delineate our city's physical environment and are more in line with the residents' eye-level perceptions than other spatial–temporal data, such as remote sensing images (Helbich et al. 2019). Therefore, street view images have aroused widespread concern in urban studies (Li et al. 2015a, Li et al. 2015b, Yin and Wang 2016, Li et al. 2017, Zhang et al. 2017, Zhang et al. 2018b, Gong et al. 2018, Tang and Long 2018).

Regarding street view images as the proxy of urban places, some researchers have successfully assessed human perceptions hidden in these images' visual elements using computer vision and machine learning technology (Yao et al. 2019, Zhang et al. 2018a). Salesses (2013) first proposed a comparative study of human perceptions through pairwise street view images. Dubey et al. (2016) expanded the research area and collected the volunteers' perceptions to construct a street view perception dataset. Zhang et al. (2018a) proposed a perception measurement method for large-scale urban regions based on the MIT Place Pulse dataset (Ordonez and Berg 2014, Porzi et al. 2015, Naik et al. 2017, Zhang et al. 2018a). To solve the problem that people from different cities and cultural backgrounds may perceive the same urban environment differently, Yao et al. (2019) proposed the human–machine adversarial scoring framework. This framework has the advantages of having a low cost, high throughput, and low deviation in the process of assessing human perceptions using street view images. Besides, it has been successfully used and achieved fruitful results in urban environments, public health, and other related fields (Wang et al. 2019a, Wang et al. 2019b). Research has mainly focused on six human perception types: wealthy, safety, lively, beautiful, boring, and depressing.

Numerous studies have confirmed the urban environment's impact reflected by the street view on urban residents' health, both physically and mentally. For instance, street views of green and blue spaces positively affect geriatric depression (Helbich et al. 2019). The degree of visual openness of a street affects urban residents' walkability and mental health (Wang et al. 2019c). Additionally, Wang et al. (2019a) and Wang et al. (2019b) reveal the quantitative impact of environmental perception on residents' physical and mental health with the human–machine adversarial scoring method. The above studies all regard residential communities as the primary research unit. By creating a buffer around the community center, the corresponding indicators of the street scene sampling points within a community can be obtained, demonstrating that community quality is an essential factor affecting the residents' physical and mental health.

Due to the lack of appropriate methods, research on residents' perceptions has not yet been carried out at the community scale. Previous studies on human perceptions focused only on a specific geographical location (Dubey et al. 2016, Yao et al. 2019, Zhang et al. 2018a, Wang et al. 2019a, Wang et al. 2019b). Does the distribution of human perceptions in continuous urban space exhibit specific characteristics (such as parcel-level aggregation or spatial variation)? How can these phenomena be quantitatively described? These are the foci of this study.

Cities are complex systems, and urban road networks, as the city's backbone, shape the city's traffic, landscape, and functional structure (Hong and Yao 2019, Michael 2008, Wang et al. 2012). Road networks naturally divide successive urban spaces into disjointed communities (sub-regions) and shape a regional pattern. Typically, using a community detection algorithm (e.g., Infomap algorithm (Rosvall and Bergstrom 2008)), the hidden community structures that are internally well connected but externally less in an urban road network can be revealed. Most street views are distributed along urban roads (Cheng et al. 2017), reflecting the physical form and properties of a cities' interior space (Zhang et al. 2018b, Gebru et al. 2017). The human perception information of roads depicts the spatial similarity of subgroups of street networks in terms of perception. Naturally, it is worthwhile to combine road perception information with a community detection algorithm to facilitate understanding human perceptions' spatial distribution.

This study proposed the geographically homogeneous perception domain, which characterized the homogeneity and purity from the residents' perception. In this study, the human–machine adversarial scoring framework was adopted to obtain the perception scores at the street view scale, and each road was assigned certain psychological awareness information. By combining the topology information of the urban road network with human perceptions, a graph network was constructed to identify human perceptions' spatial distribution compared with the community detection results obtained using Euclidian distance. Finally, the impact of urban functions on the perception regions' spatial distribution was revealed. A case study in Beijing confirmed the validity of the proposed method.

2. Methodology

The flow chart of the study is illustrated in Fig. 1. This study can be divided into four steps: 1) Assess each sampling point and road's human perception scores using the human–machine adversarial framework with street view images; 2) Establish the directed perception graph models of the road network and the Euclidean distance network models with the average perception score and Euclidean length of each road as weights, respectively; 3) Detect the multilevel homogeneous perception communities in the network using the Infomap algorithm and quantitatively analyze the spatial homogeneity from human perceptions at the community scale; and 4) Explore the relationship between the spatial distribution of human perceptions and urban functions at the community scale.

2.1. *Measuring human perceptions of the built environment based on the human–machine adversarial framework*

We focus on six types of psychological perception among the residents: safety, lively, beautiful, wealthy, depressing, and boring, which are in line with previous studies (Dubey et al. 2016, Yao et al. 2019, Zhang et al. 2018a). Assessing both positive and negative perception types can help understand residents' psychological feelings about the built environment.

We take street view images as the proxy of urban places and physical settings. The human–machine adversarial scoring method (Yao et al. 2019) is used to quantitatively describe the residents' perceptions of the built environment reflected by street view images. This method bridges the gap between urban environment, visual scenery, and residential perceptions. It combines advanced computer vision and machine learning technology to provide an accurate, low-cost, and highly efficient assessment of large-scale human perceptions. The primary process of the human–machine adversarial scoring method for perceiving the urban environment is shown in the first step of Fig. 1.

(1) Semantic segmentation of street view images. The human–machine adversarial scoring method first extracted the visual elements closely related to perception, that is, using the fully convolutional

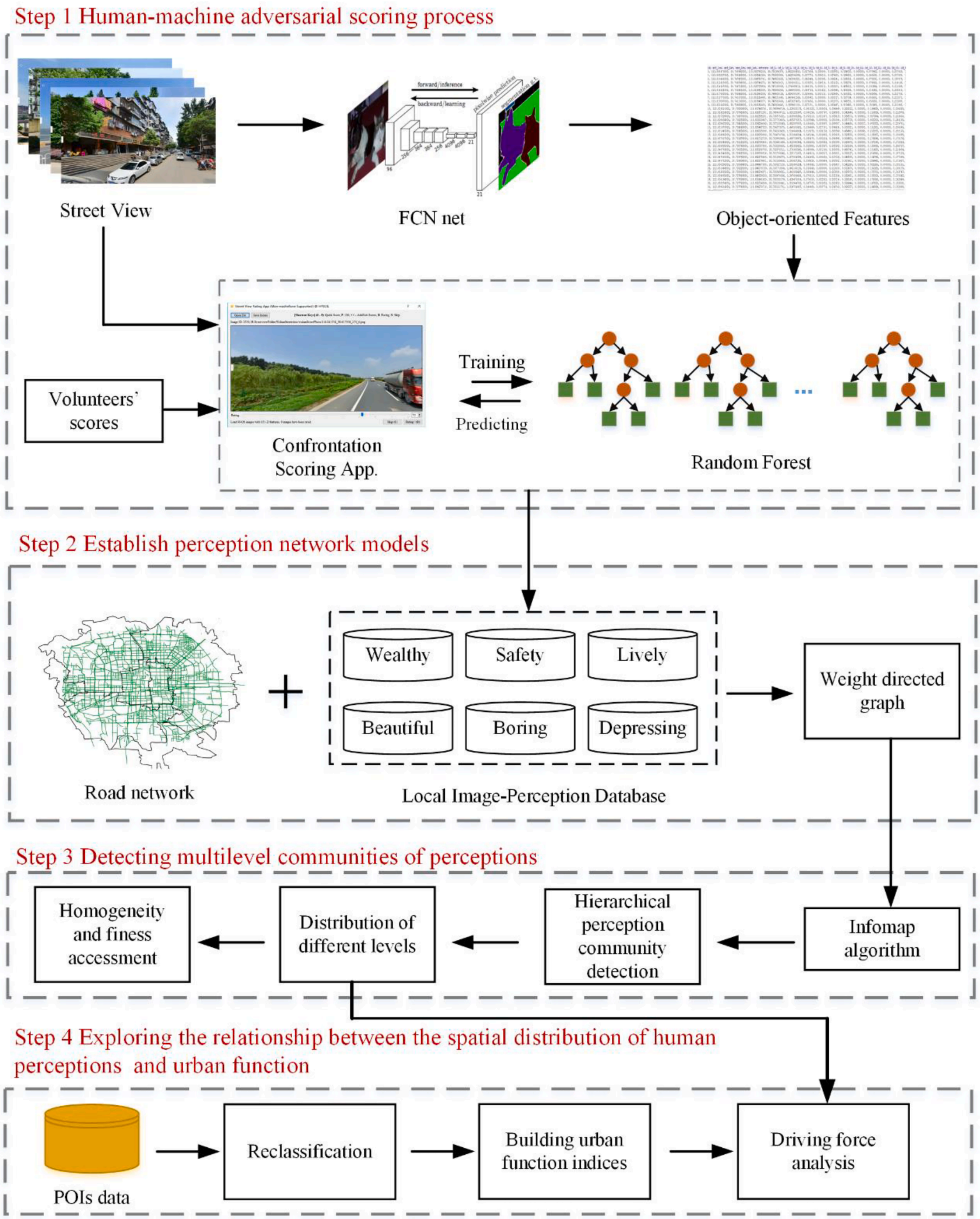

Fig. 1. Workflow of discovering the homogeneous geographic domain of human perceptions from street view images.

network (i.e., the FCN-8 s) (Long et al. 2015) to segment the street views into 151 ground objects (e.g., trees, roads, and buildings). The image segmentation results represent the physical setting and visual elements of a place. The MIT ADE20K website (http://groups.csail.mit.edu/vision/datasets/ADE20K/) contains a complete description of all categories. We used an annotated image from the ADE20K scene parsing and segmentation database to train the FCN-8 s network (Zhou et al. 2016, Zhou et al. 2017). Each street view can be represented as a 151-dimensional feature vector by counting each feature type's pixel proportion.

(2) Volunteer recruitment. After obtaining the image segmentations, this study developed a human–machine adversarial scoring system to collect the ground truth of people's perceptions and measure each street

view's perception. The human–machine adversarial scoring system requires that the volunteers should have an excellent knowledge of the urban environment to avoid perception assessment bias (Wang et al. 2019b, Yao et al. 2019). Therefore, we recruit local volunteers aware of China's regional socioeconomic background to score a dataset of street scenes from Chinese cities. Detailed information about volunteers can be found in section 4.1.

(3) Human-machine adversarial scoring. The human–machine adversarial scoring system randomly displays a street view image from 500 thousand street view photos of major Chinese cities (e.g., Beijing, Shenzhen, Guangzhou, Shanghai, Wuhan, and Hangzhou) for annotation and receives the volunteer's perception score. Each perception type's score field is 0 to 100 points, representing the lowest and highest levels, respectively. When volunteers score>50 photos, the system automatically fits the proportions of 151 elements in the image segmentation with the inputted rating scores using the random forest algorithm (Breiman 2001, Fern A Ndez-Delgado et al. 2014) for automatic rating and shows the recommended scores for following new images. If more than one person scores the same image, the image's final score will be set as the median value to avoid extreme scores.

(4) Calibration of the automatic scoring system. In the following scoring process, the trained scoring model automatically recommends rating scores for new images and refers the images to volunteers who subsequently correct the system's recommended scores. The human–machine adversarial scoring system tends to calibrate its parameters of the fitted model and provide more accurate perception scores. The calibration process stops when the volunteers' scores and the recommended scores match given a threshold (when random forest fitting accuracy is over 90% on average and the root-mean-square errors between the recommended scores and volunteers' scores are<5 for the last scored 100 images).

At the end of the rating process, the human–machine adversarial scoring system can automatically generate the perception scores for any street view image. With the help of recruited local volunteers and carefully defined stopping criterion, we believe that the obtained evaluation scores could represent real perceptions.

2.2. Constructing the human perception network models

After simplifying and checking the OSM roads' topology in the study area, we abstract the OSM road map into a directed graph (Hong and Yao 2019, Yao et al. 2018a). A directed graph $G \equiv (V, E, W)$ is composed of vertexes (V), edges (E), and the weight of each edge (W). In this study, the starting point and ending points of a road and the intersection between road segments constitute the vertex V; the road segments connecting these points form the edge set E. The "one-way" field specifies each edge's direction in the OSM road property description. In our urban perception network models, the weight W is calculated as:

$$W_{i,j} = 100 - \frac{\sum_{k=1}^{n} score_k}{n_i} \tag{1}$$

Where $W_{i,j}$ represents the weight of the type j perception of the $i-th$ road segment; n_i is the number of street view sampling points falling on the $i-th$ road; and $\frac{\sum_{k=1}^{n} score_k}{n_i}$ stands for the average perception score of the type j perception for the $i-th$ road segment.

Our work intends to discover and reveal regional homogeneous distribution patterns of human perceptions. We employ a network clustering method that considers the road network's ability to shape the urban regionalization pattern from an urban network division perspective. The division of a network depends mainly on the spatial variable similarity and topological connectivity of the nodes. Spatial nodes with higher spatial variable similarity and connectivity will be more easily clustered into the same community (Hong and Yao 2019). Because we intend to cluster graph nodes that have similar perception conditions with their neighboring nodes, we use Equation (1) as a spatial similarity metric of homogenous perception. The higher the score of urban perception between road nodes (i.e., the stronger the perception), the smaller the value of formula (1), the easier the nodes are grouped into one category.

Additionally, we carry out a comparison experiment with weight W set as the Euclidean length of each road segment (i.e., Euclidean distance network model, which implies closer things are related to each other, elaborated by Tobler's first law of geography (Tobler 1970)). With the Euclidean distance as a spatial similarity metric of perceptions, dividing the Euclidean distance network can be seen as a natural urban regionalization. Later, we show that the perception network models' advantage in finding homogeneous communities.

2.3. Discovering the homogeneous geographic domains of human perception using the Infomap algorithm

The Infomap algorithm is adopted to reveal the network's hidden spatial structure (Rosvall and Bergstrom 2008), thereby identifying the perception communities' spatial distribution. Infomap, regarded as one of the best performing nonoverlapping clustering models, can meticulously identify a network's hidden hierarchical structure by combining information theory with random walks (Lancichinetti and Fortunato, 2009). Our network's attributes depend entirely on the connectivity of the actual road network, which ensured that our network could meet and reflect the city's functional structure. The Infomap does not require preset parameters so that the network's nature and topology determine the result of the community discovery. We do not need to explicitly consider or specify any transport-related attributes such as centrality and multi-modality when modeling a network. With the constructed directed graph of the urban road network, we employ the Infomap algorithm to divide the network into multiple levels and reveal the hierarchical spatial patterns implied in the network.

The Infomap algorithm classifies each node in the graph network into a specific community. The spatial coverage of each community can be determined by constructing Thiessen polygons. Relative percentage variance (RPV) is used to describe the homogeneity of residents' perceptions within a community. RPV is defined as the ratio of the mean of perception scores' variance of the $k-th$ level communities to the perception scores' variance of all sampling points in the study area:

$$RPV_k = \frac{\sum_{i=0}^{n} var_{k,i}/n_k}{var} \times 100\% \tag{2}$$

where n_k is the number of communities in the $k-th$ level and $var_{k,i}$ is the perception variance of the sample points within the $i-th$ community at the $k-th$ level. var is the perception scores' variance at all sampling points in the study area. An RPV value greater than or equal to 1 indicates that the method has not detected a more homogeneous community. A lower RPV suggests that a community has a more uniform and consistent human perception.

We also define the average area $AvgArea_k$ and standard deviation $AreaSDE_k$ of all the communities in the kth level to describe the scale of the communities:

$$AvgArea_k = \frac{\sum_{i=0}^{n} area_{k,i}}{n_k} \tag{3}$$

$$AreaSDE_k = \sqrt{\frac{\sum_{i=0}^{n} \left(area_{k,i} - AvgArea_k\right)^2}{n_k}} \tag{4}$$

Where $area_{k,i}$ is the area of the $i-th$ community at the $k-th$ level. Relatively low $AvgArea$ and $AreaSDE$ values represent a more refined community discovery result. By calculating RPV, $AvgArea$ and $AreaSDE$ in different community levels, we can quantitatively evaluate the hierarchical community identification method's effectiveness with the perception network models.

2.4. Exploring the relationship between the spatial distribution of human perceptions and urban function

This section first considers the spatial autocorrelation of average perception scores for each community (Li et al., 2010). This study uses two criteria to quantitatively reveal each community's urban functional patterns: the point-of-interest (POI) density and mixing entropy (Hong and Yao 2019). The POI density Equation (5) reflects the urban functions' absolute density in a particular community (Li et al., 2011). In the $i-th$ community, the ratio of the number of the $j-th$ type of POI $Count_{i,j}$ and $Area_i$ of the $i-th$ community is defined as the POI $Density_{i,j}$:

$$Density_{i,j} = \frac{Count_{i,j}}{Area_i} \tag{5}$$

The POI mixing entropy Equation (6) is used to characterize the extent of mixed land use (Frank et al., 2004), revealing each particular community's urban functional mixing degree. For community i, the mixing entropy $Entropy_i$ is defined as:

$$Entropy_i = -\sum_{j=1}^{n} \left(p_{i,j} \times lnp_{i,j}\right) \tag{6}$$

Where n represents the number of all the types of POIs in the study area (we excluded POIs belonging to the "other facilities", so $n = 7$); and $p_{i,j}$ represents the proportion of the $j-th$ type of POI in all POIs within community i.

Random-forest-based regression analysis is conducted to investigate the nonlinear relationship between the spatial distribution of human perceptions and urban function. Each community's average perception score is used as the dependent variable, and the density of different type POIs, mixing entropy, and road network density are regarded as explanatory variables. By analyzing the fitting accuracy and weight parameters, we can quantitatively explore the importance of different urban functions in human perceptions' spatial distribution.

3. Case study and data

We select the urban areas within the fifth ring road in Beijing as our

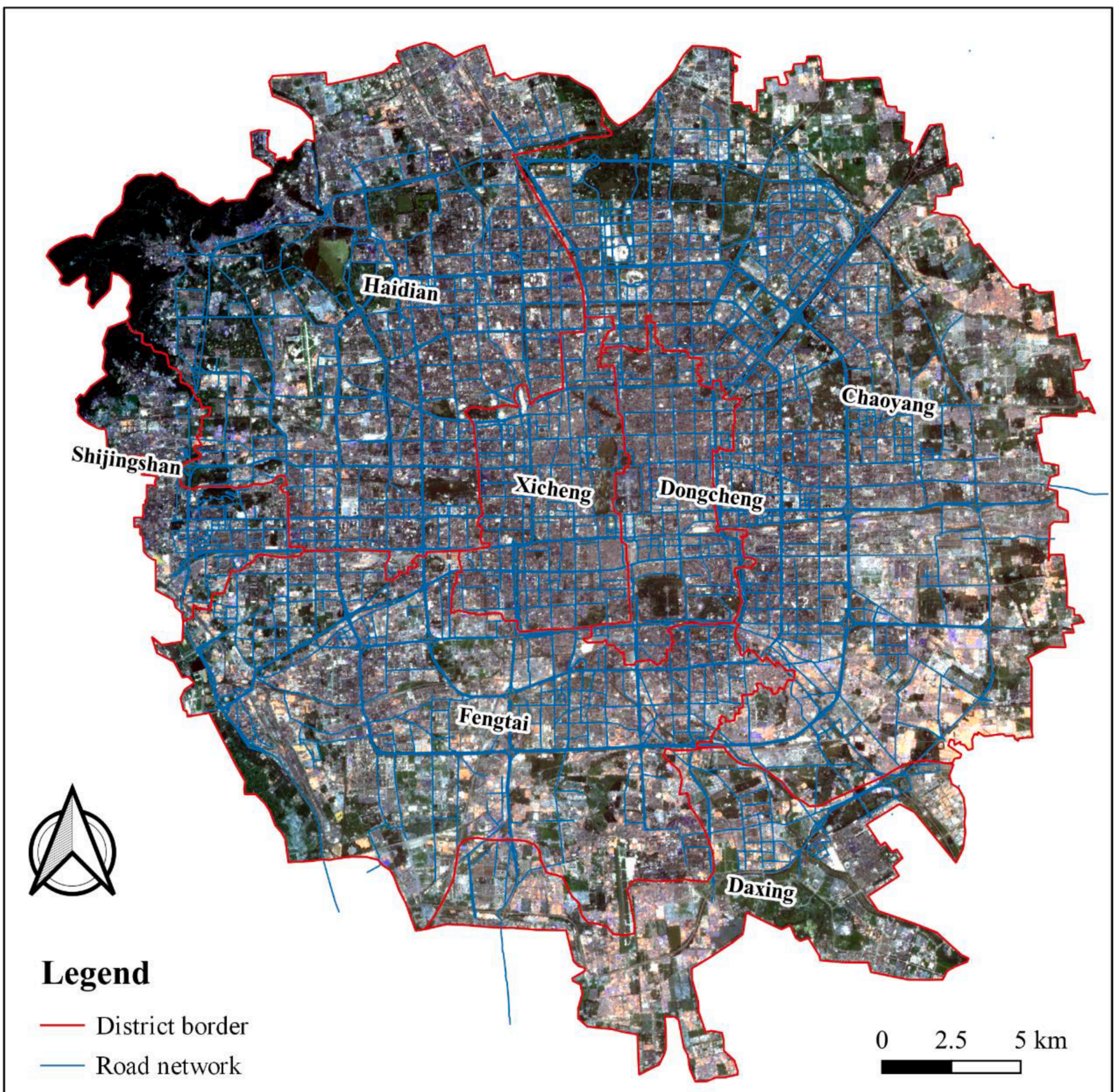

Fig. 2. Case study area: Beijing inner city. The blue lines represent the main road, and the base map is a high-resolution remote sensing image. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

study area. The fifth ring road area has the most outstanding economic, political, and cultural development of Beijing's populated areas. It has commonly taken it as a study area in previous studies (Zhang et al. 2018a, Wang et al. 2019b). As shown in Fig. 2, this study selects seven administrative units in Beijing as our research areas: Haidian district, Chaoyang district, Daxing district, Shijingshan district, Fengtai district, Dongcheng district, and Xicheng district.

The road network data of the study area are acquired from the OpenStreetMap (OSM) website (http://www.openstreetmap.org), an open geographical database of the world (Zhou, 2018; Bahrehdar et al., 2020). Studies have shown that the study area's road location and topological relationships are highly accurate (Yao et al. 2018a). After simplifying, merging, and checking the topology, 6,716 road vertexes and 8,784 road segments are extracted from the original OSM road data, shown in Figure 2.

Tencent Map, one of the largest online map service providers in China, provides an application interface for capturing street view images. Based on the road data, we sample points 100 m apart along the roads mentioned above. For each sampling point, four horizontal street view images with angles of 0°, 90°, 180°, and 270° are obtained. We collected a total of 128,592 street view images of 32,148 sampling points in the study area, as illustrated in Fig. 3.

In this study, the POI is used to identify the community's urban functions, further exploring the relationship between human perceptions and urban functional land use (Liu et al., 2017; Yao et al., 2018b). We collect a total of 1,052,852 POI in the study area from Gaode Map (https://www.amap.com/). According to the category code recorded in the POI, we reclassify the POI data into eight categories (each POI may belong to more than one category), as shown in Table 1. The POIs classified into 'Other facilities' are included in the land use density analysis. Still, they are excluded from the study of mixed land use because they do not reflect a particular area's functional characteristics.

Table 1
Types, abbreviations, counts, and proportions of the Gaode POIs after reclassification.

Type	Abbreviation	Count	Proportion
Life services	LS	430,943	0.409
Other facilities	OTH	319,785	0.304
Office building/space	OBS	128,771	0.122
Medical/Education	ME	64,058	0.061
Entertainment	ENT	38,877	0.037
Government	GOV	29,158	0.028
Residential communities	RC	26,779	0.025
Financial service	FS	18,819	0.018

4. Results

4.1. Human perception result based on the human–machine adversarial scoring framework

Twenty local volunteers with balanced gender were invited to give scores to the human–machine adversarial scoring system. The volunteer whose ages ranged from 20 to 50 years old consist of college students and staff, and they are aware of the study area's socioeconomic background. After each volunteer scored 1,000 – 2,000 images via the human–machine adversarial system, the recommended scores and volunteers' scores reached a high agreement. All street view images in the study area were fed into the system, and the human perception scores of the sampled images were obtained.

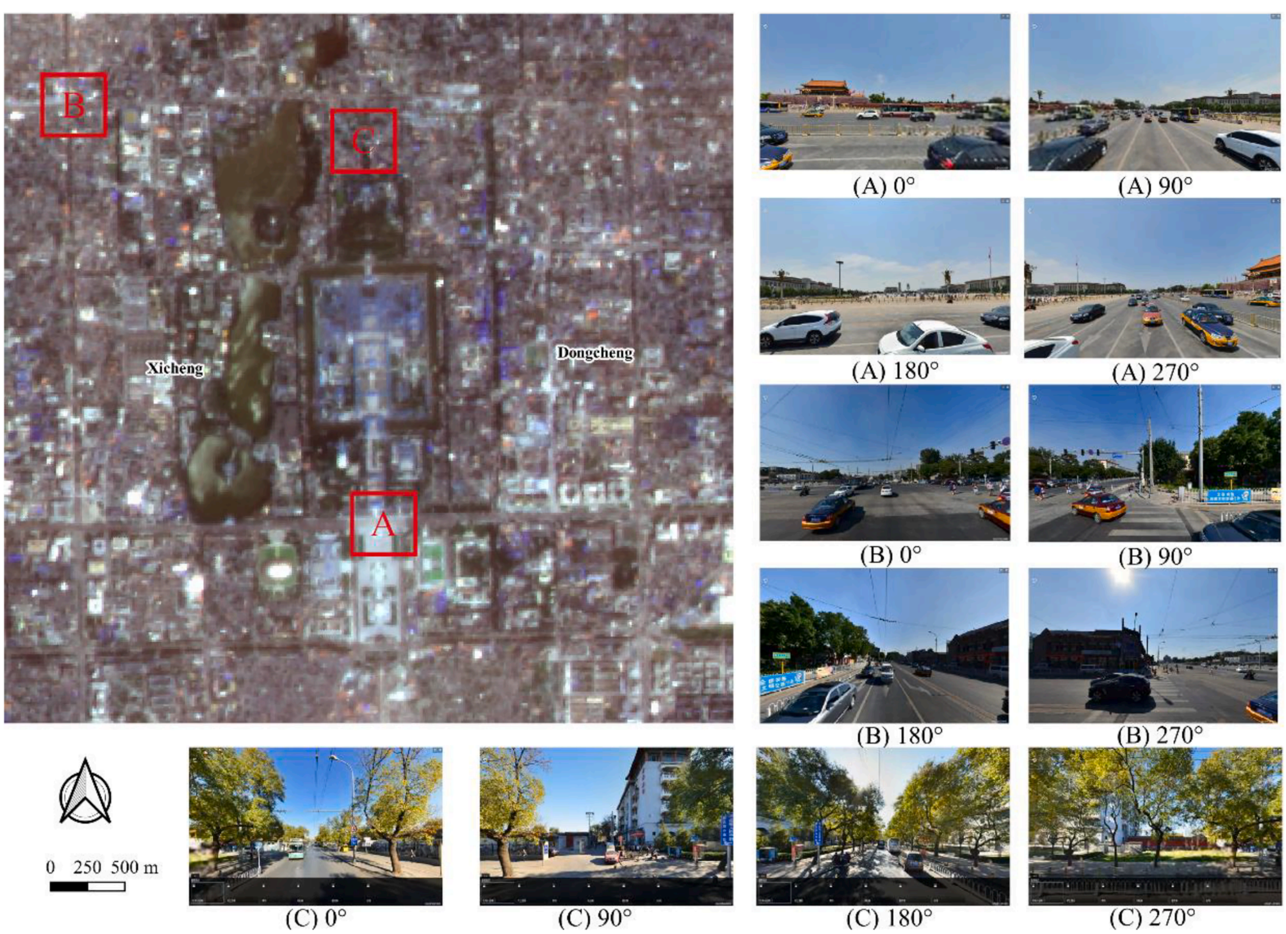

Fig. 3. Tencent street view image data. Case study areas: (A) Tiananmen Square, (B) at an intersection, and (C) Jingshan Park.

The distribution of six types of human perception along the roads in the study area is shown in Fig. 4. In general, the downtown areas are more “wealthy”, “safety”, “lively” and “depressing” than the outer suburbs (e.g., the fourth ring and the fifth ring roads). High-level roads such as ring roads and highways are more “boring” and “beautiful” than short and low-level roads. This result is reasonable because central urban areas tend to be densely populated, with well-developed commercial facilities and a lack of vegetation and natural scenery. We have given two sets of example street view images at the bottom of Fig. 4. The two on the left are from Zhongguancun (center of science and education), and the two on the right are found in the suburbs. More trees can be found in the suburbs, but with low buildings, fewer vehicles and people, making them more “beautiful” and “boring”. Noteworthy, several areas outside the third ring road are regarded as “wealthy”, “safety” and “lively”, which is related to universities, parks, and high-grade residential areas, indicating that human perceptions may associate with the actual functional area of the city. We further analyze the role of urban function in shaping the distribution of human perceptions in section 4.3.

The frequency distribution histogram of six types of human perception scores for all sampling points in the study area is shown in Fig. 5. According to the mean and median values, the study area is relatively “wealthy”, “boring”, and “depressing” (Mean $>$ 52 and Median $>$ 52), with a low degree of “safety” and “beautiful” (Mean $<$ 45 and Median $<$ 45). The frequency distribution histograms of Wealthy, Safety, and Lively scores show skew distribution patterns. Their standard deviation (Stdev $>$ 8.8) is significantly larger than others, which indicates that the three perceptions in the study area have a higher degree of heterogeneity. Perception of “boring” and “beautiful” show a normal distribution pattern with minor variances (Stdev $<$ 7.7), indicating that our study area is “boring” and “beautiful” in general.

4.2. Hierarchical community detection result of human perceptions

The 6,716 road nodes and 8,784 road segments in the study area are classified into hierarchical community structures. *PRV* for the different community levels in the two types of networks is shown in Fig. 6.

The *AvgArea* and *AreaSDE* of the different community levels for the two types of networks are shown in Fig. 7.

Fig. 6 compares the perception homogeneity of communities obtained by two network models (i.e., traditional Euclidean distance network model and perception network model, respectively). As shown in Fig. 6, the identified communities’ perception homogeneity increases with increasing community level. Compared with the traditional Euclidean distance network model, the perception network model is better at identifying homogeneous communities at a fine scale, as shown from the lower *RPV* value from the beginning of level 2. The difference in *RPV* is the largest in the third level communities, with the *RPV* of the perception network model 9.86%-20.21% lower than that of the Euclidean distance network model.

Fig. 7 compares the spatial size of areas (i.e., *AvgArea* and *AreaSDE*) identified by two network models. At the same level, the community

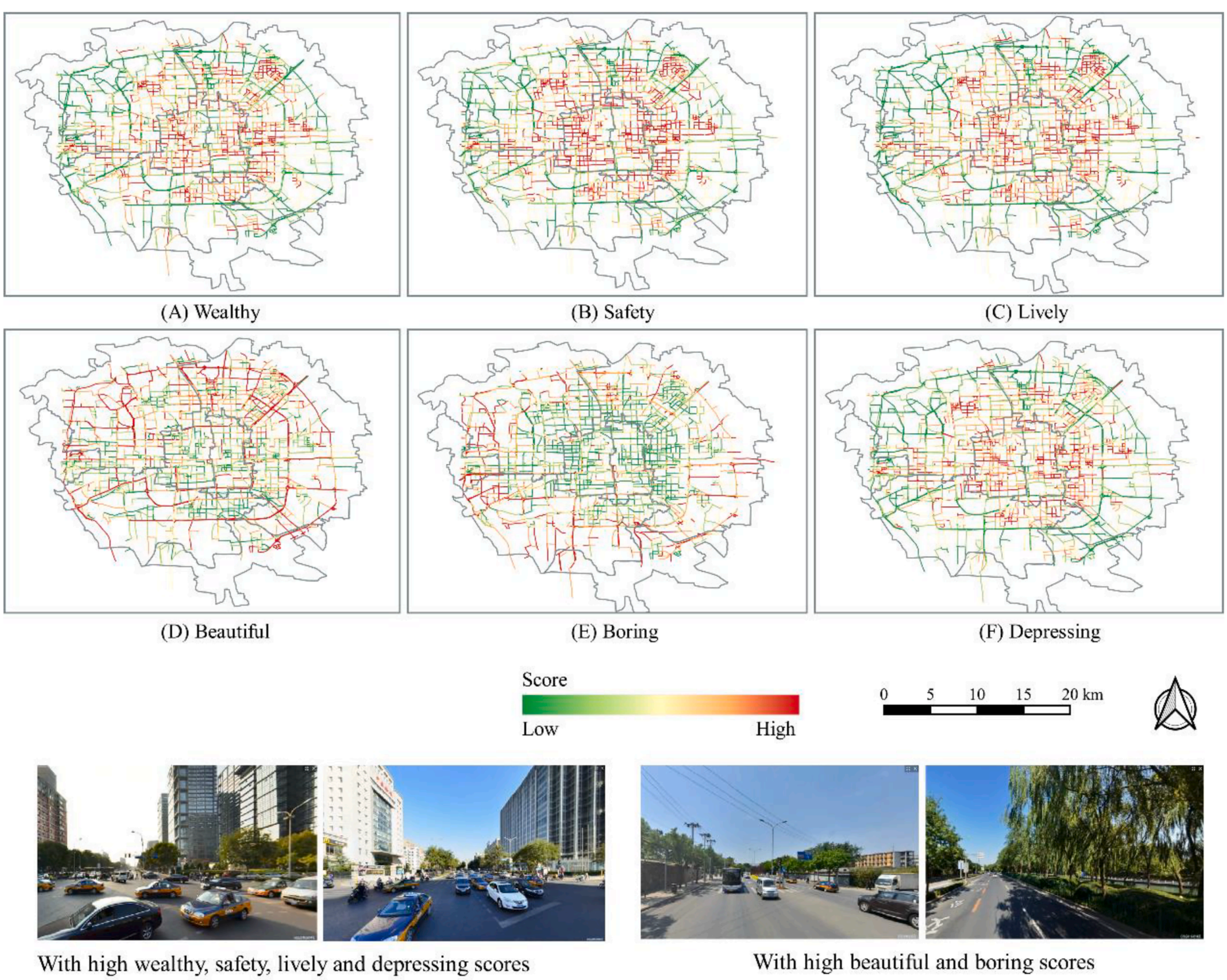

Fig. 4. The distribution of six types of human perception along the roads: (A) wealthy, (B) safety, (C) lively, (D) beautiful, (E) boring and (F) depressing.

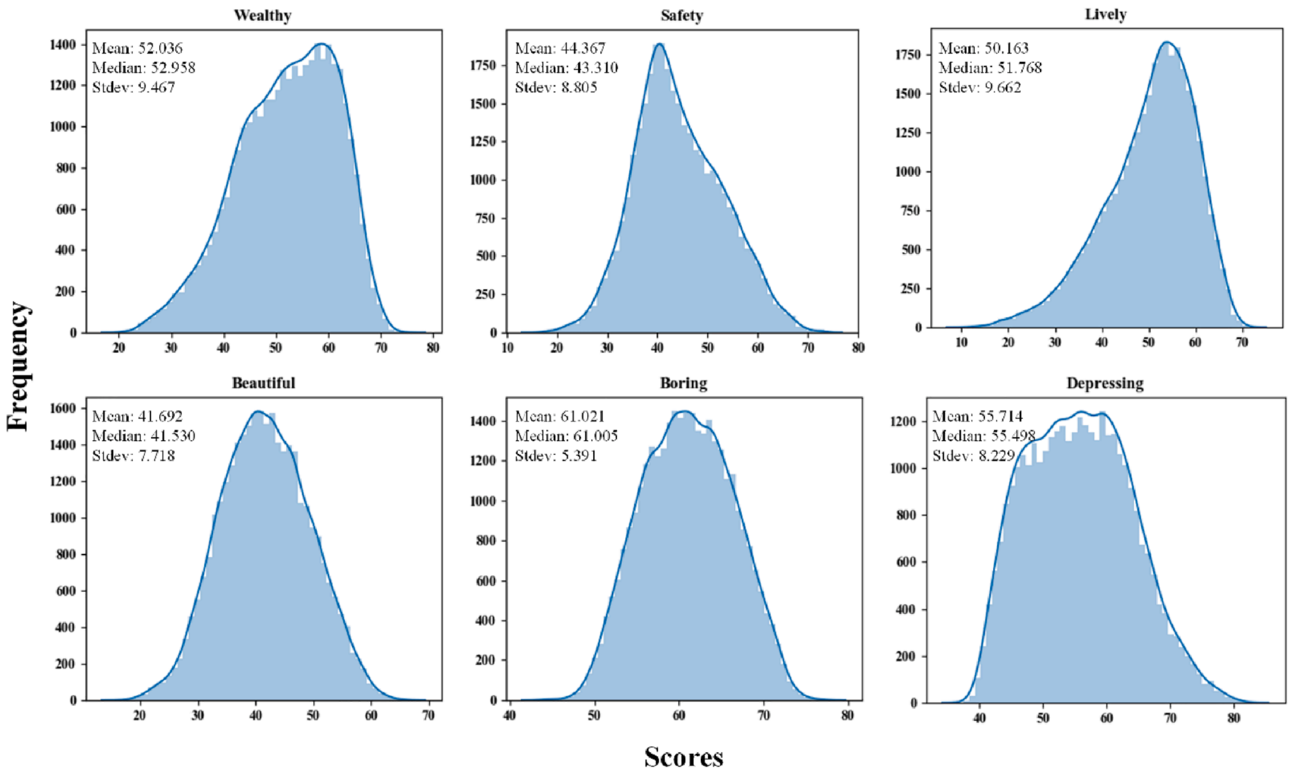

Fig. 5. The frequency distribution histogram of six types of human perception scores in the study area.

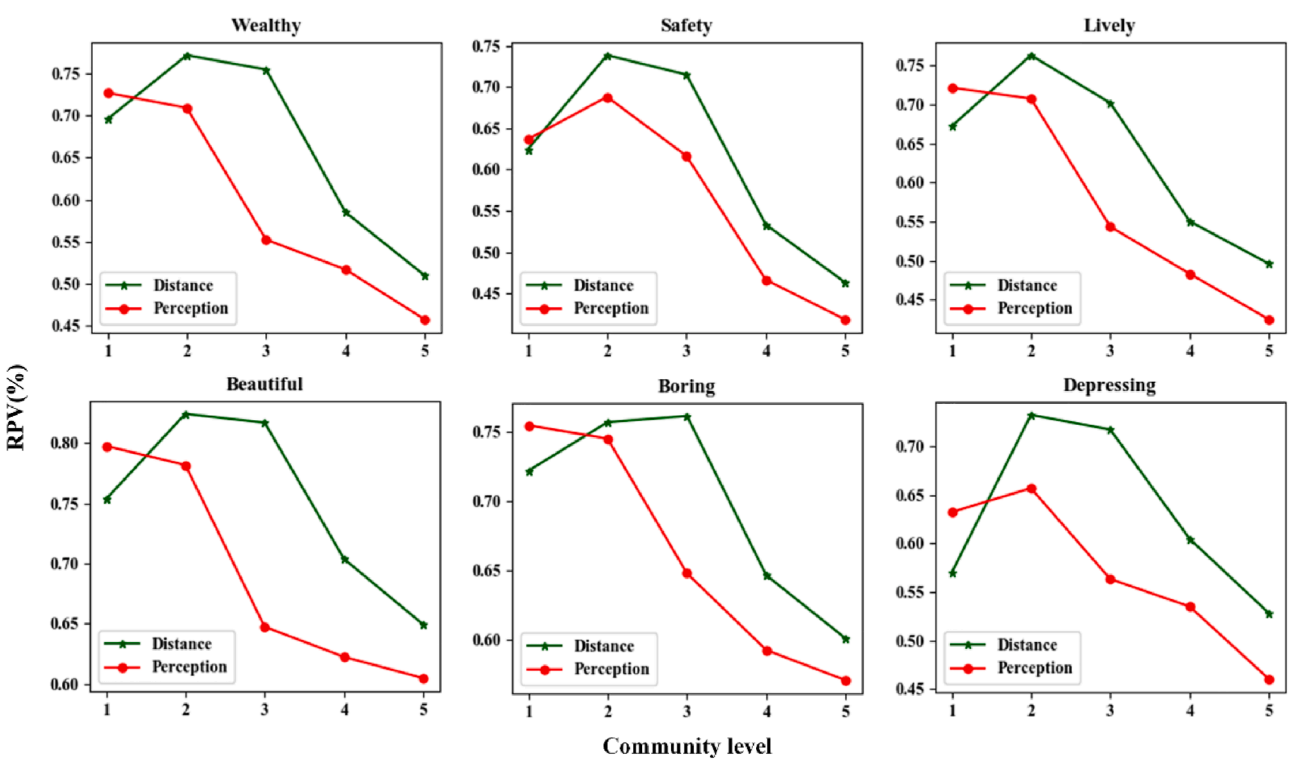

Fig. 6. The *RPV* of the wealthy, safety, lively, beautiful, boring and depressing perceptions at different community levels. The green line represents the result based on the Euclidean distance network models, and the red line is the result based on the perception network models. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

sizes determined by the Euclidean distance network model are larger than those determined by the perception network model. These results could verify the effectiveness of the perception network model to identify homogeneous communities at fine-grained scales. The division of a network depends mainly on the spatial location and spatial topological relationship of the nodes. Spatial nodes with more substantial

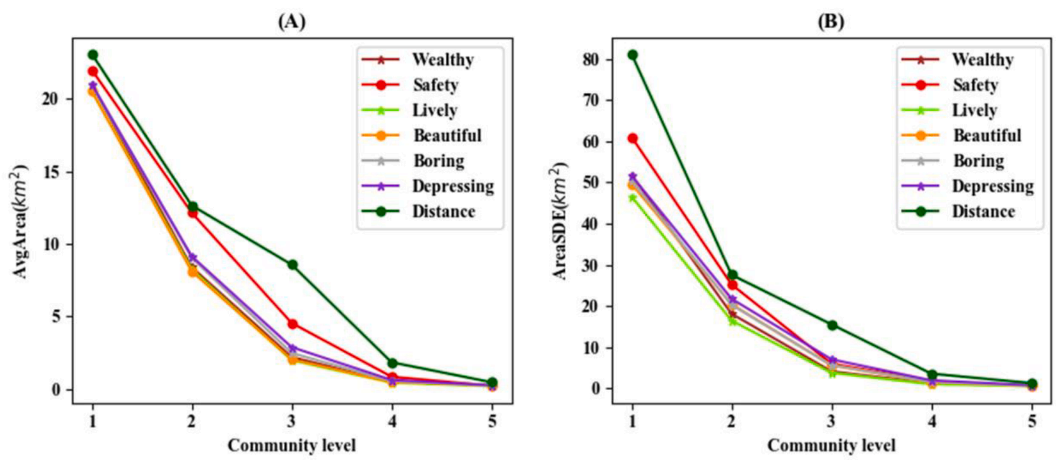

Fig. 7. (A) The *AvgArea* of identified communities at different levels. (B) The *AreaSDE* of identified communities at different levels.

spatial variable similarity and connection will be more easily clustered into the same community. Our results indicate that the hierarchical community discovery results based on the perception network models can identify homogeneous human perception domains at a fine scale while maintaining the street network's spatial topology.

The previous study has proved Infomap's ability to reveal urban space's partition at different scales (Hong and Yao 2019). Typically, the top level expresses the connection between the city center and the surrounding area; the second level reflects the spatial relationships within the core downtown area; the third level results can identify an area's single urban function. Hence, we choose the third level community detection result for exploring the relationship between the distribution of human perceptions and urban functions. We calculate the mean score of six human perceptions in the third level community, representing the community's human perception. The spatial distribution of the human perception communities based on the perception network models and the traditional Euclidean distance network models in the third level community is shown in Figs. 8 and 9.

We mark the corresponding positions in Fig. 8(A) and Fig.9(A) for comparing the outputs of the two models. The areas numbered 1 to 5 are the city's science and education, economy, transportation, recreation, and residential areas in Beijing. We could find that the identified communities align well with the actual functional area of the city. For example, results of both network models show area 2 a "wealthy" region. Area 5, which is a desolate residential area that lacks commercial facilities, is identified as an area of less "wealthy". Moreover, the

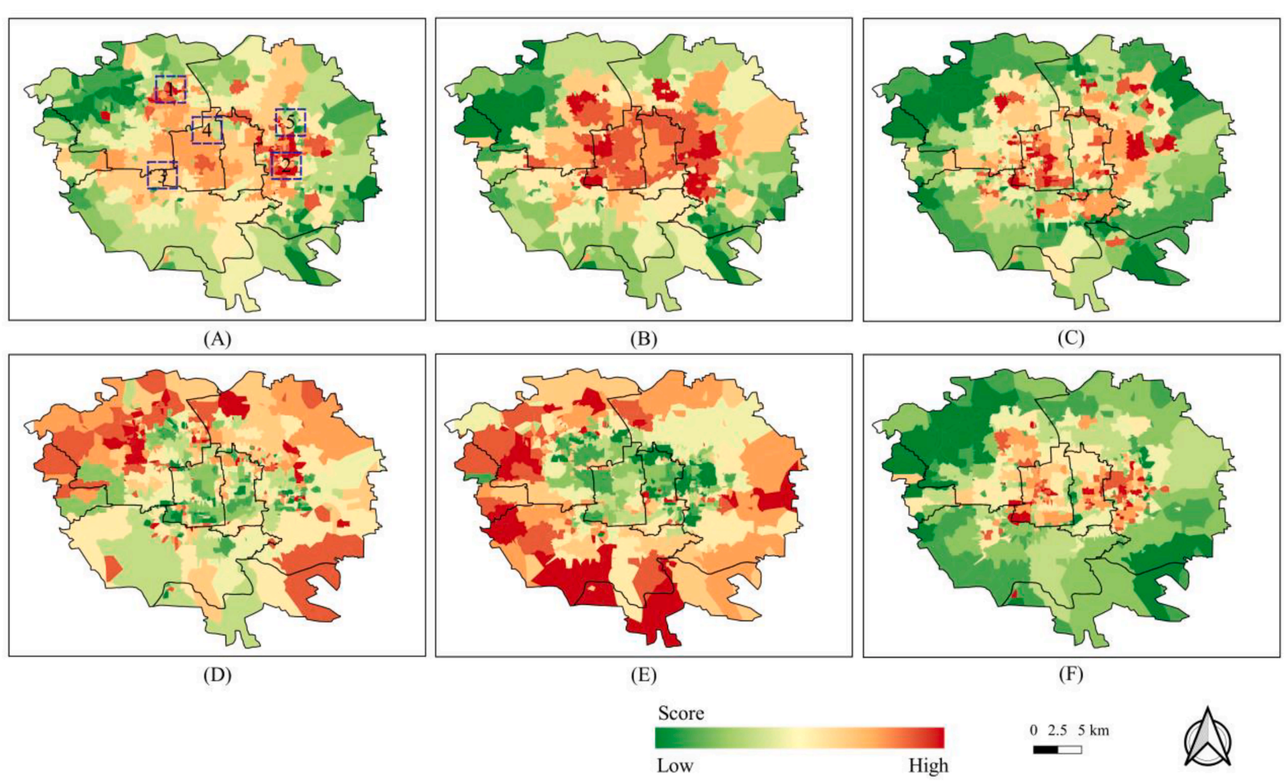

Fig. 8. Community discovery results based on the perception network models, where (A) shows the perception of wealthy, (B) shows the perception of safety, (C) shows the perception of lively, (D) shows the perception of beautiful, (E) shows the perception of boring and (F) shows the perception of depressing. Some identified areas are also shown in (A): (1) Zhongguancun, (2) financial center named Guomao, (3) Beijing west railway station, (4) Houhai park, and (5) Liangma bridge residential area.

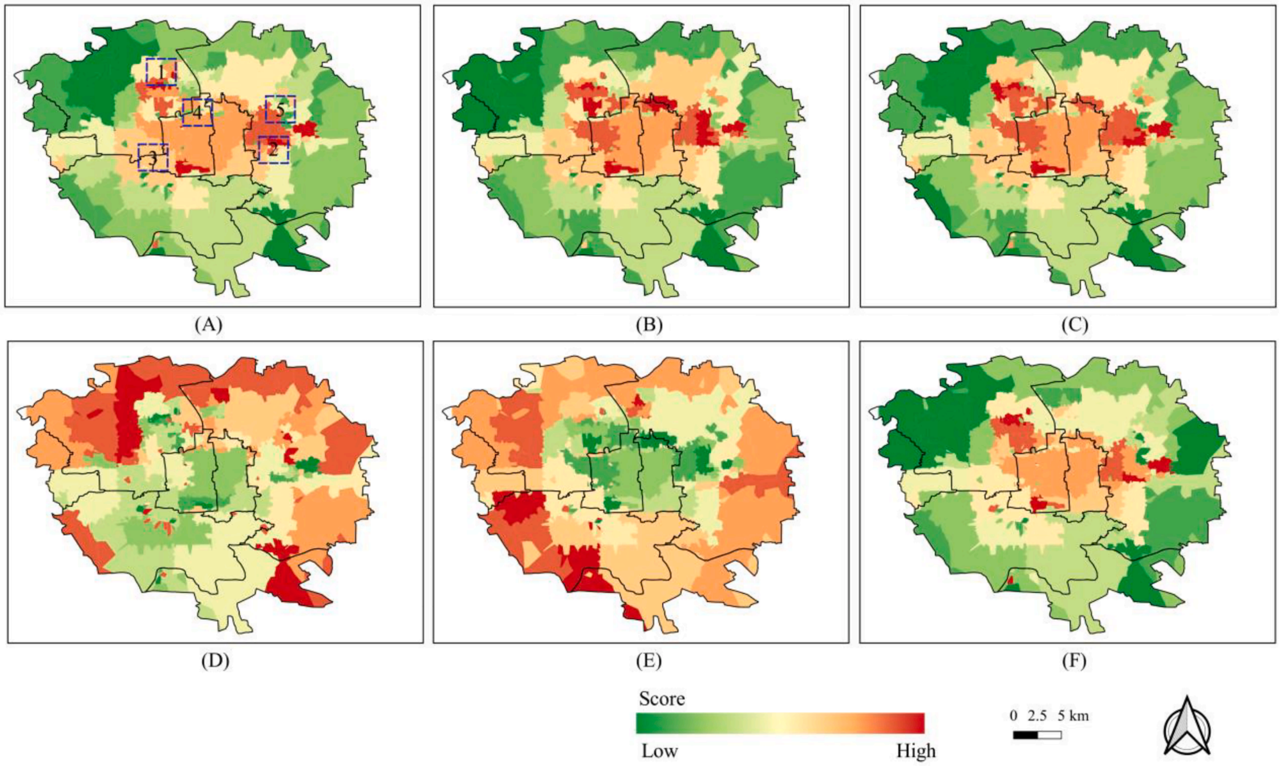

Fig. 9. Community discovery results based on Euclidean distance network models, where (A) shows the perception of wealthy, (B) shows the perception of safety, (C) shows the perception of lively, (D) shows the perception of beautiful, (E) shows the perception of boring and (F) shows the perception of depressing. Some identified areas are also shown in (A): (1) Zhongguancun, (2) financial center named Guomao, (3) Beijing west railway station, (4) Houhai park, and (5) Liangma bridge residential area.

resulting land parcels of the perception network also exhibit a more detailed spatial structure and reflect human perception's spatial heterogeneity.

4.3. The relationship between the spatial distribution of human perceptions and urban function

Based on the mean scores in the third level community, the Moran's I index value and Z-values are calculated, as shown in Table 2. The results indicate that all types of human perceptions show a positive spatial correlation in space, and they exhibit different degrees of spatial aggregation. In particular, three types of urban perception types, namely, wealthy, safety, and lively, are strongly clustered and are mainly concentrated in the center of the city, which is in line with the human perception distribution shown in Fig. 4.

We construct a nonlinear random forest regression model to explore the different perception types' driving factors from urban function. The POI density (i.e., FS, ME, ENT, GOV, RC, LS, OBS), the POI mixing entropy (i.e., Entropy), and the density of the road network (i.e., Roads) are treated as the explanatory variables. The human perception scores of the corresponding communities are treated as the dependent variables.

Table 2
Moran's I index and z-value of different types of urban perception at the community level.

	Moran's I	Z-value
Wealthy	0.34	26.90
Safety	0.24	7.31
Lively	0.25	21.49
Beautiful	0.05	4.70
Boring	0.05	5.13
Depressing	0.17	13.18

The model-fitting results are shown in Table 3. All the perception types achieve good model fitting results for the training dataset ($R^2 > 0.80$, Pearson $R > 0.90$). Moreover, all the perception types except "beautiful" and "boring" demonstrate high generalization ability based on the test dataset ($R^2 > 0.50$, Pearson $R > 0.70$).

The weights of the different explanatory variables in the model fitting procedure are shown in Table 4. We could find that different urban functions have distinct impacts on each type of human perception. In general, entertainment, medical and educational services, government departments, life services, and residential communities are the most critical factors affecting human perceptions in the study area. The density of the road networks and office space is less critical in shaping human perceptions' spatial patterns. For example, entertainment facilities are the most important for "depressing", probably because entertainment facilities help relieve stress. It also suggests that the environment's human perception can be made more friendly by optimizing the types and improving the mixing degree of urban functions.

5. Discussion

This study proposed a novel methodology for detecting the homogeneous geographical domain of human perceptions based on street view images and OSM road networks. We used the proposed method to obtain the irregular geographic environment at a fine-scale in large urban areas. Urban residents have strong homogenous perceptions of the obtained sub-regions about the regional physical settings and the visual information, thereby providing a unique perspective for observing the residents' perceptions of the urban environment. The qualitative and quantitative analysis of homogenous perception community detection results verified the feasibility of our method.

This study first proposed the concept of the human perception homogeneity domain in the urban area. Also, we extended the research

Table 3
Fitting accuracy of the human perceptions based on the POI indices of urban function and OSM road density according to the RF model.

	Training dataset				Testing dataset			
Perception	RMSE	MAE	R^2	Pearson R	RMSE	MAE	R^2	Pearson R
Wealthy	2.210	1.784	0.904	0.957	3.700	2.972	0.657	0.810
Safety	1.117	0.873	0.949	0.979	2.734	2.082	0.765	0.884
Lively	1.981	1.548	0.893	0.965	3.568	2.745	0.654	0.810
Beautiful	1.560	1.192	0.875	0.968	3.249	2.495	0.344	0.594
Boring	1.250	0.973	0.830	0.938	1.974	1.597	0.490	0.710
Depressing	1.397	1.053	0.926	0.972	3.282	3.282	0.583	0.776

Table 4
The RF's fitted weights between the human perceptions and the POI density and OSM road density. Values are given for the financial services (FS), medical/education services (ME), entertainment (ENT), government departments (GOV), residence communities (RC), life services (LS), office building/space (OBS), the POI mixing entropy (Entropy), and the road density (Roads).

Perceptions	FS	ME	ENT	GOV	RC	LS	OBS	Entropy	Roads
Wealthy	0.146	0.139	0.121	0.074	0.136	0.168	0.088	0.063	0.067
Safety	0.087	0.097	0.111	0.181	0.172	0.097	0.065	0.110	0.047
Lively	0.183	0.111	0.098	0.150	0.117	0.144	0.065	0.082	0.053
Beautiful	0.102	0.086	0.115	0.109	0.074	0.235	0.086	0.111	0.085
Boring	0.087	0.100	0.112	0.168	0.150	0.136	0.081	0.085	0.088
Depressing	0.084	0.123	0.220	0.090	0.138	0.111	0.069	0.106	0.063

scale of human perception of the urban environment from geographical point to geographical polygon unit. Previous studies on human perceptions focused only on a specific geographical location (Dubey et al. 2016, Yao et al. 2019, Zhang et al. 2018a). Through the street view images and establishing a model, a location's visual environment and the residents' psychological perception could be linked. However, they cannot provide an answer to the residents' perception of a continuous geographical area. Taking a qualitative and quantitative measurement of the regionalization of homogeneous perception at a fine-scale is still a problem that urgently needs to be solved.

This study used the human–machine adversarial scoring method to assess the city-scale human perception of the urban environment in a cost-efficient and accurate way based on massive street view images. Considering the road network's ability to shape a city's regional pattern, we employed the community detection algorithm to reveal the hidden spatial distribution of homogenous perception domain from established perception network models. In this study, these homogenous perception polygons with clear geographical boundaries indicate that the residents' psychological perceptions have a heterogeneous spatial distribution. The obtained communities are internally topologically well-connected but externally less so. The residents' perceptions of the same community's urban environment have a relatively high homogeneity level (lower *RPV* value). Therefore, the detected geographic perception area can be used as a basic unit in other urban planning and public health studies.

Few studies focused on homogeneous perception regions, and we still lack an indicator to measure the degree of homogeneity. Here, we proposed the *RPV* value as such an indicator. Still, questions like what *RPV* could be considered highly homogeneous and selecting the appropriate level of communities should be further explored according to actual research and application needs. For example, communities ranked in the top 10% *RPV* could be taken as high-perception homogeneous. We consider it as a noteworthy direction and could be further explored in subsequent studies.

At a community scale, we investigated the relationship between the spatial distribution of human perception and urban function based on POIs. We found that different urban functions may cause a community to be perceived differently. The result quantitatively identified the impact of urban functions on human perceptions of the urban environment, demonstrating that cities have specific functions and carry the psychological and emotional attachment of urban residents to their living environment (Dubey et al. 2016, Goodchild 2011, Zhang et al. 2018a).

Several limitations of this study also deserve to be paid more attention to in future works. First, this study used the human–machine adversarial scoring framework, which only considered the urban environment's visual factors to obtain the residents' psychological perceptions. However, human perceptions result from the interaction of environment, economy, culture, psychology, and other factors. Therefore, future studies need to introduce more socioeconomic factors as input. Second, when exploring the relationship between the spatial distribution of human perceptions and urban function, other spatial data (such as vehicle trajectory and mobile phone) combined with POIs can better assist in identifying urban functions. Third, for urban areas where street view coverage is low, combining images from multiple sources, such as social media images and freely available Mapillary images, could assess urban regionalization of homogenous human perceptions more accurately. Fourth, our research only recruited twenty volunteers. In future research, the scoring system can be developed into a web service. Crowd-sourced scoring methods (for example, the MIT Place Pulse data collection platform, as mentioned above) can be used to evaluate street-view perception. Last, this study quantifies the urban perception as scores, with no obvious criteria to further distinguish different scores, such as "not beautiful", "generally beautiful", "relatively beautiful", and "very beautiful", etc. Future studies would try to design appropriate evaluation metrics for human perception assessment based on street view images.

6. Conclusion

Human perceptions have great significance in urban planning and public health. However, taking a quantitative measurement of the regionalization of homogeneous perception at a fine-scale is challenging. In this study, we proposed the concept of the human perception homogeneity domain in cities. This study employed the Infomap method to reveal human perception's spatial distribution by combining the road network's topological structure with the residents' psychological perceptions. The qualitative and quantitative results verified our approach's great potential for capturing the homogeneous geographical domain of human perceptions, both in terms of fineness and high homogeneous. The random forest models' fitted weight between human perceptions and urban function indicated that different urban functions play distinct roles in shaping urban perception.

Our work is a constructive attempt to explore the urban region's human perception by combining geo-spatial data with the urban residents' psychological perception. This study first focuses on the spatial homogeneity of human perceptions at the land parcel scale. We also provide a unique perspective for understanding urban spatial heterogeneity and helps researchers understand the underlying urban structure and reveal urban function impacts. With the improvement of acquisition technology leading to the increase of street view coverage, the proposed method can explore the differences in the distribution of homogeneous perception areas between different cities. We believe that the study of human perception's spatial distribution will bring new inspiration to urban planning and public health studies.

Funding

This work was supported by the National Key R&D Program of China [2019YFB2102903]; National Natural Science Foundation of China [41801306, 41671408, 41901332]; Fundamental Research Founds for National University, China University of Geosciences(Wuhan) [CUG190606]; Open Fund of State Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University [18S01]; the Natural Science Foundation of Hubei Province [2017CFA041].

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Bahrehdar, A. R., et al. (2020). Streets of London: Using Flickr and OpenStreetMap to build an interactive image of the city. *Computers, Environment and Urban Systems, 84*, Article 101524.

Breiman, L. (2001). Random Forests. *MACHINE LEARNING, 45*(1), 5–32.

Cheng, L., et al. (2017). Use of Tencent Street View Imagery for Visual Perception of Streets. *ISPRS International Journal of Geo-Information, 6*(9), 265.

Cresswell, T. (1992). *In place-out of place: geography, ideology, and transgression*. U of Minnesota Press.

Dadvand, P., et al. (2016). Green spaces and General Health: Roles of mental health status, social support, and physical activity. *Environment International, 91*, 161–167.

Dubey, A., et al. (2016). *Deep Learning the City: Quantifying Urban Perception at a Global Scale* (pp. 196–212). Cham: Springer International Publishing.

Ndez-Delgado, F. A., M.,, et al. (2014). Do we need hundreds of classifiers to solve real world classification problems. *The Journal of Machine Learning Research, 15*(1), 3133–3181.

Frank, L. D., Andresen, M. A., & Schmid, T. L. (2004). Obesity relationships with community design, physical activity, and time spent in cars. *American Journal of Preventive Medicine, 27*(2), 87–96.

Gebru, T., et al. (2017). Using deep learning and Google Street View to estimate the demographic makeup of neighborhoods across the US. *Proceedings of the National Academy of Sciences, 114*(50), 13108–13113.

Gong, F., et al. (2018). Mapping sky, tree, and building view factors of street canyons in a high-density urban environment. *Building and Environment, 134*, 155–167.

Goodchild, M. F. (2011). Formalizing Place in Geographic Information Systems. In *Communities, Neighborhoods, and Health: Expanding the Boundaries of Place* (pp. 21–33). Springer.

Helbich, M., et al. (2019). Using deep learning to examine street view green and blue spaces and their associations with geriatric depression in Beijing, China. *Environment International, 126*, 107–117.

Hong, Y., & Yao, Y. (2019). Hierarchical community detection and functional area identification with OSM roads and complex graph theory. *International Journal of Geographical Information Science, 33*(8), 1569–1587.

Kabisch, N., Qureshi, S., & Haase, D. (2015). Human–environment interactions in urban green spaces — A systematic review of contemporary issues and prospects for future research. *Environment Impact Assessment Review, 50*(50), 25–34.

Lancichinetti, A., & Fortunato, S. (2009). Community detection algorithms: A comparative analysis. *Physical. Review. E, 80*(5), Article 056117.

Li, H., Calder, C. A., & Cressie, N. (2010). Beyond Moran's I: Testing for Spatial Dependence Based on the Spatial Autoregressive Model. *Geographical Analysis, 39*(4), 357–375.

Li, Q., et al. (2011). Dynamic accessibility mapping using floating car data: A network-constrained density estimation approach. *Journal of Transport Geography, 19*(3), 379–393.

Li, X., et al. (2015a). Assessing street-level urban greenery using Google Street View and a modified green view index. *Urban Forestry & Urban Greening, 14*(3), 675–685.

Li, X., et al. (2015b). Who lives in greener neighborhoods? The distribution of street greenery and its association with residents' socioeconomic conditions in Hartford, Connecticut, USA. *Urban Forestry & Urban Greening, 14*(4), 751–759.

Li, X., et al. (2017). Building block level urban land-use information retrieval based on Google Street View images. *GIScience & Remote Sensing, 54*(6), 819–835.

Liu, X., et al. (2017). Classifying urban land use by integrating remote sensing and social media data. *International Journal of Geographical Information Science, 31*(8), 1675–1696.

Liu, Y., et al. (2015). Social Sensing: A New Approach to Understanding Our Socioeconomic Environments. *Annals of the Association of American Geographers, 105* (3), 512–530.

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3431–3440).

Michael, B. (2008). The size, scale, and shape of cities. *Science, 319*(5864), 769–771.

Naik, N., et al. (2017). Computer vision uncovers predictors of physical urban change. *Proceedings of the National Academy of Sciences, 114*(29), 7571–7576.

Ordonez, V., & Berg, T. L. (2014). Learning High-Level Judgments of Urban Perception. In *European conference on computer vision* (pp. 494–510). Cham: Springer.

Porzi, L., et al., 2015. Predicting and understanding urban perception with convolutional neural networks. 139–148.

Rosvall, M., & Bergstrom, C. T. (2008). Maps of random walks on complex networks reveal community structure. *Proceedings of The National Academy of Sciences of The United States of America, 105*(4), 1118–1123.

Salesses, P. A. S. K. (2013). The Collaborative Image of The City: Mapping the Inequality of Urban Perception. *PLoS One, 8*(7), 1–12.

Tang, J., & Long, Y. (2018). Measuring visual quality of street space and its temporal variation: Methodology and its application in the Hutong area in Beijing. *Landscape and Urban Planning, 191*, Article 103436.

Tuan, Y. F. (2013). *Landscapes of fear*. University of Minnesota Press.

Wang, P., et al. (2012). Understanding road usage patterns in urban areas. *Scientific Reports, 2*(12), 1001.

Wang, R., et al. (2019a). The linkage between the perception of neighbourhood and physical activity in Guangzhou, China: Using street view imagery with deep learning techniques. *International Journal of Health Geographics, 18*(1), 18.

Wang, R., et al. (2019b). Perceptions of built environment and health outcomes for older Chinese in Beijing: A big data approach with street view images and deep learning technique. *Computers, Environment and Urban Systems, 78*, Article 101386.

Wang, R., et al. (2019c). The relationship between visual enclosure for neighbourhood street walkability and elders' mental health in China: Using street view images. *Journal of Transport & Health, 13*, 90–102.

Wolch, J. R., Byrne, J., & Newell, J. P. (2014). Urban green space, public health, and environmental justice: The challenge of making cities' just green enough'. *Landscape and Urban Planning, 125*, 234–244.

Yao, Y., et al. (2018a). Estimating the effects of "community opening" policy on alleviating traffic congestion in large Chinese cities by integrating ant colony optimization and complex network analyses. *Computers, Environment and Urban Systems, 70*, 163–174.

Yao, Y., et al. (2018b). Mapping fine-scale urban housing prices by fusing remotely sensed imagery and social media data. *Transactions in GIS, 22*(2), 561–581.

Yao, Y., et al. (2019). A human-machine adversarial scoring framework for urban perception assessment using street-view images. *International Journal of Geographical Information Science, 33*(12), 2363–2384.

Yin, L., & Wang, Z. (2016). Measuring visual enclosure for street walkability: Using machine learning algorithms and Google Street View imagery. *Applied Geography, 76*, 147–153.

Zhang, F., et al. (2018a). Measuring human perceptions of a large-scale urban region using machine learning. *Landscape and Urban Planning, 180*, 148–160.

Zhang, F., et al. (2018b). Representing place locales using scene elements. *Computers Environment & Urban Systems, 71*, 153–164.

Zhang, W., et al. (2017). Parcel-based urban land use classification in megacity using airborne LiDAR, high resolution orthoimagery, and Google Street View. *Computers, Environment and Urban Systems, 64*, 215–228.

Zhou, B., et al. (2016). Semantic Understanding of Scenes Through the ADE20K Dataset. *International Journal of Computer Vision, 127*(3), 302–321.

Zhou, B., et al. (2017). Scene parsing through ADE20K dataset. In *In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 633–641).

Zhou, Q. (2018). Exploring the relationship between density and completeness of urban building data in OpenStreetMap for quality estimation. *International Journal of Geographical Information Science, 32*(2), 257–281.