Graphical Abstract

Urban Data Perception: Challenges, Trends, and Applications

Felipe Moreno-Vera, Bruno Brandoli, Jorge Poco

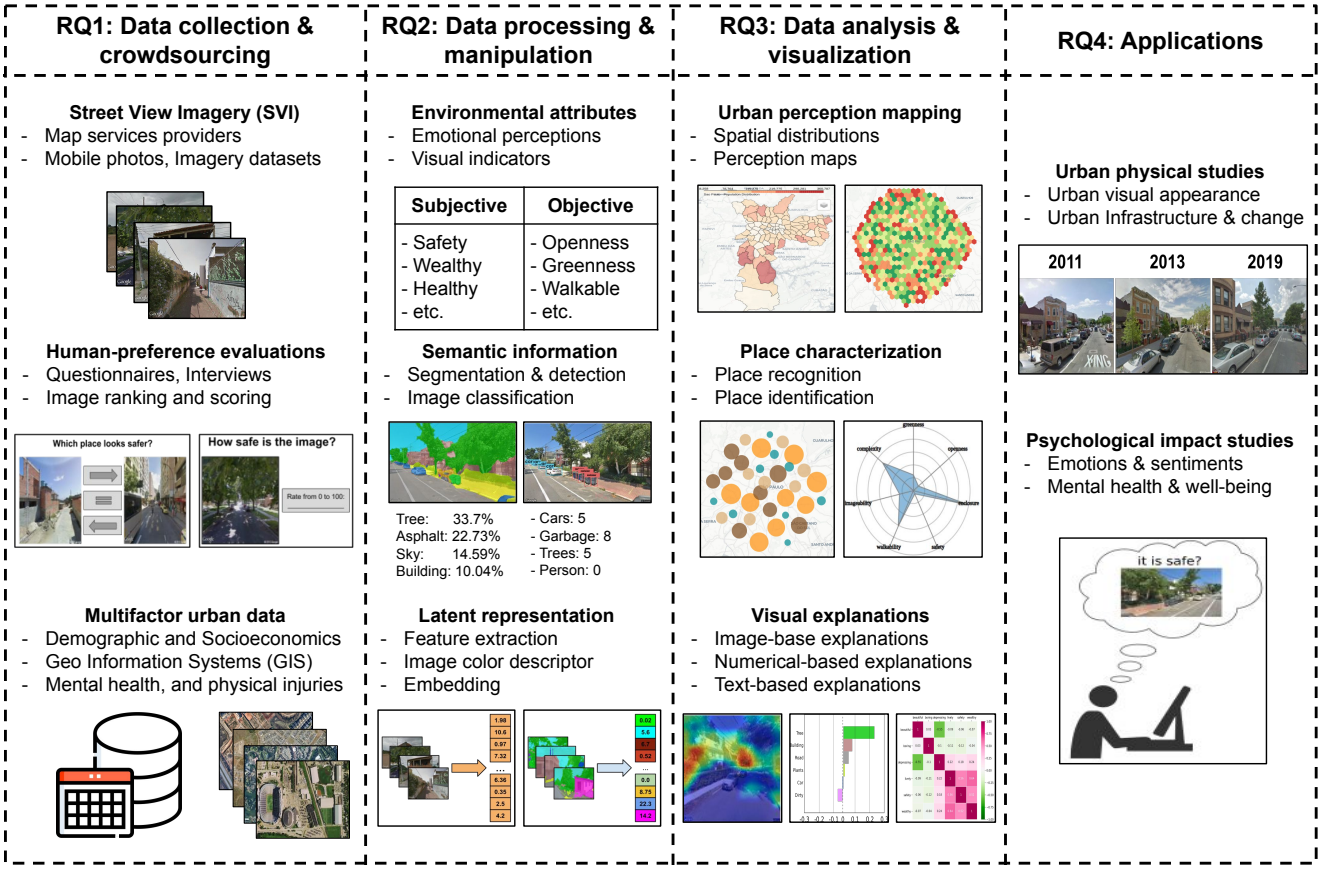

Figure 1: The overview of our systematic review taxonomy on urban perception using SVI across data collection and crowdsourcing, processing and manipulation, analysis and visualization, and application.

Highlights

Urban Data Perception: Challenges, Trends, and Applications

Felipe Moreno-Vera, Bruno Brandoli, Jorge Poco

- A comprehensive review of 207 studies focused on SVI-based urban perception analysis.
- We identified and categorized urban perception studies based on data-driven collection, process, analysis, and applications.
- We highlight challenges in urban perception data, future trends, and recommendations.

Urban Data Perception: Challenges, Trends, and Applications

Felipe Moreno-Vera*, Bruno Brandoli and Jorge Poco

Getulio Vargas Foundation (FGV), Rio de Janeiro, Brazil

ARTICLE INFO

Keywords:
Machine learning
Urban data science
Street view images
Urban design
Urban studies
Deep learning
Street human perception
Urban streetscape

ABSTRACT

Over the past four decades, urban perception has emerged as a significant research domain, spanning various disciplines such as urban perception map, urban mental well-being, and urban planning, with notable theories like the *Broken Window* theory. Its goal is to examine and explain how perception influences behavior in urban environments. While previous studies have introduced various methodologies to explore urban perception using street images alongside supplementary datasets, such as health statistics, crime rates, housing prices, demographics, and more, a recurring pattern of adopting similar pipelines and methodologies has emerged. Recognizing the need for a data-driven organization within this field, we propose a systematic review focusing on using street view imagery (SVI) and complementary relevant data in urban perception. This review provides a comprehensive overview of 207 studies published up to August 2024, selecting those that applied street images and supplementary data (e.g., crime rates, housing prices) to analyze urban perception. We then classify these studies into four key components: data collection, processing, analysis, and applications. This review not only serves as a valuable resource for urban planners, policymakers, and researchers, but also aims to benefit communities by enhancing our understanding of how urban data can be harnessed to inform decisions that improve public safety, well-being, and overall quality of life.

1. Introduction

Cities are designed to shape and influence the lives of their inhabitants (Lindal and Hartig, 2013). Numerous studies have shown that the visual appearance of a city plays a key role in human perception and reactions to the environment (Hao, Li, Han and Nie, 2024; Wei, Cao, Kwan, Jiang and Feng, 2024). For example, the *Broken Window Theory* (Wilson and Kelling, 1982) suggests that signs of environmental disorder, such as shattered windows, abandoned cars, and graffiti, can lead to negative social outcomes and increased crime. This theory has shaped public policies, including strict police tactics and guide studies comparing Well-maintained areas (e.g., clean shopping centers) with neglected areas (e.g., graffiti-covered streets) (Keizer, Lindenberg and Steg, 2008). Similarly, some studies proposed that the appearance of the city affects the psychological states of residents when evaluating green spaces (Ulrich, 1979), urban disorder (Sampson, Morenoff and Gannon-Rowley, 2002), graffiti and neglected buildings (Schroeder and Anderson, 1984).

The pivotal research *The Image of the City* (Lynch, 1984) correlated the visual appearance of the city and additional information such as crime records and demographics factors, among others, aiming to create a mental map of the city. In *The Importance of Death and Life of Great American Cities* (Wendt, 2009; De Nadai, Staiano, Larcher, Sebe, Quercia and Lepri, 2016), urban elements were analyzed to identify what characteristics can shape the identity of a city. A trend of studying urban perception using visual appearance and non-visual characteristics emerged to identify pleasant areas (Nasar, 1998), disorder-focused points (Skogan, 1992), to understand how people perceive various city designs (Rapoport and Hawkes, 1970), or the mental well-being of citizens and their city perception (Rzotkiewicz, Pearson, Dougherty, Shortridge and Wilson, 2018; Kang, Zhang, Gao, Lin and Liu, 2020).

Over the past years, the utilization of SVI services allows researchers to collect data and information on human evaluations in urban perception using online surveys through websites such as Place Pulse (Salesses, Schechtner and Hidalgo, 2013), Wmodi (Acosta and Camargo, 2018a), UrbanGems (Quercia, O'Hare and Cramer, 2014a), ELO Rating system (Liu, Chen, Wang, Huang, Thomas, Rahimi and Mamouei, 2023), City-SAFE (Costa, 2019), among others. Moreover, the increase in machine learning methods has led several studies exploring the relationship between urban appearance through SVI and crime records to create predictive models (Stalidis, Semertzidis and Daras., 2018; Glaeser, Kominers, Luca and Naik, 2018; Andersson, Birck and Araujo, 2017; He, Wang, Xie, Wu and Chen, 2022), SVI and urban green spaces (Li, Zhang and Li, 2015a; Hao et al., 2024; Li, Zhang, Li, Ricard, Meng and Zhang, 2015b; Seiferling, Naik, Ratti and Proulx, 2017), SVI and graffiti presence (Santani, Ruiz-Correa and Gática-Pérez, 2015; Tokuda, Silva and Jr., 2019), SVI and deep learning to compare images (Dubey, Naik, Parikh, Raskar and Hidalgo, 2016; Min, Mei, Liu, Wang and Jiang, 2020; Koch, Zemel and Salakhutdinov, 2015), SVI and demographic factors (Liu, Chen, Zhu, Xu and Lin, 2017b; Bai, Lam and Li, 2020; Beaucamp, Leduc, Tourre and Servieres, 2022; Yuan, Mu, Jiao, Li and Li, 2024), SVI and computer vision methods (Naik, Philipoom, Raskar and Hidalgo, 2014; Zhang, Zhou, Liu, Liu, Fung, Lin and Ratti, 2018c; Acosta and Camargo, 2018b), SVI and emotional perceptions (Moreno-Vera, Lavi and Poco, 2021a,b), and recently, a few works have conducted literature reviews on the use of Street View Imagery (SVI) and computer vision

*Corresponding author:
felipe.moreno@fgv.br (F. Moreno-Vera); bruno.brandoli@fgv.br (B. Brandoli); jorge.poco@fgv.br (J. Poco)
ORCID(s): 0000-0002-2477-9624 (F. Moreno-Vera); 0000-0001-6167-8104 (B. Brandoli); 0000-0001-9096-6287 (J. Poco)
[1]Received ...

in urban perception (Zhang, Salazar-Miranda, Duarte, Vale, Hack, Chen, Liu, Batty and Ratti, 2024a; Ito, Kang, Zhang, Zhang and Biljecki, 2024).

This paper systematically reviews urban perception focused on data-driven approaches, analyzing the merge of SVI and additional data (e.g., crime rates, house prices, mental health reports). We investigate four questions: 1) What essential information is needed to collect and define a city's perception of its urban environment? 2) What are the key algorithms used to analyze and quantify urban perception data? 3) What are the primary data-driven analyzes used to understand and interpret urban perception behavior? 4) What are the applications of urban perception data and their impacts on citizens? These questions were answered in four components: data collection, data processing, data analysis, and applications.

This work is structured as follows: literature in Section 2, systematic review in Section 3, results in Section 4, review in Section 5, discussion and future directions are described in Section 6, and conclusion in Section 7.

2. Literature

In recent years, a few SVI-based reviews in urban studies have introduced methodologies, algorithms, and research on topics such as city perception, mental health, and deep learning for urban analytics. However, none explicitly adopts a data-driven approach that covers collection, processing, analysis, and applications. This paper categorizes the reviews into three main groups: (i) urban analytics, (ii) computer vision and AI, and (iii) mental health and well-being, while also summarizing general SVI applications in urban studies (Table 1).

2.1. Urban analytics

Charreire, Mackenbach, Ouasti, Lakerveld, Compernolle, Ben-Rebah, McKee, Brug, Rutter and Oppert (2014) systematically reviews the application of remote sensing (e.g., Google Earth, SVI, and Bing maps) in defining environmental characteristics associated with physical activity and environmental factors such as infrastructure, transport, aesthetics, safety, and park equipment. He and Li (2021) performed a comprehensive analysis of research trends to assess urban neighborhood environments by applying SVI processing, exploring various methodologies and applications in urban planning, environmental analysis, and social science. Moreover, Biljecki and Ito (2021) comprehensively examined the applications of SVI to urban analytics and Geographic Information Systems (GIS). Their review explored various applications in urban analytics using SVI and aerial images, such as urban planning, transportation studies, environmental monitoring, transportation, and social sciences.Xu, Jin, Chen and Li (2021) reviewed the commonly used SVI service providers in urban studies, as organized in the China National Knowledge Infrastructure (CNKI). They also discussed the integration of machine learning techniques to measure various urban elements (e.g., greenery, buildings, vehicles), highlighting a broad range of applications, including urban planning, public health, and socio-environmental studies.

The review published by Li, Peng, chong Wu and Zhang (2022b) provides an in-depth analysis of SVI in the built environment and the urban planning context. Such as review explored the benefits and challenges of quantifying the built environment and emotional urban perception based on physical aspects, technical costs, quality, and privacy issues of working with SVI-based services and additional metadata such as graphs, city networks, and a temporal record of images per street. In the paper of Dai, Li, Stein, Yang and Jia (2024), a detailed analysis of built environment auditing tools (BEA) that use SVI in visual perception was provided, evaluating their effectiveness in various aspects of the built environment, such as street morphology, land use, transportation infrastructure, and environmental characteristics applied to urban neighbor perceptions.

Our work builds on previous approaches and extends them by not only analyzing Street View Imagery (SVI), the types of information extracted, and the processing methodologies employed, but also by incorporating various data sources such as aerial imagery, Geographic Information Systems (GIS), Virtual and Augmented Reality, as well as demographic, socioeconomic, and health-related factors. These additions are crucial to improving urban perception analysis.

2.2. Computer vision and AI

Tian, Chen, Xiong, Li, Dai, Chen, Xing, Chen, Wu, Hu et al. (2017) offers a comprehensive analysis of the ever-evolving landscape of what is termed Artificial Intelligence (AI) 2.0, encompassing images and unstructured data such as audio, video, and other formats. It specifically emphasizes the progression from human-like perception to the potential of transhuman capabilities, providing consolidated and insightful applications in areas such as computer vision, natural language processing (NLP), lensless cameras, and speech recognition techniques. In addition, Ibrahim, Haworth and Cheng (2020) explored the applications of computer vision in urban analytics, discussing the specific deep learning models employed, their performance in urban contexts, and their contributions to urban analytics (urban research that takes advantage of the new data resources captured, such as aerial images, SVI, etc.) highlighting the intersection of deep learning algorithms, satellite images, SVI, and urban analytics.

The study of Yin, Peng, Li, Shi, Yang and Jia (2023) provided an overview of the relationship and applications between SVI and computer vision methods to the 2030 Agenda for the 17 Sustainable Development Goals (SDGs). This review describes computer vision methods applied to SVI, which are related to four main objectives: health and well-being, sustainable cities and communities, climate change and its impacts, and zero hunger, highlighting the potential of this technology to inform decision-making processes aligned with the Sustainable Development Goals.

Table 1
The summary of relevant review papers on urban data science.

Urban analytics				
Source	# papers	Time frame	Data providers	Short description and topics
Charreire et al. (2014)	13	2010-2013	Google Earth, Google Maps, Bing Maps	Encompass various environmental aspects captured by remote sensing: •land use patterns; •green spaces; •accessibility to healthy food options.
He and Li (2021)	288	2000-2020	BSV, GSV, TSV, LiDAR, and POIs	Evaluates the strengths and limitations of the urban street environment: •solar radiation; •socioeconomic; •street canyon morphology; •environmental perception; •thermal environment; •building and facade;
Biljecki and Ito (2021)	250	2018-2020	GSV, KartaView, Mapillary, TSV, BSV, and GIS	Classify studies based on their applications: •spatial data infrastructure; •urban perception; •health and well-being; •urban morphology; •greenery; •walkability; •socioeconomic •transportation and mobility; •real estate;
Xu et al. (2021)	337	2005-2021	GSV, BSV, TSV,	Analyze data collection technologies for urban studies: •land use; •street walking; •landscape environment; •urban atmosphere; •street design; •built environment;
Li et al. (2022b)	263	2007-2022	GSV, KartaView, Mapillary, TSV, BSV, and AppleMap	Study applications of SVI in the Built Environment: •element identification; •community safety; •physical environment assessment; •public health; •environmental behavior; •spatial semantic speculation
Dai et al. (2024)	96	2010-2023	ANC, CANVAS, EGA-Cycling, PEDS, S-VAT, and V-STEPS	Classify building environment attributes: •social environments; •visual perceptions; •land use; •street-related features; •physical activities; •traffic-related features;
Computer vision and AI				
Source	# papers	Time frame	Methods	Short description and topics
Tian et al. (2017)	51	2006-2017	Multi-modal learning Classification, Regression Generative models Speech recognition Segmentation	Expand the concept of human perception including other data in addition of SVI: •visual perception; •auditory perception; •speech perception; •learning engines; •perceptual information processing.
Ibrahim et al. (2020)	641	2010-2020	Classification; Object Localization; Action recognition; Generative models; Decision-making Clustering; Segmentation	Computer vision applications in street-level or aerial images: •built environment; •human interaction; •transportation and traffic; •natural environment; •infrastructure
Yin et al. (2023)	147	2015-2023	Classification; Detection; Localization; Regression; Segmentation Decision-making	Study the applications of SVI in the 2030 Agenda for Sustainable Development Goals (SDGs): •SDG 11:sustainable cities and communities; •SGD 3: good health and well-being; •SGD 13: climate change, •SGD 2: zero hunger.

Marasinghe, Yigitcanlar, Mayere, Washington and Limb (2023) highlighted the opportunities and limitations of applying computer vision in urban planning, exploring various aspects of urban planning processes and computer vision applications, including data collection, analysis, issue identification, prioritization, land use classification, transportation analysis, plan design, space modeling, and infrastructure development. Song (2024) reviewed related papers on SVI and AI methods and applications, summarizing the acquisition, storage, and common data sources of SVI. It also detailed three aspects of AI-based SVI applications: quantification of the physical space, urban perception, and spatial semantic speculation.

Table 1
The summary of relevant review papers on urban data science (continued).

Computer vision and AI				
Source	# papers	Time frame	Methods	Short description and topics
Marasinghe et al. (2023)	87	2012-2023	Segmentation Classification; Regression Detection and tracking, 3D modeling Scene recognition	Study how is CV applied in the urban planning process: •visualize and model space; •augment and automate planning and design; •monitor and evaluate designs; •enhance collaborative planning; •analyze, measure, map, and predict spaces.
Song (2024)	16	2017-2022	Classification; Regression Scene recognition Detection; Localization Segmentation	Present AI-based SVI applications: •quantification of physical space; •urban perception; •spatial semantic speculation
Liu and Sevtsuk (2024)	146	2018-2022	Classification; Regression Segmentation Detection; Localization Scene recognition	Explores planning-relevant attributes obtained from CV: •visual dominants: sidewalk presence, greenery, sky, etc. •micro-Level details: cars, persons, traffic lights, etc. •composite metrics: Floor Area Ratio (FAR), Height-to-Width (H/W) and Transparency Ratio (TR) •deep features: last dense layer vectors
Zhang et al. (2024a)	180	2010-2024	Classification; Regression Scene recognition Detection; Localization Segmentation	Outlines key aspects related to the urban physical environment: •street-level imagery; •image understanding; •place characterization; •human-place relationship;
Ito et al. (2024)	393	2000-2023	Classification; Regression; Segmentation; Detection Localization; Feature extraction	Highlight and categorize subjective responses of the built environment using visual data (VR, SVI, aerial, and geo-tagged photos): •greenery and water; •street design; •public space; •building design; •landscape; •city as a whole.
Mental health and well-being				
Source	# papers	Time frame	Health issue	Short description and topics
Rzotkiewicz et al. (2018)	54	2007-2017	Injury/recovering, Injury prevention, obesogenic, Alcohol and smoking	Explores SVI applications in mental well-being: •built environment assessment; •health policy compliance; •study site selection; •disaster preparedness
Kang et al. (2020)	64	2007-2019	Physical injury Obesity, walk activities cycling, depression stress, mental disorders	Study mental impacts of urban physical environment: •walkable environment; •cycling behaviors; •obesity and food environment; •physical injury; •mental health; •perception and sentiment

Liu and Sevtsuk (2024) investigated the planning-relevant attributes extracted from SVI-based datasets using computer vision (CV) and explored their research applications. Besides, their work categorized the attributes into three main types, *categorical type* corresponding with classification attributes and labels such as socioeconomic factors and land use types; and *continous type* derived from the pixel ratios obtained through segmentation models, and *discrete type* sourced from object detection models. Zhang et al. (2024a) presented a framework to study the intersection of AI and urban studies, focusing on the use of SVI to gain insight into urban physical environments using visual information from the streets. In addition, their review categorized the level of application of AI with SVI to describe and represent the urban physical environment, starting from the visual appearance of the city to the implications for city policymakers. Ito et al. (2024) conducted a systematic review using NLP and GPT-4 to automate the filtering, extraction, and identification of papers on urban visual data, focusing on SVI, virtual reality, geo-tagged photos, and aerial imagery for urban perception. Besides, the reviewed works were categorized into greenery, public space, building design, and landscape, highlighting their scope, evolution, and future opportunities.

Regardless of the focus on SVI in previous reviews, our work extends the scope of visual data and processing by incorporating complementary data sources such as demographic factors, socioeconomic factors, geoinformation systems, and mental health data. We classify these sources according to their level of representation, also known as the

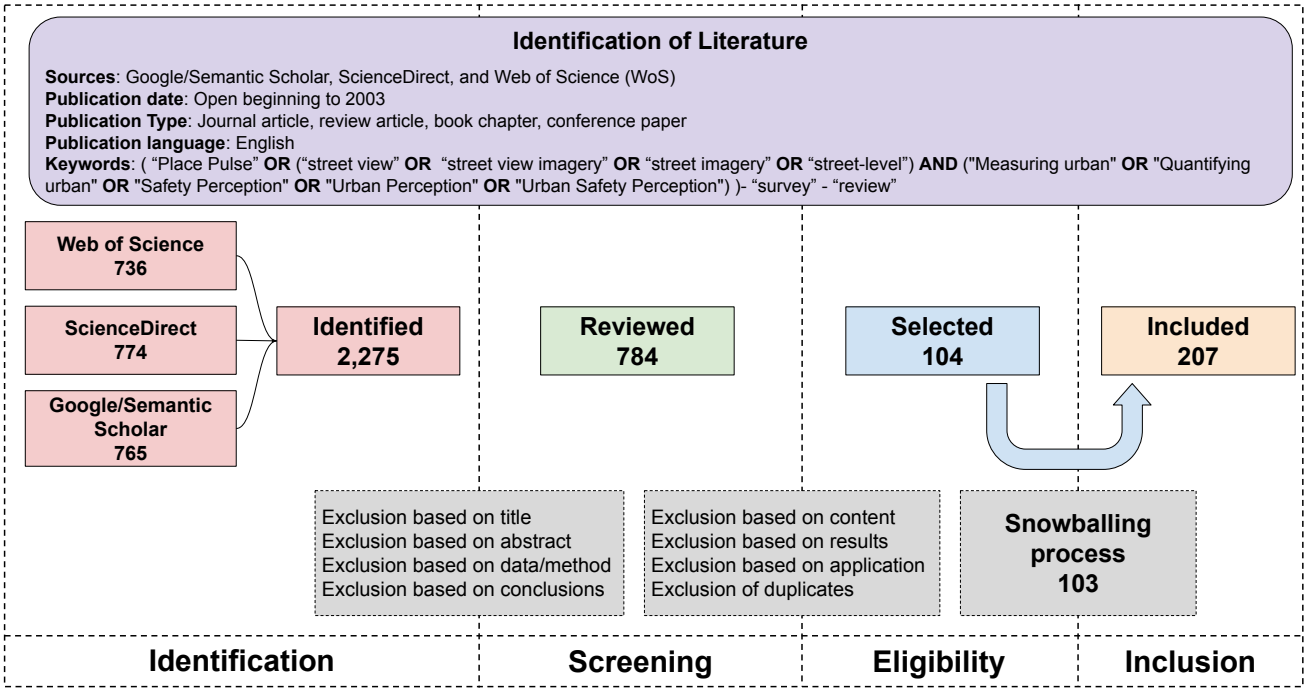

Figure 1: Flow of the systematic review following the PRISMA protocol, including snowballing in the final step to identify and capture 103 additional relevant papers for this review.

feature level (Bengio, 2007). In addition, we extend the analysis of AI-based methods and define a new *environmental perception attributes* feature level.

2.3. Mental health and well-being

Rzotkiewicz et al. (2018) presented a comprehensive analysis providing potential directions for future studies in health-related research using GSV. This review aims to categorize health outcomes and delineate the applications of GSV in health-related studies by examining key themes, strengths, and weaknesses such as the identification of sources of air, soil, or water pollution, park design and usage, and neighborhood conditions. Furthermore, Kang et al. (2020) provided a comprehensive review of the integration of SVI into public health studies to detect and analyze the physical environment of cities. Their review covers the methodologies used to detect, quantify, and characterize urban environments, such as extracting relevant characteristics such as green spaces, built environments, and other physical attributes that impact public health.

Despite numerous reviews, the key question of *Which specific urban data is relevant to planners and urban designers?* is not adequately addressed by previous works. To go beyond the previous literature, our work extends the analysis of urban perception by exploring the relationship between SVI, demographic factors, socioeconomic factors, emotional perception, visual perceptions, and human evaluations with physical and psychological aspects and applications. By expanding these applications, our research contributes to a deeper understanding of SVI's potential across multiple domains, enriching the evolving research landscape in this field.

3. Methodology

Street View Imagery (SVI) is an emerging form of remote sensing data that captures ground-level environmental perspectives, offering diverse benefits, including applications in social research, urban perception, and urban planning, among others (Wen, Liu and Yu, 2022; Ewing, Clemente, Neckerman, Purciel-Hill, Quinn and Rundle, 2013). Despite its growing use, the literature lacks a systematic review of SVI data-driven studies and applications focused on urban perception, making it challenging to monitor trends, gain insight, and study key applications effectively. This paper aims to summarize existing studies in urban perception research that: (1) employ SVIs and complementary city data; (2) use computer vision, machine learning, and deep learning methods to process information; (3) analyze and visualize insights obtained to define applications. Given these goals, we define the following four *research questions (RQ)*:

- **RQ1 - What essential information is needed to define a city's perception of its urban environment?**
- **RQ2 - What are the key algorithms used to analyze and quantify urban perception data?**
- **RQ3 - What are the primary data-driven analyzes used to understand and interpret urban perception behavior?**

- **RQ4 - What are the applications of urban perception data and their impacts on citizens?**

To address RQ1, we explored SVI providers, APIs, and open datasets, along with crowd-sourced data to enrich urban information. For RQ2, we classified studies by feature extraction algorithms and feature levels. For RQ3, we grouped studies by analysis type and visualization methods. For RQ4, we categorized urban perception applications into physical and psychological domains.

3.1. Overview and time frame

This systematic review follows the Preferred Reporting Items for Systematic Reviews and Meta-analyzes (PRISMA) protocol (Regona, Yigitcanlar, Xia and Li, 2022)[1]. PRISMA is a widely recognized framework that enhances the transparency, consistency, and thoroughness of systematic reviews. It has been used in related studies (Charreire et al., 2014; Yin et al., 2023; Marasinghe et al., 2023; He and Li, 2021). The method is divided into four steps: identification, screening, eligibility, and inclusion.

Figure 1 illustrates this review process. We first selected relevant keywords to identify an initial set of papers, which were screened to filter out irrelevant works. We then reviewed the content of the remaining papers, focusing on studies related to urban perception using SVI from 2003 to 2024. The starting year of 2003 was chosen due to the release of BSV-2005, GSV-2007, and TSV-2011. During the eligibility and inclusion stages, we employed snowballing (Wohlin, 2014) to discover additional relevant research not identified in the initial search. Further details of this process are described in the following sections.

3.2. Identification

This step, also known as the *search criteria*, aimed to identify an initial set of papers. We searched ScienceDirect, Google/Semantic Scholar, and Web of Science from January 2003 to August 2024. Following prior research (He and Li, 2021; Biljecki and Ito, 2021; Yin et al., 2023), we focused on papers that contain terms such as *street view*, *street view imagery*, *street imagery*, and *street level* (see Figure 1). Additional keywords like *Measuring urban*, *Quantifying urban*, *Safety perception*, *Urban Perception*, and *Urban Safety Perception* were included to capture studies specifically examining urban perception using SVI. We also incorporated the term *Place Pulse* to target relevant studies in urban perception.

Although terms like *street view* or *street* are generic, they help identify papers that mention services such as *Google Street View*, *Tencent Street View*, *Baidu Street View*, and *Open Street Map*. We expanded the search to include other mapping services. For example, searching *mapping* alongside *street view* and *urban perception* yielded 39 results, while excluding *urban perception* gave us 1,470 results. Papers irrelevant to our scope were excluded. The search began in February 2023 and ended in August 2024, identifying a total of **2,275 papers**.

[1] http://prisma-statement.org/

3.3. Screening

This step, also called *selection criteria*, involved reviewing the abstracts, titles, data used, methods, and conclusions of the initial 2,275 papers to compile a relevant subset, adhering to the following criteria: (1) the paper is in English; (2) the study focuses on an urban street-level context; (3) the primary method is not only computer vision, but includes other approaches such as statistical analysis and deep learning; and (4) the study examines urban perception using SVI. We excluded papers that used SVI for topics like agricultural monitoring, city pollution, traffic, city simulator, or sound analysis, as well as those that focused exclusively on computer vision or deep learning tasks (e.g., image segmentation) without applying them to urban contexts. We ultimately retained a total of **784 papers**.

3.4. Eligibility

In this step, we reviewed 784 papers by extracting key characteristics such as SVI service, publication year, keywords, publication venue, author affiliations, city studied, applied machine learning task, references, and citations. We excluded papers based on the following criteria: (1) duplicates; (2) non-peer-reviewed materials like abstracts, presentations, books, or case reports; (3) research unrelated to urban perception. This resulted in the inclusion of **104 papers**. We then applied snowballing to identify additional relevant papers.

Snowballing process

Snowballing, or citation chaining, is a systematic method to expand relevant literature beyond initial search results (Wohlin, 2014). It involves iteratively exploring references cited in identified papers and discovering new papers that contribute to the research. This process includes guidelines for performing *Backward* and *Forward* steps.

- **Backward snowballing** starts with the references of identified papers and works backward.
- **Forward snowballing** explores the papers that have cited the identified papers moving forward in time.

Here, we used Semantic Scholar [2] to identify *highly influential* references and *highly influenced* citations. Thus, Semantic Scholar strives to recognize the highly influential references in a study, specifically those references that are utilized or expanded upon in the analyzed research. This recognition is achieved by considering factors such as the location, context, and frequency with which a work is referenced. After performing this analysis on the first set composed of 104 papers, we filtered and selected an **additional 103**, completing our total of **207 papers to be included in this review**.

3.5. Inclusion

In the previous steps, we selected 207 papers that met the criteria. The papers were then analyzed and categorized

[2] https://www.semanticscholar.org/

Table 2
Top papers with over 300 citations included in this review as of the end of August 2024.

Title	Journal/conference	Year	Citations
What makes Paris look like Paris? (Doersch, Singh, Gupta, Sivic and Efros, 2012)	ACM Transactions on Graphics	2015	720
Assessing street-level urban greenery using Google Street View and a modified green view index (Li et al., 2015b)	Urban Forestry & Urban Greening	2015	586
The Shortest path to happiness: recommending beautiful, quiet, and happy routes in the city (Quercia, Schifanella and Aiello, 2014b)	ACM Conference on Hypertext and social media	2014	407
The Collaborative Image of The City: Mapping the Inequality of Urban Perception (Salesses et al., 2013)	PLoS ONE	2012	386
Measuring human perceptions of a large-scale urban region using machine learning (Zhang et al., 2018c)	Landscape and Urban Planning	2018	374
Streetscore: Predicting the Perceived Safety of One Million Streetscapes (Naik et al., 2014)	Computer Vision and Pattern Recognition Conference Workshops	2014	356
Deep Learning the City: Quantifying Urban Perception at a Global Scale (Dubey et al., 2016)	European Conference on Computer Vision	2016	323

into four main groups according to the data-driven processes commonly used in the work. This term typically refers to the end-to-end process of managing and processing data from collection to final application. By considering the number and diversity of papers reviewed, we are confident that our review minimizes bias and effectively captures the prevailing trends in urban perception.

3.6. Taxonomy

We identified and clustered four main categories based on identified data-driven pipeline methodologies, divided into four steps: (1) data collection and crowdsourcing, (2) processing and manipulation, (3) analysis and visualization, and (4) applications.

4. Results

The following section summarizes some insights obtained from the literature review, including keywords, publications and citations, countries and institutions, and learning problems and tasks.

4.1. Publication and citations

After analyzing their publication dates, we found that all included works were published since 2012, which coincides with the origin of convolutional neural networks (CNN) studies with AlexNet (Krizhevsky, Sutskever and Hinton, 2012) and the beginning of research on quantifying urban perception using websites and online surveys(Salesses et al., 2013). Despite that, published papers before 2014 used classical methods such as Gist, Histograms, and Shift-vectors (Doersch et al., 2012; Wilson, Kelly, Schootman, Baker, Banerjee, Clennin and Miller, 2012; Lindal and Hartig, 2013; Ordonez and Berg, 2014; Naik et al., 2014); while more recent studies have applied complex techniques like reinforcement learning and deep CNN (Zhang et al., 2018c; Alzate, Tabares and Vallejo, 2021; Moreno-Vera et al., 2021a; Wang, Zeng and Zhao, 2022c).

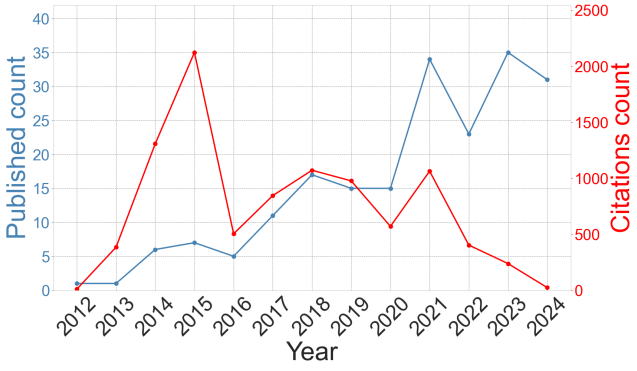

Figure 2: The number of published papers (blue) and citations (red) by year indicates a decline in citations from 2014 to 2023 (2024 is close to zero because those studies are the newest), especially for earlier works. However, the number of publications on urban perception using SVI has increased since 2016.

Figure 2 shows a yearly increase in the number of publications (blue line) and the number of citations (red line) per year. We note that the number of publications remained regular between 2014 and 2016, and 2018 and 2020, with a similar number of publications in both periods. Besides, since 2020, the field has grown significantly, with researchers

Table 3
Keywords and the number of papers found, selected, unique (after removing duplicates), and those added via snowball process.

Keywords	Found	Selected	Unique	Snowball	Total
Street View	1,090	7	5		
Measuring urban	715	11	9		
Quantifying urban	254	13	13		
Urban perception	184	69	61	103	
Place Pulse	26	12	12		
Safety perception	6	4	4		
Total	2,275	116	104	103	207

focusing on extracting relevant information from SVI for urban perception and integrating complementary data sources (e.g., crime rates and social media).

However, the citations of articles published between 2012 and 2015 show a trend of increase, but it dropped dramatically in 2016, nearly a three-fold decrease. We also have a similar behavior in 2021. In addition, all these works have been cited 9,511 times as of August 2024. Table 2 lists the top works with more than 300 citations. In total, the works included in this review were published in 113 different conferences and journals, where the most preferred journals/conferences are *Cities* and *Landscape and Urban Planning* with 11 papers, *ISPRS International Journal of Geo-Information* with 10 papers, *Computers, Environment, and Urban Systems* with 9 papers, *Land* and *International Journal of Applied Earth Observation and Geoinformation* with 6 papers, etc.

4.2. Keywords

As outlined in Section 3, we began by searching research papers using keywords such as *street view*, *street view imagery*, *street imagery*, and *street level*, later adding terms like *measuring urban*, *quantifying urban*, *safety perception*, *urban Perception*, and *urban safety perception*. Table 3 shows the number of papers found and reviewed. While more papers were initially identified under *Street View*, most were excluded as they did not focus on urban perception. Besides, *Urban perception* yielded the most relevant for this review.

Figure 3 shows the wordcloud of the most used keywords in the 207 papers included. We note that the top 5 keywords most cited are *deep learning*, *urban perception*, *computer vision*, *street view images*, and *safety perception*.

4.3. Countries and institutions

The leading countries in this research area are China (31.8%), the USA (18.36%), and the UK (9.18%). Regarding research institutions, the approaches have similarities and differences between institutions. The investigations of the institutions of the USA focus on improving the accuracy of the model for the comparison and prediction of the perception of SVI. Chinese institutions focus on fine-tuned pre-trained models using the Place Pulse 2.0 dataset (Dubey et al., 2016) and extend urban perception studies to cities around Beijing (Zhang, Pei, Wang, Wu, Song, Guo and Chen, 2020b). UK institutions are focused on studying the relationship between urban landscape and visual appearance (Seresinhe, Preis and Moat, 2017; Law, Seresinhe, Shen and Gutierrez-Roig, 2018b; Law, Paige and Russell, 2018a), panoramic perception (Liu et al., 2023; Muller, Gemmell, Choudhury, Nathvani, Metzler, Bennett, Denton, Flaxman and Ezzati, 2022), and beauty and aesthetic perception (Joglekar, Quercia, Redi, Aiello, Kauer and Sastry, 2020; Quercia et al., 2014a; Kumakoshi, Onoda, Takahashi and Yoshimura, 2021).

Although China and the USA conducted the most studies, several studies were conducted in 39 other countries, including Spain, Sweden, Ukraine, Mexico, and Scotland, among others. However, the main countries where the studies were applied and performed are China (8.31%), the USA (6.21%), and the UK (4.54%).

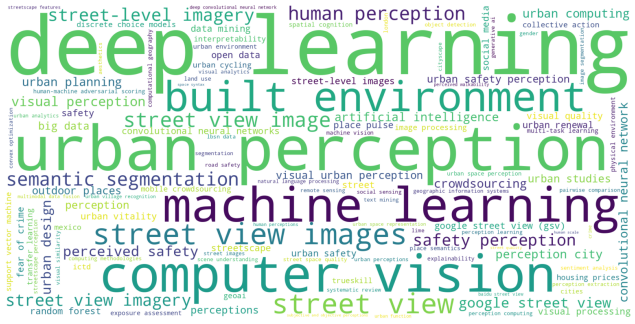

Figure 3: Word cloud depicting the most frequently occurring keywords from 207 studies related to urban perception using street view imagery. The size of each word corresponds to its frequency across the studies, with larger words indicating higher prominence. Common themes include *urban perception*, *deep learning*, *street view images*, *safety perception*, and *built environment*, highlighting the interdisciplinary nature of this research field and its focus on combining visual data with computational methods for urban analysis.

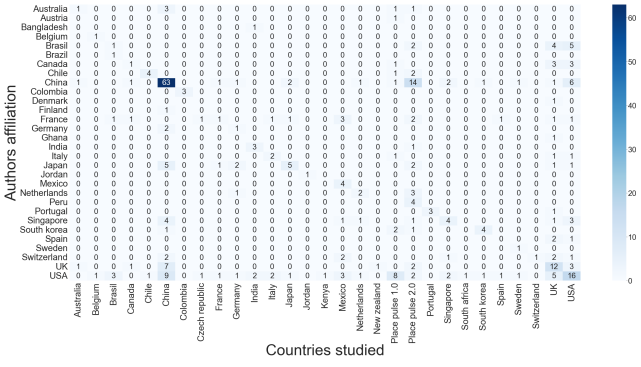

Figure 4: The number of author affiliations (red) versus country applications (blue) shows that most research is conducted in China, the USA, and the UK. Additionally, the majority of authors' affiliations are from China and USA.

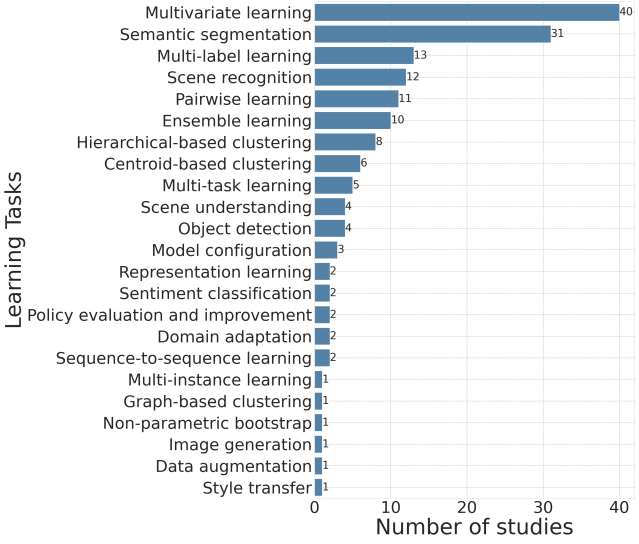

Figure 5: The number of studies per task category shows that regression (multivariate learning) is the most common method, followed by image segmentation. Some studies perform more than one task, which falls into multiple categories.

Table 4
Learning type, learning problem, and the number of papers found.

Learning type	Learning problem	Total
Supervised Learning	Classification	67
	Regression	46
	Rank learning	8
Unsupervised Learning	Clustering	16
	Generative	2
Semi and Self-Supervised Learning	Generative	1
	GPT/LLM	3
Reinforcement Learning	Policy learning	2

4.4. Learning problems and tasks

We identified and organized studies based on the learning problem focused on: (i) supervised learning (classification, regression, and rank learning); (ii) unsupervised learning (clustering and generative models); (iii) semi-supervised and self-supervised (generative models and GPT); and (iv) reinforcement learning (policy learning). Table 4 shows the distribution, with 121 methods based on supervised learning, 18 on unsupervised, 4 on semi-supervised and self-supervised learning, and 2 on reinforcement learning. We also explored the learning tasks performed, observing that *semantic segmentation* and *multivariate learning* are the most common techniques used in urban perception studies related to the exploration of the relationship between the presence of objects and data-driven analysis (see Figure 5). On the other hand, the less used methods correspond to learning tasks used at least in one single paper, such as *style transfer* and *graph-based clustering*.

Table 5
Summary of the most used SVI service provider and the number of papers found.

Provider	Coverage	Resolution	Geo scale	Total
GSV	135 countries and regions	2048×2048	global	57
			country	1
			city	91
BSV	652 cities in China	1024×512	global	7
			country	3
			city	41
TSV	296 cities in China	1680×1200	global	5
			city	16
Others	Many countries	Depends on the provider	global	24
			country	8
			city	102

5. Taxonomy

This taxonomy organizes the key components identified in our systematic review and highlights the commonalities among the selected works, encapsulating the main concepts of our systematic review of the relevant data for urban perception analysis. Figure 6 presents the proposed taxonomy that answers the RQs along the data pipeline process.

5.1. RQ1: Data collection and crowdsourcing

Here, we present the description and intrinsic process of data collection, a foundational stage in urban perception research, divided into SVI, human-preference evaluation, and complementary multifactor urban data.

Street View Imagery (SVI)

SVI is a revolutionary technology that makes the world's streets easy to access, allowing anyone to explore and navigate virtually real-world streets, experiencing a dynamic and immersive representation of cityscapes, landscapes, and neighborhoods (Li et al., 2022b). The most well-adopted services identified in this review are Google Street View (GSV), Tencent Street View (TSV), and Baidu Street View (BSV). GSV is a globally recognized mapping service known for its extensive coverage and immersive panoramic imagery from more than 100 countries, while BSV and SV provide detailed Chinese city maps, street views, satellite imagery, real-time traffic information, location-based services, and business listings restricted only to users in China. We also identify other SVI providers, such as Naver Street View (NSV), restricted to South Korean areas. Also, Mapillary and KartaView are services that cover about 1.5 billion images around the world. However, the former two services provide images uploaded by users.

In this review, we classified the level of geographical analysis of the papers into three scales: *city scale*, means that either a dataset was created or an analysis was performed in a particular city (Costa, 2019; Gao, Hou, Gao, Zhao and Jia, 2023; Colombo, Pincay, Lavrovsky, Iseli, Van Wezemael and Portmann, 2021; Xu, Xiong, Jing, Xing, An, Tong, Liu and Liu, 2023b; Rossetti, Lobel, Rocco and

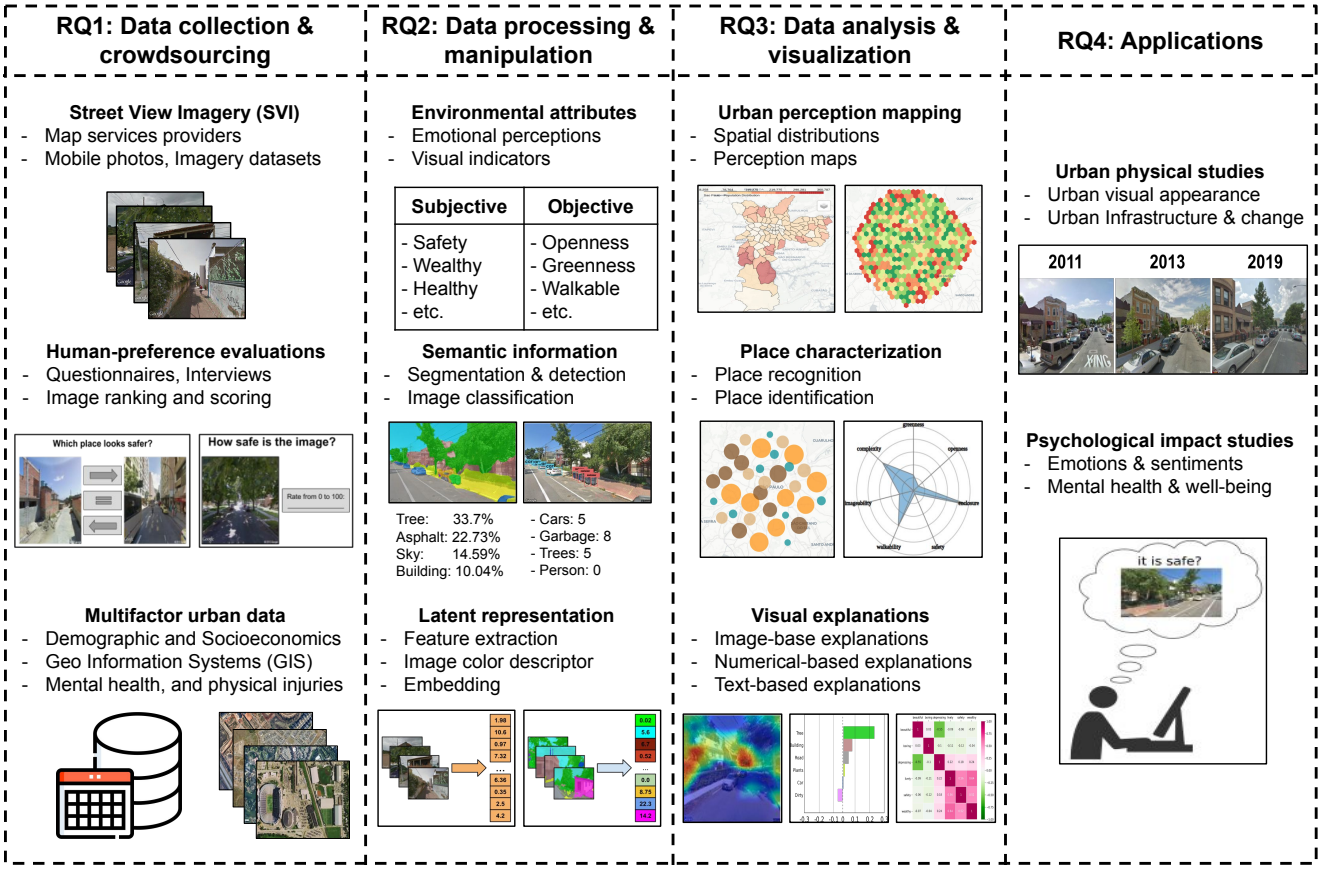

Figure 6: The overview of our systematic review taxonomy on urban perception using SVI across data collection and crowdsourcing, processing and manipulation, analysis and visualization, and applications.

Hurtubia, 2019a). *Country scale*, one or several cities of the same country (Lee, Grosz, Uzkent, Zeng, Burke, Lobell and Ermon, 2021a; Lee, Grosz, Zeng, Uzkent, Burke, Lobell and Ermon, 2021b). *Global scale*, one or more cities from different countries (Salesses et al., 2013; Dubey et al., 2016; Huang, Wang and Cong, 2024c; Liu et al., 2017b; Suel, Muller, Bennett, Blakely, Doyle, Lynch, Mackenbach, Middel, Mizdrak, Nathvani, Brauer and Ezzati, 2023).

Table 5 summarizes the information on the most used SVI services, giving information on coverage, resolution, and geographic scale of analysis. We grouped into *others* all other SVI services and datasets, and road network services used in some studies, such as Open Street Map (OSM), Naver Street View (NSV), Kakao Street View (KSV), Mapillary, Geogaph-UK, ArcGIS, KartaView, GeoGraph, Gaode, Cityscapes, FourSquare, and self-captured methods.

Human-preference evaluations

In the past few years, urban perception studies have used SVI service providers to collect human perception information from volunteers using websites, online surveys, and interviews. Volunteers were asked to choose between two images answering one question such as *Which place looks safer?* (Salesses et al., 2013; Costa, 2019; Kang, Abraham, Ceccato, Duarte, Gao, Ljungqvist, Zhang, Näsman and Ratti, 2023a) or rate the image's safety on a scale of 0 to 100 (Dai, Zheng, Dong, Yao, Wang, Zhang, Ren, Zhang, Song and Guan, 2021; Cui, Zhang, Yang, Huang and Chen, 2023b; Liu, Silva, Wu and Wang, 2017a; Hu and Chen, 2018). This category evaluated is also called *emotional perception* (e.g., safety, wealthy, lively, danger). These online surveys process gathered information about SVI such as latitude, longitude, rate/score, winner (comparison between two images), emotional perception evaluated, and, in some cases, information about volunteers such as age, nationality, and gender.

The dataset used most often is Place Pulse [3] that contains about 110,000 images from 56 cities evaluated in six categories: wealthy, safety, boring, depressing, beauty, and lively. This work inspired other studies such as StreetScore (Naik et al., 2014; Naik, Raskar and Hidalgo, 2016), which mapped the emotional perceptions in New York and Boston; Wmodi (Acosta and Camargo, 2018a,b) studied Bogotá-Colombia [4]; SubjectivityClient (Milias, Sharifi Noorian, Bozzon and Psyllidis, 2023), studied textual and emotional perceptions in Frankfurt; City-Wide (Muller et al., 2022), analyzed pedestrian walkability and preference in London [5]; City-Safe (Costa, Soares and Marques, 2019; Costa, 2019), explored perception of safety in Lisbon, Amadora, Cascais identifying the most relevant aspects of the streets and how

[3] http://pulse.media.mit.edu/
[4] http://wmodi.com/
[5] https://emilymuller1991.github.io/urban-perceptions/

to correlate them with safety [6]; SmallCity (Wang et al., 2022c), evaluated perception of safety of the age and sex of different users in Dujiangyan; Aesthetic Capital (Quercia et al., 2014a), quantified the level of Beautiful, Quiet and Happiness in London city [7]; Quali-streets (Ye, Zeng, Shen, Zhang and Lu, 2019), studied preferences about where is better to live in Shanghai; Scenic-or-not (Seresinhe et al., 2017), analyzed the preferences of habitats from UK about relax and safety perception in several landscapes [8]; SenseCityVity (Santani and Gática-Pérez, 2014; Santani et al., 2015; Santani, Ruiz-Correa and Gática-Pérez, 2018), investigated 12 emotional perceptions (e.g., dangerous, dirty, pleasant) in Guanajuato, Leon, Silao - Mexico; ELO Rating system (Liu et al., 2023) explored 6 emotional perceptions in London areas adding a ranking system of SVIs; and Places for play (Kruse, Kang, Liu, Zhang and Gao, 2021), investigated the human perception of playability in Boston, Seattle, and San Francisco.

Collecting human evaluation preferences in online surveys or websites, where users are asked to "choose between two images", is an effective method for gathering subjective judgments in different environments. This approach is commonly used in urban perception studies to assess how people perceive safety, liveliness, or other emotional aspects of a space based on visual cues. This allows researchers to create datasets with relevant data on cities and emotional perceptions at the street level.

Multifactor urban data

Since urban perception involves how individuals or groups perceive, experience, and interpret the physical, social, and environmental characteristics of urban environments, relying solely on SVI and human-based evaluations may not provide a comprehensive analysis. It is mandatory to add demographic factors such as age, gender, ethnicity, and population density; socioeconomic factors such as income, education, employment status, house pricing, and crime rates; Geoinformation Systems (GIS), such as remote sensing data, aerial images, latitude-longitude, elevations, buildings, routes, and addresses; and health-related factors.

Consequently, some studies analyze the relationship between SVI visual appearance and demographic factors (Cui et al., 2023b; Suel, Boulleau, Ezzati and Flaxman, 2018; Yao, Wang, Hong, Qian, Guan, Liang, Dai and Zhang, 2021; Andersson et al., 2017; He et al., 2022; Zhang, Wu, Zhu and Liu, 2019; Jing, Liu, Zhou, Song, Wang, Zhou, Wang and Ma, 2021; Zhu, Gong, Liu, Du, Song, Chen and Pei, 2023); demographic factors and graffiti presence (Tokuda et al., 2019; Lavi., Tokuda., Moreno-Vera., Nonato., Silva. and Poco., 2022; Diniz and Stafford, 2021; Alzate et al., 2021); GIS, SVI and emotional perceptions (Bai et al., 2020; Law et al., 2018a; Larkin, Gu, Chen and Hystad, 2021; Larkin, Krishna, Chen, Amram, Avery, Duncan and Hystad, 2022; Huang, Zhang, Gao, Tu, Duarte, Ratti, Guo and Liu, 2023c); SVI visual appearance and healthy behavior of the population (Dai et al., 2021; Wang, Yuan, Liu, Zhang, Liu, Lu and Yao, 2019b); SVI emotional perception and gender influence (Cui et al., 2023b; Zu, Gao and Wang, 2023; Jiang, Mak, Larsen and Zhong, 2017), mental stress (Suel, Polak, Bennett and Ezzati, 2019; Wang, Liu, Lu, Zhang, Liu, Yao and Grekousis, 2019a); walk and cycling activities (Biswas and Roy, 2023; Ye, Jia and Winter, 2024; Wei et al., 2024; Blečić, Cecchini and Trunfio, 2018; Rui, 2023); and physical injuries (Xu, Liu, Liu, An and Tong, 2023a; Thackway, Ng, Lee and Pettit, 2023); it is also possible to use these multifactor urban data to infer the employment and average income (Gong, Ma, Kan and Qi, 2019; Suel, Bhatt, Brauer, Flaxman and Ezzati, 2021; Suel et al., 2018), playability of the city (Kruse et al., 2021), population density (Buil-Gil, Solymosi and Moretti, 2019), crime rates (Andersson et al., 2017; Liu et al., 2023; He et al., 2022), socioeconomic (Ji, Qing, Han, Wang, Cheng and Peng, 2021; Suel et al., 2023), and house pricing (Arietta, Efros, Ramamoorthi and Agrawala, 2014; Zhang, Wang, Dong, Deng, Fu, Huang, Niu and Chen, 2023; Xu, Qiu, Li, Liu, Zhang, Li and Luo, 2022).

The inclusion of multifactor urban data with SVI and human preferences enhances the understanding of cities on a larger scale. This approach enables detailed spatial analyses that reveal how different groups perceive and interact with urban spaces, and it helps establish connections between urban perception and the specific characteristics of a place or city.

5.2. RQ2: Data processing and manipulation

Urban perception studies with SVI typically use feature extraction techniques. These features can be classified into low-level and high-level (Bengio, 2007). In this paper, we add an urban data-based level named *emotional perception features*. In addition, Yin et al. (2023) and (Kang et al., 2020) discussed the insights and fine-grained characteristics that the environment can provide, such as street elements and scenes. The top and bottom right of Figure 7 presents our classification of the SVI-based features into objective and subjective features (Ewing and Handy, 2009). In addition, we identified the type of feature used in the included studies and grouped them into latent representations, semantic information, and attributes of environmental perception.

Latent representations with low-level features

Low-level features refer to fundamental visual elements or characteristics that are directly extracted from raw pixel data, including *abstract features* and *element features*. Early works in computer vision and image processing performed feature extraction (*abstract features*) using classical methods such as GIST (Oliva and Torralba, 2001), SIFT + Fisher vectors (Perronnin, Sánchez and Mensink, 2010), *Geometric Probability Map* (Hoiem, Efros and Hebert, 2007), *Texton Histograms* (Martin, Fowlkes, Tal and Malik, 2001), *Color Histograms* (Novak, Shafer et al., 1992; Chakravarti and Meng, 2009), *Geometric Color Histograms* (Rao, Srihari and Zhang, 1999), HOG (Dalal and Triggs, 2005), Dense SIFT (Lazebnik, Schmid and Ponce, 2006), LBP (Ojala,

[6] https://smartcity.isr.tecnico.ulisboa.pt/CitySAFE/
[7] https://urbangems.org/
[8] https://scenicornot.datasciencelab.co.uk/

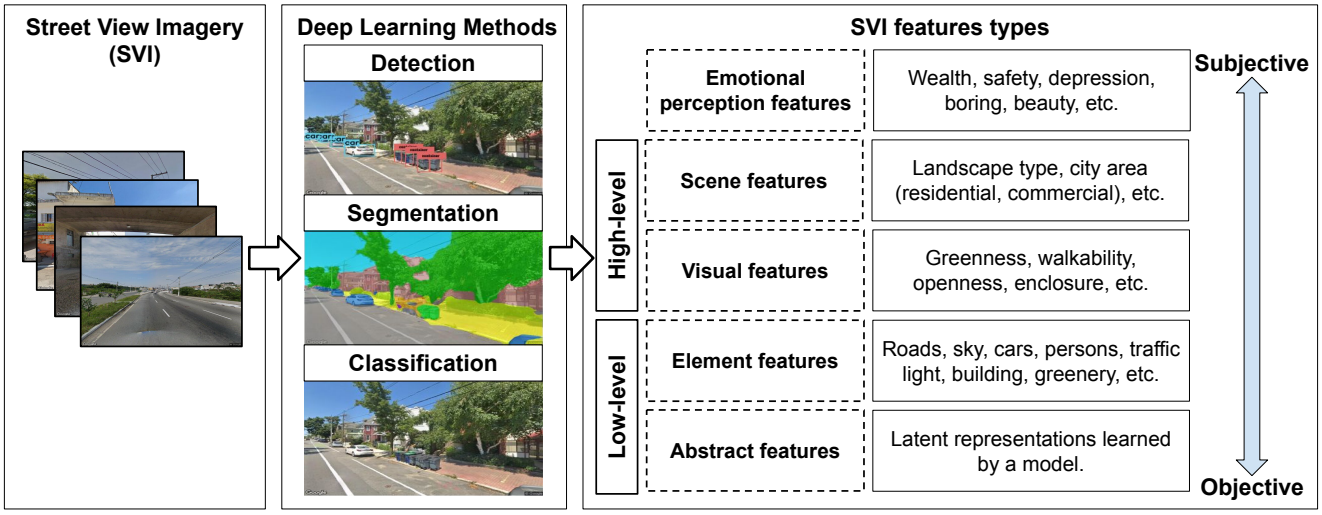

Figure 7: SVI-based features can be categorized into: (a) objective features that refer to quantifiable information and (b) subjective features that refer to personal feelings, and tastes, or opinions. In addition, we classify deep learning outputs into five groups: (i) *emotional perception features* are emotional perceptions of people; (ii) *scene features* reflect people's assessments of street elements or the semantics of street scenes; (iii) *visual features* are the *visual indicators* of streets attributes; (iv) *element features* referring to the basic, raw elements extracted within image based on their pixel value, and (v) *abstract features* are latent representations learned by deep learning methods.

Pietikainen and Maenpaa, 2002), *Sparse SIFT histograms* (Sivic and Zisserman, 2004), and SSIM (Matas, Chum, Urban and Pajdla, 2004). Then, some studies performed features extraction from CNNs such as AlexNet (Beaucamp et al., 2022; Kang and Kang, 2016; Li, Chen, Zheng, Oh and Nguyen, 2021a; Nadai, Vieriu, Zen, Dragicevic, Naik, Caraviello, Hidalgo, Sebe and Lepri, 2016; Dubey et al., 2016; Joglekar et al., 2020; Liu et al., 2017b; Porzi, Rota Bulò, Lepri and Ricci, 2015; Fu, Chen and Lu, 2018), PlacesNet (Santani et al., 2018; Kruse et al., 2021; Zhang, Zu, xiu Hu, Zhu, Kang, Gao, Zhang and Huang, 2020a; Zhang, Hu, Che, Lin and Fang, 2018a; Hu, Xu, Wu, Wu, Wang, Zhang, Lu and Mao, 2021), Xception (Muller et al., 2022; Law et al., 2018b), DenseNet (Zhang et al., 2020a), GoogleNet (Ji et al., 2021; Liu, Guo and Wu, 2022a), or the most used network, ResNet (Costa, 2019; Muller et al., 2022; Zhang, Wu, Zhang, Zhao, Wang, Cui and Yin, 2021a; Sangers, van Gemert and van Cranenburgh, 2022; Kruse et al., 2021; Tokuda et al., 2019).

Recently, several studies have performed extraction of object pixel ratios (*element features*) using segmentation models such as FPN (Liu et al., 2023; Zhang et al., 2020b), FCN (Yao, Liang, Yuan, Liu, Bie, Zhang, Wang, Wang and Guan, 2019; Jing et al., 2021; Wang, Ren, Zhang, Yao, Wang and Guan, 2021; Zhang, Li, Dong, Deng, Fu, Wang, Yu, Jia and zhu Zhao, 2021c; Hua and Yi, 2021), PSPNet (Zhang et al., 2018a,c; Verma, Jana and Ramamritham, 2019; Gong et al., 2019; Min et al., 2020; Li, Liu, Shi and Xing, 2021b; Zhang, Zhang, Liu and Lin, 2018b; Moreno-Vera et al., 2021b; Cai, Li and Ratti, 2019), DeepLabV3 (Muller et al., 2022; Wang, Zeng, Li and Deng, 2022b; Llaguno-Munitxa, Edwards, Grade, Meulen, Letesson, Sierra, Altomonte, Lacroix, Bogosian, Kris and Macagno, 2022; Liu, Ma, Hu, Lu, Ye, You, Tan and Li, 2022b; Liu et al., 2022a; Kumakoshi et al., 2021; Wen et al., 2022), SegNet (Ma, Ma, Wu, Xi, Yang, Peng, Zhang and Ren, 2021; Hu et al., 2021; Joglekar et al., 2020; Hua and Yi, 2021; Ye et al., 2019), SegFormer (Wang et al., 2022b,c), or Mask2Former (Li, Beaucamp, Tourre, Leduc and Servieres, 2023).

After extraction, the correlation between low-level features, demographic and socioeconomic factors, and urban emotional perception can be analyzed using classification or regression models. The most widely used methods are RandomForest (Yao et al., 2019; Ji et al., 2021; Santos, Silva, Loureiro and Villas, 2020; Gong et al., 2019; Liu et al., 2023; Santani et al., 2018, 2015), SVM, SVR, RankSVM, or any other SVM-based method (Kang and Kang, 2016; Li et al., 2023; Ordonez and Berg, 2014; Arietta et al., 2014; Sangers et al., 2022; Zhang et al., 2018c; Porzi et al., 2015; Fu et al., 2018; Ye et al., 2019; Moreno-Vera, 2021; Doersch et al., 2012; Kim, Jeon, Noh and Woo, 2024), or nonparemetric models (Buil-Gil and Solymosi, 2020; Buil-Gil et al., 2019). In addition, these methods investigate the correlation between emotional perception and graffiti presence (Tokuda, César Júnior and Silva, 2018; Tokuda et al., 2019; Lavi. et al., 2022), garbage presence (Patel, Patel, Patel, Patel, Shah and Patel, 2021; Wu, Shen, Liu, Xiao and Li, 2021; Sharma, Keshri, Kumar and Yadav, 2023; Alzate et al., 2021; Cai et al., 2019), tree presence (Grondin, Fortin, Pomerleau and Giguère, 2023; Choi, Lim, Chang, Jeong, Kim, Park and Ko, 2022; Jodas, Yojo, Brazolin, Velasco and Papa, 2022; Xie, Li, Yu, Zhou and Wang, 2020; Velasquez-Camacho, Etxegarai and de Miguel, 2023; Ooi, Valdez, Rogers, Ababou, Zhao and Delmas, 2023), or detection of pavement distress and deterioration (Lei, Liu, Li and Wang, 2020; Gagliardi,

Table 6
The nine most employed *visual indicators*, the definition, and their equation. Here, $VI_{element}$ means *View Index* of the respective *element*; which also refers to the pixel ratio of that *element* in relation to the total pixels within the image.

Indicator	Definition	Equation
Greenness	Urban green-spaces including tree, grass, and greenbelts	$G_i = VI_{vegetation} + VI_{terrain}$
Blueness	Urban blue-spaces including ocean, sea, rivers, lakes, and pools	$B_i = VI_{pool} + VI_{river} + VI_{sea} + VI_{lake}$
Openness	Urban visibility fraction of the street such as sky, and light day.	$O_i = VI_{sky}$
Enclosure	Degree of visual obstructions that encloses pedestrians.	$E_i = \frac{VI_{building}+VI_{wall}+VI_{vegetation}}{VI_{road}+VI_{sidewalk}+VI_{fence}}$
Safety	Guardianship of vulnerable groups and well-functioned protective equipment.	$S_i = VI_{person} + VI_{sign} + VI_{light} + VI_{fence} + VI_{guardrail}$
Traffic flow	Measure vehicles move along roads or streets and pedestrian movement	$T_i = VI_{car} + VI_{person}$
Walkability	Width of the pedestrian walkway in streets, roads, and pavements.	$W_i = \frac{VI_{sidewalk}+VI_{fence}}{VI_{road}}$
Imageability	Measure the street's uniqueness, identifiability, and memorability.	$I_i = VI_{building} + VI_{sign}$
Complexity	Street's complexity based on buildings, landscape, and street furniture	$C_i = \frac{VI_{person}+VI_{sign}+VI_{light}+VI_{vegetation}}{VI_{building}+VI_{road}}$

Giammorcaro, Bella and Sansonetti, 2023; Sarmiento, 2021; Kong, Zhong, Mai, Zhang, Chen and Lv, 2022).

Abstract features are the most objective features and non-legible for humans, extracted from SVI using feature extraction techniques such as classical image processing methods or from CNN layers. However, *Elements features* are road, buildings, trees, and sky, among others. Both features are essential for enabling models to recognize patterns and objects within the urban environment in a more abstract and sophisticated manner.

Semantic information with high-level features

High-level features are sophisticated and complex representations created by combining low-level features, including *scene features* and *visual features*. *Scene features* are related to the scene recognition task to classify landscape and scene attributes such as enclosed area, man-made, or social area; they help to understand and study the scenic aspects of the SVI. The most employed datasets are SUN397 (Xiao, Hays, Ehinger, Oliva and Torralba, 2010; Xiao, Ehinger, Hays, Torralba and Oliva, 2014) [9] and Places365 (Zhou, Lapedriza, Khosla, Oliva and Torralba, 2017a) [10], providing scene attributes divided into indoor scenes, outdoor landscapes, and urban environments. Some studies apply these models to segment a city into regions (e.g. residential and commercial areas) (Zhang et al., 2018a; Hu et al., 2021; Law et al., 2018b; Seresinhe et al., 2017); correlated the SVI scene attributes with socioeconomic factors (e.g. crime, price of houses, rate of roubbery) (Fu et al., 2018; Arietta et al., 2014; Bai et al., 2020; Ji et al., 2021); explored the relationship between emotional perception and different areas of the city (residential, highways, edge of town) (Li et al., 2021a; Nadai et al., 2016; Santos et al., 2020; Dubey et al., 2016; Min et al., 2020; Liu et al., 2017b; Porzi et al., 2015; Liu et al., 2022b); or studied the density of the population, emotional perceptions, and SVI scene attributes (Rios, Ruiz-Correa, Santani and Gática-Pérez, 2021; Kruse et al., 2021; Zhang et al., 2020a).

Moreover, *Visual features*, also called *visual indicators*, are specific visual cues that are used to represent environmental or spatial attributes such as greenness, openness, walkability, complexity, blueness, and enclosure. These *visual indicators* are calculated using the *element features*.

[9] https://vision.princeton.edu/projects/2010/SUN/

[10] http://places2.csail.mit.edu/demo.html

Table 6 summarizes the most used visual features in urban perception studies, definitions, and their respective formulation, we define the view of index (VI) of a specific *element* as $VI_{element}$, which denotes the proportion of pixels of that *element* within the SVI (e.g., VI_{fence} is the proportion of pixel corresponding to fence). These *visual features* provide a better understanding of street characteristics that allows investigation of their relationship with emotional perception (Muller et al., 2022; Dai et al., 2021; Ma, 2023; Tian, Han, Xu, Liu, Qiu and Li, 2021; Thackway et al., 2023; Hou and Chen, 2024; Lei, Zhou, Xue, Yuan, Liu, Wang and Wang, 2024), demographic and socioeconomic factors (Lavi. et al., 2022; Law et al., 2018b; He et al., 2022; Andersson et al., 2017), or mental health (e.g., stress, depression, and fear) (Wang et al., 2019b,a; Suel et al., 2019; Kang, Kim, Park and Lee, 2023b).

High-level features, including both *scene features* and *visual features*, are the semantic information obtained from SVI through deep learning models. It can be scene recognition to identify and categorize the landscape or city area, or object detection and segmentation to identify objects pixel ratios within the SVI and calculate *visual indicators*. Both methods enable researchers to understand and segregate the city areas by type or by object presence.

Environmental perception features

Emotional perception features are the most subjective aspects extracted from SVI. These features help to analyze how people understand, interpret, and mentally perceive SVI appearances. Several works focused on analyzing the emotional perception of online volunteers through online surveys using ranting or comparison method. Figure 8 shows the two main emotional perception collection approaches.

Moreover, after the volunteers completed their tasks (choosing or rating), there are two ways to calculate the emotional score from image comparisons: (i) *TrueSkill* (Minka, Cleven and Zaykov, 2018) to convert volunteers' preferences into emotional ranking scores; (ii) *Strength of Schedule* (Park and Newman, 2005) and the psychological scale (Nasar, 1998) to convert preferences into emotional perceptual scores. In addition, these emotional scores can be labeled by setting a threshold (see Equation 1), where $y_{i,k}$ corresponds to the emotional label, $q_{i,k}$ is the emotional score , i is the image$_i$, δ is a number, and k is the emotional perception.

$$y_{i,k} = \begin{cases} 1 & \text{if } (q_{i,k}) \text{ in the top } \delta\% \\ \text{-1} & \text{if } (q_{i,k}) \text{ in the bottom } \delta\% \end{cases} \quad (1)$$

Urban studies usually use human evaluation preferences to quantify the perception of cities (Moreno-Vera et al., 2021b; Buil-Gil and Solymosi, 2020; Li et al., 2021a; Muller et al., 2022; Porzi et al., 2015; Ordonez and Berg, 2014), determine what visual aspects within SVI are relevant to some perception (Zhang et al., 2018c; Min et al., 2020; Li et al., 2021b; Beaucamp et al., 2022; Kang and Kang, 2016; Yao et al., 2019; Ji et al., 2021; Zhang et al., 2021a; Li et al., 2015a, 2023; Sangers et al., 2022; Guan, Chen, Feng, Liu and Nie, 2021; Xu, Yang, Cui, Shi, Song, Han and Yin, 2019), correlate the visual appearance of SVI with demographic and socioeconomic factors (Buil-Gil et al., 2019; Liu et al., 2023; Nadai et al., 2016; Naik et al., 2016; Liu et al., 2017b; Zhang et al., 2021c; Kruse et al., 2021), and correlate SVI visual appearance with social media and blogs (Santos et al., 2020; Zhang et al., 2020a; Liu et al., 2022a; Santos, Silva, Loureiro and Villas, 2018; Santos, Silva and Villas, 2024).

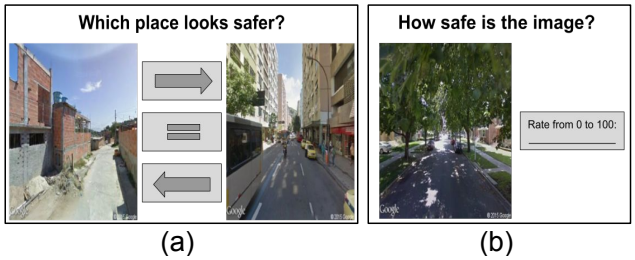

Figure 8: There are two main collection approaches for emotional perceptions: (a) gather volunteer information through surveys, interviews, or questionnaires asking e.g., *Which place looks safer?*, which volunteers choose between two images; or (b) ask for volunteers to rate an image from 0 to 100 asking *How safe is the image?*.

Emotional perception features are typically collected through online surveys aimed at assessing the emotional perception of a city's streets. This information is crucial for urban planners and policymakers to quantify and map emotional perceptions within a city.

5.3. RQ3: Data Analysis and visualization

Here, we highlight how the data is analyzed after preprocessing and what visualization techniques were used. We identify three categories based on urban studies: (i) urban perception mapping, (ii) place characterization, and (iii) visual explanations. They are detailed in the following sections.

Urban perception mapping

Most of the studies included in this review focused on analyzing *emotional perception features* and linking them to multifactor urban data and population preferences for activities like walking, sightseeing, and cycling. A few studies performed a similar methodology doing a custom SVI collection process, extracting features using segmentation models, train classification and regression models, and analyzing results in a specific cities, such as Bogotá (Acosta and Camargo, 2018b), Gaoxin (Wang et al., 2021), Dujiangyan (Wang et al., 2022b,c; Tang, Zhang, Chen, Wan and Li, 2020), Wuhan (Hu et al., 2021; Yao et al., 2019; Liu et al., 2022a), Beijing (Ji et al., 2021; Gong et al., 2019; Zhang et al., 2020b,a), Hong Kong (Ma and Wu, 2023), Singapore (Chen and Biljecki, 2023), Guangzhou (Jing et al., 2021), London (Law, Shen and Seresinhe, 2017; Quercia et al., 2014a; Muller et al., 2022; Law et al., 2018b; Santos et al., 2018; Seresinhe et al., 2017; Cai et al., 2019; Law et al.,

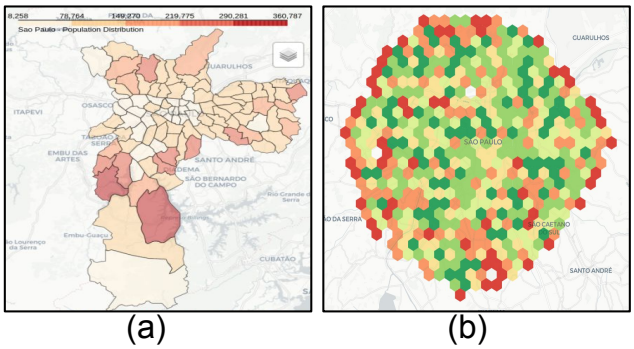

Figure 9: Two main map-based visualization charts using the population distribution and emotional perception in Sao Paulo city as example: (a) choropleth map, which uses color shading to represent data values focused on demographics and economics visualizing how these variables change across geographic areas; and (b) heat map using hexagonal bins that shows data density or intensity using color gradients, often used to represent the spatial distribution of emotional perceptions or object presences in different areas.

2018a; Suel et al., 2021; Rita, Peliteiro, Bostan, Tamagusko and Ferreira, 2023), Rome (Nadai et al., 2016), New York (Li et al., 2015a,b; Bai et al., 2020; Miranda, Hosseini, Lage, Doraiswamy, Dove and Silva, 2020), San Francisco (Arietta et al., 2014), Xiaoying (Liu et al., 2022b), Sweileh (Alatta and Momani, 2021), Mumbai (Verma et al., 2019), Shanghai (Zhang et al., 2018c; Ye et al., 2019), Shenzen (Ma et al., 2021), Brussels (Llaguno-Munitxa et al., 2022), Sao Paulo (Tokuda et al., 2018; Lavi. et al., 2022), Yau Tsim Mong (Zhang et al., 2018b), Washington (Fu et al., 2018; Suel et al., 2023), Gualajara (Izquierdo and leticiai, 2020), Guanajuato (Santani and Gática-Pérez, 2014; Santani et al., 2015), Paris (Doersch et al., 2012), Amsterdam (Liang, Chang, Gao, Zhao and Biljecki, 2024; Alpherts, Ghebreab, Hsu and Noord, 2024), Tokyo (Kumakoshi et al., 2021), Xiamen (Wen et al., 2022), New Delhi (Sengupta, Vaidya and Evans, 2023), Santiago (Cox, Rossetti and Hurtubia, 2019; Rossetti et al., 2019a; Ramírez, Hurtubia, Lobel, Lobel and Rossetti, 2021), Los Angeles (Zhao, Liu, Kuang, Chen and Yang, 2018), and Lisbon (Costa, 2019).

However, other studies focused on improving previous results in the Place Pulse dataset using comparison-based models such as Siamese networks (Andersson et al., 2017; Beaucamp et al., 2022; Li et al., 2021b,a; Zhang et al., 2021a; Li et al., 2023; Min et al., 2020; Liu et al., 2022a; Guan et al., 2021; Bai et al., 2020; Dubey et al., 2016; Xu et al., 2019; Ogawa, Oki, Zhao, Sekimoto and Shimizu, 2024), performing regression or classification to infer emotional perception with ensemble models (Zhao, Luo, Li, Xu, Zhu, He and Li, 2021; Ordonez and Berg, 2014; Wei, Yue, Li and Gao, 2022; Santos et al., 2020; Muller et al., 2022; Li et al., 2015a; Joglekar et al., 2020; Zhao, Lu and Lin, 2024; Shi and Hao, 2023; Shi, Yan, Li and Zhou, 2024; Qiu, Li, Zhang, Li, Liu and Huang, 2021; Huang, Oki, Muto and Ogawa, 2024a), or mapping objects present within the streets and correlate them with demographic and socioeconomic factors with multilinear models (Moreno-Vera et al., 2021a; Zhang et al., 2018c; Min et al., 2020; Li et al., 2021b; Beaucamp et al., 2022; Guan et al., 2021; Buil-Gil et al., 2019; Moreno-Vera et al., 2021a).

Both approaches aim to quantify human preferences and map emotional perception by using emotional perceptual scores, generating map-based visualizations. Figure 9 shows the most popular visualizations: (a) choropleth map shows the percentage of a variable of interest by using varying shading patterns within geographic boundaries; and (b) heat map that does not correspond to geographic boundaries.

This analysis aims to understand the spatial distribution of urban emotional perception within street images obtained from online surveys. These insights were obtained by training or fine-tuning a model to infer perceptual scores and emotional perceptions or determine the winner between two images. These map-based visualization tools help urban researchers summarize information across different regions, supporting decision-making and policymakers.

Place characterization

We identify studies focused on analyzing the uniqueness of the city; this analysis involves identifying and determining the specific characteristics and identity of places. We categorize the compiled works into two types based on the analysis performed: (a) *place recognition* refers to the unique identity of a particular geographic location based on its physical characteristics, and (b) *place identification* refers to the localization of a place based on demographic, socioeconomic, or specific city factors such as cultural places.

To recognize places, some studies develop custom pipelines to obtain image characteristics and cluster them or capture human preferences through online surveys. Notable studies are *What makes Paris look Paris?* (Doersch et al., 2012), *what makes London beautiful and quiet* (Quercia et al., 2014b,a), and *what makes London scenic or not?* (Seresinhe et al., 2017). Other approaches not directly related to urban perception developed frameworks to identify uniqueness in street-level images (Miranda et al., 2020; Law et al., 2018b; Shen, Zeng, Ye, Arisona, Schubiger, Burkhard and Qu, 2017), performed clustering techniques to group similar streets (Izquierdo and leticiai, 2020; Liu et al., 2017b, 2022b; Li et al., 2021a; Hu et al., 2021; Ma and Wu, 2023; Yao et al., 2021; Lee et al., 2021b; Zhanga, Lia, Fukudab and Wang, 2024; Li, 2024; Gao, Guo, Liu, Zeng, Liu, Liu and Xing, 2024), or extracted emotional perception features to describe the visual attractiveness of a street (Yao et al., 2019; Ji et al., 2021; Zhang et al., 2021a; Li et al., 2015a, 2023; Sangers et al., 2022; Min et al., 2020; Zhang, Wang, Hu, Zhang and Liu, 2024b).

Moreover, to identify places, several studies correlated the *visual indicators* and map locations (Jing et al., 2021; Naik et al., 2016; Gong et al., 2019; Li et al., 2015a; Kruse et al., 2021; Law et al., 2018b; Suel et al., 2023; Sengupta et al., 2023; Gong, Huang, White and Langenheim, 2023); analyzed the relationship between *scene features* and crime

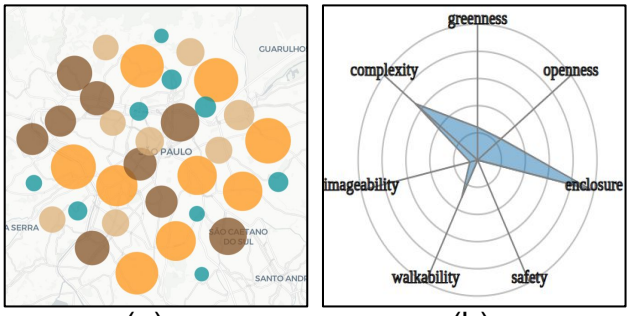

Figure 10: Two main visualization charts using clusters of similar street facades or appearance and the visual indicators of streets in different intensities. (a) Bubble map, this map uses clusters of points to group instances of similar street facades or appearance. Each cluster represents a group of streets with similar characteristics. (b) Spider or Radar chart displays multivariate data, allowing a rapid visual comparison of several variables (e.g., emotional perceptions, *visual indicators*, landscape type). Each axis represents one of the variables, and the length of the axis corresponds to the value of the variable.

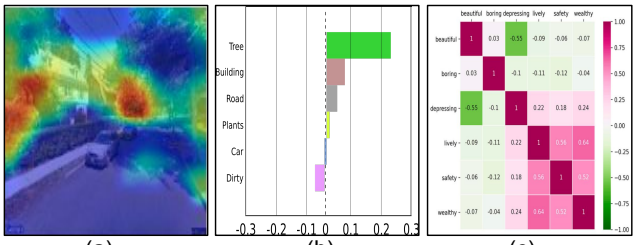

Figure 11: Three main visual methods to explain CNN-based models, linear-based models, and multifactor urban data. (a) CAM and grad-CAM provide visual explanations by highlighting the regions in an input image that contribute most to a model's decision, particularly in convolutional neural networks (CNNs). (b) LIME and SHAP assign importance scores to individual features, indicating how each feature influences the final prediction. This helps users understand which features are most relevant for the model's decision-making process. (c) correlation matrix displays the pairwise correlation coefficients between variables in a dataset. It provides a compact view of the relationships between variables, helping to identify patterns, dependencies, and potential multicollinearity.

rates (Cui et al., 2023b; Zhao and Guo, 2022; Yao et al., 2021; He et al., 2022; Zhang et al., 2019; Jing et al., 2021; Lavi. et al., 2022; Wang et al., 2019a,b), house prices (Suel et al., 2018, 2019; Buil-Gil et al., 2019; Nadai et al., 2016), and robbery (Ji et al., 2021; Liu et al., 2023; Alatta and Momani, 2021; Liu et al., 2017b; Fu et al., 2018; Law et al., 2018a; Suel et al., 2021); and developed siamese network-based frameworks to compare and extract unique characteristics of the image (Andersson et al., 2017; Beaucamp et al., 2022; Li et al., 2021b,a; Zhang et al., 2021a; Li et al., 2023; Min et al., 2020; Liu et al., 2022a; Guan et al., 2021; Bai et al., 2020; Dubey et al., 2016; Xu et al., 2019; Ogawa et al., 2024). Figure 10 shows the most commonly used visualizations for this type of analysis: (a) bubble maps to show cluster of similar images plotted on a map, and (b) spider or radar chart to show values in different dimensions such as emotional perceptions or *visual indicators*.

Place characterization allows researchers to explore a city's identity by examining distinct visual characteristics linked to demographics, emotional perceptions, and visual features over time, thereby assessing urban transformations. These studies analyze elements such as street aesthetics and temporal changes, highlighting key attributes that distinguish one location from another and providing insights into the identity of urban environments.

Visual explanations

Visual explanations, crucial for model interpretability and explainability, bridge the gap between the complex workings of machine learning models and human understanding. In urban perception, machine learning increasingly analyzes SVI and multifactor urban data (e.g., demographics, socioeconomic factors, GIS, health) to understand how people perceive different areas. We identified three types of visualizations to aid in this understanding. From this, we identified two main explanation methods: Image-based explanation methods such as LIME (Ribeiro, Singh and Guestrin, 2016), CAM (Zhou, Lapedriza, Xiao, Torralba and Oliva, 2014), and Grad-CAM (Selvaraju, Cogswell, Das, Vedantam, Parikh and Batra, 2017); and Linear approximation-based explanation methods such as LIME (Ribeiro et al., 2016) and SHAP (Lundberg and Lee, 2017).

Image-based explanation techniques interpret how deep learning models make predictions from visual data, offering transparency and insights into which specific features of an image influence the model's output. In this context, CAM was used to understand the relevance regions to classify scenes (Verma, Jana and Ramamritham, 2020; Zhang et al., 2019; Zhao et al., 2021; Zhang et al., 2019; Zhao et al., 2024). In addition, Grad-CAM was used to highlight regions for specific emotional perceptions predictions (Li et al., 2021b; Zhang et al., 2021a; Xu et al., 2019), or to visualize CNN activation maps (Min et al., 2020; Sangers et al., 2022; Law et al., 2018b; Cai et al., 2019).

Linear approximation-based explanation methods (e.g., LIME and SHAP) were used to analyze the relevance of elements (or feature importance) such as trees, fences, or crosswalks to influence perceived safety. Common visualization is a bar chart showing how their removal could reduce safety perceptions (Moreno-Vera et al., 2021b), examine the impact of demographic and socioeconomic factors on emotional perception inferences (Ma, 2023; Zhao et al., 2024), or investigate the correlation between objects presences and emotional perception features (Moreno-Vera, 2021; Kim and Lee, 2023; Hao et al., 2024; Hou and Chen, 2024; Zhanga et al., 2024; Xu et al., 2023b; Yu, Ma, Tang, Yang and Jiang, 2024). We also consider a correlation matrix not as an explanation method, but as a visualization method to observe

variable dependence. Figure 11 illustrates the most frequent methods for visualizing information: (a) CAM-based heat map highlighting key prediction regions, (b) a bar chart showing feature importance from LIME and SHAP applied to linear and ensemble models, and (c) a correlation matrix revealing variable dependencies.

Visual explanations of models are vital for urban perception researchers using machine learning. They reveal how models interpret complex urban data, making decision-making more informed and transparent. By visualizing these insights, urban planners and designers can identify key urban features that shape human perceptions, leading to data-driven improvements in urban environments.

5.4. RQ4: Applications

For several decades, researchers have focused on exploring the intricate interplay between urban perception with the built environment and various health outcomes. In addition, several studies conducted an in-depth analysis of the appearance of SVI and the mental well-being of residents, uncovering connections between the visual characteristics of urban spaces and the physical and psychological health of individuals. These studies emphasize how aspects like greenery, street aesthetics, and urban density influence mental well-being, shaping not only the experience of daily life but also long-term health outcomes.

Urban physical studies

Physical-related applications involve analyzing how the physical environment shapes urban perception. This includes assessing variables like green spaces, street walkability, enclosure, openness, and proximity to health-promoting amenities. We identified two main applications: (i) urban visual appearance, focusing on structural conditions, city aesthetics, and cultural identity, and (ii) urban infrastructure, covering transportation, mobility, walkability, and urban planning. Studies on physical visual appearance and urban perception investigate how city environments affect individuals. Research shows that visual cues, such as order or disorder, influence perceptions of chaos, risk of injury, and crime rates (Keizer et al., 2008; Wilson and Kelling, 1982; Diniz and Stafford, 2021). SVI provides insights into environmental factors such as street illumination, parks, and overall well-being (Gong et al., 2019; Hu, Zhang, Gong, Ratti and Li, 2020; Yu et al., 2024); allowing a researcher to correlate the built environment and emotional perceptions (Zhang et al., 2020b; Liu, Long, Zhang, Yang and Dong, 2024b; Meng, Sun, Lyu, Niu and Fukuda, 2024), explore street qualities and crime rates (Fu et al., 2018; Su, Li and Qiu, 2023; Zhang, Fan, Kang, Hu and Ratti, 2021b), investigate impact on socioeconomic factors (Naik et al., 2016; Freitas, Berreth, Chen and Jhala, 2023; Kim, Lee, Hipp and Ki, 2021), graffiti presence and human development indexes (Tokuda et al., 2018, 2019; Lavi. et al., 2022; Alzate et al., 2021), identify similar places with similar conditions (Zhang et al., 2021c; Hu et al., 2021), and evaluate the spatial distribution of relevant objects mapping correlations with urban perception (Tang et al., 2020; Yao et al., 2019; Wang et al., 2021; Hua and Yi, 2021; Verma et al., 2019; Li et al., 2021b; Zhang et al., 2018b).

The urban physical infrastructure approach examines the relationship between urban perception and city infrastructure, including transportation, mobility, walkability, cycling, and urban planning. These factors influence perception of city quality and can cause stress for pedestrians (Liu et al., 2017a; Gong et al., 2023; Rui and Cheng, 2023; Li, Xin, Xi, Tarkoma, Hui and Li, 2022a), study pedestrian preferences (Zhang et al., 2019; Alatta and Momani, 2021; Llaguno-Munitxa et al., 2022; Biswas and Roy, 2023), explore urban activities such as walking and cycling (Ma, 2023; Wu, Ye, Gao and Ye, 2022; Huang, Yu, Lyu, Sun, Zeng and Bart, 2023a; Li, Yabuki and Fukuda, 2022c; Rita et al., 2023; Rossetti, Saud and Hurtubia, 2019b; Rossetti, Guevara, Galilea and Hurtubia, 2018), and analyze urban gentrification and change (Rossetti et al., 2019a; Zhou, Wang and Wilson, 2022; He, Zhang, Yao and Li, 2023; Tian et al., 2021; Ilic, Sawada, Zarzelli and Zarzelli, 2019; Verma et al., 2020; Wang, Ito and Biljecki, 2024; Naik, Kominers, Raskar, Glaeser and Hidalgo, 2017).

As a consequence, SVI allows researchers, urban planners, and designers to explore how the physical characteristics of a city impact people's perceptions, focusing on how transportation, cultural diversity, population activities, and urban activities shape urban perception.

Psychological impact studies

Psychological-related applications study how urban landscapes affect mental health and well-being. This includes analyzing SVI visual representations, built environment, mental disorders, urban vitality, safety perception, and social connectivity. We categorized two main applications: (i) emotional perception, which focuses on sentiments about streets (e.g., safety, relaxation); and (ii) mental well-being and health, which addresses mental issues, crime fear, and stress. Emotional and sentimental perceptions of streets are closely related to urban behavior and overall perception (Lindal and Hartig, 2013; Ulrich, 1979). Which indicates that urban disorder can cause psychological distress and increased stress (Sampson et al., 2002; Han, Wang, Seo, He and Jung, 2022). Poor street conditions, such as graffiti and litter, can exacerbate feelings of insecurity or fear of walking alone (Buil-Gil and Solymosi, 2020; Buil-Gil et al., 2019). Besides, SVI provides valuable information on psychologically significant urban characteristics (Zhang et al., 2018c; Naik et al., 2014; Santani and Gática-Pérez, 2014; Santani et al., 2018). In addition, through online surveys, it is possible to analyze the relationship of urban perception and age, gender, ethnic, and nationality (Cui, Gong, Yang, Zhang, Huang, Shen, Wei and Chen, 2023a; Izquierdo and leticiai, 2020; Izquierdo, Palomo, Grignard, Alonso, Siller and Larson, 2021; Sengupta et al., 2023; Gong et al., 2023; Jiang et al., 2017; Cui et al., 2023b; Zu et al., 2023; Liu, Yu and Yang, 2024a). Furthermore, *visual indicators* such as walkability, openness, safety, imageability, among others;

allows researchers and urban planners to analyze the relationship between psychological impact of street appearance (Ma et al., 2021; Li, 2024; Dai et al., 2021; Wang, Han, He and Jung, 2022a).

Mental health and well-being on the streets have been explored since the *The Broken Window Theory* (Wilson and Kelling, 1982). Early studies used in-person surveys to assess mental health and well-being (Lynch, 1984), studying the presence or absence of green spaces and parks (Nasar, 1998; Skogan, 1992). Recent studies use SVI to analyze urban visual appearance and their impact on well-being, studying factors such as population density (too crowded cities) (Santani et al., 2015, 2018; Zhang et al., 2019; van Veghel, Dane, Agugiaro and Borgers, 2024), poverty (Suel et al., 2023, 2019), obesity and cardiovascular problems (Huang et al., 2023a; Hu et al., 2021; Xu et al., 2023b), or distress by cycling activity (Rita et al., 2023; Ma et al., 2021).

Psychology is essential to understanding the interactions between people and urban environments. The use of SVI allows researchers to obtain information that helps design cities that promote well-being, social connections, and inclusive experiences.

6. Discussion

In this section, we outline the challenges and highlight emerging trends in urban perception data research.

6.1. Challenges

Most of the studies included in this survey have focused on using similar methodologies aimed at using pre-trained deep-learning models and making predictions. Despite this, to our knowledge, only a few studies have focused on improving data preprocessing, but data-driven challenges persist.

Data quality and quantity

SVI data often suffer from limitations in coverage and image quality, varying between locations (Marasinghe et al., 2023; Zhou et al., 2022). Processing large volumes of SVI data requires significant computational resources (Kim and Lee, 2023; Wang et al., 2019a; Wei et al., 2024), efficient algorithms, and expertise, which pose challenges to researchers. For example, high-resolution image-based models are essential, and advanced approaches such as Vision Transformer (ViT) models offer a promising solution to this challenge by utilizing patch-based techniques to efficiently process detailed images (Dosovitskiy, Beyer, Kolesnikov, Weissenborn, Zhai, Unterthiner, Dehghani, Minderer, Heigold, Gelly, Uszkoreit and Houlsby, 2020; Huang, Qing, Han, Liao, Guo and Peng, 2023b).

Data integration and standarization

SVI data can be collected from various sources, including GSV, BSV, TSV, NSV, Mapillary, KartaView, Bing maps, and self-captured images. These images can be taken from different perspectives, such as flat on a spherical surface, panoramic views, frontal perspective of a person, or various directional angles (Hou, Quintana, Khomiakov, Yap, Ouyang, Ito, Wang, Zhao and Biljecki, 2024; Li et al., 2022b; Dai et al., 2024; Kim et al., 2021). This lack of standardization poses a challenge for urban studies (Liu and Sevtsuk, 2024; Rundle, Bader, Richards, Neckerman and Teitler, 2011). In addition, SVI combined with aerial imagery improves urban perception analysis but introduces challenges related to data integration, cleanliness, interoperability, preparation, and privacy concerns (Son, Weedon, Yigitcanlar, Sanchez, Corchado and Mehmood, 2023; Bai et al., 2020). Although SVI provides detailed street-level insight into city infrastructure, aerial imagery offers broader spatial coverage. Integrating these sources with additional demographic and socioeconomic datasets (e.g., crime rates, house prices, employment) also faces obstacles such as data alignment, resolution differences, and temporal inconsistencies, which require advanced processing techniques (Biljecki and Ito, 2021; Law et al., 2018a).

Diversity in individual perceptions

Individual perceptions present significant challenges due to the subjective nature of perception, which can vary depending on personal experiences, citizenship, nationality, age, gender, cultural background, and socioeconomic factors (Moreno-Vera et al., 2021a; Cui et al., 2023b). Urban features such as safety, aesthetics, and usability are interpreted differently among diverse populations, making it difficult to capture a unified understanding of urban perception (Quercia et al., 2014b; Izquierdo and leticiai, 2020). To mitigate this particular challenge, data enrichment is necessary, including metadata such as age, gender, nationality, or current city.

Ethical and social considerations

Ethical and social considerations are paramount in the study of urban perception and the application of machine learning. These technologies often rely on large datasets, such as SVI, which raises privacy concerns, particularly when capturing sensitive or identifying information without consent (Suel et al., 2019). There is also the risk of algorithmic bias, where machine learning models, trained in non-representative distribution data, may reinforce existing social inequalities, leading to unfair or inaccurate representations of urban areas (Kang et al., 2023a). Collaboration between technologists, urban planners, and policymakers is essential to address these ethical dilemmas and to ensure that the benefits of machine learning in urban perception are widely shared without harming specific communities.

6.2. Future trends and recommendations

The field of urban perception, focused on how people perceive and interact with urban environments, is evolving due to technological advancements, particularly the integration of SVI and other cutting-edge tools. While new methods have emerged to outperform previous works and tasks, selecting the right objectives, models, and strategies for each scenario remains a challenge.

IoT, Mobile Sensing, and Crowd-Sourced Data

The widespread use of high-resolution pictures captured by smartphones, vehicles equipped with cameras, and other sensors has facilitated the collection of large-scale imagery, city sounds, real-time data on human mobility, activity patterns, and subjective experiences in urban analysis (Zhao, Liang, Tu, Huang and Biljecki, 2023). This data enables researchers to analyze the physical characteristics of streets, buildings, noise, and public spaces with unprecedented accuracy, leading to deeper insights into how these factors influence human behavior and perception (Zhuang, Kang, Fei, Bian and Du, 2024).

Virtual Reality (VR) and Augmented Reality (AR)

VR and AR are revolutionizing the way people experience and perceive urban environments (Rui and Cheng, 2023; Rui and Li, 2023). These technologies offer novel ways to represent cities, allowing researchers to explore and interact with urban spaces (Li et al., 2022c; van Veghel et al., 2024; Yang, Deng, Hu, Guan, Chao and Wan, 2024b). Mixed Reality (MR) is a blend of VR and AR, typically involving the user in physically interacting within controlled and customizable environments through an eye-level device. Such devices at the eye level enable the study of tracking eye movement for the determination of the relevant characteristics of these urban scenarios (Llaguno-Munitxa et al., 2022; Zhang et al., 2018a; Yang, Deng, Hu, Chao, Wan, Guan and Wei, 2024a).

Temporal Analysis

Temporal analysis in urban perception using SVI provides a powerful approach to understanding how urban environments evolve and how these changes affect public perception over time (Huang, Wu, Wu, Hwang and Rajagopal, 2024b). Temporal SVI images allow researchers to track the transformation of neighborhoods, monitor construction developments, and observe changes in land use, traffic patterns, and pedestrian activity (Stalder, Volpi, Büttner, Law, Harttgen and Suel, 2023; Kang and Kang, 2015). With temporal SVI data, urban planners and policymakers can make informed decisions that account for the evolving needs and behaviors of urban populations, ensuring that the built environment adapts to the changing demands of its citizens.

Transfer learning and multimodal models

Advances in machine learning and computer vision have opened new possibilities to analyze and interpret SVI on a scale, allowing automated detection and classification of urban characteristics such as sky, buildings, facades, green spaces, and architectural styles (He et al., 2022). However, developing these models is costly, especially when handling large datasets and complex urban environments (Zhanga et al., 2024; Malekzadeh, Willberg, Torkko and Toivonen, 2024). A fast solution is to adopt transfer learning; this technique learns from these large datasets such as ImageNet (Deng, Dong, Socher, Li, Li and Fei-Fei, 2009) or ADE20K (Zhou, Zhao, Puig, Fidler, Barriuso and Torralba, 2017b) and improves performance, making it easier to adapt to specific urban contexts (Moreno-Vera, 2021; Wang et al., 2022b). Furthermore, integration with pre-trained multimodal models, such as GPT, offers new perspectives in the identification of relevant features of the SVI, enhancing the robustness and generalization across diverse urban settings (Manvi, Khanna, Mai, Burke, Lobell and Ermon, 2023; Li, Xia, Tang, Xu, Shi, Xia, Yin and Huang, 2024; Huang et al., 2024c). Future efforts should focus on fine-tuning pre-trained models for SVI to further optimize them for urban perception tasks.

Looking forward, the convergence of these challenges and future trends is set to fuel innovation in urban perception research. Leveraging advanced technologies and interdisciplinary approaches, researchers can deepen their understanding of human-urban interactions, contributing to the creation of more livable, sustainable, and inclusive cities.

7. Conclusion

In this review, we performed a detailed analysis of 207 papers, defining a taxonomy that categorizes how urban data are utilized, the models and algorithms applied, the methods for examining and quantifying urban data, and the main applications in urban perception studies. This taxonomy captures the essence of data-driven methodologies while providing a well-organized systematic review that guides stakeholders from data acquisition through analysis to practical applications.

This paper contributes to the field of urban perception through three key takeaways. It highlights the use of diverse attributes extracted from street view imagery (SVI), complemented by other data sources, for a more robust urban perception analysis. Additionally, our review provides essential insights and methodologies for integrating street view imagery, AI, and supplementary data, emphasizing their combined value for a more comprehensive understanding of urban environments. It also presents trends and recommendations for future research, addressing challenges such as data quality, eye movement analysis, temporal analysis, and the potential of large language models, including multimodal models.

References

Acosta, S.F., Camargo, J.E., 2018a. City safety perception model based on visual content of street images. 2018 IEEE International Smart Cities Conference (ISC2) , 1–8URL: https://api.semanticscholar.org/CorpusID:71150568.

Acosta, S.F., Camargo, J.E., 2018b. Predicting city safety perception based on visual image content. ArXiv abs/1902.06871. URL: https://api.semanticscholar.org/CorpusID:67749786.

Alatta, R.T.A., Momani, H.M., 2021. Integrating 3d game engines in enhancing urban perception: A case study of students' visualization of urban space. ACE: Architecture, City and Environment URL: https://api.semanticscholar.org/CorpusID:245322303.

Alpherts, T., Ghebreab, S., Hsu, Y.C., Noord, N.V., 2024. Perceptive visual urban analytics is not (yet) suitable for municipalities. Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency URL: https://api.semanticscholar.org/CorpusID:270286850.

Alzate, J.R., Tabares, M.S., Vallejo, P., 2021. Graffiti and government in smart cities: a deep learning approach applied to medellin city,

colombia, in: International Conference on Data Science, E-learning and Information Systems 2021, pp. 160–165.

Andersson, V.O., Birck, M.A., Araujo, R.M., 2017. Investigating crime rate prediction using street-level images and siamese convolutional neural networks, in: Latin American Workshop on Computational Neuroscience, Springer. pp. 81–93.

Arietta, S.M., Efros, A.A., Ramamoorthi, R., Agrawala, M., 2014. City forensics: Using visual elements to predict non-visual city attributes. IEEE transactions on visualization and computer graphics 20, 2624–2633.

Bai, R., Lam, J.C.K., Li, V.O.K., 2020. Siamese-like convolutional neural network for fine-grained income estimation of developed economies. IEEE Access 8, 162533–162547. URL: https://api.semanticscholar.org/CorpusID:221847628.

Beaucamp, B., Leduc, T., Tourre, V., Servieres, M.C.J., 2022. The whole is other than the sum of its parts: Sensibility analysis of 360° urban image splitting. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences URL: https://api.semanticscholar.org/CorpusID:248896063.

Bengio, Y., 2007. Learning deep architectures for ai. Found. Trends Mach. Learn. 2, 1–127. URL: https://api.semanticscholar.org/CorpusID:207178999.

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

Biswas, G., Roy, T.K., 2023. Measuring objective walkability from pedestrian-level visual perception using machine learning and gsv in khulna, bangladesh. Geomatics and Environmental Engineering .

Blečić, I., Cecchini, A., Trunfio, G.A., 2018. Towards automatic assessment of perceived walkability, in: Computational Science and Its Applications–ICCSA 2018: 18th International Conference, Melbourne, VIC, Australia, July 2–5, 2018, Proceedings, Part III 18, Springer. pp. 351–365.

Buil-Gil, D., Solymosi, R., 2020. Using crowdsourced data to study crime and place. SocArXiv URL: https://api.semanticscholar.org/CorpusID:226747992.

Buil-Gil, D., Solymosi, R., Moretti, A., 2019. Nonparametric bootstrap and small area estimation to mitigate bias in crowdsourced data. Big Data Meets Survey Science URL: https://api.semanticscholar.org/CorpusID:225045498.

Cai, B.Y., Li, X., Ratti, C., 2019. Quantifying urban canopy cover with deep convolutional neural networks. ArXiv abs/1912.02109. URL: https://api.semanticscholar.org/CorpusID:208617487.

Chakravarti, R., Meng, X., 2009. A study of color histogram based image retrieval, pp. 1323 – 1328. doi:10.1109/ITNG.2009.126.

Charreire, H., Mackenbach, J.D., Ouasti, M., Lakerveld, J., Compernolle, S., Ben-Rebah, M., McKee, M., Brug, J., Rutter, H., Oppert, J.M., 2014. Using remote sensing to define environmental characteristics related to physical activity and dietary behaviours: a systematic review (the spotlight project). Health & place 25, 1–9.

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

Choi, K., Lim, W., Chang, B., Jeong, J., Kim, I., Park, C.R., Ko, D.W., 2022. An automatic approach for tree species detection and profile estimation of urban street trees using deep learning and google street view images. Isprs Journal of Photogrammetry and Remote Sensing 190, 165–180.

Colombo, M., Pincay, J., Lavrovsky, O., Iseli, L., Van Wezemael, J., Portmann, E., 2021. Streetwise: Mapping citizens' perceived spatial qualities, in: Proceedings of the 23rd international conference on enterprise information systems, SciTePress. pp. 810–818.

Costa, G., 2019. City-safe: Estimating urban safety perception. URL: https://api.semanticscholar.org/CorpusID:218562162.

Costa, G., Soares, C., Marques, M., 2019. Finding common image semantics for urban perceived safety based on pairwise comparisons. 2019 27th European Signal Processing Conference (EUSIPCO) , 1–5URL: https://api.semanticscholar.org/CorpusID:208209177.

Cox, T., Rossetti, T., Hurtubia, R., 2019. Validating street-level perceptual attributes with location choice models .

Cui, Q., Gong, P., Yang, G., Zhang, S., Huang, Y., Shen, S., Wei, B., Chen, Y., 2023a. Women-oriented evaluation of perceived safety of walking routes between home and mass transit: A case study and methodology test in guangzhou. Buildings 13, 715.

Cui, Q., Zhang, Y., Yang, G., Huang, Y., Chen, Y., 2023b. Analysing gender differences in the perceived safety from street view imagery. International Journal of Applied Earth Observation and Geoinformation URL: https://api.semanticscholar.org/CorpusID:264892628.

Dai, L., Zheng, C., Dong, Z., Yao, Y., Wang, R., Zhang, X., Ren, S., Zhang, J., Song, X., Guan, Q., 2021. Analyzing the correlation between visual space and residents' psychology in wuhan, china using street-view images and deep-learning technique. City and Environment Interactions 11, 100069.

Dai, S., Li, Y., Stein, A., Yang, S., Jia, P., 2024. Street view imagery-based built environment auditing tools: a systematic review. International Journal of Geographical Information Science , 1–22.

Dalal, N., Triggs, B., 2005. Histograms of oriented gradients for human detection, in: 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05), IEEE. pp. 886–893.

De Nadai, M., Staiano, J., Larcher, R., Sebe, N., Quercia, D., Lepri, B., 2016. The death and life of great italian cities: a mobile phone data perspective, in: Proceedings of the 25th international conference on world wide web, pp. 413–423.

Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L., 2009. Imagenet: A large-scale hierarchical image database. 2009 IEEE Conference on Computer Vision and Pattern Recognition , 248–255.

Diniz, A.M.A., Stafford, M.C., 2021. Graffiti and crime in belo horizonte, brazil: The broken promises of broken windows theory. Applied Geography 131, 102459.

Doersch, C., Singh, S., Gupta, A.K., Sivic, J., Efros, A.A., 2012. What makes paris look like paris? ACM Transactions on Graphics (TOG) 31, 1 – 9.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N., 2020. An image is worth 16x16 words: Transformers for image recognition at scale. International Conference on Learning Representations .

Dubey, A., Naik, N., Parikh, D., Raskar, R., Hidalgo, C.A., 2016. Deep learning the city: Quantifying urban perception at a global scale, in: Leibe, B., Matas, J., Sebe, N., Welling, M. (Eds.), Computer Vision – ECCV 2016, Springer International Publishing, Cham. pp. 196–212.

Ewing, R., Clemente, O., Neckerman, K.M., Purciel-Hill, M., Quinn, J.W., Rundle, A.G., 2013. Measuring urban design: Metrics for livable places. Measuring Urban Design URL: https://api.semanticscholar.org/CorpusID:107673264.

Ewing, R., Handy, S.L., 2009. Measuring the unmeasurable: Urban design qualities related to walkability. Journal of Urban Design 14, 65–84. URL: https://api.semanticscholar.org/CorpusID:89607311.

Freitas, F., Berreth, T., Chen, Y.C., Jhala, A., 2023. Characterizing the perception of urban spaces from visual analytics of street-level imagery. Ai & Society 38, 1361–1371.

Fu, K., Chen, Z., Lu, C.T., 2018. Streetnet: preference learning with convolutional neural network on urban crime perception. Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems URL: https://api.semanticscholar.org/CorpusID:53305895.

Gagliardi, V., Giammorcaro, B., Bella, F., Sansonetti, G., 2023. Deep neural networks for asphalt pavement distress detection and condition assessment, in: Remote Sensing. URL: https://api.semanticscholar.org/CorpusID:264373942.

Gao, M., Guo, H., Liu, L., Zeng, Y., Liu, W., Liu, Y., Xing, H., 2024. Integrating street view imagery and taxi trajectory for identifying urban function of street space. Geo-spatial Information Science , 1–23.

Gao, W., Hou, J., Gao, Y., Zhao, M., Jia, M., 2023. Quantifying the spatial ratio of streets in beijing based on street-view images. ISPRS International Journal of Geo-Information 12, 246.

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

Economic Inquiry 56, 114–137.

Gong, W., Huang, X., White, M., Langenheim, N., 2023. Walkability perceptions and gender differences in urban fringe new towns: A case study of shanghai. Land .

Gong, Z., Ma, Q., Kan, C., Qi, Q., 2019. Classifying street spaces with street view images for a spatial indicator of urban functions. Sustainability URL: https://api.semanticscholar.org/CorpusID:208877310.

Grondin, V., Fortin, J.M., Pomerleau, F., Giguère, P., 2023. Tree detection and diameter estimation based on deep learning. Forestry 96, 264–276.

Guan, W., Chen, Z., Feng, F., Liu, W., Nie, L., 2021. Urban perception: Sensing cities via a deep interactive multi-task learning framework. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM) 17, 1 – 20.

Han, X., Wang, L., Seo, S.H., He, J., Jung, T.Y., 2022. Measuring perceived psychological stress in urban built environments using google street view and deep learning. Frontiers in Public Health 10. URL: https://api.semanticscholar.org/CorpusID:248672503.

Hao, N., Li, X., Han, D., Nie, W., 2024. Quantifying the impact of street greening during full-leaf seasons on emotional perception: Guidelines for resident well-being. Forests 15, 119.

He, J., Zhang, J., Yao, Y., Li, X., 2023. Extracting human perceptions from street view images for better assessing urban renewal potential. Cities .

He, N., Li, G., 2021. Urban neighbourhood environment assessment based on street view image processing: A review of research trends. Environmental Challenges 4, 100090.

He, Z., Wang, Z., Xie, Z., Wu, L., Chen, Z., 2022. Multiscale analysis of the influence of street built environment on crime occurrence using street-view images. Comput. Environ. Urban Syst. 97, 101865. URL: https://api.semanticscholar.org/CorpusID:251203532.

Hoiem, D., Efros, A.A., Hebert, M., 2007. Recovering surface layout from an image. International Journal of Computer Vision 75, 151–172.

Hou, X., Chen, P., 2024. Analysis of road safety perception and influencing factors in a complex urban environment—taking chaoyang district, beijing, as an example. ISPRS International Journal of Geo-Information URL: https://api.semanticscholar.org/CorpusID:271626750.

Hou, Y., Quintana, M., Khomiakov, M., Yap, W., Ouyang, J., Ito, K., Wang, Z., Zhao, T., Biljecki, F., 2024. Global streetscapes—a comprehensive dataset of 10 million street-level images across 688 cities for urban science and analytics. ISPRS Journal of Photogrammetry and Remote Sensing 215, 216–238.

Hu, C.B., Zhang, F., Gong, F.Y., Ratti, C., Li, X., 2020. Classification and mapping of urban canyon geometry using google street view images and deep multitask learning. Building and Environment 167, 106424.

Hu, M., Chen, R., 2018. A framework for understanding sense of place in an urban design context. Urban Science 2, 34.

Hu, S., Xu, Y., Wu, L., Wu, X., Wang, R., Zhang, Z., Lu, R., Mao, W., 2021. A framework to detect and understand thematic places of a city using geospatial data. Cities 109, 103012. URL: https://api.semanticscholar.org/CorpusID:229440341.

Hua, Y., Yi, D., 2021. Synthetic to realistic imbalanced domain adaption for urban scene perception. IEEE Transactions on Industrial Informatics 18, 3248–3255. URL: https://api.semanticscholar.org/CorpusID:239642394.

Huang, G., Yu, Y., Lyu, M., Sun, D., Zeng, Q., Bart, D., 2023a. Using google street view panoramas to investigate the influence of urban coastal street environment on visual walkability. Environmental Research Communications 5.

Huang, J., Qing, L., Han, L., Liao, J., Guo, L., Peng, Y., 2023b. A collaborative perception method of human-urban environment based on machine learning and its application to the case area. Eng. Appl. Artif. Intell. 119, 105746. URL: https://api.semanticscholar.org/CorpusID:255216213.

Huang, L., Oki, T., Muto, S., Ogawa, Y., 2024a. Unveiling the non-linear influence of eye-level streetscape factors on walking preference: Evidence from tokyo. ISPRS International Journal of Geo-Information 13, 131.

Huang, T., Wu, Z., Wu, J., Hwang, J., Rajagopal, R., 2024b. Citypulse: Fine-grained assessment of urban change with street view time series, in: AAAI Conference on Artificial Intelligence. URL: https://api.semanticscholar.org/CorpusID:266725739.

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

Huang, Y., Zhang, F., Gao, Y., Tu, W., Duarte, F., Ratti, C., Guo, D., Liu, Y., 2023c. Comprehensive urban space representation with varying numbers of street-level images. Computers, Environment and Urban Systems 106, 102043.

Ibrahim, M.R., Haworth, J., Cheng, T., 2020. Understanding cities with machine eyes: A review of deep computer vision in urban analytics. Cities 96, 102481.

Ilic, L., Sawada, M., Zarzelli, A., Zarzelli, A., 2019. Deep mapping gentrification in a large canadian city using deep learning and google street view. PLoS ONE 14.

Ito, K., Kang, Y., Zhang, Y., Zhang, F., Biljecki, F., 2024. Understanding urban perception with visual data: A systematic review. Cities .

Izquierdo, L., leticiai, 2020. Urban safety perception metrics in conflict-sensitive informal settlements: A case study of lomas del centinela, guadalajara (mexico).

Izquierdo, L., Palomo, G., Grignard, A., Alonso, L., Siller, M., Larson, K., 2021. An agent-based model to evaluate the perception of safety in informal settlements. Proceedings of the 2020 Conference of The Computational Social Science Society of the Americas URL: https://api.semanticscholar.org/CorpusID:245719758.

Ji, H., Qing, L., Han, L., Wang, Z., Cheng, Y., Peng, Y., 2021. A new data-enabled intelligence framework for evaluating urban space perception. ISPRS Int. J. Geo Inf. 10, 400. URL: https://api.semanticscholar.org/CorpusID:236233311.

Jiang, B., Mak, C.N.S., Larsen, L., Zhong, H., 2017. Minimizing the gender difference in perceived safety: Comparing the effects of urban back alley interventions. Journal of Environmental Psychology 51, 117–131.

Jing, F., Liu, L., Zhou, S., Song, J., Wang, L., Zhou, H., Wang, Y., Ma, R., 2021. Assessing the impact of street-view greenery on fear of neighborhood crime in guangzhou, china. International Journal of Environmental Research and Public Health 18. URL: https://api.semanticscholar.org/CorpusID:230820071.

Jodas, D.S., Yojo, T., Brazolin, S., Velasco, G.D.N., Papa, J.P., 2022. Detection of trees on street-view images using a convolutional neural network. International Journal of Neural Systems 32, 2150042.

Joglekar, S., Quercia, D., Redi, M., Aiello, L.M., Kauer, T., Sastry, N.R., 2020. Facelift: a transparent deep learning framework to beautify urban scenes. Royal Society Open Science 7. URL: https://api.semanticscholar.org/CorpusID:210172759.

Kang, H.W., Kang, H.B., 2015. A new context-aware computing method for urban safety, in: New Trends in Image Analysis and Processing–ICIAP 2015 Workshops: ICIAP 2015 International Workshops, BioFor, CTMR, RHEUMA, ISCA, MADiMa, SBMI, and QoEM, Genoa, Italy, September 7-8, 2015, Proceedings 18, Springer. pp. 298–305.

Kang, H.W., Kang, H.B., 2016. Urban safety prediction using context and object information via double-column convolutional neural network. 2016 13th Conference on Computer and Robot Vision (CRV) , 399–405URL: https://api.semanticscholar.org/CorpusID:16158253.

Kang, Y., Abraham, J., Ceccato, V., Duarte, F., Gao, S., Ljungqvist, L., Zhang, F., Näsman, P., Ratti, C., 2023a. Assessing differences in safety perceptions using geoai and survey across neighbourhoods in stockholm, sweden. Landscape and Urban Planning 236, 104768.

Kang, Y., Kim, J., Park, J., Lee, J., 2023b. Assessment of perceived and physical walkability using street view images and deep learning technology. ISPRS Int. J. Geo Inf. 12, 186.

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

Keizer, K., Lindenberg, S., Steg, L., 2008. The spreading of disorder. Science (New York, N.Y.) 322, 1681–5. doi:10.1126/science.1161405.

Kim, J.H., Lee, S., Hipp, J.R., Ki, D., 2021. Decoding urban landscapes: Google street view and measurement sensitivity. Comput. Environ. Urban Syst. 88, 101626. URL: https://api.semanticscholar.org/CorpusID:233548120.

Kim, S., Jeon, J., Noh, Y., Woo, A., 2024. Impacts of streetscape features on individual social capital: Applying korea's neighborhood data to street view images to improve lives of the socially vulnerable. Land URL: https://api.semanticscholar.org/CorpusID:269673310.

Kim, S., Lee, S., 2023. Nonlinear relationships and interaction effects of an urban environment on crime incidence: Application of urban big data and an interpretable machine learning method. Sustainable Cities and Society 91, 104419.

Koch, G., Zemel, R., Salakhutdinov, R., 2015. Siamese neural networks for one-shot image recognition, in: ICML deep learning workshop.

Kong, W., Zhong, T., Mai, X., Zhang, S., Chen, M., Lv, G., 2022. Automatic detection and assessment of pavement marking defects with street view imagery at the city scale. Remote. Sens. 14, 4037. URL: https://api.semanticscholar.org/CorpusID:251709043.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks, in: Pereira, F., Burges, C.J.C., Bottou, L., Weinberger, K.Q. (Eds.), Advances in Neural Information Processing Systems 25, Curran Associates, Inc.. pp. 1097–1105. URL: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf.

Kruse, J., Kang, Y., Liu, Y.N., Zhang, F., Gao, S., 2021. Places for play: Understanding human perception of playability in cities using street view images and deep learning. Comput. Environ. Urban Syst. 90, 101693. URL: https://api.semanticscholar.org/CorpusID:237477202.

Kumakoshi, Y., Onoda, S., Takahashi, T., Yoshimura, Y., 2021. Quantifying urban streetscapes with deep learning: focus on aesthetic evaluation. ArXiv abs/2106.15361. URL: https://api.semanticscholar.org/CorpusID:235669798.

Larkin, A., Gu, X., Chen, L., Hystad, P., 2021. Predicting perceptions of the built environment using gis, satellite and street view image approaches. Landscape and Urban Planning 216, 104257.

Larkin, A., Krishna, A., Chen, L., Amram, O., Avery, A.R., Duncan, G.E., Hystad, P., 2022. Measuring and modelling perceptions of the built environment for epidemiological research using crowd-sourcing and image-based deep learning models. Journal of Exposure Science & Environmental Epidemiology 32, 892 – 899. URL: https://api.semanticscholar.org/CorpusID:253479825.

Lavi., B., Tokuda., E., Moreno-Vera., F., Nonato., L., Silva., C., Poco., J., 2022. 17k-graffiti: Spatial and crime data assessments in são paulo city, in: Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 4: VISAPP,, INSTICC. SciTePress. pp. 968–975. doi:10.5220/0010883300003124.

Law, S., Paige, B., Russell, C., 2018a. Take a look around. ACM Transactions on Intelligent Systems and Technology (TIST) 10, 1 – 19. URL: https://api.semanticscholar.org/CorpusID:221658610.

Law, S., Seresinhe, C.I., Shen, Y., Gutierrez-Roig, M., 2018b. Street-frontage-net: urban image classification using deep convolutional neural networks. International Journal of Geographical Information Science 34, 681 – 707. URL: https://api.semanticscholar.org/CorpusID:59349850.

Law, S., Shen, Y., Seresinhe, C., 2017. An application of convolutional neural network in street image classification: The case study of london, in: Proceedings of the 1st Workshop on Artificial Intelligence and Deep Learning for Geographic Knowledge Discovery, pp. 5–9.

Lazebnik, S., Schmid, C., Ponce, J., 2006. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories, in: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), IEEE. pp. 2169–2178.

Lee, J., Grosz, D., Uzkent, B., Zeng, S., Burke, M., Lobell, D., Ermon, S., 2021a. Predicting livelihood indicators from community-generated street-level imagery, in: AAAI Conference on Artificial Intelligence.

Lee, J., Grosz, D., Zeng, S., Uzkent, B., Burke, M., Lobell, D., Ermon, S., 2021b. Predicting livelihood indicators from crowdsourced street level images, in: Proceedings of the AAAI Conference on Artificial Intelligence.

Lei, X., Liu, C., Li, L., Wang, G., 2020. Automated pavement distress detection and deterioration analysis using street view map. IEEE Access 8, 76163–76172. URL: https://api.semanticscholar.org/CorpusID:218494462.

Lei, Y., Zhou, H., Xue, L., Yuan, L., Liu, Y., Wang, M., Wang, C., 2024. Evaluating and comparing human perceptions of streets in two megacities by integrating street-view images, deep learning, and space syntax. Buildings URL: https://api.semanticscholar.org/CorpusID:270598377.

Li, K., 2024. Research on the factors influencing the spatial quality of high-density urban streets: A framework using deep learning, street scene images, and principal component analysis. Land .

Li, T., Xin, S., Xi, Y., Tarkoma, S., Hui, P., Li, Y., 2022a. Predicting multi-level socioeconomic indicators from structural urban imagery. Proceedings of the 31st ACM International Conference on Information & Knowledge Management .

Li, X., Beaucamp, B., Tourre, V., Leduc, T., Servieres, M.C.J., 2023. Evaluation of urban perception using only image segmentation features, in: International Conference on Geographical Information Systems Theory, Applications and Management. URL: https://api.semanticscholar.org/CorpusID:258411276.

Li, X., Zhang, C., Li, W., 2015a. Does the visibility of greenery increase perceived safety in urban areas? evidence from the place pulse 1.0 dataset. ISPRS Int. J. Geo Inf. 4, 1166–1183. URL: https://api.semanticscholar.org/CorpusID:14862449.

Li, X., Zhang, C., Li, W., Ricard, R.M., Meng, Q., Zhang, W., 2015b. Assessing street-level urban greenery using google street view and a modified green view index. Urban Forestry & Urban Greening 14, 675–685.

Li, Y., Peng, L., chong Wu, C., Zhang, J., 2022b. Street view imagery (svi) in the built environment: A theoretical and systematic review. Buildings URL: https://api.semanticscholar.org/CorpusID:251424332.

Li, Y., Yabuki, N., Fukuda, T., 2022c. Measuring visual walkability perception using panoramic street view images, virtual reality, and deep learning. SSRN Electronic Journal .

Li, Z., Chen, Z., Zheng, W., Oh, S., Nguyen, K., 2021a. Ar-cnn: an attention ranking network for learning urban perception. Science China Information Sciences 65. URL: https://api.semanticscholar.org/CorpusID:245623602.

Li, Z., Liu, P., Shi, J., Xing, Y., 2021b. Research on street space quality combined with attention multi-task deep learning. 2021 2nd International Conference on Big Data Economy and Information Management (BDEIM) , 434–441URL: https://api.semanticscholar.org/CorpusID:246945182.

Li, Z., Xia, L., Tang, J., Xu, Y., Shi, L., Xia, L., Yin, D., Huang, C., 2024. Urbangpt: Spatio-temporal large language models, in: Knowledge Discovery and Data Mining. URL: https://api.semanticscholar.org/CorpusID:268230972.

Liang, X., Chang, J.H., Gao, S., Zhao, T., Biljecki, F., 2024. Evaluating human perception of building exteriors using street view imagery. Building and Environment URL: https://api.semanticscholar.org/CorpusID:271539395.

Lindal, P.J., Hartig, T., 2013. Architectural variation, building height, and the restorative quality of urban residential streetscapes. Journal of Environmental Psychology 33, 26–36.

Liu, C., Yu, Y., Yang, X., 2024a. Perceptual evaluation of street quality in underdeveloped ethnic areas: A random forest method combined with human–machine confrontation framework provides insights for improved urban planning—a case study of lhasa city. Buildings URL: https://api.semanticscholar.org/CorpusID:270351362.

Liu, L., Sevtsuk, A., 2024. Clarity or confusion: A review of computer vision street attributes in urban studies and planning. Cities .

Liu, L., Silva, E.A., Wu, C., Wang, H., 2017a. A machine learning-based method for the large-scale evaluation of the qualities of the urban environment. Comput. Environ. Urban Syst. 65, 113–125.

Liu, S., Long, Y., Zhang, L., Yang, J., Dong, W., 2024b. Quantitative measurement of urban spatial vitality by integrating physical built

environment and subjective perception dimensions. Environment and Planning B: Urban Analytics and City Science URL: https://api.semanticscholar.org/CorpusID:269966766.

Liu, W., Guo, W., Wu, R.X., 2022a. Understanding urban wealth perception using a hybrid dataset and ranking-scoring framework. Transactions in GIS 26, 2366 – 2382. URL: https://api.semanticscholar.org/CorpusID:250323453.

Liu, X., Chen, Q., Zhu, L., Xu, Y., Lin, L., 2017b. Place-centric visual urban perception with deep multi-instance regression. Proceedings of the 25th ACM international conference on Multimedia URL: https://api.semanticscholar.org/CorpusID:35385937.

Liu, Y., Chen, M., Wang, M., Huang, J., Thomas, F., Rahimi, K., Mamouei, M., 2023. An interpretable machine learning framework for measuring urban perceptions from panoramic street view images. iScience 26. URL: https://api.semanticscholar.org/CorpusID:256586135.

Liu, Z., Ma, X., Hu, L.H., Lu, S., Ye, X., You, S., Tan, Z., Li, X., 2022b. Information in streetscapes - research on visual perception information quantity of street space based on information entropy and machine learning. ISPRS Int. J. Geo Inf. 11, 628. URL: https://api.semanticscholar.org/CorpusID:254917707.

Llaguno-Munitxa, M., Edwards, M.L., Grade, S., Meulen, M.V., Letesson, C., Sierra, E.A., Altomonte, S., Lacroix, E., Bogosian, B., Kris, M., Macagno, E., 2022. Quantifying stress level reduction induced by urban greenery perception. IOP Conference Series: Earth and Environmental Science 1122. URL: https://api.semanticscholar.org/CorpusID:254995704.

Lundberg, S.M., Lee, S.I., 2017. A unified approach to interpreting model predictions, in: Neural Information Processing Systems.

Lynch, K., 1984. Reconsidering the image of the city , 151–161.

Ma, H., Wu, D., 2023. A natural language processing-based approach: mapping human perception by understanding deep semantic features in street view images. ArXiv abs/2311.17354. URL: https://api.semanticscholar.org/CorpusID:265498828.

Ma, X., Ma, C., Wu, C., Xi, Y., Yang, R., Peng, N., Zhang, C., Ren, F., 2021. Measuring human perceptions of streetscapes to better inform urban renewal: A perspective of scene semantic parsing. Cities 110, 103086. URL: https://api.semanticscholar.org/CorpusID:233787178.

Ma, Z., 2023. Deep exploration of street view features for identifying urban vitality: A case study of qingdao city. Int. J. Appl. Earth Obs. Geoinformation 123, 103476.

Malekzadeh, M.S., Willberg, E.S., Torkko, J., Toivonen, T., 2024. Urban visual appeal according to chatgpt: Contrasting ai and human insights. ArXiv abs/2407.14268.

Manvi, R., Khanna, S., Mai, G., Burke, M., Lobell, D.B., Ermon, S., 2023. Geollm: Extracting geospatial knowledge from large language models. ArXiv abs/2310.06213. URL: https://api.semanticscholar.org/CorpusID:263831484.

Marasinghe, R., Yigitcanlar, T., Mayere, S., Washington, T., Limb, M., 2023. Computer vision applications for urban planning: A systematic review of opportunities and constraints. Sustainable Cities and Society , 105047.

Martin, D., Fowlkes, C., Tal, D., Malik, J., 2001. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics, in: Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001, IEEE. pp. 416–423.

Matas, J., Chum, O., Urban, M., Pajdla, T., 2004. Robust wide-baseline stereo from maximally stable extremal regions. Image and vision computing 22, 761–767.

Meng, Y., Sun, D., Lyu, M., Niu, J., Fukuda, H., 2024. Measuring human perception of residential built environment through street view image and deep learning. Environmental Research Communications 6. URL: https://api.semanticscholar.org/CorpusID:269939329.

Milias, V., Sharifi Noorian, S., Bozzon, A., Psyllidis, A., 2023. Is it safe to be attractive? disentangling the influence of streetscape features on the perceived safety and attractiveness of city streets. AGILE: GIScience Series 4, 1–12.

Min, W., Mei, S., Liu, L., Wang, Y., Jiang, S., 2020. Multi-task deep relative attribute learning for visual urban perception. IEEE Transactions on Image Processing 29, 657–669. URL: https://api.semanticscholar.org/CorpusID:199518188.

Minka, T., Cleven, R., Zaykov, Y., 2018. TrueSkill 2: An improved Bayesian skill rating system. Technical Report MSR-TR-2018-8. Microsoft. URL: https://www.microsoft.com/en-us/research/publication/trueskill-2-improved-bayesian-skill-rating-system/.

Miranda, F., Hosseini, M., Lage, M., Doraiswamy, H., Dove, G., Silva, C.T., 2020. Urban mosaic: Visual exploration of streetscapes using large-scale image data. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems URL: https://api.semanticscholar.org/CorpusID:218483535.

Moreno-Vera, F., 2021. Understanding safety based on urban perception, in: International Conference on Intelligent Computing, Springer. pp. 54–64.

Moreno-Vera, F., Lavi, B., Poco, J., 2021a. Quantifying urban safety perception on street view images, in: International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT). URL: http://www.visualdslab.com/papers/UrbanPerceptionQuantify.

Moreno-Vera, F., Lavi, B., Poco, J., 2021b. Urban perception: Can we understand why a street is safe?, in: Mexican International Conference on Artificial Intelligence, Springer. pp. 277–288.

Muller, E., Gemmell, E., Choudhury, I., Nathvani, R.S., Metzler, A.B., Bennett, J.E., Denton, E.L., Flaxman, S., Ezzati, M., 2022. City-wide perceptions of neighbourhood quality using street view images. ArXiv abs/2211.12139. URL: https://api.semanticscholar.org/CorpusID:253760879.

Nadai, M.D., Vieriu, R.L., Zen, G., Dragicevic, S., Naik, N., Caraviello, M., Hidalgo, C.A., Sebe, N., Lepri, B., 2016. Are safer looking neighborhoods more lively?: A multimodal investigation into urban life. Proceedings of the 24th ACM international conference on Multimedia URL: https://api.semanticscholar.org/CorpusID:15699725.

Naik, N., Kominers, S.D., Raskar, R., Glaeser, E.L., Hidalgo, C.A., 2017. Computer vision uncovers predictors of physical urban change. Proceedings of the National Academy of Sciences 114, 7571 – 7576. URL: https://api.semanticscholar.org/CorpusID:6875964.

Naik, N., Philipoom, J., Raskar, R., Hidalgo, C., 2014. StreetScore: predicting the perceived safety of one million streetscapes. 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops URL: http://streetscore.media.mit.edu/static/files/streetscore_paper.pdf.

Naik, N., Raskar, R., Hidalgo, C.A., 2016. Cities are physical too: Using computer vision to measure the quality and impact of urban appearance. The American Economic Review 106, 128–132. URL: https://api.semanticscholar.org/CorpusID:55790645.

Nasar, J.L., 1998. The evaluative image of the city .

Novak, C.L., Shafer, S.A., et al., 1992. Anatomy of a color histogram., in: CVPR, pp. 599–605.

Ogawa, Y., Oki, T., Zhao, C., Sekimoto, Y., Shimizu, C., 2024. Evaluating the subjective perceptions of streetscapes using street-view images. Landscape and Urban Planning URL: https://api.semanticscholar.org/CorpusID:268844882.

Ojala, T., Pietikainen, M., Maenpaa, T., 2002. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on pattern analysis and machine intelligence 24, 971–987.

Oliva, A., Torralba, A., 2001. Modeling the shape of the scene: A holistic representation of the spatial envelope. International Journal of Computer Vision 42, 145–175. doi:10.1023/A:1011139631724.

Ooi, M., Valdez, D.A.S., Rogers, M., Ababou, R., Zhao, K., Delmas, P., 2023. Construction of a novel data set for pedestrian tree species detection using google street view data, in: Advanced Concepts for Intelligent Vision Systems Conference. URL: https://api.semanticscholar.org/CorpusID:265255275.

Ordonez, V., Berg, T.L., 2014. Learning high-level judgments of urban perception. European Conference on Computer Vision (ECCV) .

Park, J., Newman, M., 2005. A network-based ranking system for us college football. Journal of Statistical Mechanics: Theory and Experiment 2005,

P10014 – P10014. URL: https://api.semanticscholar.org/CorpusID:15120571.

Patel, D., Patel, F., Patel, S., Patel, N., Shah, D., Patel, V., 2021. Garbage detection using advanced object detection techniques, in: 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), IEEE. pp. 526–531.

Perronnin, F., Sánchez, J., Mensink, T., 2010. Improving the fisher kernel for large-scale image classification, in: European conference on computer vision, Springer. pp. 143–156.

Porzi, L., Rota Bulò, S., Lepri, B., Ricci, E., 2015. Predicting and understanding urban perception with convolutional neural networks. doi:10.1145/2733373.2806273.

Qiu, W., Li, W., Zhang, Z., Li, X.F., Liu, X.L., Huang, X.D., 2021. Subjective and objective measures of streetscape perceptions: Relationships with property value in shanghai. Cities URL: https://api.semanticscholar.org/CorpusID:233688580.

Quercia, D., O'Hare, N., Cramer, H., 2014a. Aesthetic capital: what makes london look beautiful, quiet, and happy? Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing URL: https://api.semanticscholar.org/CorpusID:4541733.

Quercia, D., Schifanella, R., Aiello, L.M., 2014b. The shortest path to happiness: recommending beautiful, quiet, and happy routes in the city. Proceedings of the 25th ACM conference on Hypertext and social media .

Ramírez, T., Hurtubia, R., Lobel, H., Lobel, H., Rossetti, T., 2021. Measuring heterogeneous perception of urban space with massive data and machine learning: An application to safety. Landscape and Urban Planning 208, 104002. URL: https://api.semanticscholar.org/CorpusID:233072976.

Rao, A., Srihari, R.K., Zhang, Z., 1999. Geometric histogram: A distribution of geometric configurations of color subsets, in: Internet Imaging, International Society for Optics and Photonics. pp. 91–101.

Rapoport, A., Hawkes, R., 1970. The perception of urban complexity. Journal of the American Institute of Planners 36, 106–111.

Regona, M., Yigitcanlar, T., Xia, B., Li, R.Y.M., 2022. Opportunities and adoption challenges of ai in the construction industry: a prisma review. Journal of Open Innovation: Technology, Market, and Complexity 8, 45.

Ribeiro, M.T., Singh, S., Guestrin, C., 2016. Why should i trust you?: Explaining the predictions of any classifier, in: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, ACM. pp. 1135–1144.

Rios, L.E.M., Ruiz-Correa, S., Santani, D., Gática-Pérez, D., 2021. Who sees what? examining urban impressions in global south cities. Human Perception of Visual Information URL: https://api.semanticscholar.org/CorpusID:235329791.

Rita, L., Peliteiro, M., Bostan, T.C., Tamagusko, T., Ferreira, A., 2023. Using deep learning and google street view imagery to assess and improve cyclist safety in london. Sustainability .

Rossetti, T., Guevara, C.A., Galilea, P., Hurtubia, R., 2018. Modeling safety as a perceptual latent variable to assess cycling infrastructure. Transportation Research Part A-policy and Practice 111, 252–265. URL: https://api.semanticscholar.org/CorpusID:56426242.

Rossetti, T., Lobel, H., Rocco, V., Hurtubia, R., 2019a. Explaining subjective perceptions of public spaces as a function of the built environment: A massive data approach. Landscape and Urban Planning .

Rossetti, T., Saud, V., Hurtubia, R., 2019b. I want to ride it where i like: measuring design preferences in cycling infrastructure. Transportation 46, 697–718. URL: https://api.semanticscholar.org/CorpusID:158265231.

Rui, J., 2023. Measuring streetscape perceptions from driveways and sidewalks to inform pedestrian-oriented street renewal in düsseldorf. Cities URL: https://api.semanticscholar.org/CorpusID:259571138.

Rui, J., Li, X., 2023. Decoding vibrant neighborhoods: Disparities between formal neighborhoods and urban villages in eye-level perceptions and physical environment. Sustainable Cities and Society URL: https://api.semanticscholar.org/CorpusID:266352693.

Rui, Q., Cheng, H., 2023. Quantifying the spatial quality of urban streets with open street view images: A case study of the main urban area of fuzhou. Ecological Indicators .

Rundle, A.G., Bader, M.D., Richards, C.A., Neckerman, K.M., Teitler, J.O., 2011. Using google street view to audit neighborhood environments. American journal of preventive medicine 40, 94–100.

Rzotkiewicz, A., Pearson, A.L., Dougherty, B.V., Shortridge, s., 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.

Salesses, P., Schechtner, K., Hidalgo, C.A., 2013. The collaborative image of the city: Mapping the inequality of urban perception. PLOS ONE URL: https://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0068400&type=printable.

Sampson, R.J., Morenoff, J.D., Gannon-Rowley, T., 2002. Assessing "neighborhood effects": Social processes and new directions in research. Annual review of sociology 28, 443–478.

Sangers, R., van Gemert, J.C., van Cranenburgh, S., 2022. Explainability of deep learning models for urban space perception. ArXiv abs/2208.13555. URL: https://api.semanticscholar.org/CorpusID:251903811.

Santani, D., Gática-Pérez, D., 2014. Loud and trendy: Crowdsourcing impressions of social ambiance in popular indoor urban places. Proceedings of the 23rd ACM international conference on Multimedia URL: https://api.semanticscholar.org/CorpusID:1536820.

Santani, D., Ruiz-Correa, S., Gática-Pérez, D., 2015. Looking at cities in mexico with crowds. Proceedings of the 2015 Annual Symposium on Computing for Development URL: https://api.semanticscholar.org/CorpusID:16331424.

Santani, D., Ruiz-Correa, S., Gática-Pérez, D., 2018. Looking south. ACM Transactions on Social Computing 1, 1 – 23.

Santos, F.A., Silva, T.H., Loureiro, A.A.F., Villas, L.A., 2018. Uncovering the perception of urban outdoor areas expressed in social media. 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI) , 120–127.

Santos, F.A., Silva, T.H., Loureiro, A.A.F., Villas, L.A., 2020. Automatic extraction of urban outdoor perception from geolocated free texts. Social Network Analysis and Mining 10. URL: https://api.semanticscholar.org/CorpusID:222310236.

Santos, F.A., Silva, T.H., Villas, L.A., 2024. Real-up: Urban perceptions from lbsns helping moving real-estate market to the next level. Companion Proceedings of the ACM on Web Conference 2024 URL: https://api.semanticscholar.org/CorpusID:269762587.

Sarmiento, J.A.R., 2021. Pavement distress detection and segmentation using yolov4 and deeplabv3 on pavements in the philippines abs/2103.06467. URL: https://api.semanticscholar.org/CorpusID:232185162.

Schroeder, H.W., Anderson, L.M., 1984. Perception of personal safety in urban recreation sites. Journal of leisure research 16, 178–194.

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.

Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D., 2017. Grad-cam: Visual explanations from deep networks via gradient-based localization, in: Proceedings of the IEEE International Conference on Computer Vision, pp. 618–626.

Sengupta, N., Vaidya, A., Evans, J., 2023. In her shoes: Gendered labelling in crowdsourced safety perceptions data from india. Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency URL: https://api.semanticscholar.org/CorpusID:259139823.

Seresinhe, C.I., Preis, T., Moat, H.S., 2017. Using deep learning to quantify the beauty of outdoor places. Royal Society Open Science 4. URL: https://api.semanticscholar.org/CorpusID:21795600.

Sharma, A., Keshri, A., Kumar, A., Yadav, R., 2023. Garbage classification with deep learning techniques, in: 2023 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES), IEEE. pp. 406–411.

Shen, Q., Zeng, W., Ye, Y., Arisona, S.M., Schubiger, S., Burkhard, R., Qu, H., 2017. Streetvizor: Visual exploration of human-scale urban forms based on street views. IEEE transactions on visualization and computer

graphics 24, 1004–1013.

Shi, J., Hao, K., 2023. Measuring human perception and negative elements of public space quality using deep learning: A case study of area within the inner road of tianjin city, in: International Conference on Human-Computer Interaction, Springer. pp. 593–606.

Shi, J., Yan, Y., Li, M., Zhou, L., 2024. Measuring the convergence and divergence in urban street perception among residents and tourists through deep learning: A case study of macau. Land 13, 345.

Sivic, J., Zisserman, A., 2004. Video data mining using configurations of viewpoint invariant regions, in: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004., IEEE. pp. I–I.

Skogan, W.G., 1992. Disorder and decline: Crime and the spiral of decay in American neighborhoods. Univ of California Press.

Son, T.H., Weedon, Z., Yigitcanlar, T., Sanchez, h., Corchado, J.M., Mehmood, R., 2023. Algorithmic urban planning for smart and sustainable development: Systematic review of the literature. Sustainable Cities and Society , 104562.

Song, H., 2024. Street view imagery: Ai-based analysis method and application. Applied and Computational Engineering URL: https://api.semanticscholar.org/CorpusID:267956904.

Stalder, S., Volpi, M., Büttner, N., Law, S., Harttgen, K., Suel, E., 2023. Self-supervised learning unveils change in urban housing from street-level images. Comput. Environ. Urban Syst. 112, 102156. URL: https://api.semanticscholar.org/CorpusID:262053609.

Stalidis, P., Semertzidis, T., Daras., P., 2018. Examining deep learning architectures for crime classification and prediction. arXiv URL: https://arxiv.org/pdf/1812.00602.pdf.

Su, N., Li, W., Qiu, W., 2023. Measuring the associations between eye-level urban design quality and on-street crime density around new york subway entrances. Habitat International .

Suel, E., Bhatt, S., Brauer, M., Flaxman, S., Ezzati, M., 2021. Multimodal deep learning from satellite and street-level imagery for measuring income, overcrowding, and environmental deprivation in urban areas. Remote Sensing of Environment 257. URL: https://api.semanticscholar.org/CorpusID:233470324.

Suel, E., Boulleau, M., Ezzati, M., Flaxman, S., 2018. Combining street imagery and spatial information for measuring socioeconomic status. URL: https://api.semanticscholar.org/CorpusID:53349742.

Suel, E., Muller, E., Bennett, J.E., Blakely, T., Doyle, Y., Lynch, J., Mackenbach, J.D., Middel, A., Mizdrak, A., Nathvani, R.S., Brauer, M., Ezzati, M., 2023. Do poverty and wealth look the same the world over? a comparative study of 12 cities from five high-income countries using street images. Epj Data Science 12. URL: https://api.semanticscholar.org/CorpusID:259093335.

Suel, E., Polak, J.W., Bennett, J.E., Ezzati, M., 2019. Measuring social, environmental and health inequalities using deep learning and street imagery. Scientific Reports 9. URL: https://api.semanticscholar.org/CorpusID:121306746.

Tang, X., Zhang, L., Chen, Z., Wan, J., Li, L., 2020. Urban street landscape analysis based on street view image recognition. 2020 International Conference on Urban Engineering and Management Science (ICUEMS) , 145–150URL: https://api.semanticscholar.org/CorpusID:220888358.

Thackway, W., Ng, M.K.M., Lee, C.L., Pettit, C., 2023. Implementing a deep-learning model using google street view to combine social and physical indicators of gentrification. Comput. Environ. Urban Syst. 102, 101970.

Tian, H., Han, Z., Xu, W., Liu, X., Qiu, W., Li, W., 2021. Evolution of historical urban landscape with computer vision and machine learning: a case study of berlin. J. Digit. Landsc. Archit 16, 436–445.

Tian, Y.h., Chen, X.l., Xiong, H.k., Li, H.l., Dai, L.r., Chen, J., Xing, J.l., Chen, J., Wu, X.h., Hu, W.m., et al., 2017. Towards human-like and transhuman perception in ai 2.0: A review. Frontiers of Information Technology & Electronic Engineering 18, 58–67.

Tokuda, E.K., César Júnior, R.M., Silva, C., 2018. Identificaçao automática de pichaçao a partir de imagens urbanas. SIBGRAPI Digital Library Archive .

Tokuda, E.K., Silva, C.T., Jr., R.M.C., 2019. Quantifying the presence of graffiti in urban environments. CoRR abs/1904.04336. URL: http://arxiv.org/abs/1904.04336, arXiv:1904.04336.

Ulrich, R.S., 1979. Visual landscapes and psychological well-being. Landscape research 4, 17–23.

van Veghel, J., Dane, G., Agugiaro, G., Borgers, A.W.J., 2024. Human-centric computational urban design: optimizing high-density urban areas to enhance human subjective well-being. Computational Urban Science URL: https://api.semanticscholar.org/CorpusID:270106179.

Velasquez-Camacho, L., Etxegarai, M., de Miguel, S., 2023. Implementing deep learning algorithms for urban tree detection and geolocation with high-resolution aerial, satellite, and ground-level images. Computers, Environment and Urban Systems 105, 102025.

Verma, D., Jana, A., Ramamritham, K., 2019. Machine-based understanding of manually collected visual and auditory datasets for urban perception studies. Landscape and Urban Planning URL: https://api.semanticscholar.org/CorpusID:198253380.

Verma, D., Jana, A., Ramamritham, K., 2020. Predicting human perception of the urban environment in a spatiotemporal urban setting using locally acquired street view images and audio clips. Building and Environment 186, 107340.

Wang, L., Han, X., He, J., Jung, T.Y., 2022a. Measuring residents' perceptions of city streets to inform better street planning through deep learning and space syntax. ISPRS Journal of Photogrammetry and Remote Sensing URL: https://api.semanticscholar.org/CorpusID:250104109.

Wang, R., Liu, Y., Lu, Y., Zhang, J., Liu, P., Yao, Y., Grekousis, G., 2019a. Perceptions of built environment and health outcomes for older chinese in beijing: A big data approach with street view images and deep learning technique. Comput. Environ. Urban Syst. 78. URL: https://api.semanticscholar.org/CorpusID:202185958.

Wang, R., Ren, S., Zhang, J., Yao, Y., Wang, Y., Guan, Q., 2021. A comparison of two deep-learning-based urban perception models: which one is better? Computational Urban Science 1. URL: https://api.semanticscholar.org/CorpusID:236303632.

Wang, R., Yuan, Y., Liu, Y., Zhang, J., Liu, P., Lu, Y., Yao, Y., 2019b. Using street view data and machine learning to assess how perception of neighborhood safety influences urban residents' mental health. Health & place 59, 102186. URL: https://api.semanticscholar.org/CorpusID:199527372.

Wang, Y., Zeng, Z., Li, Q., Deng, Y., 2022b. A complete reinforcement-learning-based framework for urban-safety perception. ISPRS Int. J. Geo Inf. 11, 465. URL: https://api.semanticscholar.org/CorpusID:251969078.

Wang, Y., Zeng, Z., Zhao, Q., 2022c. Evaluating the perceived safety of urban city via maximum entropy deep inverse reinforcement learning, in: Asian Conference on Machine Learning. URL: https://api.semanticscholar.org/CorpusID:253734684.

Wang, Z., Ito, K., Biljecki, F., 2024. Assessing the equity and evolution of urban visual perceptual quality with time series street view imagery. Cities 145, 104704.

Wei, J., Yue, W., Li, M., Gao, J., 2022. Mapping human perception of urban landscape from street-view images: A deep-learning approach. International Journal of Applied Earth Observation and Geoinformation 112, 102886.

Wei, Z., Cao, K., Kwan, M.P., Jiang, Y., Feng, Q., 2024. Measuring the age-friendliness of streets' walking environment using multi-source big data: A case study in shanghai, china. Cities 148, 104829. URL: https://www.sciencedirect.com/science/article/pii/S026427512400043X, doi:https://doi.org/10.1016/j.cities.2024.104829.

Wen, D., Liu, M., Yu, Z., 2022. Quantifying ecological landscape quality of urban street by open street view images: A case study of xiamen island, china. Remote. Sens. 14, 3360. URL: https://api.semanticscholar.org/CorpusID:250547729.

Wendt, M., 2009. The importance of death and life of great american cities (1961) by jane jacobs to the profession of urban planning. New Visions for Public Affairs 1, 1–24.

Wilson, J.Q., Kelling, G.L., 1982. Broken windows. Atlantic monthly 249, 29–38.

Wilson, J.S., Kelly, C.M., Schootman, M., Baker, E.A., Banerjee, A., Clennin, M.N., Miller, D.K., 2012. Assessing the built environment using omnidirectional imagery. American journal of preventive medicine 42 2, 193–9. URL: https://api.semanticscholar.org/CorpusID:25602191.

Wohlin, C., 2014. Guidelines for snowballing in systematic literature studies and a replication in software engineering, in: Proceedings of the 18th international conference on evaluation and assessment in software engineering, pp. 1–10.

Wu, C., Ye, Y., Gao, F., Ye, X., 2022. Using street view images to examine the association between human perceptions of locale and urban vitality in shenzhen, china. Sustainable Cities and Society .

Wu, Y., Shen, X., Liu, Q., Xiao, F., Li, C., 2021. A garbage detection and classification method based on visual scene understanding in the home environment. Complexity 2021, 1–14.

Xiao, J., Ehinger, K.A., Hays, J., Torralba, A., Oliva, A., 2014. Sun database: Exploring a large collection of scene categories. International Journal of Computer Vision 119, 3–22.

Xiao, J., Hays, J., Ehinger, K.A., Oliva, A., Torralba, A., 2010. Sun database: Large-scale scene recognition from abbey to zoo. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition , 3485–3492.

Xie, Q., Li, D., Yu, Z., Zhou, J., Wang, J., 2020. Detecting trees in street images via deep learning with attention module. IEEE Transactions on Instrumentation and Measurement 69, 5395–5406. URL: https://api.semanticscholar.org/CorpusID:209334499.

Xu, F., Jin, A., Chen, X., Li, G., 2021. New data, integrated methods and multiple applications: A review of urban studies based on street view images. 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS , 6532–6535.

Xu, J., Liu, Y., Liu, Y., An, R., Tong, Z., 2023a. Integrating street view images and deep learning to explore the association between human perceptions of the built environment and cardiovascular disease in older adults. Social science & medicine 338, 116304.

Xu, J., Xiong, Q., Jing, Y., Xing, L., An, R., Tong, Z., Liu, Y., Liu, Y., 2023b. Understanding the nonlinear effects of the street canyon characteristics on human perceptions with street view images. Ecological Indicators 154, 110756.

Xu, X., Qiu, W., Li, W., Liu, X., Zhang, Z., Li, X., Luo, D., 2022. Associations between street-view perceptions and housing prices: Subjective vs. objective measures using computer vision and machine learning techniques. Remote. Sens. 14, 891.

Xu, Y., Yang, Q., Cui, C., Shi, C., Song, G., Han, X., Yin, Y., 2019. Visual urban perception with deep semantic-aware network, in: International Conference on Multimedia Modeling, Springer. pp. 28–40.

Yang, N., Deng, Z., Hu, F., Chao, Y., Wan, L., Guan, Q., Wei, Z., 2024a. Urban perception by using eye movement data on street view images. Transactions in GIS URL: https://api.semanticscholar.org/CorpusID:269705741.

Yang, N., Deng, Z., Hu, F., Guan, Q., Chao, Y., Wan, L., 2024b. Urban perception assessment from street view images based on a multifeature integration encompassing human visual attention. Annals of the American Association of Geographers URL: https://api.semanticscholar.org/CorpusID:271075674.

Yao, Y., Liang, Z., Yuan, Z., Liu, P., Bie, Y., Zhang, J., Wang, R., Wang, J., Guan, Q., 2019. A human-machine adversarial scoring framework for urban perception assessment using street-view images. International Journal of Geographical Information Science 33, 2363 – 2384. URL: https://api.semanticscholar.org/CorpusID:199589510.

Yao, Y., Wang, J., Hong, Y., Qian, C., Guan, Q., Liang, X., Dai, L., Zhang, J., 2021. Discovering the homogeneous geographic domain of human perceptions from street view images. Landscape and Urban Planning 212, 104125. URL: https://api.semanticscholar.org/CorpusID:235513379.

Ye, Y., Jia, C., Winter, S., 2024. Measuring perceived walkability at the city scale using open data. Land .

Ye, Y., Zeng, W., Shen, Q., Zhang, X., Lu, Y., 2019. The visual quality of streets: A human-centred continuous measurement based on machine learning algorithms and street view images. Environment and Planning B: Urban Analytics and City Science 46, 1439 – 1457. URL: https://api.semanticscholar.org/CorpusID:203206693.

Yin, C., Peng, N., Li, Y., Shi, Y., Yang, S., Jia, P., 2023. A review on street view observations in support of the sustainable development goals. International Journal of Applied Earth Observation and Geoinformation 117, 103205.

Yu, X., Ma, J., Tang, Y., Yang, T., Jiang, F., 2024. Can we trust our eyes? interpreting the misperception of road safety from street view images and deep learning. Accident Analysis & Prevention 197, 107455.

Yuan, W., Mu, X., Jiao, J., Li, D., Li, J., 2024. How to enhancing urban space renewal through visual landscape perception? an approach from street view image recognition. Social Indicators Research URL: https://api.semanticscholar.org/CorpusID:271367632.

Zhang, C., Wu, T., Zhang, Y., Zhao, B., Wang, T., Cui, C., Yin, Y., 2021a. Deep semantic-aware network for zero-shot visual urban perception. International Journal of Machine Learning and Cybernetics 13, 1197 – 1211. URL: https://api.semanticscholar.org/CorpusID:238697226.

Zhang, F., Fan, Z., Kang, Y., Hu, Y., Ratti, C., 2021b. "perception bias": Deciphering a mismatch between urban crime and perception of safety. Landscape and Urban Planning 207, 104003.

Zhang, F., Hu, M., Che, W., Lin, H., Fang, C., 2018a. Framework for virtual cognitive experiment in virtual geographic environments. ISPRS Int. J. Geo Inf. 7, 36. URL: https://api.semanticscholar.org/CorpusID:3343257.

Zhang, F., Salazar-Miranda, A., Duarte, F., Vale, L., Hack, G., Chen, M., Liu, Y., Batty, M., Ratti, C., 2024a. Urban visual intelligence: Studying cities with artificial intelligence and street-level imagery. Annals of the American Association of Geographers , 1–22.

Zhang, F., Wu, L., Zhu, D., Liu, Y., 2019. Social sensing from street-level imagery: A case study in learning spatio-temporal urban mobility patterns. ISPRS Journal of Photogrammetry and Remote Sensing URL: https://api.semanticscholar.org/CorpusID:164553442.

Zhang, F., Zhang, D., Liu, Y., Lin, H., 2018b. Representing place locales using scene elements. Comput. Environ. Urban Syst. 71, 153–164. URL: https://api.semanticscholar.org/CorpusID:52073263.

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

Zhang, F., Zu, J., xiu Hu, M., Zhu, D., Kang, Y., Gao, S., Zhang, Y., Huang, Z., 2020a. Uncovering inconspicuous places using social media check-ins and street view images. Comput. Environ. Urban Syst. 81, 101478. URL: https://api.semanticscholar.org/CorpusID:216298157.

Zhang, L., Pei, T., Wang, X., Wu, M., Song, C., Guo, S., Chen, Y., 2020b. Quantifying the urban visual perception of chinese traditional-style building with street view images. Applied Sciences URL: https://api.semanticscholar.org/CorpusID:225211348.

Zhang, T., Wang, L., Hu, Y., Zhang, W., Liu, Y., 2024b. Measuring urban green space exposure based on street view images and machine learning. Forests 15, 655.

Zhang, Y., Li, S., Dong, R., Deng, H., Fu, X., Wang, C., Yu, T., Jia, T., zhu Zhao, J., 2021c. Quantifying physical and psychological perceptions of urban scenes using deep learning. Land Use Policy URL: https://api.semanticscholar.org/CorpusID:240552539.

Zhang, Y., Wang, L., Dong, R., Deng, H., Fu, X., Huang, B., Niu, Z., Chen, F., 2023. Understanding the effects of urban perceptions on housing rent using big data and machine learning. International Journal of Sustainable Development & World Ecology 30, 964 – 980.

Zhanga, J., Lia, Y., Fukudab, T., Wang, B., 2024. Revolutionizing urban safety perception assessments: Integrating multimodal large language models with street view images. ArXiv abs/2407.19719.

Zhao, J., Guo, Q., 2022. Intelligent assessment for visual quality of streets: Exploration based on machine learning and large-scale street view data. Sustainability URL: https://api.semanticscholar.org/CorpusID:250327637.

Zhao, J., Liu, X., Kuang, Y., Chen, Y.V., Yang, B., 2018. Deep cnn-based methods to evaluate neighborhood-scale urban valuation through street scenes perception, in: 2018 ieee third international conference on data science in cyberspace (dsc), IEEE. pp. 20–27.

Zhao, L., Luo, L., Li, B., Xu, L., Zhu, J., He, S., Li, H., 2021. Analysis of the uniqueness and similarity of city landscapes based on deep style learning. ISPRS International Journal of Geo-Information 10, 734.

Zhao, T., Liang, X., Tu, W., Huang, Z., Biljecki, F., 2023. Sensing urban soundscapes from street view imagery. Comput. Environ. Urban Syst. 99, 101915.

Zhao, X., Lu, Y., Lin, G., 2024. An integrated deep learning approach for assessing the visual qualities of built environments utilizing street view images. Engineering Applications of Artificial Intelligence 130, 107805.

Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., Torralba, A., 2017a. Places: A 10 million image database for scene recognition. IEEE transactions on pattern analysis and machine intelligence 40, 1452–1464.

Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., Oliva, A., 2014. Learning deep features for scene recognition using places database, in: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N.D., Weinberger, K.Q. (Eds.), Advances in Neural Information Processing Systems 27, Curran Associates, Inc.. pp. 487–495.

Zhou, B., Zhao, H., Puig, X., Fidler, S., Barriuso, A., Torralba, A., 2017b. Scene parsing through ade20k dataset, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 633–641.

Zhou, H., Wang, J., Wilson, K., 2022. Impacts of perceived safety and beauty of park environments on time spent in parks: Examining the potential of street view imagery and phone-based gps data. Int. J. Appl. Earth Obs. Geoinformation 115, 103078.

Zhu, J., Gong, Y., Liu, C., Du, J., Song, C., Chen, J., Pei, T., 2023. Assessing the effects of subjective and objective measures on housing prices with street view imagery: A case study of suzhou. Land 12, 2095.

Zhuang, Y., Kang, Y., Fei, T., Bian, M., Du, Y., 2024. From hearing to seeing: Linking auditory and visual place perceptions with soundscape-to-image generative artificial intelligence. Comput. Environ. Urban Syst. 110, 102122. URL: https://api.semanticscholar.org/CorpusID:269515136.

Zu, X., Gao, C., Wang, Y., 2023. Interpretation of gender divergence in consumption places based on machine learning and equilibrium index - a case study of the main urban area of beijing, china. Int. J. Appl. Earth Obs. Geoinformation 122, 103428.