Sustainable Cities and Society 134 (2025) 106906

Contents lists available at ScienceDirect

Sustainable Cities and Society

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

Who gets to use the street? Evaluate the utilization and inclusiveness using crowdsourced videos and vision-language models

Xiamengwei Zhang [a], Mingze Chen [b,*], Yongming Huang [c]

[a] College of Fine Arts, Capital Normal University, Beijing 100048, PR China
[b] Faculty of Forestry, University of British Columbia, Vancouver, BC V6T 1Z4, Canada
[c] Landscape Planning Laboratory, Graduate School of Horticulture, Chiba University, B-Building, Matsudo Campus 648 Matsudo, Chiba, 271-8510, Japan

ARTICLE INFO

Keywords:
Pedestrian behavior
Vulnerable groups
Public street spaces
Social justice
Computer vision
Crowdsourced data

ABSTRACT

Equitable access to public street spaces (PSS) represents a fundamental prerequisite for sustainable urban development, yet systematic evaluation frameworks remain inadequately developed. Existing studies suffer from demographic homogenization assumptions, and insufficient data resolution. The deficiency undermines the effective utilization and full inclusiveness of PSS, ultimately constraining street vitality, social cohesion, and progress toward sustainable development goals. We introduce a crowdsourced data collection method based on food delivery riders, significantly enhancing the detail and coverage of the dataset. Therefore, we collect over 11,000 h of video and construct the "EgoCity Dataset". Subsequently, we develop a powerful vision-language model (VLM) capable of identifying demographic groups and behaviors (F1-score: 0.9583), addressing critical gaps in existing computer vision applications for urban analysis. Finally, we propose a framework for evaluating PSS, which quantifies the actual level of urban vitality and spatial equity through two key indicators: utilization and inclusiveness. The main findings are: (1) At the individual level, pedestrian distribution shows significant demographic disparities, with adults dominating both volume and activity diversity. (2) At the street level, there is a spatial mismatch between pedestrian flow and residential population. (3) At the regional level, spatial advantages and resources are concentrated in select areas at the expense of broader equity. Our study proposes a low-cost data collection method, and is the first to integrate crowdsourced street video data with vision-language models for pedestrian group identification, as well as to introduce a novel framework for evaluating utilization and inclusiveness in public street spaces. This study offers practical tools for urban planning and advances the sustainable development goals.

1. Introduction

In the context of global urbanization and social diversification, the concept of the just city, as developed by Susan Fainstein (2014)), emphasizes equity, democracy, and diversity to ensure fair and inclusive urban development for all social groups (Biernacka et al., 2022; Dutta et al., 2025; Fang et al., 2021; S. Guo et al., 2019; Lu et al., 2024; Ruoyu Wang et al., 2022; Xiao et al., 2017), which aligns with Sustainable Development Goal (SDG) 10 (Reduced Inequalities) and SDG 16 (Peace, Justice, and Strong Institutions). Urban land is classified into various functional zones, such as residential, commercial, and industrial areas, each fulfilling specific role (Song et al., 2013). Public street spaces (PSS) as the connective infrastructure that integrates these discrete zones into a coherent urban entity (Salazar-Miranda et al., 2023). Streets serve not only as pathways for daily human activities—contributing to good health and well-being (SDG 3)—but also provide a framework for linking and coordinating various social interactions (Karndacharuk et al., 2014; Su et al., 2022; Wang et al., 2022), ensuring the smooth and organized operation of society (Jacobs, 1961).

1.1. *Core concepts: utilization and inclusiveness*

Balancing urban vitality and social equity is essential for achieving sustainable urban development (Campbell, 1996; Development, 1996; Shirazi & Keivani, 2019). On one hand, an excessive focus on economic growth and urban vitality often leads to what Fainstein (2014)) criticizes as a "growth machine model", which systematically marginalizes vulnerable populations. As Harvey (2010) notes, urban space reinforces

* Corresponding author.
E-mail addresses: 2231202078@cnu.edu.cn (X. Zhang), mingze.chen@ubc.ca (M. Chen), 25hd0402@student.gs.chiba-u.jp (Y. Huang).

https://doi.org/10.1016/j.scs.2025.106906
Received 29 May 2025; Received in revised form 29 July 2025; Accepted 15 October 2025
Available online 16 October 2025

social inequality. Conversely, if social equity is overemphasized at the expense of economic competitiveness, it tends to result in economic decline, weakening the city's capacity to provide public services and social welfare, ultimately harming the long-term interests of all groups (Le Grand & Bartlett, 1993; Y. Tang & Zhang, 2017). Previous studies have highlighted the significance of vitality and equity. However, they lack concrete indicators to measure them in practice. Therefore, we propose two quantitative indicators—utilization and inclusiveness—to access the levels of vitality and equity in PSS.

The utilization of PSS is reflected in various activities such as walking, staying, exercise, socializing and so on (H. He et al., 2020; Salazar-Miranda et al., 2023, 2024; J. Zhao et al., 2024). Utilization evaluates the intensity of space usage(Sun et al., 2020), and diversity of activities (Su et al., 2022), revealing whether the space is being fully and effectively utilized. Subramanian and Jana (2018) investigates how elderly people use recreational open spaces in three Indian cities. Liu et al. (2024)explored the impact of green space accessibility on actual usage frequency. Wang and He (2023) used computer vision to evaluate how people of different age groups use community-level square, fitness, and leisure spaces at different times. Current research largely focuses on public spaces with specific functions, while neglecting the most fundamental and ubiquitous unit of urban public spaces: streets (Karndacharuk et al., 2014; Su et al., 2022). Although these studies provide valuable insights into specific types of street space usage, this single-perspective approach limits our comprehensive understanding of PSS.

Inclusiveness refers to whether a street space can meet the needs of different demographic and ensure that everyone can access and use the space equally and conveniently (Shan & He, 2025). Although children and the elderly are potentially more vulnerable to unfair treatment due to their age and limited mobility (Boone et al., 2009; S. Guo et al., 2019; La Rosa et al., 2018), many studies still treat the population as a homogeneous entity (i.e., horizontal spatial equity)(Delbosc & Currie, 2011; Rad & & Alimohammadi, 2024; Xiao et al., 2017), ignoring the differences across different age groups (R. Guo et al., 2024; S. Guo et al., 2019).

1.2. Evolution of methods and data sources

Previous studies have employed various methods and tools to study public spaces, which can generally be divided into three phases based on technological advancements. The first phase, characterized by traditional on-site observation, involves direct observation and surveys (Gehl, 2011; Whyte, 1980). Aghaabbasi et al. (2017) combined questionnaire surveys with field observations to collect residents' evaluations of the importance and condition of sidewalk design elements. However, these methods can only collect data in a time-consuming and labor-intensive manner within a small scope, it's difficult to make large-scale researches and unable to draw universal conclusions. The second phase relies on GPS data such as cellphone location signals. Jiang et al. (2017) demonstrated how mobile phone data infers human activity networks. But it fails to provide accurate pedestrian information and detailed descriptions of activities (S. Guo et al., 2019; Salazar-Miranda et al., 2023). The third phase involves computer vision data. Early studies used social media data to analyze urban activity (Samiul Hasan, 2013). However, the quality of it is inconsistent. Later, Qi et al. (2020) applied deep convolutional neural networks to detect visual indicators of street vitality using street view images (SVI). Computer vision algorithms like Mask R-CNN and YOLO were also used to detect pedestrian behaviors via fixed surveillance cameras (Sun et al., 2020; Wang & He, 2023). Salazar-Miranda et al. (2023) further proposed a more advanced method by installing mobile devices on buses to collect data and using computer vision to measure street activities. Existing data collection methods face significant limitations, particularly in street view imagery (SVI) and fixed-camera. SVI-based approaches suffer from several key drawbacks (Rzotkiewicz et al., 2018; Seiferling et al., 2017). First, acquiring and processing SVI often requires trained personnel which limits scalability. Second, the imagery is typically updated once every 6 to 12 months, and exhibits a significant shooting bias which limits its timeliness and representativeness. Third, access to SVI is increasingly being restricted. Furthermore, most images are collected from vehicle-mounted cameras which may not be representative of pedestrian views (Chen et al., 2025). Lastly, trees, vehicles, and street furniture frequently cause occlusions, reducing the visibility of human activities. Meanwhile, Fixed-cameras provide real-time monitoring capabilities, but they are also constrained in important ways (Wang & He, 2023). The cost of deployment and maintenance is high, and operation relies on the stable infrastructure and technical support. Moreover, these cameras typically cover specific location and cannot capture the full spectrum of street space usage.

The previous phases of methodologies, although collect some parts of data, still unable to capture comprehensive and fine-grained data. Pedestrian activities in PSS are highly dynamic and dispersed, occurring across different times and locations, so measuring utilization and inclusiveness facing significant challenges. To overcome these limitations, we develop a low-cost and scalable framework for data collection through the daily mobility of delivery riders. According to reports in early 2025, the total number of food delivery riders nationwide has exceeded 10 million, and they work 8 to 12 h daily (based on field interviews and backend platform data). The reasons behind we selecting them are that 1) They are active in PSS for extended periods and across broad areas; 2) Their daily movement makes it possible to collect large amounts of real-time street video without hiring extra staff, greatly reducing costs; 3) They capture data from a pedestrian's perspective.

1.3. Research motivation and objectives

Existing studies lack attention to streets and often overlook differences across age groups in concept, and lack comprehensive coverage and fine-grained details in data. The deficiency will hinder effectively usage and fully inclusiveness of PSS (D. Liu et al., 2024; Lu et al., 2024; L. Zhang et al., 2019). Poorly designed streets significantly reduce individuals' engagement in outdoor activities, adversely affecting both well-being and traffic efficiency (de Vries et al., 2013; Hyejin Jung & Sae-young Lee, 2017; Zepp et al., 2018). Unequal access to PSS leads to weakened community cohesion and a diminished sense of belonging (S. Guo et al., 2019; M. Li et al., 2024; Lu et al., 2024). On a societal level, such disparities reinforce inequality, restrict the vitality and equity of PSS, and hinder sustainable development (Harvey, 2010; B. Tang, 2017). As Agboola et al. (2015) emphasize, sustainable public space governance hinges on integrating scientific knowledge with participatory planning processes.

In our study, we conduct population categorization and propose a novel data collection method to ensure broader spatial coverage and capture more fine-grained street activities. This research aims to (1) Propose a low-cost and scalable framework for data collection and analysis, and developed a corresponding high-accuracy hybrid vision-language model (VLM). (2) Analyse pedestrian volume and activities distribution from street videos using VLM. (3) Evaluate the utilization and inclusiveness of PSS based on different age groups.

2. Methodology

2.1. Study area

We select the Haidian District and its surrounding areas as the study area (Fig. 1). Haidian features diverse street types including residential, industrial, commercial, office, parks, research institutes, and universities. According to the Seventh National Population Census (2020), its population structure consists of 11.83 % children, 68.53 % adults, and 19.64 % elderly, similar to the overall population structure of Beijing. Therefore, the findings of this study can broadly reflect the trends across

(b) Street View Images of Rider Start Point (c) Detailed Map of Study Area

Fig. 1. The Map of the Study Area.

the Beijing.

2.2. Research flow

This study constructs a systematic framework to assess the inclusiveness and utilization of PSS in Beijing. As illustrated in Fig. 2, the framework comprises three stages.

First, we designate Haidian District and its surrounding area as the study area. Eight delivery riders (The specific locations of the riders are shown in Fig. 1) were employed and equipped with portable cameras to collect street-level videos over three weeks. >220 h of video footage was recorded. During the preprocessing phase, we extracted frames from videos, resulting in over 200,000 images. These images were filtered based on six exclusion criteria. Over 115,000 valid images were retained, forming the EgoCity dataset used in this study.

In the second stage, we developed the "LLaVA-PI (Pedestrian Identify)" model, a hybrid vision-language model (VLM). Which integrates DETA (used for pedestrian detection and cropping) and LLaVA-OV (demographic classification and activity recognition). 800 representative street images were selected for annotation, which constructed the labeled training set and supervising the training of LLaVA-OV. As a result, the LLaVA-PI model takes images as input and outputs the number of different demographic groups present in the scene as well as their corresponding activities. achieving strong performance with an F1-score of 0.9583 for pedestrian identification and 0.8689 for activity classification.

In the third stage, two core questions were proposed to evaluate public street space: (1) "Is the PSS fully and diversely used?" and (2) "Does the PSS serve all age groups fairly?" Correspondingly, two indicators were developed: Utilization and Inclusiveness. Utilization was calculated by weighting pedestrian volume and the diversity of activity types. Inclusiveness was computed based on the discrepancy between the model-detected demographic composition and the estimated residential population structure (mobile signaling data).

2.3. EgoCity dataset and model training

2.3.1. Street-level videos collection with crowdsourced

We recruited and compensated eight food delivery riders, affiliated with two major Chinese food delivery platforms (Ele.me and Meituan). They participated in the data collection process with full informed consent. These riders were strategically selected from various subdistricts across the study area, with portable cameras mounted behind the windshields of their scooters. All devices were pre-configured by technicians prior to deployment, and riders only needed to turn on the camera to begin street-level recording. This setup eliminated the need for technical staff to follow or for riders to perform complex operations. During food delivery operations, on-vehicle devices continuously captured real-time (timestamps, dates, and geographic coordinates) street data.

To address ethical and privacy concerns, we implemented several preventive measures throughout the workflow. Prior to data collection, communication with the local police department was established, and necessary registration and approval procedures were completed. We also obtained formal institutional ethics approval. During data collection, all devices operated in silent recording mode and were restricted to capturing footage in public street environments, with no entry into private or sensitive spaces. In the post-processing stage, all video footage was anonymized by masking visible faces, numbers, and textual elements (Fig. 3). Furthermore, only aggregated pedestrian data were used in the analysis, and no individual tracking was conducted.

The MALUCP portable camera was selected based on four key considerations: (1) its compact, portable design with dust-, water-, and shock-resistant feature; (2) its 4 K resolution capability enhances model recognition accuracy; (3) the integrated BeiDou-GPS dual-mode satellite positioning ensures accurate geolocation even in signal-marginal areas; and (4) the supercapacitor-powered system supports continuous recording for up to 23 h.

The data were collected from January 14 to 18, January 23 to 27, and February 23 to 27, 2025, from 7:00 a.m. to 6:00 pm. daily, covering both

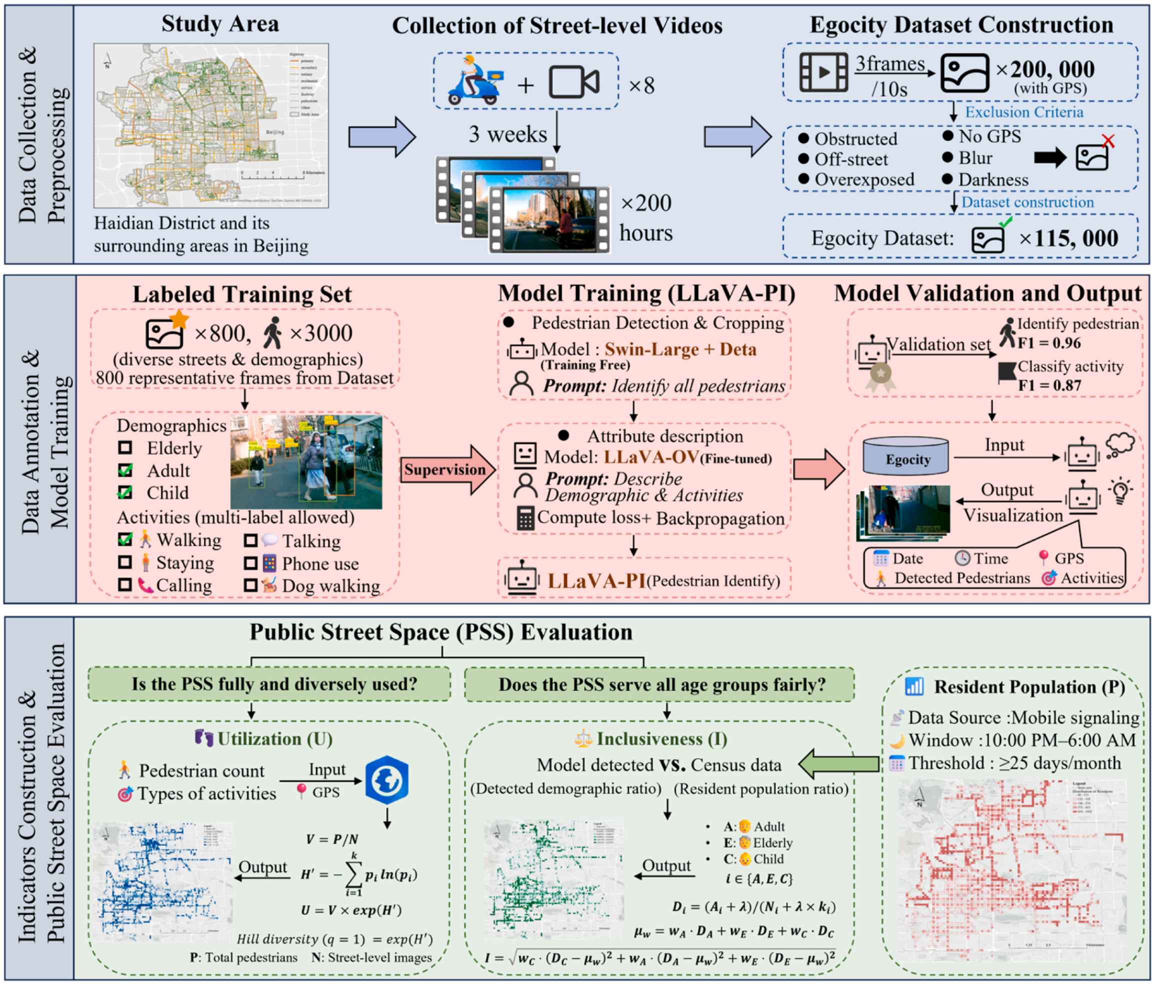


Fig. 2. Research Method and Framework.

Fig. 3. Examples of Images in the Training Dataset.

weekdays and weekends over three weeks, including periods before and after the Spring Festival, ensuring temporal representativeness. Finally, 221.3 h of street-level video footage were collected.

2.3.2. Steet-level images preprocessing

We extracted 3 frames per 10 s from the collected street activity videos and conducted a review process on the resulting image frames. 239, 004 street-level images were extracted through frame sampling, and two annotators were employed to assess the validity of the images. Any image marked as "invalid" by either annotator was discarded. Images were filtered based on the following criteria: 1) substantial occlusion by obstacles, 2) non-street scenes, 3) overexposure, 4) missing GPS information, 5) image blur, and 6) insufficient lighting. After data preprocessing, 118, 240 valid images were obtained, forming the EgoCity Dataset.

2.3.3. Data annotation phase

Since existing models cannot effectively and accurately identify diverse demographics or their activities, this study involves training a model. We selected 800 typical street scene images for annotation from the EgoCity Dataset, covering various street types and demographic groups. Two annotators were recruited to manually label each individual in the images using a custom annotation tool. The annotations included demographic categories (children, adults, elderly) and their activities (walking, staying, talking, phone use, calling, dog walking). Note that each individual could be engaged in multiple activities simultaneously. Specifically, individuals were labeled as elderly if they exhibited visibly aged facial features. In cases where facial details were unclear, secondary cues such as silver or white hair were considered, followed by mobility aids like canes or wheelchairs. Children were identified primarily through youthful facial features. When facial recognition was insufficient, indicators such as significantly shorter height compared to nearby pedestrians, school bags, toys, or scooters were used. All other individuals who did not meet the criteria for elderly or children were classified as adults by default.

To minimize subjective bias in the labeling process, both annotators were required to agree on the same labels for an image to be considered "valid." For images with inconsistent annotation results, multiple rounds of annotation were conducted until both annotators achieved complete agreement. At the end of the annotation process, the author reviewed all images to ensure accuracy.

2.3.4. Model training phase

The LLaVA-PI model is a VLM designed for pedestrian and activity classification, the data collection and model training workflow illustrated in Fig. 4. Beyond simply localizing pedestrians, the model classifies each detected individual into categories such as child, adult, or elderly, and further distinguishes their status as either walking or lingering. Additionally, the model recognizes auxiliary activities, such as chatting or using a phone, as part of a multi-label classification task. We employed the advanced detection model Deta, built on the Swin-Large backbone, to detect and crop pedestrians.

After the initial detection phase, the multi-label classification model is fine-tuned using the advanced LLaVA-OV vision-language model, which contains 7 billion parameters. The fine-tuning process of our multimodal model was conducted under a set of carefully selected hyperparameters to ensure training stability and performance. We adopted the Qwen/Qwen2–7B-Instruct language backbone and the SigLIP-SO400M-Patch14–384 vision encoder from Google. The training was executed using BFloat 16 (BF16) mixed-precision and DeepSpeed ZeRO-3 optimization for memory efficiency and optimized distributed training. The base learning rate was set to 1e-5 with a cosine learning rate scheduler and a warmup ratio of 0.03. We used the AdamW optimizer with no weight decay and applied gradient accumulation with two steps. The training lasted for one epoch, using a per-device batch size of 1 for training and 4 for evaluation. To improve memory efficiency and computational speed, both gradient checkpointing and PyTorch compilation (with the inductor backend) were enabled. The maximum input sequence length was 32,768 tokens. During training, model checkpoints were saved every 1000 steps, retaining only the single most recent checkpoint. For visual inputs, we adopted the `anyres_max_9` aspect ratio setting and a spatial unpad patch merge strategy, with image grid pinpoints set at intervals from 1×1 up to 6×6. The fine-tuning process targeted the `mm_vision_tower`, `mm_mlp_adapter`, and `mm_language_model` components, with a lower learning rate of 2e-6 assigned to the vision tower. Training was distributed across multiple GPUs using `torchrun`, with `NCCL` backend configured (`NCCL_IB_DISABLE=0`, `NCCL_SOCKET_IFNAME=eth0`, `NCCL_IB_GID_INDEX=3`) for optimal performance on InfiniBand-enabled GPU clusters. The training data consisted of image inputs, and the model was

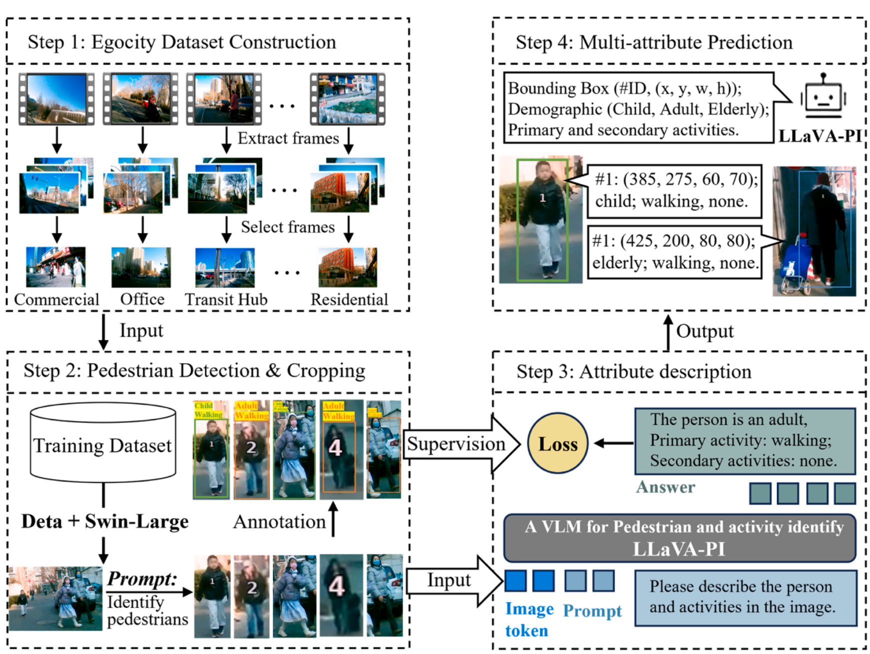


Fig. 4. EgoCity Dataset Construction and LLaVA-PI Training Framework.

trained without intermediate evaluation (`evaluation_strategy="no"`), with logs recorded at every training step (`logging_steps=1`). To ensure reproducibility and transparency, the training process was also tracked using Weights & Biases.

During training, each cropped pedestrian is queried with the prompt: "Please describe the person in the image." The model is expected to generate a response in the form of a sentence with a specified format, such as "The person is an adult: walking. Other activities: none." In cases where a person is driving, their status is classified as "driving," and they are excluded from the pedestrian category. If a cropped region does not contain a person, the VLM generates the response "no person."

For each image, the VLM processes all cropped pedestrian regions from the first stage, and a parsing mechanism is employed to convert the output into labels for statistical analysis. Note that the first stage is free of training, as it utilizes pre-trained detection models for the cropping task. The training dataset for the VLM consists of 800 images and >3000 cropped pedestrian instances, which provide supervision for the model's fine-tuning. The rationale for using vision-language models lies in their strong generalizability, which enables robust performance in pedestrian description tasks compared to traditional classifier models. By combining the detection and multi-label classification stages in a hybrid architecture, we achieve highly accurate pedestrian identification, categorization, and activity recognition. This hybrid approach significantly enhances the model's ability to provide detailed descriptions and improve overall performance in complex real-world scenarios.

2.3.5. Model validation

To evaluate performance, we tested LLaVA-PI on 105 manually annotated street images and compared it with two general-purpose multimodal models: GPT-4o and Gemini-1.5-pro (Fig. 5). In only pedestrian identification, LLaVA-PI achieved an F1 score of 0.9583, outperforming GPT-4o (0.7260) and Gemini-1.5-pro (0.8034). In the more complex pedestrian and activity task, it again led with an F1 score of 0.8689, compared to GPT-4o (0.5787) and Gemini-1.5-pro (0.6299). These results highlight its superior accuracy and strong adaptability to the complexity of PSS.

2.4. Evaluate the utilization and inclusiveness of PSS

The EgoCity dataset was processed with the LLaVA-PI model to extract timestamps, dates, GPS coordinates, pedestrian counts, demographic categories, and activity types. The data were georeferenced and imported into the Geographic Information System (GIS). All measures were spatially resampled into 100 m × 100 m grids, generating 2888 grids. To ensure data reliability, we retained grids containing at least 5 street-level images and finally resulted in 2106 valid grids.

2.4.1. . utilization measurement

Jane Jacobs (1961) systematically articulated the mechanisms underlying urban street vitality, emphasizing that diversity and continuous pedestrian flows are fundamental to the safety and vibrancy of city streets. Current studies focusing on how people utilize PSS can be broadly categorized into two types. Firstly, researches (Ewing & Cervero, 2010; Hyejin Jung & Sae-young Lee, 2017) confirmed the strong association between the public spaces and pedestrian volume. The second group (H. He et al., 2020; S. Jiang et al., 2017; Sun et al., 2020) focused on analyzing human activity patterns and their spatial distributions in public spaces. Based on these findings, this study adopts pedestrian volume and activity types as key indicators to measure the utilization of PSS.

To quantify the density of pedestrians, we calculated the average pedestrian volume (V) (Eq. (1)). Next, based on the detected activities of pedestrians in each image, we extracted and counted six distinct activity types: walking, staying, talking, phone use, calling, and dog walking. To evaluate the diversity of activities, we computed the *Hill diversity* $q = 1$, which corresponds to the exponential of the Shannon entropy (Eq. (2)). The final utilization value was derived by multiplying activity diversity with the average pedestrian volume (Eq. (3)).

$$V = P/N \tag{1}$$

$$H' = -\sum_{i=1}^{k} p_i \ln(p_i),\ \textit{Hill diversity} = \exp(H') \tag{2}$$

$$U = V \times \exp(H') \tag{3}$$

Here, P is the total number of pedestrians in the grid, N is the number of street-level images in the grid, and p_i represents the proportion of the i th term of activity type within the grid.

The Utilization Index is a relative metric without a physical unit. It is calculated as the product of pedestrian volume (people per image) and activity diversity (Hill diversity). Importantly, this index is not normalized. We deliberately retain its original numerical scale to enhance visual differentiation across grids. This allows the resulting heatmaps to more clearly reflect spatial disparities in street-level pedestrian activity and diversity, rather than compressing them into a restricted value range. This computed value (U) is referred to as the Utilization Index, which is used for mapping and spatial classification in subsequent figures.

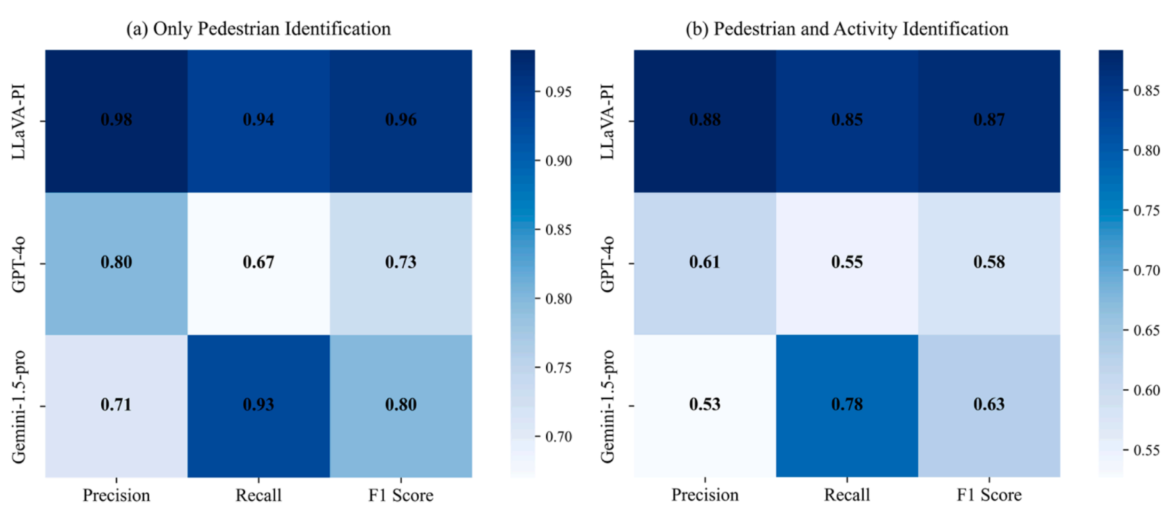


Fig. 5. Comparison of Detection Performance Across Models.

2.4.2. . inclusiveness measurement using the smartphone GPS data

This paper aims to assess the inclusiveness of PSS by comparing pedestrian demographic composition to resident population composition. Prior studies demonstrated the importance of census data when evaluating spatial equity. For example, Delbosc and Currie (2011) utilized Census and public transport service data to assess the equity between transport supply and population distribution in Melbourne. Xiao et al. (2017) employed census and urban park distribution data to investigate the equity of park accessibility in Shanghai. Similarly, Fang et al. (2021) integrated social media data with census to evaluate the spatial accessibility of attractions in China.

This study developed a standard deviation-based inclusiveness index (I) that measures how equitably street spaces serve different age groups. The approach integrates resident population data derived from census and mobile signaling data with street-level pedestrian recognition data from our LLaVA-PI model, employing Laplace smoothing to mitigate zero-frequency issues.

As census data are aggregated at the administrative district level and cannot provide the age-specific resident population data at the street scale, this study integrates census data with mobile signaling data to precisely estimate the demographic composition within each grid. Mobile signaling data recorded between 10:00 pm. and 6:00 a.m. were analyzed. If a user was located within the same grid during nighttime hours on >25 out of 30 days, the user was classified as a resident of that grid. Specifically, mobile signaling data were used to quantify the total resident count within each grid, and calculated the demographic structure of each administrative district using census data. Subsequently, the number of individuals across different age groups within each grid was then estimated by multiplying the total population of each grid by the corresponding demographic proportions of its district. The resident population counts of adults, elderly and children within a grid denote as N_A, N_E, N_C respectively. Similarly, the average pedestrian counts of per grid for three demographic groups—adults (A), elderly individuals (E), and children (C)—were identified using the LLaVA-PI model, denoted as P_A, P_E, P_C respectively. However, considering the prevalence of zero values in pedestrian counts (particularly among children and elderly groups), we apply Laplace smoothing to prevent extreme results. The observation density (D_i) for each group is calculated as:

$$D_i = (A_i + \lambda)/(N_i + \lambda \times k_i),\ i \in \{A, E, C\} \tag{4}$$

Where λ is a small smoothing constant set at 0.1 to match the scale of our observation data, and k_i is an adjustment factor defined as the non-zero median of P_i/N_i across all grids.

Next, to account for significant differences in population structure, we introduce population weight coefficients (w). w_A, w_E, and w_C represent the proportions of adults, elderly and children in the total resident population, respectively. We calculate the weighted average observation density (μ_w) as:

$$\mu_w = w_A{\cdot}D_A + w_E{\cdot}D_E + w_C{\cdot}D_C \tag{5}$$

Based on this, the Inclusiveness Index (I) is defined as the weighted standard deviation of observation densities across demographic groups:

$$I = \sqrt{w_C{\cdot}(D_C - \mu_w)^2 + w_A{\cdot}(D_A - \mu_w)^2 + w_E{\cdot}(D_E - \mu_w)^2} \tag{6}$$

This index I measures the deviation in observation densities across demographic groups. A smaller value of I indicates that the observed densities of children, adults, and elderly individuals are more similar, meaning that each group appears on the PSS at a comparable rate relative to their population size. This suggests a higher level of inclusiveness. Under the condition of sufficient observation counts, an index value of $I = 0$ implies that all three groups exhibit identical observed densities, signifying an ideally inclusive street environment where spatial representation is fully balanced.

However, this index may yield misleading results in certain edge cases. For instance, when all values of P_A, P_E, P_C are near zero, the densities D_i also approach zero. This leads to a low standard deviation and hence a falsely high inclusiveness score, even though the street is scarcely used by any group. In contrast, grids with very small total pedestrian counts but uneven group distribution may exhibit artificially high standard deviations, causing falsely low inclusiveness values due to random fluctuations.

To address these edge cases, we define an exclusion threshold based on total observed pedestrian. Through sensitivity analysis, we found that when the total number of detected pedestrians $P_A + P_E + P_C \leq 2$, the inclusiveness index I is highly concentrated below 0.2, dominated by the smoothing term and lacking real discriminative power. When the total observed count exceeds 2, the index exhibits broader distribution and more meaningful differentiation. Therefore, we set the exclusion condition as: when $P_A + P_E + P_C \leq 2$, the grid is excluded from inclusiveness calculation.

2.5. Spatial statistics

To comprehensively evaluate the spatial distribution patterns of utilization and inclusiveness of PSS, as well as their interrelationship, this study employed a series of spatial statistical analysis methods. All analyses were based on standardized 100 m × 100 m grid units, ensuring consistency across variables. Heatmap and Local Indicators of Spatial Association (LISA) were used to detect spatial patterns of key indicators including pedestrian volume, activity diversity, utilization index, and inclusiveness index, revealing both typical and anomalous spatial structures (Anselin, 1995; Páez & Scott, 2004). We introduced Local Bivariate Moran's I to further explore the interaction between pedestrian volume and activity diversity (Rey, 2001; Wartenberg, 1985). All heatmaps in the manuscript apply the quantile classification method to divide the index values into five levels. And we performed a spatial overlay analysis using the Intersect to examine the coupling relationship between utilization and inclusiveness.

3. Results

3.1. Utilization analysis

3.1.1. Pedestrian volume

Based on our model's performance of pedestrian identification (F1 = 0.96, recall = 0.94, precision = 0.98), we further estimated the confidence intervals for each demographic group. Given the total number of detected individuals (n = 91,524), the model identified 87,866 adults, 1302 children, and 2356 elderly individuals. Using the recall-based estimation method, we calculated that the true number of adults likely falls between 91,527 and 95,506, children between 1356 and 1415, and elderly individuals between 2454 and 2561. These intervals provide a quantitative sense of the model's detection uncertainty and reinforce the robustness of our demographic classification. As shown in Table 1, while the average resident population per grid is relatively high (Mean = 296.10), the average pedestrian volume remains low (Mean = 0.472), indicating a substantial gap between residential presence and actual street usage. Fig. 6 illustrates the average pedestrian volume heatmap along with its corresponding spatial clustering pattern, and the residential population heatmap. Fig. 6a shows that high pedestrian volume grids exhibit distinct radial patterns, predominantly concentrated within Shuguang Subdistrict, Balizhuang Subdistrict, and Haidian Subdistrict.

Table 1
Descriptive Statistics of Average Pedestrian Volume and Resident Population.

	Mean	Median	Std	Max	Min
Average Pedestrian Volume	0.472	0.359	0.467	3.970	0.0
Resident Population	296.102	234.0	230.795	1892.0	49.0

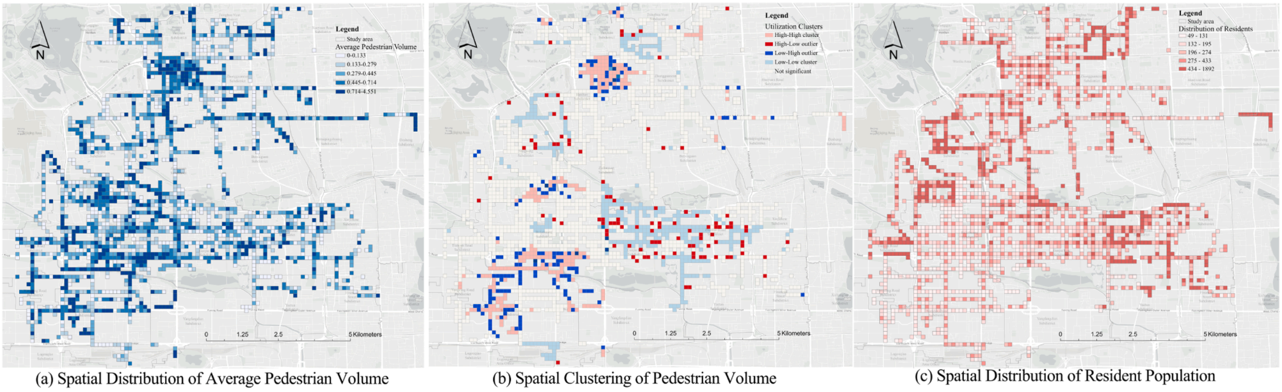


Fig. 6. Spatial Patterns of Pedestrian Volume and Resident Population.

In contrast, other subdistricts such as Financial Street Subdistrict and Zhanlan Road Subdistrict display scattered high-volume grids with isolated distributions. This uneven distribution reveals clear spatial disparities, with high pedestrian activity restricted to a limited portion of the street network. In Fig. 6b, three prominent clusters can be observed. Two high-high clusters (located in the Haidian, Wanshou Road, and Balizhuang subdistricts) indicate strong spatial aggregation of pedestrian, which also aligns with the high-volume subdistricts in heatmap. In

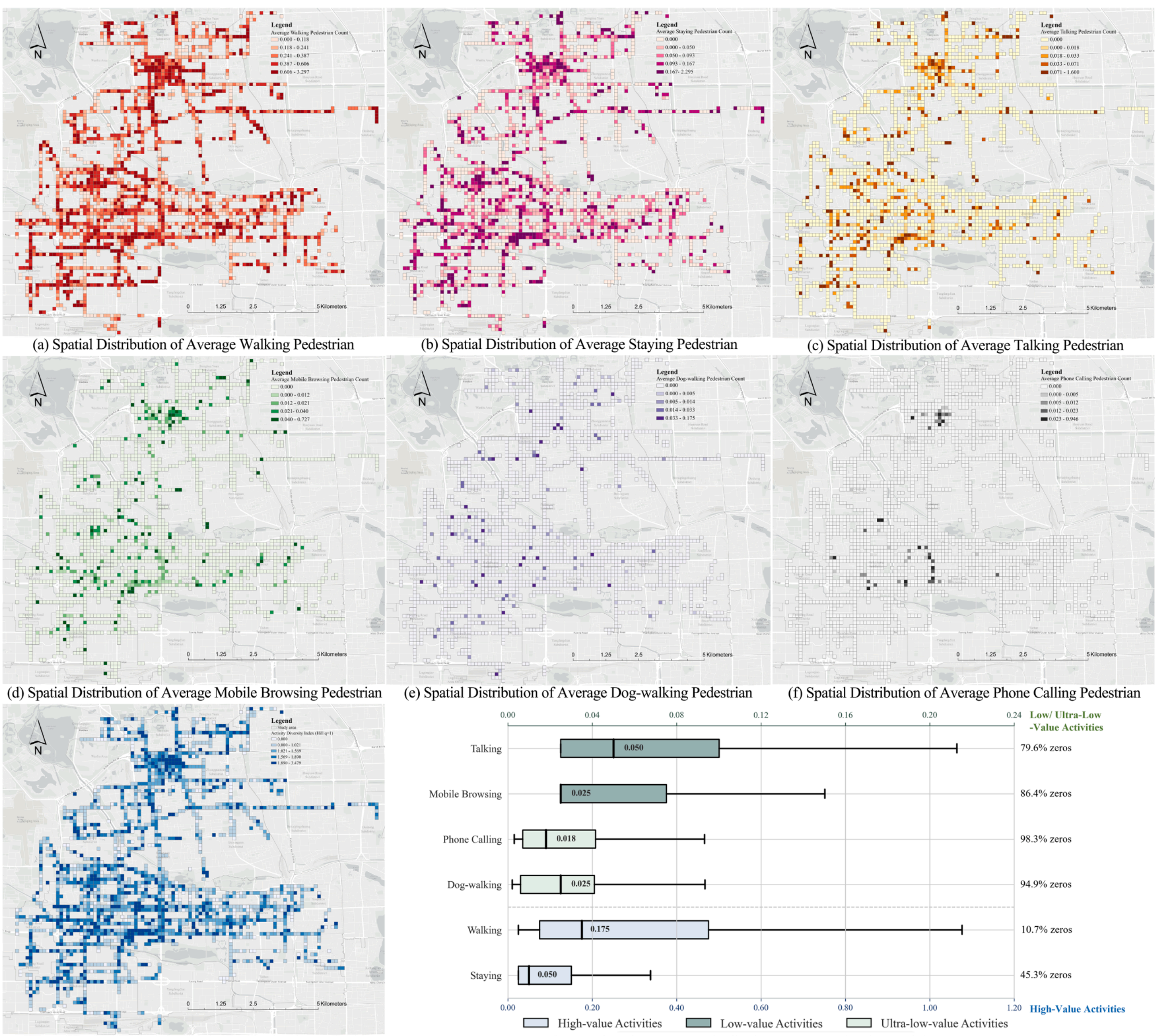


Fig. 7. Six Types of Pedestrian Activity Distribution and Activity Diversity Heatmap.

contrast, low-low clusters (such as the Zhanlan Road and Ganjia Kou subdistricts) reflect relatively spatial isolation. Other scattered clusters are sparsely distributed across other subdistricts. The remaining grids are categorized as “not significant”. Overall, the LISA results suggest that the distribution of pedestrian volume exhibits a non-random spatial pattern with notable local concentrations. Fig. 6c shows the residential population distribution, derived from mobile signaling data. Notably, the comparison between pedestrian volume and resident population reveals limited spatial correspondence. Several densely populated residential streets (such as Zhanlan Road Subdistrict) exhibit surprisingly low pedestrian volume. Conversely, some streets (such as Zhongguancun Subdistrict) with moderate population density present high pedestrian volum. This mismatch suggests that high population density alone does not guarantee high pedestrian volume.

3.1.2. *Activity diversity*

Fig. 7 presents a comprehensive analysis of pedestrian activity types, including the spatial distributions of six activities; the spatial pattern of activity diversity measured by the Hill Diversity Index; and the frequency distributions of activities through the boxplot. Fig. 7h illustrates that walking constitutes the predominant activity, with the highest median value (0.175) and lowest zero-value rate (10.7 %), followed by staying (median: 0.050, 45.3 % zeros) and talking (median: 0.050, 79.6 % zeros). Mobile browsing, dog-walking, and phone calling exhibit progressively diminishing presence, with phone calling showing the maximum zero-value rate (98.3 %). The spatial distributions of the six activity types (Fig. 7a–f) reveal varying degrees of clustering and dispersal. Walking (Fig. 7a) is widely distributed across the study area. High-intensity walking is observed along major street such as Zhongguancun subdistrict. Staying (Fig. 7b) is more localized. High values for staying are concentrated in specific nodes rather than evenly spread. Talking and mobile browsing (Fig. 7c–d) show fragmented yet widespread patterns. In contrast, dog-walking (Fig. 7e) demonstrates a highly isolated spatial distribution, while phone calling (Fig. 7f) exhibits a highly localized pattern. Fig. 7g depicts the spatial distribution of the Activity Diversity Index (Hill $q = 1$), measuring the richness and evenness of pedestrian activities across the study area. High-diversity zones are primarily concentrated in central Zhongguancun Subdistrict, forming a cohesive cluster with radial extensions along major thoroughfares. Additionally, a considerable number of high-diversity grids are distributed within the Balizhuang Subdistrict. Lower diversity areas typically appear in peripheral regions. The spatial variability in activity diversity highlights the difference of PSS in accommodating pedestrian activities.

3.1.3. *Utilization index*

Fig. 8 presents a comprehensive analysis of PSS utilization patterns, encompassing the local bivariate relationship between pedestrian volume and activity diversity, the utilization heatmap, the spatial clustering of utilization, as well as the bubble chart and flow relationships among them. Fig. 8a illustrates the majority of grids exhibit a positive linear relationship. This indicates that higher pedestrian volumes in these areas are generally associated with greater activity diversity. Concave relationships are scattered around the Shuguang and Financial Street Subdistricts, indicating behavioral convergence in grids with high volume. Undefined complex patterns are primarily concentrated in the Ganjiakou Subdistrict, indicating unstable or mixed associations between the two variables. In contrast, negative linear and convex relationships are rare and spatially scattered. Overall, strong coupling between pedestrian volume and activity diversity is evident in the core corridors, whereas marginal areas exhibit weaker or inconsistent relationships. Fig. 8b indicates that high-utilization grids are primarily concentrated in the western part of Zhongguancun Subdistrict and the southern section of Balizhuang Subdistrict. In contrast, low-utilization grids are widely distributed along the northern and eastern peripheries, such as in the Zhanlan Road Subdistrict. The clustering analysis (Fig. 8c) further reveals the spatial aggregation patterns of PSS utilization. High–High clusters are primarily concentrated along the boundary between Haidian and Zhongguancun Subdistricts, and Balizhuang

(a) Local Bivariate Relationship between Volume and Activities (b) Spatial Distribution of Utilization Intensity (c) Spatial Distribution of Utilization Patterns

(d) Sankey Diagram of the Flow Relationships Among Activity Types, Utilization Levels, and Pedestrian Volume Levels (e) Joint Distribution of Volume and Diversity Across Streets

Fig. 8. Multidimensional Utilization Analysis of PSS.

Subdistrict, with a few clusters in Wanshou Road Subdistrict. These areas constitute stable and intensely used PSS cores. In contrast, Low–Low clusters are widespread across the eastern part of the study area, such as in Ganjiakou Subdistrict. Notably, the Lugouqiao Subdistrict displays a complete Low–Low clustering pattern, indicating persistently low utilization of PSS. Fig. 8d visualizes the internal flows among activity types, utilization levels, and pedestrian volume. All categorical levels in the Sankey diagram (e.g., the five utilization levels: ultra-high, high, medium, low, and ultra-low) are classified using the quantile method, with threshold values consistent with those applied in the corresponding heatmap. Walking and staying are the dominant activities, mostly linked to high or ultra-high utilization levels. And utilization is generally positively correlated with pedestrian volume. Fig. 8e displays the joint distribution of pedestrian volume and activity diversity across subdistricts. Streets in Zhongguancun, Tsinghua, and Haidian Subdistricts exhibit high volume, high diversity, and high utilization, forming vibrant street clusters.

3.2. Inclusiveness analysis

3.2.1. Demographic distribution of pedestrians and residents

As shown in Table 2, adult residents dominate the population distribution, with a mean value more than five times that of children or elderly. Fig. 9 presents the heatmaps of average pedestrian volume and resident population across three age groups: adults, children, and the elderly. As shown in Figs. 9a–c, adults dominate street presence, with high pedestrian volume grids concentrated in the western part of Zhongguancun Subdistrict and Balizhuang Subdistrict. In contrast, children and elderly pedestrians are sparsely distributed.

However, Figs. 9e-**f** clearly reveal high-density clusters of children and elderly residents (such as Zhanlan Road Subdistrict). Nevertheless, these streets do not exhibit correspondingly high pedestrian volumes for these age groups, reflecting a significant mismatch between observed street-level pedestrians and the demographic structure.

3.2.2. Spatial patterns and model residuals of inclusiveness

Fig. 10 provides an integrated assessment of spatial inclusiveness in the study area, combining intensity distribution, clustering analysis, and behavioral flow patterns. The heatmap (Fig. 10a) reveals that areas with higher inclusiveness are primarily located along the western periphery and southeastern subdistricts (such as Shuguang), while lower inclusiveness is concentrated in the northern and southwestern areas, including the western part of Zhongguancun. This spatial pattern reflects a critical mismatch—segments with the highest pedestrian volumes and utilization rates often exhibit the poorest inclusiveness values. Clustering analysis (Fig. 10b) further identifies continuous “high-inclusiveness corridors” in the central and eastern zones, notably in Haidian and Zhanlan Road Subdistricts, whereas low-inclusiveness clusters are concentrated in southwestern subdistricts like Yongding Road and Wanshou Road, where persistent demographic underrepresentation reflects pronounced spatial inequality. The Sankey diagram (Fig. 10c) offers behavioral insights, showing that walking—primarily performed by adults—is strongly associated with grids of high or ultra-high activity diversity. This indicates that the inclusiveness imbalance is largely shaped by unequal participation opportunities among demographic groups.

Table 2
Descriptive Statistics of Average Pedestrian Volume and Resident Population by Age Group.

	Mean	Median	Std	Max	Min
Average Adult Volume	0.466	0.356	0.462	3.817	0.0
Average Children Volume	0.003	0.0	0.021	0.546	0.0
Average Elderly Volume	0.006	0.0	0.041	1.638	0.0
Resident Adult Population	198.91	155.028	157.13	1290.454	29.114
Resident Children Population	37.576	27.923	32.178	345.376	2.745
Resident Elderly Population	59.615	46.63	47.079	448.37	5.426

3.3. Intersecting patterns of inclusiveness and utilization

Fig. 11 presents a comprehensive analysis of the spatial intersections and demographic dynamics between street space utilization and inclusiveness. Fig. 11a–b highlight four composite grid categories formed by intersecting the top and bottom 20 % of utilization and inclusiveness indices. A limited number of grids exhibit both high utilization and high inclusiveness, sparsely located along the northwestern edge, suggesting a scarcity of spaces that are simultaneously vibrant and equitable. In contrast, grids with high inclusiveness but low utilization are widely distributed across central and northern subdistricts, such as Zhanlan Road. Grids characterized by low inclusiveness but high utilization are concentrated in southwestern zones like Wanshou Road, reflecting socially exclusive yet highly active environments. Grids with both low utilization and low inclusiveness, mostly located at the periphery. Fig. 11c further depicts the internal flow relationships among utilization levels, demographic groups, and inclusiveness levels. It shows that ultra-high and high utilization grids are predominantly occupied by adults, suggesting that the most active areas tend to lack demographic diversity. While most grids exhibit ultra-high utilization, the majority also display ultra-low inclusiveness. This asymmetry underscores a structural mismatch: street vitality is often achieved at the cost of inclusiveness, particularly for vulnerable groups who remain underrepresented in the city’s most vibrant PSS.

4. Discussion

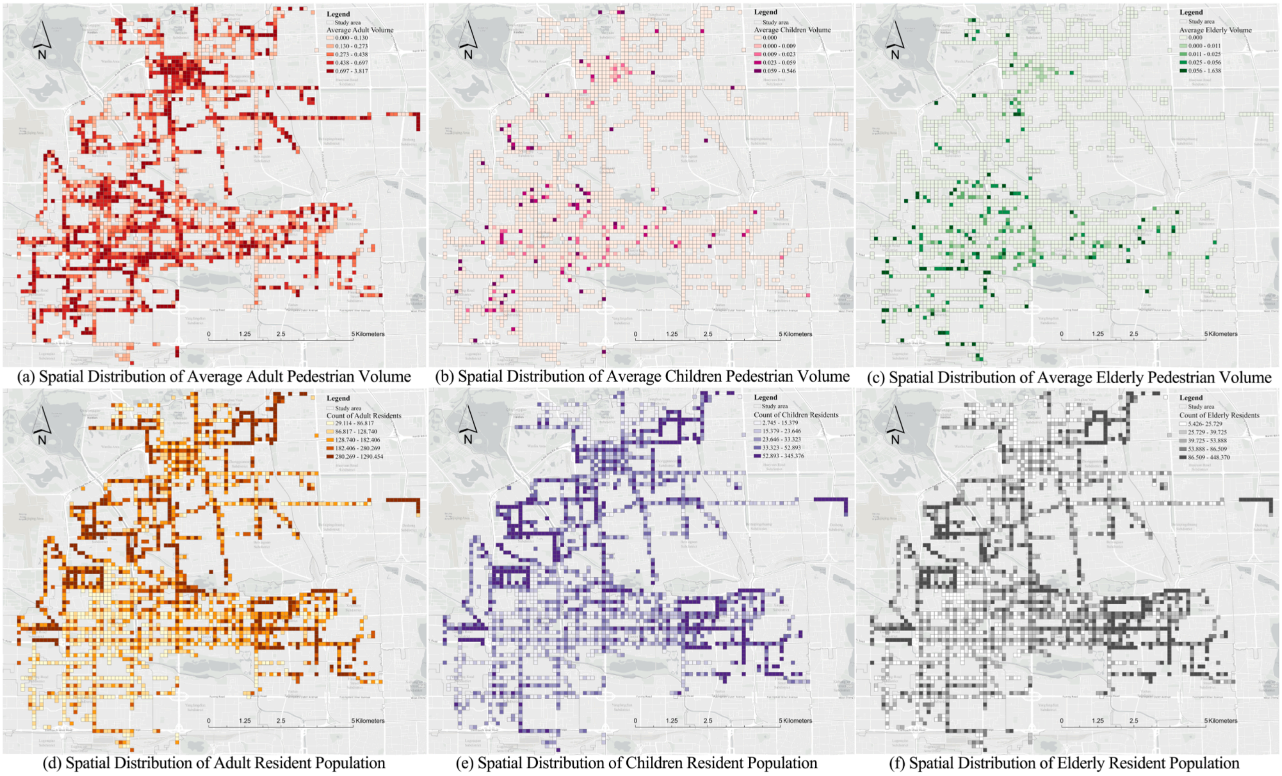
4.1. Recap of the main findings

Urban vitality and social inclusiveness have long been central concerns in urban studies (Fainstein, 2014; Harvey, 2010; Henri Lefebvre, 1967; Sennett, 2017), and are also core aspirations of the United Nations Sustainable Development Goals for "reducing inequalities (SDG 10) and building inclusive, safe, and sustainable cities (SDG 11) ". However, our research reveals a concerning phenomenon: Vulnerable groups are underrepresented in high-utilization streets.

At the individual level, pedestrian distribution reveals significant disparities among different demographic groups. Adults not only dominate numerically but also engage in a broader range of activities (Fig. 9a), while children and the elderly appearing at lower proportions in PSS (Figs. 9b–c). This unbalanced proportion appears inconsistent with the "human-centered" urban public space design principles (Gehl, 2013) and suggests a potential misalignment with "ensure safe, inclusive and accessible public spaces for all, particularly for children and older persons" (SDG 11.7).

At the street network level, our analysis reveals a spatial mismatch between pedestrian flow and residential population. Although streets such as Zhanlan Road have numbers of child and elderly residents (Figs. 9e–f), these groups are rarely seen in PSS (Figs. 9b–c). This finding aligns with the concept of "mobility injustice" (Sheller, 2018), which highlights seemingly open PSS may unintentionally discourage certain groups(Mitchell, 2003; Sibley, 2002).

At the regional level, high-utilization areas are concentrated in specific areas like Balizhuang Subdistrict, while remaining streets remain underused (Fig. 8b). This is consistent with the “uneven geographies” of urban life described by E. Soja (, 2009), where spatial advantages and resources are concentrated in select areas at the expense of broader equity. Commercial clusters such as Zhongguancun exhibit a critical spatial imbalance: extremely high utilization (Fig. 8b) coupled with extremely low inclusiveness (Fig. 10a). This aligns with the viewpoint that urban planning tends to emphasize economic competitiveness



(a) Spatial Distribution of Average Adult Pedestrian Volume (b) Spatial Distribution of Average Children Pedestrian Volume (c) Spatial Distribution of Average Elderly Pedestrian Volume

(d) Spatial Distribution of Adult Resident Population (e) Spatial Distribution of Children Resident Population (f) Spatial Distribution of Elderly Resident Population

Fig. 9. Spatial Distributions of Pedestrian Volume and Resident Population by Age Group.

rather than equity proposed by Fainstein(2014). Importantly, areas with high-utilization but low-inclusiveness constitute the largest proportion of the study area (Fig. 11b). This suggests a tendency: PSS is increasingly shaped as infrastructure for economic circulation, rather than as a platform that serves the entire population (Loukaitou-Sideris & Ehrenfeucht, 2011). Most notably, areas that achieve both high utilization and high inclusiveness account for <5 % of the entire study area (Fig. 11a). This not only contradicts the assumptions of creative city (Landry, 2012) and creative class (Florida, 2002) theories, which argues the mutual promotion of diversity, inclusiveness and urban vitality, but also suggests that "spatial equality and social justice" (SDG 10) still face systemic challenges in current urban governance practices.

4.2. The hybrid vision-language model for pedestrian automatically detection

Computer vision applications in urban analysis have evolved rapidly in recent years, from basic object detection to sophisticated recognition of human activities and demographic characteristics. Current applications in this domain generally follow three main technological pathways: (1) object detection models such as YOLO (Redmon et al., 2016) and Mask R-CNN (K. He et al., 2017), which identify, localize, and classify various objects within images; (2) behavior recognition models, including I3D (Carreira & Zisserman, 2017), SlowFast Networks (Feichtenhofer et al., 2019), and Temporal Segment Networks (Wang et al., 2016), which focus on activity recognition; and (3) emerging vision-language models (VLMs) such as CLIP (Radford et al., 2021) and LLaVA (H. Liu et al., 2023), which transform visual content into natural language descriptions, enabling more comprehensive scene understanding.

As Zhang et al. (2016) demonstrated, even advanced pedestrian detection systems face substantial challenges in authentic urban scenes. Object detection and behavior recognition typically require separately trained models for distinct tasks (W. Liu et al., 2016; Z.-Q. Zhao et al., 2019), struggling to handle the complex conditions of real-world streets characterized by variable lighting, partial occlusions, and diverse viewpoints (Dollar et al., 2012; S. Zhang et al., 2018). Our hybrid vision-language approach, LLaVA-PI, addresses these limitations by integrating precise detection capabilities of DETA with semantic comprehension of LLaVA-OV, enabling end-to-end analysis that simultaneously extracts demographic attributes and behavioral patterns from images. The model demonstrates exceptional accuracy under complex real-world conditions, significantly outperforming general-purpose multimodal models such as GPT-4o and Gemini-1.5-pro. These results highlight the model's robust adaptability to real-world conditions and its capacity to address common challenges in street environments.

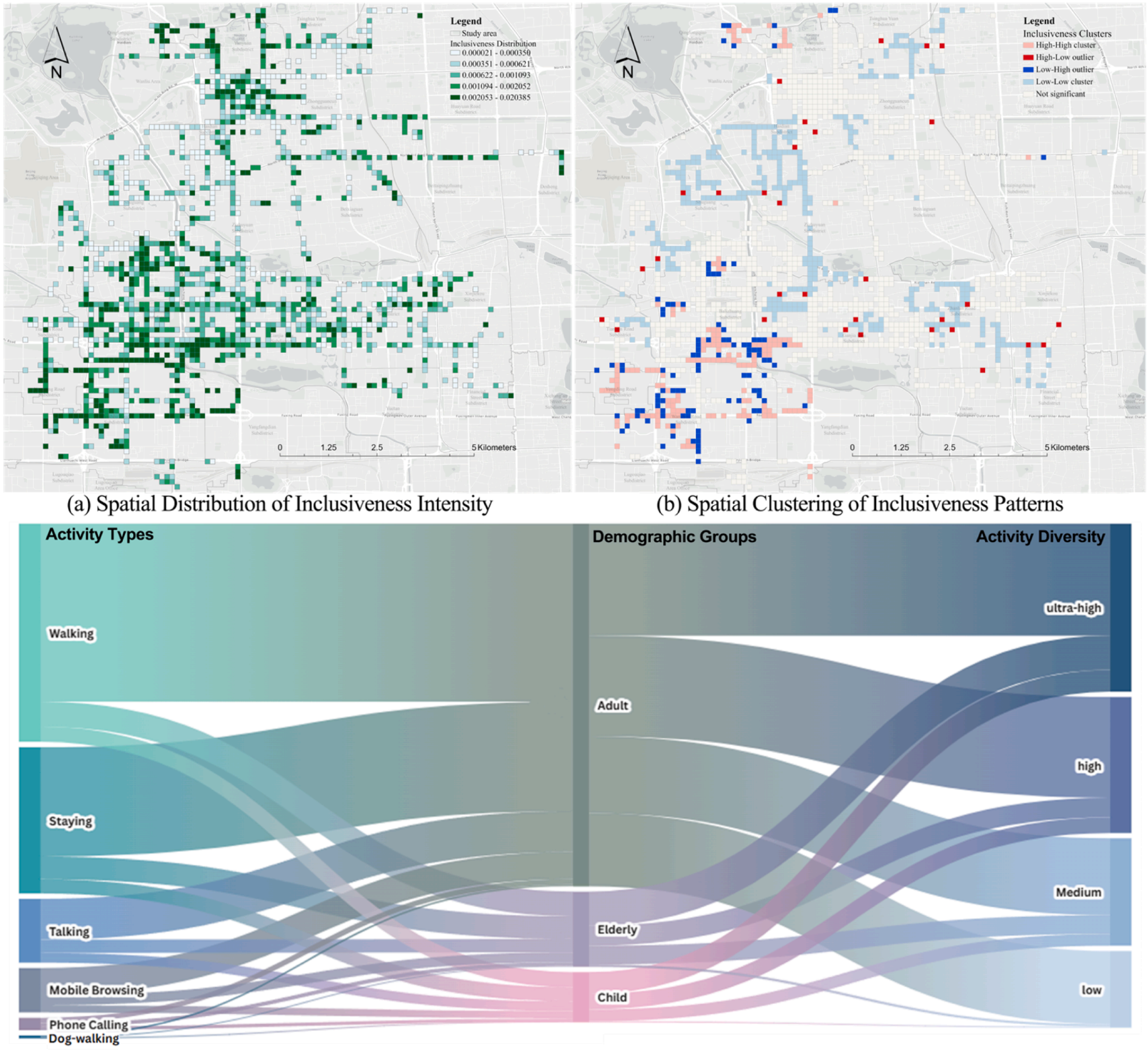
Meanwhile, this study overcomes the limitations of conventional data collection methods by proposing an ego-perspective approach to capture comprehensive and fine-grained street-level data (Yang et al., 2025). Finally, we constructed the EgoCity dataset—currently the largest known pedestrian-scale street image dataset for Beijing.

EgoCity Datasets not only capture specific types of street activities at specific times for specific demographic groups, but also provide valuable insights into the activity patterns of different demographic groups and spatial utilization in urban environments. More importantly, the LLaVA-PI framework exhibits excellent scalability and transferability. It reliably processes video data collected across diverse times, locations, and urban settings, generating structured outputs suitable for spatial analysis. This data–model integrated pipeline substantially enhances the precision, spatial coverage, and demographic granularity of PSS evaluation, establishing a new technical foundation for computational assessments of urban equity.

4.3. Global and local factors affecting the street utilization and inclusiveness

4.3.1. Local factors

A diverse functional mix may contribute to increased utilization. As Jacobs (1961) emphasizes, functional mix can attract people with diverse purposes and stimulate greater street vitality. In this study, for



(a) Spatial Distribution of Inclusiveness Intensity
(b) Spatial Clustering of Inclusiveness Patterns
(c) Sankey Diagram of the Flow Relationships Among Activity Types, Demographic Groups, and Activity Diversity Levels

Fig. 10. Spatial Patterns and Behavioral Dynamics of the Inclusiveness of PSS.

instance, the Ganjiakou Subdistrict, which combines universities, commercial centers, office buildings, and residential zones, exhibits high volume and activity diversity (Fig. 8e).

Public amenities influence both inclusiveness and activity diversity. In PSS, these amenities go beyond use value—they also serve as cultural symbols that shape spatial belonging for different age groups (Alexander, 1977; Lefebvre, 2014). Amenities like playgrounds provide opportunities for children's activities for the elderly, fitness equipments and benches signify places for rest and interaction (Gehl, 2011). In this study, we observe two notable patterns: first, areas with high inclusiveness are often located at streets with many residential communities such as Shuguang Subdistrict, which contains large residential zones and several parks (Fig. 10a); second, areas with high activity diversity (Fig. 7g) tend to have some functional nodes, such as playing zones and commercial clusters.

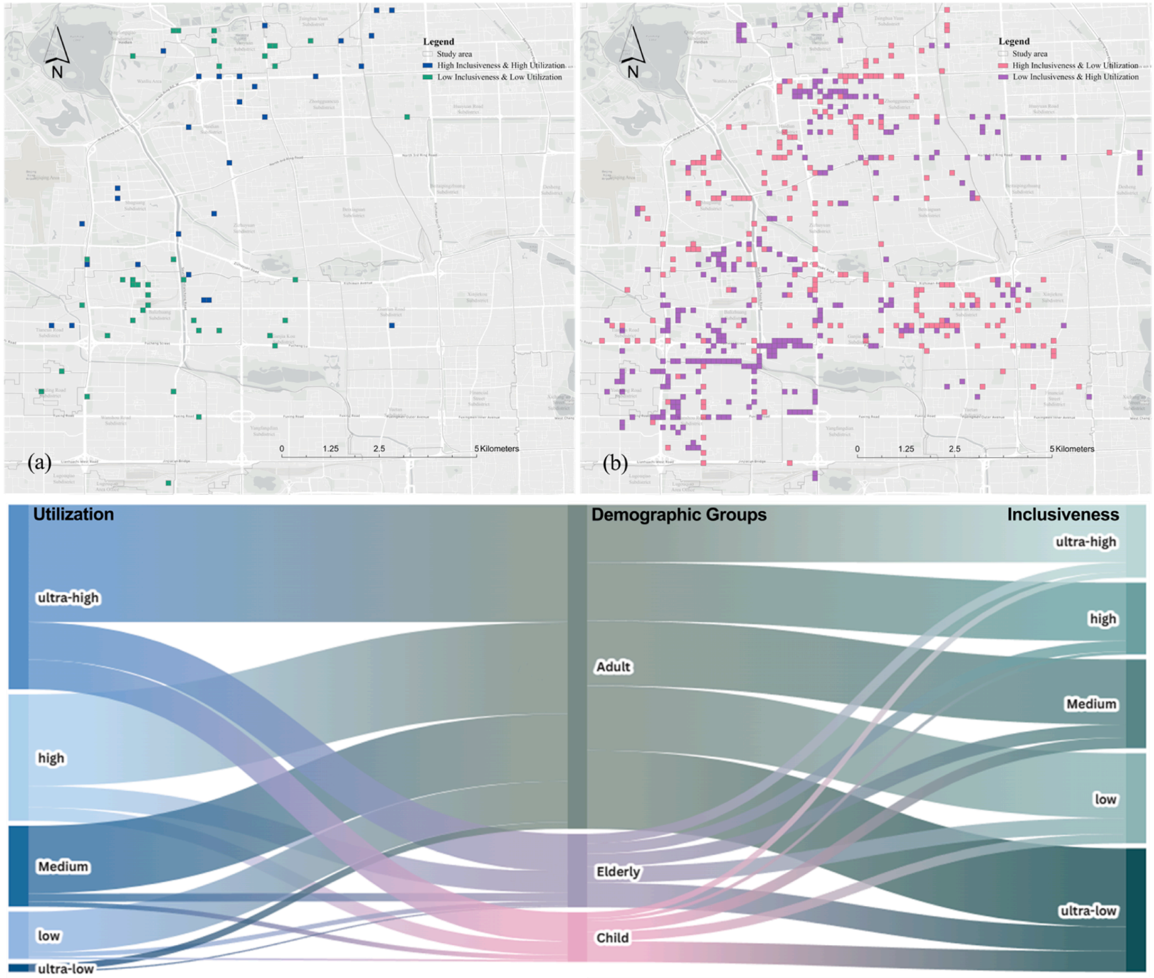
4.3.2. Global factors

In Beijing, street governance follows a territorial management principle, where PSS are often co-managed by multiple departments (urban management, traffic police, landscaping, municipal services and so on). This fragmented authority frequently has been associated with unclear responsibilities, delayed maintenance, and slow implementation of supporting facilities (Wu, 2015).

At a broader scale, China's urban development strategy emphasizes "intensive and efficient" land use. High floor area ratios often result in commercialized streets, with limited spaces for informal activities (S. He & Wu, 2009; Lin, 2007). Additionally, cultural norms confine the elderly to residential areas due to concerns about inconvenient transportation and safety (Jianxi Feng, 2017). Educational pressure also shapes children's mobility patterns: many are directly transported between home, school, and tutoring centers, reducing their engagement in PSS (S. Y. He, 2013; Yarlagadda & Srinivasan, 2008).

In fast-paced, high-density cities, street systems are primarily designed for vehicles, and PSS are implicitly governed by efficiency. As a result, streets are often reduced to mere transportation corridors, compressing pedestrian spaces and marginalizing non-utilitarian functions such as leisure (Appleyard et al., 1981; Gehl, 2013). Meanwhile, globally dense cities generally adopt a development approach that prioritizes economic growth, aiming to attract capital and enhance competitiveness (Brenner & Theodore, 2002; Harvey, 2007). Under this model, PSS are often viewed as settings that drive consumption, tourism, and real estate development, equating vitality with pedestrian flow simply rather than demographic diversity (Mitchell, 2003; Zukin, 1995). Within such development patterns, street design may favor producers over everyone (Loukaitou-Sideris & Ehrenfeucht, 2011).

From an environmental justice perspective, these structural inequalities classify in two dimensions: distributional justice and procedural justice (Schlosberg, 2007; Walker, 2012). Distributional justice



(c) Sankey Diagram of the Flow Relationships Among Utilization Levels, Demographic Groups, and Inclusiveness Levels

Fig. 11. Spatial Coupling and Demographic Flow Between Utilization and Inclusiveness in PSS.

refers to the uneven access to high-quality walking environments, typically concentrated in well-planned neighborhoods, while marginalized areas often suffer from poor street conditions (Talen, 2010). Procedural justice highlights the lack of voice among vulnerable groups in decision-making processes, leading to spatial planning that favors dominant populations (Brinkley & Wagner, 2024; Fainstein, 2014; Rot et al., 2025; Young, 2002).

4.4. Implications for urban vitality and social justice

The implications of these findings go far beyond simple demographics and may reveal a phenomenon: urban development often prioritizes economic gain over social justice (Fainstein, 2014; Loukaitou-Sideris & Ehrenfeucht, 2011; Harvey, 2010).

For government, integrating the goals of utilization and inclusiveness is crucial to achieving sustainable urban development (Florida, 2017; Yigitcanlar et al., 2018). Urban vitality can only be sustained if it is shared by all demographics; otherwise, superficial prosperity only masks deeper inequalities (Harvey, 2012; Sassen, 2014). Our research provides a quantitative evaluation approach that can be incorporated into routine urban governance frameworks (Batty, 2013; Kitchin et al., 2015). For example, indicators like inclusiveness could be used in the assessments of urban renewal projects to evaluate whether these measures have truly improved fairness in PSS (Banerjee, 2001; Talen, 2010).

At the individual level, increasing utilization often stimulates greater diversity of behavioral activities, which contributes to enhancing residents' physical and mental well-being (SDG 3). At the community level, our methods and findings also serve as advocacy tools—making public to effectively voice their needs to decision-makers (Arnstein, 1969; Cornwall, 2008). Such efforts help promote procedural justice and ensure that the voices of marginalized groups are heard throughout the planning process (Healey, 2003). This bottom-up supervision and feedback mechanism fosters more equitable urban governance (Fung & Wright, 2001; Gaventa, 2006).

From a global development perspective, this study aligns with the core concepts of "Reducing Inequality" (SDG 10) and "Building Inclusive, Safe, Resilient, and Sustainable Cities" (SDG 11).. The methods we present offer new ways to measure and support these goals: by using technology and data to monitor the extent to which PSS serves all population (Kummitha & Crutzen, 2017; Townsend, 2013).

The policy and practical implications of this study extend beyond Beijing, offering valuable insights for other rapidly urbanizing regions (Angel et al., 2011; Un-Habitat, 2016). Although some cities have begun to promote policies such as "child-friendly cities" and "age-friendly communities" (WTO, 2007; UNICEF, 2019), gaps remaining between policy intentions and actual outcomes (Riggio, 2002). This highlights policymakers need to pay closer attention to how PSS are actually used (Skelton, 2013; Woolcock et al., 2010).

In conclusion, this study highlights the importance of balancing urban vitality with social justice (Fainstein, 2014; Mitchell, 2003), and offers practical tools for urban design and governance that support a more vibrant and equitable urban (Healey, 2020; Sandercock &

Lyssiotis, 2003). Only when PSS truly belongs to everyone can we speak of social justice (Mitchell, 2003; Purcell, 2002), this vision responding to the call for "the right to the city for all" (Henri Lefebvre, 1967).

4.5. Policy translation and planning applications

To facilitate practical implementation, the proposed indicators (utilization and inclusiveness) can serve as actionable tools in urban planning and street design. The Utilization Index, defined as the ratio between actual pedestrian flow and the registered residential population of each spatial unit, helps planners detect underutilized or overcrowded street segments to promote walkability and balanced use (Soja, 2009; Walker, 2012). The Inclusiveness Index measures the deviation between the observed proportion of demographic groups (e.g., children, elderly) in the street space and their expected proportion based on the surrounding residential composition. Areas with significant gaps may signal potential accessibility barriers or age-specific environmental discomforts. This index could be integrated into performance checklists for age-friendly or child-friendly urban design guidelines (Skelton, 2013; UNICEF, 2019).

These indicators support the move toward data-informed governance and can be integrated into routine evaluation frameworks for public space equity (Kitchin et al., 2015). Beyond diagnosis, they offer planners a way to track longitudinal improvements, assess intervention outcomes, and establish evidence-based priorities in street-level investments (Campbell, 1996; Schlosberg, 2007).

4.6. The limitations and future study

While this study proposes a scalable, data-driven framework for evaluating the utilization and inclusiveness of PSS, several limitations must be acknowledged. First, although this study employed a novel crowdsourced method and achieved relatively comprehensive spatial coverage, it remains confined to the core metropolitan region of Beijing. Furthermore, the current dataset is extensive, but does not provide full coverage of all possible urban street types. Future research should expand the geographic scope to include diverse urban contexts, which would not only facilitate cross-regional comparisons but also help validate the generalizability and robustness of the proposed framework. Second, although the current study classifies pedestrians by age and identifies the predicaments of vulnerable groups in PSS, urban equity is a complex, multifaceted issue. Future work should explore how our quantitative indicators can be combined with participatory or interpretive methods to better account for social, historical, and cultural dynamics Third, although the LLaVA-PI model demonstrates high recognition accuracy, algorithmic bias may arise from external visual factors such as clothing color or occlusion of key age-related features, potentially affecting recognition accuracy for certain demographic groups.—for example, older children may be misidentified as adults due to ambiguous physical characteristics. Moreover, the detection of phone calling behavior primarily relies on gestures of a phone near the ear, potentially underestimating instances where users calling via wireless earbuds. In addition, data collection was limited to daytime hours due to reduced model performance under night-time scenarios. Lastly, although this study reveals disparities and inequalities in PSS, it does not delve into the causal mechanisms underlying these patterns. Future studies should investigate potential drivers of spatial inequality and evaluate how different urban interventions may enhance both vitality and equity in PSS.

CRediT authorship contribution statement

Xiamengwei Zhang: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Mingze Chen:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization. **Yongming Huang:** Writing – review & editing, Supervision, Project administration.

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.

Data availability

The EgoCity dataset and associated code will be made publicly available on GitHub upon acceptance of the manuscript. The repository link will be updated accordingly.

References

Agboola, O. P., Zakka, S. D., & Zango, M. S. (2015). Towards sustainability of open spaceâ€™ s planning and management in nigeria: The role of science and technology. *Jurnal Teknologi (Sciences & Engineering), 77*(14).

Aghaabbasi, M., Moeinaddini, M., Zaly Shah, M., & Asadi-Shekari, Z. (2017). A new assessment model to evaluate the microscale sidewalk design factors at the neighbourhood level. *Journal of Transport & Health, 5*, 97–112. https://doi.org/10.1016/j.jth.2016.08.012

Alexander, C. (1977). *A pattern language: Towns, buildings, construction*. Oxford university press.

Angel, S., Parent, J., Civco, D. L., Blei, A., & Potere, D. (2011). The dimensions of global urban expansion: Estimates and projections for all countries, 2000–2050. *Progress in Planning, 75*(2), 53–107.

Anselin, L. (1995). Local Indicators of Spatial Association—LISA. *Geographical Analysis, 27*(2), 93–115. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x

Appleyard, D., Gerson, M. S., & Lintell, M. (1981). *Livable streets*. University of California Press.

Arnstein, S. R. (1969). A ladder of citizen participation. *Journal of the American Institute of Planners, 35*(4), 216–224. https://doi.org/10.1080/01944366908977225

Banerjee, T. (2001). The future of public space: *Beyond invented streets and reinvented places*. *Journal of the American Planning Association, 67*(1), 9–24. https://doi.org/10.1080/01944360108976352

Batty, M. (2013). *The new science of cities*. MIT press.

Biernacka, M., Łaszkiewicz, E., & Kronenberg, J. (2022). Park availability, accessibility, and attractiveness in relation to the least and most vulnerable inhabitants. *Urban Forestry & Urban Greening, 73*, Article 127585. https://doi.org/10.1016/j.ufug.2022.127585

Boone, C.G., Buckley, G.L., Grove, J.M., & Sister, C. (2009). *Parks and People: An environmental justice Inquiry in Baltimore, Maryland*.

Brenner, N., & Theodore, N. (2002). Cities and the geographies of "actually existing neoliberalism. *Antipode, 34*(3), 349–379. https://doi.org/10.1111/1467-8330.00246

Brinkley, C., & Wagner, J. (2024). Who is planning for environmental justice—And how? *Journal of the American Planning Association, 90*(1), 63–76. https://doi.org/10.1080/01944363.2022.2118155

Campbell, S. (1996). Green cities, growing cities, just cities. *American Planning Association. Journal of the American Planning Association, 62*, 3.

Carreira, J., & Zisserman, A. (2017). *Quo Vadis, Action recognition? A new model and the Kinetics dataset*. 6299–6308. https://openaccess.thecvf.com/content_cvpr_2017/html/Carreira_Quo_Vadis_Action_CVPR_2017_paper.html.

Chen, M., Liu, Y., Liu, F., Chadha, T., & Park, K. (2025). Measuring pedestrian-level street greenery visibility through space syntax and crowdsourced imagery: A case study in London, UK. *Urban Forestry & Urban Greening, 105*, Article 128725.

Cornwall, A. (2008). Unpacking 'participation': Models, meanings and practices. *Community Development Journal, 43*(3), 269–283.

de Vries, S., van Dillen, S. M. E., Groenewegen, P. P., & Spreeuwenberg, P. (2013). Streetscape greenery and health: Stress, social cohesion and physical activity as mediators. *Social Science & Medicine, 94*, 26–33. https://doi.org/10.1016/j.socscimed.2013.06.030

Delbosc, A., & Currie, G. (2011). Using Lorenz curves to assess public transport equity. *Journal of Transport Geography, 19*(6), 1252–1259. https://doi.org/10.1016/j.jtrangeo.2011.02.008

Development, P. C.on S (1996). *Sustainable america: A new consensus for prosperity, opportunity, and a healthy environment for the future*. President's Council on Sustainable Development.

Dollar, P., Wojek, C., Schiele, B., & Perona, P. (2012). Pedestrian detection: An evaluation of the State of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 34*(4), 743–761. https://doi.org/10.1109/TPAMI.2011.155

Dutta, M., Herreros-Cantis, P., McPhearson, T., Mustafa, A., Palmer, M. I., Tosca, M., Ventrella, J., & Cook, E. M. (2025). New York City 2100: Environmental justice implications of future scenarios for addressing extreme heat. *Landscape and Urban Planning, 254*, Article 105249. https://doi.org/10.1016/j.landurbplan.2024.105249

Ewing, R., & Cervero, R. (2010). Travel and the built environment. *Journal of the American Planning Association*. https://doi.org/10.1080/01944361003766766

Fainstein, S. S. (2014). The just city. *International Journal of Urban Sciences*. https://www.tandfonline.com/doi/abs/10.1080/12265934.2013.834643.
Fang, L., Zhang, D., Liu, T., Yao, S., Fan, Z., Xie, Y., Wang, X., & Li, X. (2021). A multi-level investigation of environmental justice on cultural ecosystem services at a national scale based on social media data: A case of accessibility to five-A ecological attractions in China. *Journal of Cleaner Production, 286*, Article 124923. https://doi.org/10.1016/j.jclepro.2020.124923
Feichtenhofer, C., Fan, H., Malik, J., & He, K. (2019). *SlowFast Networks for Video Recognition*, 6202–6211. https://openaccess.thecvf.com/content_ICCV_2019/html/Feichtenhofer_SlowFast_Networks_for_Video_Recognition_ICCV_2019_paper.html.
Florida, R. (2002). *The rise of the creative class (Vol. 9)*. Basic books New York. https://papers.ssrn.com/sol3/Delivery.cfm?abstractid=925464.
Florida, R. (2017). *The new urban crisis: How our cities are increasing inequality, deepening segregation, and failing the middle class-and what we can do about it*. Basic Books.
Fung, A., & Wright, E. O. (2001). Deepening democracy: Innovations in empowered participatory governance. *Politics & Society, 29*(1), 5–41. https://doi.org/10.1177/0032329201029001002
Gaventa, J. (2006). Finding the spaces for change: A power analysis. *IDS Bulletin, 37*(6), 23–33. https://doi.org/10.1111/j.1759-5436.2006.tb00320.x
Gehl. (2011a). *Life between buildings*. https://scholar.google.com/scholar_lookup?title=Life%20between%20buildings%3A%20using%20public%20space&author=J.%20Gehl&publication_year=2011.
Gehl, J. (2011b). *Life between buildings*. https://playspaces.ru/wp-content/uploads/2016/08/Geyl_zhizn_sredi_zdaniy.pdf.
Gehl, J. (2013). *Cities for people*. Island Press.
Guo, R., Ann Diehl, J., Zhang, R., & Wang, H. (2024). Spatial equity of urban parks from the perspective of recreational opportunities and recreational environment quality: A case study in Singapore. *Landscape and Urban Planning, 247*, Article 105065. https://doi.org/10.1016/j.landurbplan.2024.105065
Guo, S., Song, C., Pei, T., Liu, Y., Ma, T., Du, Y., Chen, J., Fan, Z., Tang, X., Peng, Y., & Wang, Y. (2019). Accessibility to urban parks for elderly residents: Perspectives from mobile phone data. *Landscape and Urban Planning, 191*, Article 103642. https://doi.org/10.1016/j.landurbplan.2019.103642
Harvey, D. (2007). *A brief history of neoliberalism*. Oxford university press.
Harvey, D. (2010). *Social justice and the city*. University of Georgia Press.
Harvey, D. (2012). *Rebel cities: From the right to the city to the urban revolution*. Verso books.
He, H., Lin, X., Yang, Y., & Lu, Y. (2020). Association of street greenery and physical activity in older adults: A novel study using pedestrian-centered photographs. *Urban Forestry & Urban Greening, 55*, Article 126789. https://doi.org/10.1016/j.ufug.2020.126789
He, K., Gkioxari, G., Dollar, P., & Girshick, R. (2017). *Mask R-CNN*, 2961–2969. https://openaccess.thecvf.com/content_iccv_2017/html/He_Mask_R-CNN_ICCV_2017_paper.html.
He, S., & Wu, F. (2009). China's emerging neoliberal urbanism: Perspectives from urban redevelopment. *Antipode, 41*(2), 282–304. https://doi.org/10.1111/j.1467-8330.2009.00673.x
He, S. Y. (2013). Will you escort your child to school? The effect of spatial and temporal constraints of parental employment. *Applied Geography, 42*, 116–123.
Healey, P. (2003). Collaborative planning in perspective. *Planning Theory, 2*(2), 101–123. https://doi.org/10.1177/14730952030022002
Healey, P. (2020). *Collaborative planning: Shaping places in fragmented societies*. Bloomsbury Publishing.
Henri Lefebvre. (1967). *Le droit à la ville—Persée*. https://www.persee.fr/doc/homso_0018-4306_1967_num_6_1_1063.
Jung, Hyejin, & Lee, Sae-young (2017). Does improving the physical street environment create satisfactory and active streets? Evidence from Seoul's Design Street Project. *Transportation Research Part D: Transport and Environment, 50*, 269–279. https://doi.org/10.1016/j.trd.2016.11.013
Jiang, S., Ferreira, J., & Gonzalez, M. C. (2017). Activity-based Human mobility patterns inferred from mobile phone data: A case study of Singapore. *IEEE Transactions on Big Data, 3*(2), 208–219. https://doi.org/10.1109/TBDATA.2016.2631141
Feng, Jianxi (2017). The influence of built environment on travel behavior of the elderly in urban China. *Transportation Research Part D: Transport and Environment, 52*, 619–633. https://doi.org/10.1016/j.trd.2016.11.003
Karndacharuk, A., Wilson, Douglas J., & Dunn, R. (2014). A review of the evolution of shared (Street) space concepts in urban environments. *Transport Reviews, 34*(2), 190–220. https://doi.org/10.1080/01441647.2014.893038
Kitchin, R., Lauriault, T. P., & McArdle, G. (2015). Knowing and governing cities through urban indicators, city benchmarking and real-time dashboards. *Regional Studies, Regional Science, 2*(1), 6–28. https://doi.org/10.1080/21681376.2014.983149
Kummitha, R. K. R., & Crutzen, N. (2017). How do we understand smart cities? An evolutionary perspective. *Cities (London, England), 67*, 43–52.
La Rosa, D., Takatori, C., Shimizu, H., & Privitera, R. (2018). A planning framework to evaluate demands and preferences by different social groups for accessibility to urban greenspaces. *Sustainable Cities and Society, 36*, 346–362. https://doi.org/10.1016/j.scs.2017.10.026
Landry, C. (2012). *The creative city: A toolkit for urban innovators* (2nd ed.). Routledge. https://doi.org/10.4324/9781849772945
Le Grand, J., & Bartlett, W. (1993). *Quasi-markets and social policy*. Bloomsbury Publishing.
Lefebvre, H. (2014). The production of space (1991). *The people, place, and space reader*. Routledge.
Li, M., Peng, P., Zhu, H., Ao, Y., & Zhou, Z. (2024). Exploring the equitable spatial planning on the accessibility of institutional elderly care in megacity: A study based on Chengdu, China. *Sustainable Cities and Society, 117*, Article 105998. https://doi.org/10.1016/j.scs.2024.105998
Lin, G. C. S. (2007). Chinese urbanism in question: State, society, and the reproduction of urban spaces. *Urban Geography, 28*(1), 7–29. https://doi.org/10.2747/0272-3638.28.1.7
Liu, D., Kwan, M.-P., Yang, Z., & Kan, Z. (2024). Comparing subjective and objective greenspace accessibility: Implications for real greenspace usage among adults. *Urban Forestry & Urban Greening, 96*, Article 128335. https://doi.org/10.1016/j.ufug.2024.128335
Liu, H., Li, C., Wu, Q., & Lee, Y. J. (2023). Visual instruction tuning. *Advances in Neural Information Processing Systems, 36*, 34892–34916.
Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., & Berg, A. C. (2016). SSD: Single shot MultiBox detector. In B. Leibe, J. Matas, N. Sebe, & M. Welling (Eds.), *Computer vision – eccv 2016* (pp. 21–37). Springer International Publishing. https://doi.org/10.1007/978-3-319-46448-0_2.
Loukaitou-Sideris, A., & Ehrenfeucht, R. (2011). *Sidewalks: Conflict and negotiation over public space*. MIT Press.
Lu, Y., Chen, R., Chen, B., & Wu, J. (2024). Inclusive green environment for all? An investigation of spatial access equity of urban green space and associated socioeconomic drivers in China. *Landscape and Urban Planning, 241*, Article 104926. https://doi.org/10.1016/j.landurbplan.2023.104926
Mitchell, D. (2003). *The right to the city: Social justice and the fight for public space*. Guilford Press.
Organization, W. H. (2007). *Global age-friendly cities: A guide*. World Health Organization.
Páez, A., & Scott, D. M. (2004). Spatial statistics for urban analysis: A review of techniques with examples. *GeoJournal, 61*(1), 53–67. https://doi.org/10.1007/s10708-005-0877-5
Purcell, M. (2002). Excavating Lefebvre: The right to the city and its urban politics of the inhabitant. *GeoJournal, 58*(2/3), 99–108. https://doi.org/10.1023/B:GEJO.0000010829.62237.8f
Qi, Y., Drolma, S. C., Zhang, X., Liang, J., Jiang, H., Xu, J., & Ni, T. (2020). An investigation of the visual features of urban street vitality using a convolutional neural network. *Geo-Spatial Information Science*. https://www.tandfonline.com/doi/abs/10.1080/10095020.2020.1847002.
Rad, T. G., & Alimohammadi, A. (2024). A user-based approach for assessing spatial equity of attractiveness and accessibility to alternative urban parks. *Geo-Spatial Information Science, 27*(2), 487–504. https://doi.org/10.1080/10095020.2022.2135463
Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). Learning transferable visual models from natural language supervision. In *Proceedings of the 38th International Conference on Machine Learning* (pp. 8748–8763). https://proceedings.mlr.press/v139/radford21a.html.
Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). *You only look once: Unified, real-time object detection*. 779–788. https://www.cv-foundation.org/openaccess/content_cvpr_2016/html/Redmon_You_Only_Look_CVPR_2016_paper.html.
Rey, S. J. (2001). Spatial empirics for economic growth and convergence. *Geographical Analysis, 33*(3), 195–214. https://doi.org/10.1111/j.1538-4632.2001.tb00444.x
Riggio, E. (2002). Child friendly cities: Good governance in the best interests of the child. *Environment and Urbanization, 14*(2), 45–58. https://doi.org/10.1177/095624780201400204
Rot, C., Tan, W., Tempels, B., & Patuano, A. (2025). Is it just planning? Unpacking markers of justice in spatial planning literature. *Journal of Planning Literature*. , Article 08854122251328259. https://doi.org/10.1177/08854122251328259
Rzotkiewicz, A., Pearson, A. L., Dougherty, B. V., Shortridge, A., & Wilson, N. (2018). Systematic review of the use of Google Street View in health research: Major themes, strengths, weaknesses and possibilities for future research. *Health & Place, 52*, 240–246. https://doi.org/10.1016/j.healthplace.2018.07.001
Salazar-Miranda, A., Fan, Z., Baick, M. B., Hampton, K. N., Duarte, F., Loo, B. P. Y., Glaeser, E. L., & Ratti, C. (2024). *Shifting patterns of social interaction: Exploring the social life of urban spaces through A.I. (Working paper 33185)*. National Bureau of Economic Research. https://doi.org/10.3386/w33185
Salazar-Miranda, A., Zhang, F., Sun, M., Leoni, P., Duarte, F., & Ratti, C. (2023). Smart curbs: Measuring street activities in real-time using computer vision. *Landscape and Urban Planning, 234*, Article 104715. https://doi.org/10.1016/j.landurbplan.2023.104715
Samiul Hasan. (2013). *Understanding urban human activity and mobility patterns using large-scale location-based data from online social media | proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing* (world). ACM Conferences. https://doi.org/10.1145/2505821.2505823.
Sandercock, L., & Lyssiotis, P. (2003). *Cosmopolis II: Mongrel cities of the 21st century*. A&C Black.
Sassen, S. (2014). *Expulsions: Brutality and complexity in the global economy*. Harvard University Press.
Schlosberg, D. (2007). *Defining environmental justice: Theories, movements, and nature*. OUP Oxford.
Seiferling, I., Naik, N., Ratti, C., & Proulx, R. (2017). Green streets: Quantifying and mapping urban trees with street-level imagery and computer vision. *Landscape and Urban Planning, 165*, 93–101. https://doi.org/10.1016/j.landurbplan.2017.05.010
Sennett, R. (2017). *The fall of public man*. W. W. Norton & Company.
Shan, L., & He, S. (2025). The role of peri-urban parks in enhancing urban green spaces accessibility in high-density contexts: An environmental justice perspective. *Landscape and Urban Planning, 254*, Article 105244. https://doi.org/10.1016/j.landurbplan.2024.105244
Sheller, M. (2018). *Mobility justice: The politics of movement in an age of extremes*. Verso Books.

Shirazi, M., & Keivani, R. (2019). *Urban social sustainability: Theory, policy and practice*. Routledge.
Sibley, D. (2002). *Geographies of exclusion: Society and difference in the west*. Routledge. https://doi.org/10.4324/9780203430545
Skelton, T. (2013). Children, young people, UNICEF and participation. *Global childhoods* (pp. 165–181). Routledge. https://api.taylorfrancis.com/content/chapters/edit/download?identifierName=doi&identifierValue=10.4324/9781315878676-11&type=chapterpdf.
Soja, E. (2009). The city and spatial justice. *Justice Spatiale/Spatial Justice, 1*(1), 1–5.
Song, Y., Popkin, B., & Gordon-Larsen, P. (2013). A national-level analysis of neighborhood form metrics. *Landscape and Urban Planning, 116*, 73–85. https://doi.org/10.1016/j.landurbplan.2013.04.002
Su, T., Sun, M., Fan, Z., Noyman, A., Pentland, A., & Moro, E. (2022). Rhythm of the streets: A street classification framework based on street activity patterns. *EPJ Data Science, 11*(1). https://doi.org/10.1140/epjds/s13688-022-00355-5. Article 1.
Subramanian, D., & Jana, A. (2018). Assessing urban recreational open spaces for the elderly: A case of three Indian cities. *Urban Forestry & Urban Greening, 35*, 115–128. https://doi.org/10.1016/j.ufug.2018.08.015
Sun, P., Hou, R., & Lynch, J. P. (2020). Measuring the utilization of public open spaces by deep learning: A benchmark study at the Detroit Riverfront. In *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)* (pp. 2217–2226). https://doi.org/10.1109/WACV45572.2020.9093336
Talen, E. (2010). The spatial logic of parks. *Journal of Urban Design, 15*(4), 473–491. https://doi.org/10.1080/13574809.2010.502335
Tang, B. (2017). Is the distribution of public open space in Hong Kong equitable, why not? *Landscape and Urban Planning, 161*, 80–89. https://doi.org/10.1016/j.landurbplan.2017.01.004
Tang, Y., & Zhang, R. (2017). The balance between equity and efficiency in urban planning. In Q. Jiang, L. Qian, & M. Ding (Eds.), *Fair development in china* (pp. 185–203). Springer International Publishing. https://doi.org/10.1007/978-3-319-43663-0_11.
Townsend, A. M. (2013). *Smart cities: Big data, civic hackers, and the quest for a new utopia*. WW Norton & Company.
Un-Habitat. (2016). *World cities report 2016: Urbanization and development-Emerging futures*. UN.
UNICEF. (2019). *Shaping urbanization for children: A handbook on child-responsive urban planning*. United Nations.
Walker, G. (2012). *Environmental justice: Concepts, evidence and politics*. Routledge.
Wang, L., Han, X., He, J., & Jung, T. (2022). Measuring residents' perceptions of city streets to inform better street planning through deep learning and space syntax. *ISPRS Journal of Photogrammetry and Remote Sensing, 190*, 215–230. https://doi.org/10.1016/j.isprsjprs.2022.06.011
Wang, L., & He, W. (2023). Analysis of community outdoor public spaces based on computer vision behavior detection algorithm. *Applied Sciences, 13*(19). https://doi.org/10.3390/app131910922. Article 19.
Wang, L., Xiong, Y., Wang, Z., Qiao, Y., Lin, D., Tang, X., & Van Gool, L. (2016). Temporal segment networks: Towards good practices for deep action recognition. In B. Leibe, J. Matas, N. Sebe, & M. Welling (Eds.), *Computer vision – eccv 2016* (pp. 20–36). Springer International Publishing. https://doi.org/10.1007/978-3-319-46484-8_2.
Wartenberg, D. (1985). Multivariate spatial correlation: A method for exploratory geographical analysis. *Geographical Analysis, 17*(4), 263–283. https://doi.org/10.1111/j.1538-4632.1985.tb00849.x
Whyte, W. H. (1980). *The social life of small urban spaces*. https://www.academia.edu/download/55908872/thesociallifeofsmallurbanspacesmohddanish-150407042346-conversion-gate01.pdf.
Woolcock, G., Gleeson, B., & Randolph, B. (2010). Urban research and child-friendly cities: A new Australian outline. *Children's Geographies, 8*(2), 177–192. https://doi.org/10.1080/14733281003691426
Wu, F. (2015). *Planning for growth: Urban and regional planning in china*. Routledge. https://doi.org/10.4324/9780203067345
Xiao, Y., Wang, Z., Li, Z., & Tang, Z. (2017). An assessment of urban park access in Shanghai – Implications for the social equity in urban China. *Landscape and Urban Planning, 157*, 383–393. https://doi.org/10.1016/j.landurbplan.2016.08.007
Yang, J., Liu, S., Guo, H., Dong, Y., Zhang, X., Zhang, S., Wang, P., Zhou, Z., Xie, B., Wang, Z., Ouyang, B., Lin, Z., Cominelli, M., Cai, Z., Zhang, Y., Zhang, P., Hong, F., Widmer, J., Gringoli, F., … Liu, Z. (2025). *EgoLife: Towards egocentric Life Assistant* (arXiv:2503.03803). ArXiv. https://doi.org/10.48550/arXiv.2503.03803.
Yarlagadda, A. K., & Srinivasan, S. (2008). Modeling children's school travel mode and parental escort decisions. *Transportation, 35*(2), 201–218. https://doi.org/10.1007/s11116-007-9144-6
Yigitcanlar, T., Kamruzzaman, M., Buys, L., Ioppolo, G., Sabatini-Marques, J., da Costa, E. M., & Yun, J. J. (2018). Understanding 'smart cities': Intertwining development drivers with desired outcomes in a multidimensional framework. *Cities (London, England), 81*, 145–160.
Young, I. M. (2002). *Inclusion and democracy*. OUP Oxford.
Zepp, H., Inostroza, L., Sutcliffe, R., Ahmed, S., & Moebus, S. (2018). Neighbourhood Environmental contribution and Health. A novel indicator integrating urban form and urban green. *Change and Adaptation in Socio-Ecological Systems, 4*(1), 46–51. https://doi.org/10.1515/cass-2018-0005
Zhang, L., Ye, Y., Zeng, W., & Chiaradia, A. (2019). A systematic measurement of street quality through multi-sourced urban data: A Human-oriented analysis. *Int. J. Environ. Res. Public Health.*
Zhang, S., Benenson, R., Omran, M., Hosang, J., & Schiele, B. (2016). *How Far Are We From Solving Pedestrian Detection?*, 1259–1267. https://www.cv-foundation.org/openaccess/content_cvpr_2016/html/Zhang_How_Far_Are_CVPR_2016_paper.html.
Zhang, S., Yang, J., & Schiele, B. (2018). *Occluded pedestrian detection through guided attention in CNNs*. 6995–7003. https://openaccess.thecvf.com/content_cvpr_2018/html/Zhang_Occluded_Pedestrian_Detection_CVPR_2018_paper.html.
Zhao, J., Aziz, F. A., & Ujang, N. (2024). Factors influencing the usage, restrictions, accessibility, and preference of urban neighborhood parks—A review of the empirical evidence. *Urban Forestry & Urban Greening, 100*, Article 128473. https://doi.org/10.1016/j.ufug.2024.128473
Zhao, Z.-Q., Zheng, P., Xu, S.-T., & Wu, X. (2019). Object detection with deep learning: A review. *IEEE Transactions on Neural Networks and Learning Systems, 30*(11), 3212–3232. https://doi.org/10.1109/TNNLS.2018.2876865
Zukin, S. (1995). *The cultures of cities*. Blackwell Publishers.