Understanding Urban Perception with Visual Data: A Systematic Review

Koichi Ito\textsuperscript{a}, Yuhao Kang\textsuperscript{b,c}, Ye Zhang\textsuperscript{d,a}, Fan Zhang\textsuperscript{e}, Filip Biljecki\textsuperscript{a,f,\ast}

\textsuperscript{a}Department of Architecture, National University of Singapore, 8 Architecture Drive, Singapore, 117356, Singapore
\textsuperscript{b}GISense Lab, Department of Geography, University of South Carolina, 709 Bull St, Columbia, 29208, USA
\textsuperscript{c}Department of Geography and the Environment, University of Texas at Austin, 305 E 23rd St, Austin, 78712, USA
\textsuperscript{d}School of Architecture, Tsinghua University, Haidian District, Beijing, 100084, China
\textsuperscript{e}Institute of Remote Sensing and Geographical Information System, School of Earth and Space Sciences, Peking University, 5 YiHeYuan Road, Haidian District, Beijing, 100871, China
\textsuperscript{f}Department of Real Estate, National University of Singapore, 15 Kent Ridge Drive, Singapore, 119245, Singapore

Abstract

Visual characteristics of the built environment affect how people perceive and experience cities. For a long time, many studies have examined visual perception in cities. Such efforts have accelerated in recent years due to advancements in technologies and the proliferation of relevant data (e.g., street view imagery, geo-tagged photos, videos, virtual reality, and aerial imagery). There has not been a comprehensive systematic review paper on this topic to reveal an overarching set of research trends, limitations, and future research opportunities. Such omission is plausibly due to the difficulty in reviewing a large number of relevant papers on this popular topic. In this study, we utilized machine learning techniques (i.e., natural language processing and large language models) to semi-automate the review process and reviewed 393 relevant papers. Through the review, we found that these papers can be categorized into the physical aspects of cities: greenery and water, street design, building design, landscape, public space, and the city as a whole. We also revealed that many studies conducted quantitative analyses with a recent trend of increasingly utilizing big data and advanced technologies, such as combinations of street view imagery and deep learning models. Limitations and research gaps were also identified as follows: (1) a limited scope in terms of study areas, sample size, and attributes; (2) low quality of subjective and visual data; and (3) the need for more controlled and sophisticated methods to infer more closely

\ast Corresponding author. Email: filip@nus.edu.sg
examined impacts of visual features on human perceptions. We suggest that future studies utilize and contribute to open data and take advantage of existing data and technologies to examine the causality of visual features on human perception. The approach developed to accelerate this review proved to be accurate, efficient, and insightful. Considering its novelty, we also describe it to enable replications in the future.

**Keywords:** Urban visual perception, systematic review, Natural language processing

1. **Introduction**

The built environment has had a significant impact on people’s perceptions, subsequently influencing their behaviors (Dijksterhuis and Bargh, 2001). For more than five decades, urban designers and scholars have conducted extensive research to understand how people perceive various city designs, with the ultimate goal of enhancing the overall urban experience (Lynch, 1960; Rapoport and Hawkes, 1970; Tuan, 1977). Urban perception, defined as the interpretation and experience of urban environments by individuals and communities, has been a central concept in this field (Zhang et al., 2018). Among the many factors that have been found to influence human perceptions (e.g., cultural connections to the place), urban visual perception has been viewed as a subset of urban perception that focuses on the influences of visual elements in cities on perceptions (Gordon, 1961). It has been found that urban visual perception can provide valuable insights into human subjective experiences from a human-centered perspective (Kang et al., 2021, 2023a, 2019; Huang et al., 2023b; Cinnamon and Jahiu, 2023; Yan et al., 2023). The recent development of technologies such as computer vision, coupled with the growing availability of urban visual data sources (e.g., street view imagery, geo-tagged photos, videos, virtual reality, and aerial imagery), has yielded opportunities for researchers to scale up their studies and quantify people’s visual perceptions in the built environment that were not quantifiable at a large scale before. Such technological advancements have led to a proliferation of papers. For example, we identified 3,067 papers in total as of July 19, 2023 (see Figure 1).

Previous review papers have focused on specific aspects of urban perception studies, such as urban greenery, soundscape, physical activities, neuroscience studies, and traffic safety (Dzhambov and Dimitrova, 2014; Farahani and Maller, 2018; Lionello et al., 2020; Mucci et al., 2020; Homolja et al., 2020; Bele and Chakradeo, 2021; Amiour et al., 2022). However, almost no study has conducted a comprehensive systematic review of urban visual perception studies, and there are more multi-disciplinary, comprehensive, and niche studies that do not fall into these
Figure 1: There has been a rapid increase in publications in this domain. The plot shows the number of papers with relevant keywords found in Scopus as of mid-2023. Papers published between 1960 and 2000 are aggregated into one column in this plot and colored with light blue.

specific categories and have not been covered by existing reviews (Kang et al., 2020). Thus, this study aims to bridge this research gap (Ito and Biljecki, 2021; Zhang et al., 2018; Kruse et al., 2021; Zhang et al., 2021a; Kang et al., 2021). Such a lack of comprehensive reviews can be attributed to the rapidly increasing number of papers. Many studies have reported the time- and resource-intensive aspects of systematic reviews and called for more integration of technologies to reduce the burden on researchers (Borah et al., 2017; Bullers et al., 2018; Nussbaumer-Streit et al., 2021). A recent development of machine learning models has yielded opportunities for researchers to automate the review process (Cagigas et al., 2021; Schouw et al., 2021; Rodriguez Müller et al., 2021; Nivette et al., 2022; Alhaj et al., 2022; Warren and Moustafa, 2023); moreover, large language models (LLM), such as OpenAI’s GPT-4, have shown promising accuracy in retrieving and summarizing information from documents. By leveraging such technologies, this study fills the research gap and conducts a systematic review study by utilizing natural language processing techniques and an LLM — GPT-4 model — to automate the process of filtering papers, extracting information, categorizing papers, and identifying trends in the purposes and methods of research as well as limitations and future research opportunities. This review will contribute to the urban science literature not only...
with the values of a conventional systematic review study (i.e. investigating their research directions so far and highlighting future research opportunities) but also with its novel method to improve the scalability of the reviewing process to keep up with rapidly growing fields of research.

The paper is structured as follows: related work in Section 2, method in Section 3, overview in Section 4, review in Section 5, discussion in Section 6, and conclusion in Section 7.

2. Related work

2.1. Urban perception

We found several relevant systematic/literature review papers and discussed their approaches, key findings, and research gaps below. As there have not been many review papers specifically focusing on urban visual perception, in this section, we included reviews that are broadly related to human perception in urban environments under three themes: greenery, transportation, and urban design.

Greenery: Dzhambov and Dimitrova (2014) reviewed five papers on the effectiveness of urban green spaces as a psychological buffer for negative health impacts of noise pollution and found greenery’s statistically significant effects on reducing the negative. Although this review described and evaluated the methodologies of the reviewed papers, future research opportunities were not discussed. Another study by Farahani and Maller (2018) reviewed 45 papers and proposed a framework to assess demographics, perceptions, preferences, and characteristics of green spaces as a guideline for future studies, providing specific and useful references. The most recent review by Bele and Chakradeo (2021) analyzed 43 studies to address questions such as how people gain knowledge about green space bio-diversity, correlations among socio-demographic factors and their supports for green spaces, and restorative effects of different types and densities of greenery. This comprehensive review put forward several recommendations for future work, such as analyzing socio-demographic subgroups and more different types of green spaces. However, this study included not only visual data but public perception data in general; thus, the findings are not fully focused on the context of visual perception. Khaledi et al. (2022) conducted the most comprehensive systematic review on landscape and perception with 255 studies and proposed frameworks for various concepts, such as tangible and intangible factors influencing humans and the environment and categories that influence human landscape perceptions. Although this review noted that visual perception is a predominant topic in the reviewed studies, it did not delve into the data and methods used in these studies.
**Transportation:** Another major field that previous studies have reviewed is people’s experience with transportation and mobility. Homolja et al. (2020) looked into how 24 previous studies have utilized virtual environments (e.g., virtual reality and extended reality) and objective measurement tools (e.g., electroencephalogram) for human emotions to analyze people’s perceptions when exploring cities. By looking into different experiment settings, target variables, and analysis procedures, this review made concrete recommendations for future research (e.g., visual complexity as a factor to investigate). Amiour et al. (2022) scrutinized the relationships between the built environment characteristics (e.g., infrastructure, population density, and land use) and perceived safety among school-age children by reviewing 38 papers through meta-analysis. Unlike Homolja et al. (2020), Amiour et al. (2022) only focused on extracting the existing studies’ findings without reporting any future opportunities.

**Urban design:** Based on 94 papers, Jin (2023) delineated the historical evolution of themes in the visual perception of urban design. While this review explored visual perception and touched upon technological advancements, it remained ambiguous regarding the extent to which such advancements have been applied within the field. Shynu and Suseelan (2023) also reviewed 50 papers and highlighted future opportunities for utilizing eye trackers and the lack of research in the global south. These two reviews were relatively topically comprehensive but there could be more investigation into data and methods.

The above-mentioned review papers vary in terms of the scope of their reviews and discussion. Some reviews focused on answering narrowly defined questions (e.g., Dzhambov and Dimitrova (2014) and Amiour et al. (2022)), and others had a broad coverage (e.g., Jin (2023)) and established frameworks for key concepts (e.g., Khaledi et al. (2022)). It was evident that visual perception and greener are popular topics in perception studies, but it was not clear what has been studied in other domains and whether different domains have similar challenges and opportunities. Moreover, many existing reviews discussed the evolution of methods towards more advanced techniques, but specific data and methods were not extensively investigated.

Through the analysis of the previous relevant reviews, we highlighted a need for a comprehensive, technically detailed systematic review. To fill these gaps and further facilitate urban visual perception research, our review paper aims to include not only these domains but also many other various aspects of the built environment such as building/street design and analyze more technical details of the studies. Our review will contribute to the researchers as a comprehensive review of urban visual perception; more specifically, the value of this review paper lies in its provision
of a comprehensive, one-stop overview of urban visual perception studies, as well as a detailed description of research trends in each taxonomy defined in Section 4. Furthermore, this paper identifies limitations and future research opportunities through a review of papers on a wide range of topics.

2.2. Automation of systematic reviews

Previous studies have explored the automation of systematic reviews as a means of reducing the time and resources required by researchers, which is becoming more important as recent systematic review papers review large numbers of studies (Abdelrahman et al., 2021). Several studies have conducted systematic reviews on this topic and found that machine learning techniques have been applied to automate the screening process (Mohan et al., 2021; Laynor, 2022; van Dinter et al., 2021; Khalil et al., 2022; Blaizot et al., 2022). These reviews have identified between 15 to 47 papers and tools that aimed to automate systematic review processes, and most of them are focused on the screening phase of the PRISMA method by applying machine learning techniques (van Dinter et al., 2021; Khalil et al., 2022). Among the reviewed methods and tools to conduct screening, the highest performance seems to be achieved by tools that utilize active learning, which involves human judgments until the model is sufficiently trained. One such example with a user-friendly interface is ASReview developed by van de Schoot et al. (2021) — this tool has been widely used for systematic reviews in various fields since its release (Cagigas et al., 2021; Schouw et al., 2021; Rodriguez Müller et al., 2021; Nivette et al., 2022; Alhaj et al., 2022; Warren and Moustafa, 2023).

However, the full automation of the systematic reviews has not been achieved by previous studies due to primarily two bottlenecks: 1. automation of full-text document retrieval from various databases and 2. automation of flexible data extraction and reporting phases (van Dinter et al., 2021). The recent development of LLM, especially GPT-4 models, has achieved high accuracy in retrieving and summarizing information from documents; for example, the GPT-4 model reached 97% accuracy in text summarization (Hughes and Bae, 2023) and 100% in information retrieval for a typical context length of an academic paper around 10,000-20,000 tokens (gkamradt, 2024). Considering that humans’ accuracy when conducting a systematic review is around 90%, we believe that they can be used to overcome the second bottleneck mentioned above (Wang et al., 2020). To the best of our knowledge, our paper will be one of the first to investigate the potential of overcoming the second bottleneck mentioned above, and furthermore, to utilize an LLM to achieve this goal. Our work aims to meet the growing demand for scalable systematic reviews amidst a rapidly increasing volume of publications (Khalil et al., 2022; Wang et al., 2024a).
Listing 1: This shows a query condition used to retrieve relevant papers in Scopus.

```sql
TITLE-ABS-KEY (  
   ("perceive*" OR "perception")  
   AND ("urban*" OR {built environment})  
   AND ("image*" OR "photo*" OR "video*" OR {street view})  
)  
AND (LIMIT-TO(DOCTYPE,"cp") OR LIMIT-TO(DOCTYPE,"ar"))  
AND (  
   LIMIT-TO(SUBJAREA,"SOCI") OR LIMIT-TO(SUBJAREA,"ENGI")  
   OR LIMIT-TO(SUBJAREA,"ENVI") OR LIMIT-TO(SUBJAREA,"COMP")  
   OR LIMIT-TO(SUBJAREA,"EART") OR LIMIT-TO(SUBJAREA,"PSYC")  
   OR LIMIT-TO(SUBJAREA,"DECI") OR LIMIT-TO(SUBJAREA,"MULT")  
)  
AND (LIMIT-TO(LANGUAGE,"English"))
```

3. Method

In identifying relevant papers in the field, this study combined a conventional systematic review approach with more recent machine learning and LLM methods to make the review reproducible while achieving efficiency enabled by these new techniques. Figure 2 provides an overview of the paper’s entire process, which is comprised of two major phases: 1. conventional processing, and 2. novel processing. We collected and screened papers in the first phase and extracted key information from papers to write the review in the second phase.

The first phase of the flow is similar to a conventional systematic review method called the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA), which aims to ensure the reproducibility of systematic reviews (Page et al., 2021). As for the source of papers, we used Scopus — one of the largest abstract and citation databases of peer-reviewed literature — because of its comprehensive literature coverage, with over 29,000 titles from more than 7,000 publishers, ensuring a wide-ranging and inclusive representation of global research outputs (Elsevier, 2024). Previous systematic reviews in urban studies, such as Foroughi et al. (2023) and Biljecki and Ito (2021), used Scopus as a primary source of papers because of its extensive range of scientific journals, books, and conference proceedings. In Scopus, we found 3,067 potentially relevant papers (i.e. N1 in step 1 in Figure 2) by using the searching query condition in Listing 1.

This search condition detects papers that contain words related to perception, urban environment, and visual data published in potentially relevant fields (e.g., social and multi-disciplinary studies) of journals and conferences written in English. After identifying the papers with relevant keywords, this study followed
the PRISMA method but also replaced its screening phase with a software called ASReview developed by van de Schoot et al. (2021) that automates the screening of papers after humans label between 5% and 45% of the total papers by training an NLP model as the papers are labeled by humans (see step 2 in Figure 2). In this study, to ensure the reproducibility of this process, we adopted a technique called snowballing, in which one sets a certain number of consecutive irrelevant papers as a threshold to stop labeling, and we used 10 papers as a threshold (van Haastrecht et al., 2021). As a result of this step, we reviewed 528 papers and identified 364 relevant papers in this step (i.e. N2 in step 2 in Figure 2). We excluded papers that do not involve subjective assessment (e.g., only using semantic segmentation without subjective data) or visual input (e.g., only using public perception data without using any visual data) and do not focus on urban context (e.g., only focusing on greenery in rural areas) or perception (e.g., only focusing on predictive model architectures). After the completion of labeling, ASReview outputs a list of papers ranked from most relevant to least relevant papers based on the result of the trained NLP model’s inference. In step 3, we manually labeled the rest of the unlabeled papers from the top of the list until 10 consecutive papers were labeled as irrelevant by following the threshold set above. This step added 42 papers and resulted in 406 papers in total, which corresponds to N3 in Figure 2.

The next phase of our study involves information extraction and insight identification with the aid of an LLM (i.e., GPT-4 model), and we propose it as a novel addition to the conventional PRISMA method. We selected GPT-4 as it is known for its state-of-the-art accuracy in text summarization and information retrieval — 97% accuracy in text summarization (Hughes and Bae, 2023) and 100% in information retrieval for a context length around 10,000-20,000 tokens (gkamradt, 2024). Most of the processes in this phase have conventionally been done manually by researchers and have taken a significant amount of time and resources; therefore, we propose the following method that utilizes an LLM to semi-automate it. In step 4 of the information extraction phase, we used Scopus API to download text files of papers published in journals by Elsevier and manually downloaded PDF files of papers published in other journals without API. Out of 406 papers, full texts of 393 papers were found (i.e. N4 in Figure 2) and, thus, used in the subsequent steps. From step 5 onward, we used Python to extract texts from the papers (step 5) and formulate and send prompts (steps 6 and 7).

In steps 8 and 9 of the insight identification phase, we first extracted the results of step 7 into tabular format and combined the following outputs into one text for papers in each aspect (i.e. each taxonomy category) of the built environment that the LLM classified: summary, limitations, and future opportunities. In step
Figure 2: This figure shows the overall workflow done by this study, which consists of two major steps. The first step is similar to the conventional screening process but is aided by NLP techniques proposed by van de Schoot et al. (2021). The second step utilizes an LLM to extract information about questions inputted by the users, which is a novel method to semi-automate the systematic review process proposed by this paper.
10, we then passed the combined texts to the LLM and asked to summarize the common trends in the research purpose, methods, findings, limitations, and future research opportunities, which are often discussed in conventional systematic reviews. Finally, we used the texts obtained in step 11 to discuss the findings in Section 5. The described steps only require an initial pool of papers, a few human labels, manual downloading of papers, and user input questions, and the rest of the process can be automatically run to produce the outputs in a natural language format for different categories, thereby achieving the semi-automation of the systematic review from screening to reporting. The use of LLM for information retrieval has been discussed extensively, and some ethical concerns about the generation of non-factual information (Kang et al., 2023c). However, we manually validated the accuracy of the GPT-4 model for randomly sampled 20 papers (about 5% of the total papers) and confirmed a high accuracy of 98.6%, aligning with what has been reported by other studies (Hughes and Bae, 2023; gkamradt, 2024). The GPT-4 model provided 574 correct answers out of 580 total questions (29 questions per paper), and the incorrect answers were due to ambiguously written texts; for example, the model did not give any answer to the questions about the number of participants when multiple case studies mention only a range of numbers. We further ensured the reliability of the answers by instructing the model to answer ‘not mentioned’ when there is no relevant information to prevent hallucination. The questions and prompt format used in this study can be found in the appendix (see Listing 2 and Listing 3).

4. Overview of content

In this section, we provide an overview of the findings identified through the systematic review. Most studies were found to be organized in the following format: 1. identifying aspects of the built environment to examine (e.g., street and building), 2. identifying perceptions of interests (e.g., safety and liveliness), 3. identifying specific visual data (e.g., virtual reality and street view imagery), and 4. analyzing the collected data. Further details of such a typical flow of research were illustrated in Figure 3.

Motivated by the general workflow identified above, we categorized papers based on the specific aspects of the built environment examined by the papers. The plot on the left side of the first row in Figure 5 illustrates the following six unique categories and their percentages over the total number of papers: greenery and water, street design, building design, landscape, public space, and the city as a whole. These categories were created based on the findings through the review process. Although the categorization overlaps with one another in some cases, systematic reviews such as this paper often face this issue. For example, previous reviews by
Figure 3: This figure illustrates the methodology followed by many studies. Credit: The icons used in this diagram are obtained by the Noun Project, and the image of virtual reality was captured on OneMap 3D created by the Singapore Government. The image of street view image (SVI) was downloaded from Mapillary, and the geotagged photo was obtained from Flickr. The aerial image was taken by Luo et al. (2022a), and the example of a visual perception study is Place Pulse 2.0 by Dubey et al. (2016).

Biljecki et al. (2015) and Biljecki and Ito (2021) have also faced this issue and have reported that it is not possible to create a perfect taxonomy without any overlaps when reviewing a complex and intertwined field. Further details of studies included in each category are discussed in each subsequent subsection.

To analyze the overall trends among the reviewed papers, we visualized key findings as bar plots and combined them in Figure 5, and the descriptions below refer to the figure. We selected 10 key findings as follows and discussed their importance and trends in the following paragraphs below: 1. aspects of the built environment, 2. extent of study areas, 3. visual data types, 4. purposes of computer vision models, 5. training processes for computer vision models, 6. subjective data types, 7. overall types of research, 8. detailed types of research, 9. availability of data, and 10. approval from institutional review boards.

Drawing from the categorization of built environment aspects, we devised the plot on the left side of the first row in Figure 5, which illustrates that greenery and waters are predominant in research, as evidenced by the highest number of papers in these areas. This prominence likely stems from a significant scholarly interest in the perception of natural elements within urban contexts. The hierarchy that follows — street design, building design, landscape, public spaces, and city as a whole — reveals a descending order of research focus. This sequence not only
reflects the current priorities in urban design studies but also indicates potential
gaps and opportunities for future inquiry in the less-explored areas.

The geographic distribution and scale of study areas in urban visual perception
research offer critical insights into the field’s scope and the contexts within which
studies are conducted. To elucidate this, study areas and extents were also extracted
and visualized in Figure 4 and the plot on the right side of the first row in Figure 5.
Most study areas are concentrated in North America, Europe, and East Asia, a
few studies are scattered across South America, Australia, South Asia, and the
Middle East, and very few studies exist in Africa and Central Asia. This result
and the following results might change if we changed the query condition to
include papers published in other languages and other types of documents than
conference papers and journal articles. As for the extent of the study areas, about
50% of the papers were done at the city-level extent, followed by “not applicable”
(e.g., images of urban environments without georeference and spatial analysis)
of about 20%, the neighborhood-level extent of around 20%, and district- and
building-level extents only about 10%. This result confirms that we successfully
targeted studies that focused on urban environments. Additionally, we analyzed
the countries of the first authors and study areas in Figure 8, which illustrates the
connection between the geographical origins of the first authors’ affiliations and
the locales of their research as an alluvial plot. Both China and the United States
dominate in terms of the presence of the first authors and the locations of their
studies. It is noteworthy that a majority of first authors prefer to undertake their
research within their home countries. However, the United Kingdom stands out
as being comparatively less studied by its own first authors, suggesting that their
research interests extend internationally compared to first authors in other countries.
Remarkably, it is observed that first authors from Japan, Germany, and Australia
exclusively carry out their research within their own countries and that every study
conducted in these countries is done by first authors in these countries.

The exploration of visual data types and computer vision models is crucial in
assessing the methodological landscape of urban visual perception research. By
extracting this information, we can understand the technological trends and tools
that are shaping the field. Hence, information on visual data and computer vision
models used in the reviewed papers was also extracted and depicted in the plot
on the left side of the second row in Figure 5. It depicts the percentage of papers
that used different types of visual data, where non-geo-tagged photos were used by
about 50%, and street view images (SVI) were used by about 30% of the papers,
followed by videos, virtual reality (VR), geo-tagged photos, and aerial images.
SVI and geo-tagged photos were registered as separate data sources because street
view imagery has earned its own popularity among urban studies (Biljecki and Ito, 2021). The proliferated use of SVI aligns with what has been found by another systematic review on SVI that reported an increasing number of studies with SVI due to its wide spatial coverage, and the high scalability of SVI might have led to its popularity (Biljecki and Ito, 2021). In the reviewed papers, aerial imagery was usually recorded by drones and was used to assess people’s perception from the top-down views while keeping the visual details of cities in the images (Luo et al., 2022a). Although aerial imagery is currently underutilized, it might gain more popularity in the future as a way to scalably scan cityscapes, even in areas where it is difficult to collect SVI or photos (e.g., slum areas (Sliuzas et al., 2017)).

The plot on the right side of the second row in Figure 5 shows the percentage of different purposes of computer vision models used by the reviewed papers. The ratio of papers without computer vision models was surprisingly high around 70%, potentially indicating that more studies can leverage these techniques in the future. The predominant purpose was semantic/instance segmentation with about 10% of the total papers. This interest might stem from its application in analyzing urban landscapes by categorizing each pixel into classes such as buildings, roads, and greenery, which is crucial for urban planning and environmental monitoring. Semantic segmentation offers high detail and accuracy, providing a pixel-level understanding essential for mapping urban features, alongside rich spatial information to study urban environments. However, its computational intensity and the complexity of creating detailed annotations for training datasets pose significant challenges. Image classification followed, indicating its utility in categorizing entire images into distinct classes like urban versus rural areas or types of land use, which is beneficial for broad urban studies and land use classification due to its simplicity and efficiency. However, its lack of detail and potential for over-generalization may limit its applicability for in-depth urban analysis. Object detection accounted for a similar share, underscored by its specificity in identifying and locating objects within images, such as vehicles and pedestrians, crucial for traffic monitoring and urban safety assessments. Despite its versatility, variable performance across different conditions and dependency on diverse training datasets remain disadvantages. The category “others” included tasks such as estimating specific indicators based on visual data, showing a diverse application of computer vision in urban studies.

The training process of computer vision models was also presented in the plot on the left side of the third row in Figure 5. Pre-trained with fine-tuning accounted for about 40%, and pre-trained without fine-tuning accounted for about 15%, indicating that more than half of the studies used pre-trained models. This finding was expected given the availability of various computer vision models and the computational
resources that training in computer vision models requires. In Figure 9, we also plotted the shares of visual data types and computer vision model purposes by year. One can observe that the increase in the use of computer vision models started around 2015, which corresponds to the beginning of the increase in the use of SVI. This probably confirms the synergy between computer vision models and SVI — a combination that can enhance the scalability of research significantly.

Urban visual perception research requires the examination of subjective responses to the built environment, and how they are collected can provide insights into methodological trends and the scale of the research. To that end, the data on subjective perceptions was scrutinized, as illustrated in the plot on the right side of the third row in Figure 5, which shows two types of subjective perception data sources: own collection and publicly available data. About 90% of the papers collected their own data, suggesting that publicly available data were not popular perhaps because they were not specific and diverse enough to cover various topics of urban visual perceptions. In Figure 6, the number of participants recruited for the reviewed studies is plotted as a histogram, where one can observe that most papers had participants fewer than 500.

The categorization of research methodologies employed in the field of urban visual perception is helpful for us to understand the trends in research design. Our extraction and subsequent analysis, as depicted in the plot on the left side of the fourth row in Figure 5, underscore the predominance of quantitative research, comprising over 90% of the studies. This skew towards quantitative methods reflects the field’s inclination towards empirical, data-driven approaches that facilitate objective measurement and analysis of urban visual elements. Further details of research types are also shown in the plot on the right side of the fourth row in Figure 5, which indicates that exploratory analysis was the most used method, followed by regression, model development, index construction, and others. This observation suggests that the proliferation of urban visual perception is propelled by these early-stage exploratory studies and that more novel studies as well as more confirmatory studies will potentially be published in the future.

Evaluating the data availability in urban visual perception research is a critical measure of the field’s commitment to transparency, reproducibility, and collaborative potential (Wilson et al., 2021). Thus, the availability of data was examined in the plot on the left side of the fifth row in Figure 5. The data were not available in about 80% of the papers, and only about 20% of the papers stated that the data were available on the request or via URL. This low availability of subjective data might be the cause of the low use of publicly available data shown in the plot on the right side of the third row in Figure 5. These observations might suggest missed
opportunities to enhance the reproducibility of research and to save costs of data collection for similar studies. But it is also important to note that some studies might be difficult to release the data due to the sensitivity of their personal data.

When conducting human subject research such as urban visual perception based on human feedback, it is required to obtain approval from institutional review boards (IRB) to ensure that the research is ethical. Thus, we also extracted such information and visualized it in the plot on the right side of the fifth row in Figure 5 and found that more than 60% of the papers did not mention IRB approvals. To maintain high research ethics and quality, future studies should ensure to obtain and mention IRB approvals, which can keep the validity of studies and ultimately help avoid the decline of the field due to unethical practices.

To analyze the temporal and thematic trends in the field, we illustrated the number of papers published in each aspect each year as a heatmap plot in Figure 7. The temporal dynamics of research publications within the domain of the built environment from 2000 to 2023 reveal a compelling narrative of the field’s evolution. The gradation of color intensity on the heatmap shows a robust increase in scholarly output over the years, with a notable proliferation of studies in the realms of street design and greenery and water. This pattern suggests a response to emerging urban development challenges and an increased awareness of environmental sustainability in urban planning. The ascending trajectory of publications in recent years underscores the growing urgency and relevance of these topics.

To analyze patterns of subjective and visual data types, the number of papers, subjective data types, and visual data types were analyzed by aspect in Table 1. Interestingly, the use of publicly available data was lower than 25% for many aspects, except for city as a whole. This might be due to the limited applicability of existing publicly available data, such as Place Pulse datasets published by Dubey et al. (2016), but this also indicates potential opportunities to create publicly available datasets in aspects other than the city as a whole. We also found a few intriguing patterns in the following visual data types across aspects: non-geo-tagged photos, street view imagery (SVI), video, virtual reality, geo-tagged photos, and aerial imagery. Non-geo-tagged photos had the highest share in many aspects, except for street design and the city as a whole. This is perhaps due to their synergies to conduct experiments on natural elements’ effects, which sometimes do not require locational information of the visual data. It also has the advantage of flexibility and simplicity of data collection as researchers can take photos in areas where other visual data are not available, such as SVI and aerial imagery, and the popularity of non-geo-tagged photos could remain until other data types become as flexible and simple to collect. One common pattern across most aspects except for landscape is
the use of SVI — at least about 20% of the papers utilized street view images in all aspects, and a few aspects had it for more than 40%, such as street design and city as a whole. Videos and VR were used more often in the categories that require people’s perception of motion, such as street design and building design. Both geo-tagged photos and aerial images were not used very frequently despite their scalability; however, their future trend might differ from each other. Geo-tagged photos were first used in 2015 in Figure 9, but their use did not increase. Aerial images were first used in 2020 and, by contrast, have distinct strengths over other visual data types — their capability of capturing images from remote locations. By using aerial images, researchers might be able to explore new aspects of the urban environment in the future.

We further analyzed the visual data types based on the countries of the study areas (top 10 countries and others) and study extent (i.e., scale of research) in Figure 10. Non-geo-tagged photos were highly used in the United Kingdom, Iran, and Germany, where SVI has a low share. Videos, virtual reality, geo-tagged photos, and aerial images were used the most in Canada, Singapore, Italy, and Chile, respectively. As for the share of visual data types by extent, not applicable had a high proportion of non-geo-tagged photos and videos, showing their suitability for individual image-level analysis. SVI had a higher share for neighborhood-, district-, and city-level studies, indicating their usefulness in multiple scales of analysis, while aerial images had a slightly higher share in district-level studies. Building-level studies only had non-geo-tagged photos understandably as they do not require large-scale spatial information.

In Figure 11, we present word clouds that capture the top 20 terms most commonly used across each category. This visualization provides an at-a-glance synthesis of the predominant themes and terminology characteristic of each research domain. All the categories had high frequencies for words such as ‘urban,’ ‘perception,’ and ‘images’ as they are the focus of this review; hence, these words were omitted in the description below. In the greenery and water category, words other than their own description were ‘safety,’ ‘health,’ and ‘participants,’ which highlight the popularity of these topics. For street design, the emphasis on ‘safety,’ ‘crime,’ and ‘environment’ reveals a concern for how individuals experience and navigate street spaces, with safety being a key consideration. In building design, the frequency of ‘built,’ ‘building,’ and ‘natural’ highlights a dual focus on constructed environments and their integration with natural elements. The landscape category shows words such as ‘characteristics,’ ‘design,’ and ‘preferences’ suggest a detailed interest in the attributes of landscapes and their perceptual effects on people. Papers on public space often mention ‘social,’ ‘park,’ and ‘fear,’ revealing an interest in
5. Review

In this section, we address the chronological and thematic research trends and accomplishments thus far in different categories, including greenery and water, street design, building design, landscape, public space, and the city as a whole.
Figure 5: From the top row to the bottom, the figure portrays: (1: left) A bar plot showing the distribution of research focuses on various aspects of the built environment. (1: right) A bar plot categorizing the spatial extent of study areas. (2: left) A bar plot delineating the types of visual data utilized in research. (2: right) A bar plot outlining the different purposes for which computer vision models are applied. (3: left) A bar plot representing the array of computer vision training processes adopted in the papers. (3: right) A bar plot displaying the types of subjective perception data collected. (4: left) A bar plot illustrating the proportion of studies by research type. (4: right) A bar plot detailing the various research methods employed. (5: left) A bar plot indicating the reported status of data availability. (5: right) A bar plot summarizing the extent to which IRB approval was reported.

Figure 6: The number of participants recruited by the reviewed papers is shown in this figure. Most papers had participants fewer than 1,000.
Table 1: This table shows the number of papers, subjective data types, and visual data types by aspect.

<table>
<thead>
<tr>
<th>Aspect</th>
<th># of papers</th>
<th>Subjective data types</th>
<th>Visual data types</th>
</tr>
</thead>
<tbody>
<tr>
<td>greenery and water</td>
<td>153</td>
<td></td>
<td></td>
</tr>
<tr>
<td>street design</td>
<td>90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>building design</td>
<td>53</td>
<td></td>
<td></td>
</tr>
<tr>
<td>landscape</td>
<td>42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>public space</td>
<td>28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>city as a whole</td>
<td>27</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
5.1. Greenery and water

5.1.1. Definition and scope

This category includes papers on natural and semi-natural areas with greenery and water without human-made elements within cities that provide ecological, recreational, and aesthetic benefits. We prioritized papers that focused on specific greenery and water elements in this section and included those papers that focused on landscape as a whole in the subsequent section on landscape.

5.1.2. Thematic evolution

Early inquiries as well as subsequent works, such as those by Hartig et al. (1997); Wolf (2003, 2006); Chang et al. (2008); Lange et al. (2008); Mayer et al. (2009); Wilkie and Stavridou (2013); Gatersleben and Andrews (2013); Qiu et al. (2013), focused on the restorative, visual quality of greenery in various contexts, and they evolved from simple analysis of linear effects to more nuanced analysis over the years with multiple data sources (Lin et al., 2014; Franěk and Režný, 2014; Kardan et al., 2015; Hunter and Askarinejad, 2015; Wilkie and Clouston, 2015; Lindal and Hartig, 2015; Deghati Najd et al., 2015; McAllister et al., 2017; Neale et al., 2021; Liu et al., 2021b; Qiu et al., 2021; Sztuka et al., 2022; Johnson et al., 2022; Matos Silva et al., 2023), such as its implications in the context of the COVID-19 pandemic (Zabini et al., 2020) and restorative effects of low- and high-level features of greenery images (Celikors and Wells, 2022; Menzel and Reese, 2022). Another research theme was the estimation of greenery’s economic
Figure 8: This plot shows the relationship between the locations of the first authors’ institutions and study areas.
Figure 9: These plots show shares of visual data types, computer vision model purposes, and computer vision model training processes by year. One can observe the increases in street view imagery and computer vision models from around 2015.
value and visual quality in the context of residential neighborhoods Orland et al. (1992); White and Gatersleben (2011); Hofmann et al. (2012); Southon et al. (2017); Xu et al. (2022); Wu et al. (2021b); Schmid and Säumel (2021).

Research on urban water bodies has gained attention. Gabr (2004) examined Cairo residents’ preferences for Nile waterfront design, revealing early interest in urban water aesthetics. Junker and Bucheker (2008) studied the link between ecological quality and aesthetic preferences in Swiss river landscapes. Zhao (2009) analyzed the acoustic and visual perception of Hangzhou’s waterfronts. Further studies assessed waterfront preferences (Dobbie, 2013; Poledniková and Galia, 2021), while recent research employed computer vision and SVI to examine water perceptions (Luo et al., 2022b,c). Community engagement has been crucial in greenery research, involving participatory and experience-based methods since the early years (Kaplan and Austin, 2004; Ryan, 2005; Hadavi et al., 2015; Hami and Maruthaveeran, 2018; Macintyre et al., 2019; Sun et al., 2019; Lis et al., 2022; Martens et al., 2022). Studies have assessed community preferences for urban greenery, combining ecosystem services with qualitative experiences (Heyman, 2012; Juntti et al., 2021; Ryan, 2005), and utilized mixed methods for in-depth analysis (Al-Akl et al., 2018; Gawryszewska et al., 2018; Chen et al., 2018).

As the field advanced, researchers began employing various methods to delve deeper into the more nuanced role of greenery. Numerous studies expanded the understanding of greenery beyond aesthetics to include their restorative effects for stress recovery (Wang et al., 2016, 2019d; Jin et al., 2023; Beute and de Kort, 2018; Wang et al., 2019d; Benita and Tunçer, 2019; Campagnaro et al., 2020;
Figure 11: The most frequently occurring 20 words within each category have been visualized using word clouds to illustrate their prevalence in the respective bodies of literature.
Masoudinejad and Hartig, 2020; Llaguno-Munitxa et al., 2022). Moreover, many papers have also examined mental and physical health outcomes by correlating greenery and their perceptual effects with mental well-being in various contexts (White et al., 2015; Navarrete-Hernandez and Laffan, 2019; Wang et al., 2019a; Cheesbrough et al., 2019; Jiang et al., 2020; Zhang et al., 2021d; Shan et al., 2022; Navarrete-Hernandez and Laffan, 2023). Such investigations have extended to its stress relief effects and perceptions in physical activities, such as walking, running, and cycling (Resch et al., 2020; Xiang et al., 2021a; Huang and Ouyang, 2022; Sun et al., 2022; Brancato et al., 2022; Zhang et al., 2022a; Song et al., 2022a; Wang et al., 2023b,d; Huang et al., 2023a; Zhang et al., 2023a). Spatiotemporal and seasonal variations of greenery’s perceived visual quality have also been studied by multiple papers (Eroğlu et al., 2012; Tang and Long, 2019; Verma et al., 2020).

Technological advancements have enabled sophisticated, large-scale analyses of greenery using big data and crowd-sourced surveys (Quercia et al., 2014; Li et al., 2015; Traunmueller et al., 2016). Virtual reality (VR) has been applied to study the restorative and perceptual effects of greenery (Tabrizian et al., 2018; Gao et al., 2019; Li et al., 2022h; Chen et al., 2022), while eye-tracking (Neilson et al., 2016; Franěk et al., 2019; Li et al., 2020a) and EEG devices (Olszewska-Guizzo et al., 2018; Grassini et al., 2019, 2022; Mavros et al., 2022; Zhang et al., 2023b) have assessed physiological responses. Additionally, face recognition (Qiao et al., 2021; Zhang et al., 2022c) and machine learning models (Acosta and Camargo, 2019b; Min et al., 2020; Larkin et al., 2021; Dai et al., 2021; Zhang et al., 2021e; Larkin et al., 2022; Yang et al., 2022; Liu et al., 2023b), including 3D point cloud analysis (Torkko et al., 2023; Hu et al., 2022), have advanced the understanding of greenery’s effects. Research has expanded to include sensory data beyond visuals in studying urban greenery, exploring audiovisual perceptions and their restorative effects (Carles et al., 1999; Rummukainen et al., 2014; Ali et al., 2021; Xie et al., 2022; Ren et al., 2023; Stevens et al., 2018; Deng et al., 2020). Studies have delved into the sensory and physiological impacts of greenery, examining aspects like air quality (Zhao et al., 2020a) and multi-sensory restoration (Qi et al., 2022). Additionally, the influence of green exercise on attention and mood has been investigated, linking health outcomes with environmental psychology (Zhang et al., 2023a).

Recent research has broadened to explore how various demographic groups perceive and prefer urban greenery, examining the impacts of design, safety, and restorative effects across different socio-demographic profiles (Mysyuk and Huisman, 2020; Ma et al., 2023; Bonthoux et al., 2019; Hami and Tarashkar, 2019; Jin et al., 2023; Yue et al., 2022; Veinberga et al., 2019; Jiang et al., 2017; Lis et al., 2019b,a; Jing et al., 2021; Moreno-Vera et al., 2021b,a; Luo et al., 2021; Lis et al.,
Studies have also delved into the aesthetic and perceptual qualities of greenery, their influence on urban scenic quality, and specific preferences towards greenery types and designs (Gerstenberg and Hofmann, 2016; Santosa et al., 2018; Hwang et al., 2019; Suppakittpaisarn et al., 2020; Kozamernik et al., 2020; Ma et al., 2021a; Zhuang et al., 2021; Tomitaka et al., 2021; He et al., 2023; Fikfak et al., 2022; Li et al., 2022e; Liang et al., 2023a; Sabouri et al., 2023). This evolving field has shown a sustained interest in understanding the nuanced visual and psychological effects of urban greenery, including investigations into potential negative impacts and the comparison of online versus onsite survey results (Xiang et al., 2021b; Lis and Iwankowski, 2021).

5.1.3. Future opportunities

The category of greenery and water is the most popular and has been extensively studied; however, there are still a few areas to be studied in the future. One such area is the examination of different types of greenery and water; for example, vertical greenery and urban wetlands are still under-explored. Future studies can also research practical applications, bridging theoretical insights with actionable urban planning strategies, including the strategic placement of greenery for ecological and mental health benefits. Incorporating advanced technologies, such as advanced deep learning and physiological devices, can facilitate large-scale analyses of greenery’s long-term impacts. Additionally, a focus on under-represented demographics will ensure equitable access to the benefits of greenery and water.

5.2. Street design

5.2.1. Definition and scope

We categorized studies into street design if their scopes include the arrangement of urban roadways, encompassing the layout, infrastructure, materials, and aesthetic elements of streets. This section discusses the following aspects explored by research on street design: safety, aesthetic quality, perceived level of service, advanced methodologies, health implications, and perceptions among different demographic groups.

5.2.2. Thematic evolution

The exploration began with qualitative analyses, such as an investigation of advertising signs’ influences on drivers’ safety (Smiley et al., 2005), aesthetic judgments of selected properties on streetscapes (Weber et al., 2008; Nejad and Ali, 2015), an examination of the role of roads in people’s perception (Garré et al., 2009), the influence of the built environment on people’s crossing decisions (Granié
et al., 2014), and the evaluation of streetscape complexity (Cavalcante et al., 2014). These studies laid the groundwork for understanding the multifaceted perceptions individuals hold regarding street designs.

Research in street design has long focused on the perceived level of service and safety among drivers, cyclists, and pedestrians. Studies have shown that perceptions of service and safety are influenced by both road and traveler characteristics, including travel speed, infrastructure design, and demographics (Flannery et al., 2008, 2005; Dowling et al., 2008; Kang et al., 2013; Raj and Vedagiri, 2022). Specifically, the design of cycling infrastructure, such as physical separations and wider lanes, significantly affects cyclists’ safety perceptions (von Stülpmagel and Binnig, 2022). Virtual reality (VR) has been utilized to assess cyclists’ experiences, highlighting the importance of greenery, lane width, and traffic volume on perceived safety and aesthetics (Bialkova et al., 2018; Agudelo-Vélez et al., 2021). Surveys on cyclists’ safety perceptions with varying infrastructure levels further underline the impact of design elements on urban mobility experiences (Beura and Bhuyan, 2018; Monsere et al., 2020; Desjardins et al., 2021; Fitch et al., 2022; Tian et al., 2021).

Recent years have marked a significant increase in leveraging technology for street design and mobility analysis. Virtual reality (VR) has been utilized to gauge bicycling perceptions (Nazemi et al., 2021), while wearable sensors have assessed the impact of visual street qualities on human comfort and physiology (Gorgul et al., 2019). LiDAR point cloud data has facilitated the mapping of street visual quality (Wu et al., 2021a), and 3D data has contributed to developing a walkability index through subjective visual perceptions (Boongaling et al., 2022; Liao et al., 2022). Additionally, eye-tracking technology has been employed to understand human perceptions towards specific street designs (Spanjar and Suurenbroek, 2020).

Recent studies have emphasized street safety, exploring the effects of streetscape improvements (Carlson et al., 2019), bicycle facilities in vulnerable neighborhoods (Lusk et al., 2019), and pedestrian safety perceptions (Park and Garcia, 2020). Research has also investigated the impact of street design on safety perceptions (Jiang et al., 2018) and how nighttime lighting influences fear of crime (Son et al., 2023). Several studies have utilized street view images to model perceptions of crime and traffic safety, providing insights into urban environments (Naik et al., 2014; Fu et al., 2018; Jing et al., 2023; HE et al., 2022; Wang et al., 2022a; Hollander et al., 2021; Su et al., 2023; Costa et al., 2019; Wang et al., 2022b; Acosta and Camargo, 2019a; Kwon and Cho, 2020). Further research has linked these perceptions to reality, exploring how street design, fear of crime, and actual crime incidents correlate, along with the impact of the built environment’s visual properties on safety perceptions (Zhang et al., 2021a; Su et al., 2022; Moreno-Vera,
SVI and deep learning have been applied to analyze the relationship between human activities and streetscapes (Tao et al., 2022), as well as to assess perceptions related to walkability, physical activities, urban renewal, cultural elements, street vitality, comfort, and playability (Li et al., 2023c; Zhang and Mu, 2020; Li et al., 2020c; Kang et al., 2023b; Zhou et al., 2019; Dong et al., 2023; Ma et al., 2021b; Inoue et al., 2022; Anzai et al., 2021; Li et al., 2022g; Wang et al., 2023e; Shao et al., 2023a; Kawshalya et al., 2022; Kruse et al., 2021; Li et al., 2022i; Song et al., 2022b; Sottini et al., 2021; Feng et al., 2022; Asadia et al., 2023). Notably, Liang et al. (2023b) and Wang et al. (2024b) uniquely utilized SVI’s time-series data to explore changes in street perception over time.

Studies have linked perceived visual quality of streets with physical activities, mental health (Buttazzoni and Minaker, 2022; Luo and Jiang, 2022; Buttazzoni and Minaker, 2023), and restorativeness (Zieff et al., 2018; Vera-Villarroel et al., 2016; Zhao et al., 2020b), indicating a shift towards integrating health and aesthetic considerations in street design. The impact of lighting on pedestrian perceptions has also been explored, highlighting a growing interest in psychological and sensory experiences (Hao et al., 2022). Furthermore, research has investigated the relationship between street aesthetics and economic factors, including housing prices and the impact of economic downturns on visual quality (Kang et al., 2021; Qiu et al., 2022, 2023; Song et al., 2022c; Freitas et al., 2022).

Research increasingly focuses on the subjective safety experiences of specific demographic groups. Studies have utilized photovoice to examine factors affecting female adolescents’ perceptions of traffic safety (Yang, 2023) and investigated gender’s influence on urban transportation security perceptions (Hidayati et al., 2020; Soto et al., 2022; Cui et al., 2023). Additional topics include children’s hazard perceptions in traffic (Meir and Oron-Gilad, 2020), the elderly’s street safety (Wu et al., 2020), international differences in street design preferences (Norouzian-Maleki et al., 2018), and tourists’ streetscape perceptions (Li et al., 2022a). Kang et al. (2023a) highlighted differences in safety perceptions between local and non-local residents using GeoAI.

5.2.3. Future opportunities

Future research can analyze the perceptions of various users by considering their transportation modes and demographic backgrounds. Also, VR technology can enable future studies to design experiments that offer a more tangible exploration of street designs. Furthermore, through visual perception analysis, more research can be done to analyze the impacts of street designs on people’s usage of sustainable transportation modes, such as walking and cycling. With the emergence of new mobility options, such as autonomous vehicles and micro-mobility solutions, there
is also a need for studies on how they influence people’s perceptions of street designs.

5.3. Building design

5.3.1. Definition and scope

We included papers in the category of building design if they focused on architectural planning, aesthetic styling, spatial organization, and functions and effects of buildings. This section reviews the following aspects explored by research on street design: aesthetic quality, perceived safety, psychological effects, utilization of advanced technologies, and health implications.

5.3.2. Thematic evolution

Early research in building design, such as the work by Moore et al. (2006), used multi-method approaches to explore environmental quality and aesthetics, with subsequent studies examining the impact of aesthetics on house prices and color perceptions on building exteriors (Cetintahra and Cubukcu, 2015; Cubukcu and Kahraman, 2008). Mid-2010s research shifted focus towards crime prevention, highlighting the role of maintenance and Crime Prevention Through Environmental Design (CPTED) in enhancing safety (O’Brien and Wilson, 2011; Cozens and Davies, 2013). Computer vision techniques were later used to assess safety and wealth perceptions in building design (Porzi et al., 2015; Ordonez and Berg, 2014), along with crowdsourcing safety perception surveys (Traunmueller et al., 2015). Recent studies have investigated the effects of visual exposure to environments on subjective time perception and the psychological impacts of design, including perceived density and safety, suggesting a deeper understanding of building design’s psychological effects (van den Berg et al., 2003; Berry et al., 2015; Berman et al., 2014; Valtchanov and Ellard, 2015; Martínez-Soto et al., 2014; Kang and Kim, 2019; Kühn et al., 2021; Emo et al., 2017; Li et al., 2020b, 2022f).

The advent of deep learning and AI introduced a new dimension to the field. For example, several works modeled human perceptions and the visual quality of buildings using advanced technologies (Ibarra et al., 2017; Liu et al., 2022; Zhang et al., 2018; Ye et al., 2019; Yao et al., 2019; Wu et al., 2023). This period also saw research that utilizes VR, eye-trackers, and other physiological measurement devices to evaluate people’s reactions to design (Zeile and Resch, 2018; Hollander et al., 2020; Fisher-Gewirtzman, 2019; Wang et al., 2023a; Sakhaei et al., 2023; Chinazzo et al., 2021). Geotagged images and SVI were also used by many studies; for instance, evaluation of soundscape (Zhao et al., 2023), assessment of beauty and color quality in buildings (Saiz et al., 2018; Wan et al., 2022).
In the realm of health and well-being, Wang et al. (2019b) related residents’ perceptions of the built environment to their health outcomes, while other studies (Sadeghifar et al., 2019; Ho and Chiu, 2021) looked at how urban building façades and shapes impact preferences and emotions, respectively. A more recent phase of research has delved into diverse yet specific aspects of design. Karandinos and Turner (2017) linked specific building design characteristics with brain waves detected by electroencephalography (EEG) devices, Hashemi Kashani et al. (2023) built discrete choice models of preferred building façade designs, and Biag (2014) assessed diverse youth perceptions of campus spaces. This research line was followed by Mangone et al. (2017), which focused on natural elements ideal for workplace activities, Hollander and Anderson (2020) and Oludare et al. (2021) who focused on the public perception of building façades, and Oreskovic et al. (2014), Ziani and Biara (2022), and Balasubramanian et al. (2022), who examined walkability and building designs in urban spaces. Visual perception analysis of specific building types has been coupled with other sensory data as well (Yilmaz et al., 2023). The research has also considered aesthetic harmony in historic districts as shown by Zhou et al. (2022), the perception of decayed materials by Wells (2020), people’s perception of traditional architectural styles by Zhang et al. (2020) and Mouratidis and Hassan (2020), the psychological restorative effects examined by Wang et al. (2023c) and the assessment of high-rise buildings’ aesthetic quality by Asur and Yazici (2020), indicating a nuanced approach to the visual and aesthetic evaluation of building design.

5.3.3. Future opportunities

In the face of rapid urbanization, there could be more studies on alleviating perceived density through building design interventions. Another opportunity is the use of generative artificial intelligence (AI) in designing buildings. Future research could explore whether it is possible to harness this new technology to understand how humans view and perceive different generated designs. Such research could help in leveraging new generative AI while incorporating human perception to ensure human-centric building designs.

5.4. Landscape

5.4.1. Definition and scope

Landscape encompasses the holistic design and planning of outdoor areas, blending natural and man-made elements. This category extends beyond the greenery and water’s focus on purely natural elements by integrating human-made features for functional and cohesive environments. It also differs from public space, which focuses on publicly accessible space and may not integrate these two elements. The
landscape category explores user preferences, demographic impacts, safety, and technology in creating inclusive, appealing, and safe outdoor environments.

5.4.2. Thematic evolution

Research in this category has long focused on the diverse preferences and perceptions of landscapes across various demographic groups. Early studies, such as Dearden (1984), explored how familiarity influences landscape preferences, often utilizing questionnaires to understand the impact of people’s backgrounds. Works by Chokor (1990) and Chokor and Mene (1992) delved into landscape preferences among socio-demographic groups in developing countries, emphasizing cross-cultural comparisons. The recreational value of landscape management was also examined (Tahvanainen et al., 2001). Over time, studies such as Brush et al. (2000) investigated rural roadside landscape preferences, highlighting group-specific differences. Yamashita (2002) extended this research by looking at water’s role in landscape perception across age groups, reflecting a growing interest in how specific elements affect landscape appreciation.

During the mid to late 2000s, research on landscape perception emphasized diverse groups’ views (Oku and Fukamachi, 2006). Studies explored rural residents’ landscape attachment (Walker and Ryan, 2008), differences in landscape preferences between native Dutch individuals and immigrants (Buijs et al., 2009), and contrasts between students and non-students (Tveit, 2009), highlighting the socio-cultural factors influencing landscape perception. In the early 2010s, research expanded to include a wide range of demographic factors. Studies explored landscape preferences in suburban Australia (Kurz and Baudains, 2012) and Brazilian students’ views on forest environments (Silva et al., 2010), highlighting how culture and education influence landscape perception. Research has explored factors affecting landscape preferences and perceptions, employing methods like regression analysis to link descriptive indicators with preferences and examining the interplay between soundscape and visual landscape (Sevenant and Antrop, 2010; Yao et al., 2012; Akten and Çelik, 2013; Clay and Smidt, 2004; Liu et al., 2014). Later, studies continued to assess the impact of sociocultural backgrounds (Alizadeh et al., 2015; Taylor, 2018; Hami et al., 2020), including investigations into park landscape preferences among Iranian immigrants (Yazdani, 2019), intergenerational differences in ecosystem service perceptions (Zhang et al., 2022b), and diverse user groups’ views on ecosystem services (Xia et al., 2023).

Research on perceived safety in landscapes (Mahrous et al., 2018) expanded during the COVID-19 pandemic to include studies on landscape’s impact on mental health and human perceptions (Zhang et al., 2021c; Li et al., 2022b; Suppakittpaisarn et al., 2023). Studies demonstrated how place attachment influences restorative
perceptions (Menatti et al., 2019; Liu et al., 2020) and explored landscape characteristics’ effects on perceived restoration (Li et al., 2023b).

Recent research has applied advanced methods to analyze the link between landscape visuals and preferences (Schirpke et al., 2019; Zhang et al., 2021b; Altamirano et al., 2020; Li et al., 2022d), including deep learning and big visual data for large-scale mapping of landscape perceptions (Wei et al., 2022; Song et al., 2023). Additionally, innovative technologies, such as UAVs, VR, and computer vision, have been utilized for quantitative evaluations of visual information’s impact on perception (Luo et al., 2022a). Physiological measurements, such as eye-tracking, have also been employed to assess landscape preferences (Dupont et al., 2017; Liu et al., 2021a; Wu et al., 2021c; Han and Lee, 2022).

5.4.3. Future opportunities

As urban areas evolve and climate change challenges emerge, the necessity for research on how retrofitting existing landscapes to enhance sustainability and resilience affects human perception becomes evident. Future studies can provide insights into how enhancement of existing landscape can change people’s visual perception. Moreover, the field is ripe for more concrete design experiments that harness cutting-edge technologies such as Virtual Reality (VR), unmanned aerial vehicles (UAVs), and deep learning. These tools can provide nuanced insights into human landscape perceptions, facilitating the exploration and evaluation of diverse design interventions before their real-world application.

5.5. Public space

5.5.1. Definition and scope

In the category of public space, we included studies that focus on areas that are open and accessible to all people, such as parks, town squares, and plazas, which do not necessarily include natural elements. We discuss the following aspects examined by research on public space: visual quality, human behavior, attractiveness, the use of advanced technologies, and safety.

5.5.2. Thematic evolution

Research on public spaces in the late 1990s initially focused on the subjective experiences of specific demographic groups, with Day (1999) examining women’s safety perceptions through photography and interviews. By the early 2000s, the field began utilizing technological tools, with Clay and Smidt (2004) comparing expert and public perceptions of visual quality along road corridors and Ryu et al. (2007) exploring public space design and human behavior via VR simulations.
In the mid-2010s and 2020s, research adopted an analytical lens, assessing urban public spaces’ restorative potential and attractiveness through various methods (Van den Berg et al., 2014; Hurtubia et al., 2015; Gargoum and Gargoum, 2021). This period saw a broadening of focus to include the impacts of public spaces on psychological well-being (Hadavi, 2017; Bornioli et al., 2018; Lee, 2022) and emotional responses to soundscapes (Zhang and Kang, 2020). Studies also utilized photovoice to capture diverse demographic perceptions (Ronzi et al., 2016; Gullón et al., 2019) and visual questionnaires to probe public space design perception (Jović et al., 2019).

The use of advanced data sources also became prevalent (Candeia et al., 2017; Rossetti et al., 2019; Ho and Au, 2020; Ramírez et al., 2021; Colombo et al., 2021). Bernetti et al. (2020) integrated GIS and remote sensing data to assess urban spaces, while van Renterghem et al. (2019) utilized VR to study natural sounds in parks. Bernetti et al. (2020) used SVI and social media data to identify characteristics of urban public spaces in a data-driven way. A few recent articles put a significant emphasis on the analysis of socio-demographic groups. Navarrete-Hernandez et al. (2021) used photo simulations to enhance safety perceptions, focusing on women in public spaces. Veitch et al. (2021), Rivera et al. (2021), and Gómez-Varo et al. (2023) both investigated children’s and adolescents’ perceptions of park features.

More diverse research was conducted recently. Szczepańska and Pietrzyk (2021) evaluated the seasonal impact on public space perception, while Zhao et al. (2022) integrated stated preference experiments with virtual environments to understand the cognitive and affective components of momentary experiences in public spaces. Navarrete-Hernandez et al. (2023) combined fear of crime measurements with photo simulations for urban regeneration plans. Chen and Biljecki (2023) explored public open spaces using SVI and GIS data. Palmieri (2023) analyzed the psychological impact of color perception in urban parks.

5.5.3. Future opportunities

Future research on public spaces could explore the application of augmented reality (AR) in understanding and enhancing urban visual perception. For instance, Nijholt (2021) discussed how AR can augment social activities in public spaces. Moreover, Wang and Lin (2023) studied how AR could promote public participation in urban designs in public spaces, successfully confirming positive feedback from the participants and higher engagement from the public. Future studies can also investigate how such use of AR affects people’s visual perception in public space.
5.6. City as a whole

5.6.1. Definition and scope

For the category of the city as a whole, we selected studies that examine urban environments holistically, rather than focusing on specific aspects (i.e., other categories). This section discusses the following aspects studied by research in this category: people’s view towards cities, safety, health complications, and the use of advanced technologies.

5.6.2. Thematic evolution

The research field concerning the city as a whole has developed a complex understanding of how human perceptions and urban functionalities intertwine, progressing toward a more sophisticated analysis of urban perception over time. In the latter part of the 20th century, a nascent study by Garling (1972) laid the basis for indicators used in urban perception studies with visual data, which was followed by Nasar and Hong (1999), who investigated the role of retail signage’s obtrusiveness in perception. Moving into the early 21st century, Moore et al. (2008) focused on capturing the experiences and views of residents in city centers, marking a shift towards a more human-centric view of urban research.

The mid-2010s saw an emphasis on safety and perception with Kang and Kang (2015) developing context-aware predictions for urban safety by considering objects in images, and De Nadai et al. (2016) along with Dubey et al. (2016) examining the connection between the appearance of safety and activity levels in urban environments with computer vision techniques. By the late 2010s and early 2020s, the research had evolved to include the spatial distribution of perceptions and urban functions (Yao et al., 2021), and the pathways linking neighborhood safety perception to mental health (Wang et al., 2019c). This period also saw the development of sophisticated models for predicting urban perceptions (Xu et al., 2019; Li et al., 2022; Wang et al., 2021) and more critical research on the prediction of visual perception with SVI (Beaucamp et al., 2022; Li et al., 2023a).

The most recent studies have leveraged more advanced technologies such as virtual realities and eye-tracking devices to gain more accurate and explainable insights into human perception in urban environments (Tabrizian et al., 2020; Vainio et al., 2019). Huang et al. (2023c) has contributed to this evolving field by correlating environmental perception measured from SVI with human activities.

5.6.3. Future opportunities

Future research on the city as a whole could benefit greatly from validating big data approaches (e.g., SVI and computer vision models to predict human perceptions) to better understand ground truth. Validating these methods ensures
the reliability of insights drawn from vast amounts of data. Moreover, embracing more diverse locations and demographics is crucial. To overcome the difficulty of collecting diverse data, establishing an open-source and open-data platform that facilitates continuous assessment — for example, contributors can rate images and, in return, extract data proportional to their contributions — could democratize and enrich urban perception research. Additionally, gamification of perception data collection presents an innovative strategy to engage a broader audience in urban studies. By transforming data collection into an interactive and rewarding process, researchers can tap into a wider pool of perceptions across different times, locations, and demographic profiles. Rijsdouwer and van Zoonen (2023) has demonstrated such an approach and reported that they were able to collect a larger number of responses.

6. Discussion

This study revealed a few common limitations and future opportunities in the urban visual perception literature across different aspects as well as in the field of the automation of systematic reviews.

6.1. Challenges and opportunities in urban visual perception

The review sections above highlighted some common limitations that need to be overcome by future studies across different topics. The most mentioned issue is a limited scope of study areas, sample sizes, and attributes of studies and their findings, which can also be confirmed by Figure 4, Figure 6, and the plot on the right side of the first row in Figure 5. It can be inferred that this issue is caused by the time- and resource-intensiveness of data collection, especially subjective data. An intuitive solution might be the promotion of open data to reduce the costs required to obtain similar data that have already been used by previous studies. However, as highlighted in the plot on the left side of the third row in Figure 5, studies rarely use publicly available data possibly because of a lack of high-quality instances, except for the Place Pulse dataset published by Dubey et al. (2016). Such scarcity of open data can be observed in the plot on the left side of the fifth row in Figure 5. To further advance the urban visual perception literature as a whole by reducing the time and resources required for data collection, more attention should be paid to standardization and methodology of data collection and sharing.

Another shared issue reported by the reviewed studies is the quality of visual and subjective data and different types of biases: 1. data bias and 2. perception bias. As shown in the plot on the left side of the second row in Figure 5, many studies utilized SVI as a source of visual data due to its wide coverage and high resolution;
however, some studies have raised their concerns about this dataset. Firstly, data biases have been discussed in the literature, and they can cause inaccuracy in the downstream analysis. A review of SVI studies by (Biljecki and Ito, 2021) reported various issues with the quality of the imagery, such as blurriness, poor lighting, heterogeneous weather conditions, and occlusions. These issues directly affect those studies that use SVI to assess urban perception; for example, poor lighting and heterogeneous weather conditions can cause bias in the data and may not accurately reflect what people perceive at actual locations. Another limitation of SVI is biases in perspectives. When analyzing urban perceptions by non-vehicle drivers, such as pedestrians and cyclists, a typical SVI perspective from a vehicular road does not reflect their perspectives (Ito and Biljecki, 2021). As many studies rely on SVI as a source of visual data, these issues also need to be overcome to establish the reliability of the results of their perception assessments. One possible approach is to collect and utilize Volunteered SVI (V-SVI) platforms such as Mapillary¹ and Kartaview² (Juhász and Hochmair, 2016; Mahabir et al., 2020; Ma et al., 2019). V-SVI has different coverage and perspectives from commercial SVI services such as Google Street View (GSV) because anyone and anywhere can capture images, which are often not covered by GSV, for example, sidewalks and cycle tracks (Hou and Biljecki, 2022; Leon and Quinn, 2019; Ding et al., 2021). A few studies have taken such approaches, but more studies can take advantage of its flexibility (Luo et al., 2022a,b).

The second bias is a perception bias. A few studies have reported potential bias (e.g., response bias and recall bias) and confounding factors in subjective measures of perception based on self-reported questionnaires (Oku and Fukamachi, 2006; Bonthoux et al., 2019). To mitigate these issues with data quality, future studies need to consider alternative data sources to complement current mainstream data. For visual data, more immersive visual experiences can be used for future studies; for example, VR and videos taken from first-person points of view are currently under-utilized. And subjective perception data can be combined with more objective measures, such as EEG and eye-tracker, to further analyze beyond simple regression analysis with self-reported data.

Another research gap is a lack of studies that closely examine the impact of visual features on human perceptions. Most previous studies simply conduct regression analysis without removing possible confounding factors, such as selection bias and endogeneity. A more thorough examination of the relationships between

¹https://www.mapillary.com
²https://kartaview.org
visual elements and perception can help policy-makers, urban planners, and urban designers implement appropriate interventions that can bring intended outcomes for the target population. Future research can utilize several underutilized data sources. One such data source is time-series SVI, which has only been used by a few studies (Liang et al., 2023b; Wang et al., 2024b). Another source is more carefully collected perception data in controlled environments, which have been enabled by recent technical advancements in facilitating 3D simulation and VR (Dosovitskiy et al., 2017), although it is also important to note that VR might not be the same as the real world settings. These datasets can be further coupled with physiological measurement data (e.g., brain waves from EEG, sweat amount from electrodermal activity (EDA) devices, and eye movements from eye-tracking devices) to gain deeper insights into human perception. Aerial imagery can also be a more popular source of visual input as drones become more prevalent and processing methods have been developed rapidly in recent years; for example, Gaussian Splatting has been an increasingly popular model to reconstruct 3D scenes from images and can enable researchers to collect aerial images with drones, shift perspectives to the ground level, and analyze visual perception from human eye-levels (Xiong et al., 2024). The collected data can be used for various topics by other studies by sharing them more openly while securing anonymity of the data — such collective accumulation of data can further improve the quality and speed of research by overcoming the aforementioned issues.

A lack of consideration towards local knowledge and experience is a major limitation in the current research trend as well. It has been reported that sociocultural ties and local knowledge and experience of the study areas affect the sense of place, which plays a key role in forming people’s perception of the place (Pánek et al., 2020; Kang et al., 2023a), and ignoring these factors might lead to inaccurate assessments when researchers are interested in the local population’s perceptions. This issue could be seen as a combination of limitations discussed so far — a lack of attribute-rich data on participants’ socioeconomic backgrounds and local knowledge, limitations of SVI, and failure to control for confounding factors. Thus, this limitation perhaps requires collective efforts of researchers in the field to overcome.

A risk around personal information involved in urban visual perception is another challenge. As mentioned above, future opportunities lie in collecting more attributes about the participants to increase the generalizability of the analysis; however, this raises concerns about the secure protection of personal information. As more and more studies explore such a direction, researchers need to be careful about how to handle sensitive data. Currently, only 40% of the reviewed papers
mentioned that they obtained approvals from institutional review boards, which is low given that more than 80% of the studies used their own collection of data (see the plots on the right side of the third row and fifth row in Figure 5). To enhance the overall reliability and ensure adherence to ethical practices in surveys and experiments, future studies should take more careful steps toward following guidelines and protecting sensitive data.

6.2. Challenges and opportunities in automation of systematic reviews

Despite the benefits of utilizing a large language model to automate this systematic review, several points need to be discussed. The first limitation of this study is a lack of transparency because of the use of non-open-source products. The architecture and training process of OpenAI’s GPT-4 model used in this study has not been either published as a scientific paper or released to the public. While we validated the approach to ensure reliability, this lack of transparency makes it difficult to discuss potential problems with the results and to improve them. In addition to the lack of transparency, the hallucinations (i.e., generation of non-factual information) by LLM have raised concerns among researchers (Kang et al., 2023c; Ye et al., 2023). We used a technique called retrieval-augmented generation to ensure that the generated information is based on the content of the paper, and this technique has been widely adopted by many researchers for its accuracy and reliability of the generated information (Xu et al., 2023; Shao et al., 2023b; Liu et al., 2023a; Li et al., 2022c; Chen et al., 2023).

Additionally, despite its relatively low cost — US$0.002 / 1K tokens as of March 23, 2023 — of the model, the use of a commercial product might limit the use of the method for certain researchers. These issues can potentially be overcome by using an open-source alternative such as Taori et al. (2023) and Anand et al. (2023); however, it is also important to note that it is unknown whether the open-source model performs sufficiently well to conduct the systematic review. This leads to the next issue of the difficulty of rigorous validation of the accuracy and quality of the outputs. Previous studies on the automation of systematic reviews mostly focused on the screening phase, so it was relatively straightforward to benchmark their results (van de Schoot et al., 2021). However, the assessment of data extraction and summarizing tasks involves subjective evaluation; thus, a future research opportunity can be an extensive examination of the method used in this study to check if it is close enough or better than the human performance by conducting surveys.

Our systematic review involved manual work in deriving different categories for the review papers, which could have potentially introduced subjectivity bias. Future studies can also implement more data-driven classification of papers by utilizing text
analysis techniques, for example, text embedding, clustering, and topic modeling. Such an approach can further streamline the review process, enabling researchers to conduct reviews more efficiently.

Finally, future studies need to be careful and transparent about the above-mentioned limitations when they conduct systematic reviews with the assistance of LLM and other machine-learning techniques. Despite their usefulness, efficiency, and accuracy (i.e., the GPT-4 model’s 97-100% accuracy for summarization, information retrieval, and our systematic review validation) in processing a large amount of information created by a rapidly increasing number of papers (Vectara, 2023), we believe that researchers must keep their research ethics by clarifying the extent of their LLM utilization.

7. Conclusion

Urban visual perception is an extensive and rapidly growing field of research, but there has not been any comprehensive systematic review paper to consolidate the developments and identify the current and future research directions. Our study reviewed 393 papers in the urban visual field to produce a comprehensive review in the field and demonstrated the possibility of automated systematic review for the first time by employing advanced machine learning models and large language processing techniques. This innovative approach allowed for efficient and comprehensive screening, data extraction, and reporting phases, paving the way for similar endeavors in the future when dealing with topics described by a rich body of literature.

Our review included a wide spectrum of categories, including landscape, street design, walkability, urban vitality, public space, greenery and water, building design, and several others. Notably, this study summarized and analyzed insights from 393 papers, offering an unparalleled width and depth in reviewing the urban visual perception field. Through the review, several limitations within the urban visual perception research were identified, including concerns about generalizability, data quality, and the need for causal inferences. The challenges related to the automation of systematic reviews were also dissected, such as the need for transparency, validation rigor, and ethical considerations.

In conclusion, this paper contributes to the field in two significant ways. First, it offers the most extensive review in the domain of urban visual perception, amalgamating insights from various research threads. Second, it pioneers the use of automated systematic review techniques, demonstrating their potential efficacy and highlighting areas for improvement. The insights and methodologies presented herein can serve as pivotal references for researchers, urban designers, and policy-
makers, guiding future investigations and applications in the realm of urban visual perception and beyond.

Acknowledgements

The first author is thankful for funding provided by the Singapore International Graduate Research Award scholarship. We thank the members of the NUS Urban Analytics Lab for the discussions. This research is part of the project Large-scale 3D Geospatial Data for Urban Analytics, which is supported by the National University of Singapore under the Start-Up Grant R-295-000-171-133. This work was supported by the National Natural Science Foundation of China under Grant 42371468. We thank the editor and reviewers for their comments. Statement: During the preparation of this work the author(s) used ChatGPT and GPT-4 turbo in order to proofread and summarize content as described in the paper. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.
References


Anand, Y., Nussbaum, Z., Duderstadt, B., Schmidt, B., Mulyar, A., 2023. GPT4All: Training an Assistant-style Chatbot with Large Scale Data Distillation from GPT-3.5-Turbo.


gkamradt, 2024. Gkamradt/LLMTest_NeedleInAHaystack.


Hami, A., Tarashkar, M., 2019. Women’s perception towards interrelation between plants’ attributes with restorative effects in urban parks of Tabriz, Iran. Alam Cipta 12.


Hughes, S., Bae, M., 2023. Vectara Hallucination Leaderboard.


Li, C., Yuan, Y., Sun, C., Sun, M., 2022b. The Perceived Restorative Quality of Viewing Various Types of Urban and Rural Scenes: Based on Psychological and Physiological Responses. Sustainability (Switzerland) 14. doi:10.3390/su14073799.


Li, X., Zhang, X., Jia, T., 2023b. Humanization of nature: Testing the influences of urban park characteristics and psychological factors on collegers’ perceived


Shao, Y., Yin, Y., Xue, Z., Ma, D., 2023a. Assessing and Comparing the Visual Comfort of Streets across Four Chinese Megacities Using AI-Based Image


Song, Q., Li, M., Qiu, W., Li, W., Luo, D., 2022b. The Coherence and Divergence Between the Objective and Subjective Measurement of Street Perceptions for Shanghai. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 13725 LNAI. doi:10.1007/978-3-031-22064-7_19.


Sun, F., Xiang, J., Tao, Y., Tong, C., Che, Y., 2019. Mapping the social values for ecosystem services in urban green spaces: Integrating a visitor-employed


van Renterghem, T., Sun, K., Filipan, K., Vanhecke, K., de Pessemier, T., de Coensel, B., Joseph, W., Botteldooren, D., 2019. Interactive soundscape augmentation of an urban park in a real and virtual setting. Proceedings of


Yang, Y., 2023. "Because anyone can misuse this (a gun) and they can use it for weapon": A Photovoice Exploration of Reasons Why Female Adolescents Feel Unsafe in Rural and Suburban Communities. Violence and Gender 10. doi:10.1089/vio.2021.0084.


Yazdani, N., 2019. The effects of cultural background and past usage on Iranian-Australians’ appreciation of urban parks and aesthetic preferences. Landscape Online 70. doi:10.3097/Lo.201970.


Listing 2: Questions used for this systematic review.

""

Below is a list of questions. Please refer to the example answer formatted in JSON format to see how you should respond to all the questions in one JSON object.

- Answer DOI and title of the paper in a JSON format.
- Summarize the study in 3 bullet points in a JSON format based on introduction, conclusion, and abstract. 3 points should cover: Purpose/aim/objective of the study, method, and findings.
- Choose which aspect of the built environment this study examined: Green and blue space, Street design, Building design, Landscape, Public space, City as a whole, and others. If it’s "others", please provide the appropriate aspect that the study examined after "others:"
- Answer in which country(s) and city(s) this study was conducted (i.e. the study area/site(s) that this research collected data from) in a JSON format with a Python list. The number of the list elements may vary depending on the information in the source document.
- Choose the extent/scale of the study area from the following options: individual image level, building level, neighborhood level, district level, city level, country level, or not applicable
- Answer the spatial data aggregation unit (i.e., a unit of analysis).
- Answer the type, data source(s), and data size(s) of visual data used to assess perception in a Python list of JSON objects. For types of visual data, choose from the following: street view images, other geo-tagged photos, non-geotagged photos, aerial images, video, virtual reality, and others. For the visual data source, provide sources of the images/videos: i.e. where they collected them, and specific service names if possible (e.g., Google 89
Street View, Baidu Map, Flickr). For the number/volume of images, please answer how many images or how much data volume (e.g., GB) the study used.

- Answer the sampling interval distance between each street view image used in this study (e.g., 15m interval, etc). Answer "not applicable" if the study didn’t use street view imagery. Answer "not mentioned" if the study used street view imagery but not report sampling interval distance.

- Answer the data source(s), data collection method(s), and the number of participants of the subjective perception data in a Python list of JSON objects. For the data source, choose from the following: their own collection, publicly available data, and others. If it’s publicly available data, provide their names or citations in parentheses (e.g., Place Pulse 2.0 dataset). For the subjective data collection method, choose the method of how they collected the subjective data: survey/questionnaire, observation, physiological signals, and others. For the number of participants, provide the number of participants/raters in the survey/questionnaire. Use comma (,) between answers to return multiple answers if needed.

- Answer the data sources of other sensory data in a Python list of JSON objects. Other sensory data types should include smell/olfactory data (e.g., air quality), texture (e.g., vibration), sound/auditory (e.g., noise/acoustic data), or not applicable if the study didn’t use other sensory data. For other sensory data sources, describe where and how they collected data or write "not applicable" if the study didn’t use other sensory data.

- 1. Answer if this study is quantitative or qualitative research (i.e. type of research). 2. Explain the method in a JSON format. Include the following points and be as specific as possible:
data collection, data processing, analysis (e.g., modeling)

- Answer the type of analysis in this research in a Python list of JSON objects. For the type of analysis, choose from the following options: regression, model development, index construction, exploratory analysis, and others. If it’s "others", please provide the appropriate research type after "others:"

- If the study used computer vision models, answer the model architecture name(s), purpose(s), and training procedure(s) of the computer vision model(s) used in this study in a Python list of JSON objects. For the model architecture names, provide the specific names of architectures (e.g., ResNet-50) or answer "not applicable" if the study didn’t use any computer vision model. For the purpose of the model, choose from the following options: object detection, semantic/instance segmentation, image classification, feature extraction, others, or answer "not applicable" if the study didn’t use any computer vision model. For the training procedure, choose from the following options: pre-trained without fine-tuning, pre-trained with fine-tuning, trained from scratch by themselves, or others, or answer "not applicable" if the study didn’t use any computer vision model. Use comma (,) between answers to return multiple answers if needed.

- Choose the availability of the code used in this study in a JSON format from the following options: code available via URL (e.g. GitHub), code available upon request, code available with restrictions, code is not available, not mentioned, others. If you choose "code available via URL", make sure the URL is a link to a version control repository, for example, Git Hub.
Choose the availability of the data used in this study in a JSON format from the following options: data available via URL, data available upon request, data available with restrictions, data not available, not mentioned, and others. The URL needs to be a link to the whole dataset used by the author via a data host service, such as Google Drive (https://drive.google.com/), Harvard Dataverse (https://dataverse.harvard.edu/), Figshare (https://figshare.com/). Other types of URLs (e.g., social media data providers) should not be considered.

Answer whether the study obtained research ethical approval from an institutional review board (IRB) with only "Yes" or "No" in a JSON format.

1. Explain the limitations of this study in a JSON format with a Python list based on results, discussion, and conclusion. 2. Explain future research opportunities based on the limitations and findings of this study in a JSON format with a Python list.

Example Answer:

```json
{
    "paper_details": {
        "DOI": "XXX",
        "Title": "XXX"
    },
    "study_summary": {
        "Purpose": "XXX",
        "Method": "XXX",
        "Findings": "XXX"
    },
    "built_environment_aspect": "XXX",
    "study_area": [
        {
            "Country": "XXX",
            "City": "XXX"
        }
    ]
}
```
"extent_scale": "XXX-level",
"spatial_data_aggregation_unit": "XXX",
"image_data": [
{
  "Type_of_image_data": "XXX",
  "Image_data_source": "XXX",
  "Number_Volume_of_images": "XXX"
},
...,
],
"sampling_interval_distance": "XXX",
"subjective_perception_data": [
{
  "Subjective_data_source": "XXX",
  "Subjective_data_collection_method": "XXX",
  "Number_of_participants": "XXX"
},
...,
],
"other_sensory_data": [
{
  "Other_sensory_data_type": "XXX",
  "Other_sensory_data_source": "XXX"
},
...,
],
"research_type_and_method": {
  "Type_of_research": "XXX",
  "Method": {
    "Data_collection": "XXX",
    "Data_processing": "XXX",
    "Analysis": "XXX"
  }
},
"analysis_type": [}
Listing 3: Prompt format used in this systematic review.

```python
prompt = f"""Use the following pieces of context to answer the question at the end. If you don't know
```
```
the answer, just say that you don’t know, don’t try to make up an answer.

Paper Context:
{paper_content}

Question: {questions}

Important Note:
- Please answer the question solely based on the Paper Content and follow the specified format.
- If the information needed to answer a question is not found in the document, respond with ‘NA’ to prevent misinformation.
- In the examples, *XXX* is a placeholder for the actual answer.
- In the examples, ‘...’ means there could be more than one set of answers.
- Please answer the questions based on the Paper Content only.

Answer:""