論文 / 著書情報
Article / Book Information

題目(和文)	エビデンスに基づくウォーカブルな街路デザインに向けて：人々の知覚と嗜好を反映した街路レベルのウォーカビリティのための体系的・自動的アプローチ
Title(English)	Toward an Evidence-Based Walkable Street Design: A Systematic and Automatic Approach for Street-Level Walkability Reflecting People's Perceptions and Preferences
著者(和文)	HuangLu
Author(English)	Lu Huang
出典(和文)	学位:博士(学術), 学位授与機関:東京工業大学, 報告番号:甲第12927号, 授与年月日:2024年9月20日, 学位の種別:課程博士, 審査員:沖 拓弥,藤井 晴行,大佛 俊泰,斎尾 直子,松岡 昌志
Citation(English)	Degree:Doctor (Academic), Conferring organization: Tokyo Institute of Technology, Report number:甲第12927号, Conferred date:2024/9/20, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

Doctoral Dissertation

August 2024

Toward an Evidence-Based Walkable Street Design: A Systematic and Automatic Approach for Street-Level Walkability Reflecting People's Perceptions and Preferences

Department of Architecture and Building Engineering

School of Environment and Society

Tokyo Institute of Technology

Ph.D. Candidate: HUANG Lu

Supervisor: Dr. OKI Takuya

Corresponding Published Paper (peer-reviewed)

Huang, L., Oki, T., Muto, S., Kim, H., Ogawa, Y. and Sekimoto, Y., 2023, June. Automatic Evaluation of Street-Level Walkability Based on Computer Vision Techniques and Urban Big Data: A Case Study of Kowloon West, Hong Kong. In *International Conference on Computers in Urban Planning and Urban Management* (pp. 231-259). Cham: Springer Nature Switzerland. **[Corresponds to Chapter 4]**

Huang, L., Oki, T., Muto, S. and Ogawa, Y., 2024. Unveiling the Non-Linear Influence of Eye-Level Streetscape Factors on Walking Preference: Evidence from Tokyo. *ISPRS International Journal of Geo-Information*, *13*(4), p.131. **[Corresponds to Chapter 5]**

Huang, L., Oki, T. Enhancing People's Walking Preferences in Street Design Through Generative Artificial Intelligence and Crowdsourcing Surveys: The Case of Tokyo. **[To be submitted] [Corresponds to Chapter 6]**

Oral Presentation Related

Huang, L., Oki, T., Muto, S. and Ogawa, Y. 2024. Capturing Walking-Related perceptions and willingness within Tokyo's station areas: leveraging Crowd-Sourced Methods and AI approach, *GISA & IAG'i 2023.* https://confit.atlas.jp/guide/event-img/gisa2023/E1-01/public/pdf?type=in.

Curriculum Vitae

Education Backgrounds

PhD Candidate Architecture and Building Engineering, Tokyo Institute of Technology, Tokyo, Japan 09/2021-Present

Academic Supervisor: Dr. Takuya OKI

MUD Urban Design, University of Hong Kong, Hong Kong SAR, China 09/2019-09/2020

MA Landscape Architecture, University of Sheffield, Sheffield, UK

09/2013-10/2015

BA Public Design, Beijing University of Technology, Beijing, China

09/2007-06/2011

Work Experience

Lecturer Academy of Architecture and Arts, Guangxi Arts Universality, Nanning, China 2015- 2018

Landscape Architect Architecture Design Institute of China Agricultural University, Beijing, China 2011- 2012

Contents

List of Figures

List of Tables

CHAPTER 1: INTRODUCTION

1. 1 Research Backgrounds

1.1.1 The importance and development of walkability

Walkability refers to how conducive an urban environment is to pedestrian activity (Moura, et al., 2017). A walkable environment plays a crucial role in encouraging urban residents to engage in walking activities, delivering a wide range of social, economic, and health benefits (Hu, et al., 2007; Cavill, et al., 2007; Hu, et al., 1999; Hu, et al., 2001; Williams & Thompson, 2013). Given the profound influence of walkability on everyday life, cities around the world are increasingly emphasizing eco-friendly and human-centric walking environments in their urban planning strategies. In many major Asian cities, efforts to enhance walkability have accelerated in recent years. For example, since 2020, the Tokyo Metropolitan Government (TMG) has embarked on initiatives to transform streetscapes, focusing on walkability and centering around the needs of residents (TMG, 2023). This involves reimagining and converting existing streets into pedestrian-friendly zones across various Tokyo districts, underscoring Tokyo's commitment to becoming a more accessible and pedestrian-oriented megacity (MLIT, 2021). Similarly, in Hong Kong, the Transport Department has undertaken extensive research to create planning and design standards that prioritize pedestrian needs, along with a strategic plan to transform Hong Kong into a world-class walkable city (Ng, et al., 2016).

1.1.2 The concept and focus of walkability in this dissertation

Walkability can be divided into two distinct but interrelated categories: physical walkability and perceived walkability (Jun & Hur, 2015). Physical walkability encompasses the objective, quantifiable aspects of an environment that facilitate walking, such as the presence and quality of sidewalks, density of destinations, availability of amenities, and overall street design. In contrast, perceived walkability refers to the subjective evaluation of how walkable an area feels to individuals, including factors such as the safety of the streetscape, comfort while walking, perceived interestingness (Adkins, et al., 2012; Alfonzo, 2005; Koo, et al., 2022), and whether people are inclined to walk or linger in the area (Jones & Boujenko, 2009; Jones, et al., 2008; Matthews, et al., 2004).

In the past, discussions on improving walkability primarily focused on enhancing physical attributes (Kang, et al., 2023). Urban planners and policymakers concentrated on improving the physical environment through measures such as adding and

repairing sidewalks, increasing commercial facilities along streets, and enhancing lighting and resting amenities. However, with the growing influence of human-centric approaches, the perceived aspect of walkability has received increasing attention. It has become evident that the lack of discussion regarding the public's environmental experiences and feedback is insufficient to truly enhance the walking experience. By addressing people's actual perceptions and experiences, urban design can more effectively promote healthy and sustainable walking activities, creating environments that are not only walkable but also enjoyable for pedestrians. This shift is based on the understanding that a well-perceived walking environment is crucial for increasing pedestrian activity (Alfonzo, 2005). Consequently, it is essential to grasp "what people perceive and prefer" regarding walking environments (Choi, et al., 2016).

Moreover, research on walkability has shifted from focusing on macroscale and mesoscale factors, such as the 3Ds (Density, Diversity, Design) and 5Ds (Density, Diversity, Design, Destination accessibility, Distance to transit), to a street-level scale (Cain, et al., 2012; Boarnet, et al., 2006). This scale is more closely aligned with human visual perception, which is the most critical way human sense and interact with surrounding environment. Street-level factors, such as detailed streetscapes and well-proportioned spatial structures, have increasingly captured the interest of urban planners and designers (Cain, et al., 2012). Moreover, street-level factors offer distinct advantages in terms of adaptability and cost-effectiveness. They are comparatively easier, quicker, and less expensive to modify, enabling timely interventions to bolster active transportation and physical activity (Cain, et al., 2012; Boarnet, et al., 2006). This shift highlights the growing understanding that the fine-grained details of the street environment significantly impact walkability and the overall pedestrian experience.

This dissertation focuses on street-level elements and examines both the physical and perceived aspects of walkability, with particular emphasis on the visual dimension, which is central to the design of this research.

1.1.3 The enhancement of walkability design by Artificial Intelligence (AI) technology and urban big data

For a long time, the challenge of quantifying street-level walkability through on-site surveys has hindered the widespread adoption of evidence-based design practices (Koo, et al., 2022). The perceived aspect of walkability was even considered immeasurable (Ewing & Handy, 2009). Addressing these challenges is crucial for effectively integrating evidence into street design.

The fusion of big data and AI technologies over the past decade has spurred a paradigm shift, expediting the adoption of evidence-based approaches in urban street design. Among these, recent advancements in increasing the availability of street view image datasets present a unique opportunity (Yan & Ryu, 2021; He & Li, 2021; Zeng, et al., 2018). The expansive coverage of street view data enables comprehensive fine-grained research, providing imagery of streets across major cities globally, often accessible to the public via API services. Simultaneously, advancements in computer vision (CV) technology, including segmentation, detection, and even AI-generative methods empower researchers and designers to automatically quantify or create visual elements in street scenes in a more scalable manner compared to traditional approaches (Koo, et al., 2022). Regarding the measurement of environmental perceptions and preferences, online crowdsourcing methods are emerging alongside street view and deep learning algorithms to quantify previously unmeasured perceptual dimensions, enhancing our subjective understanding of the built environment (Ordonez & Berg, 2014; Dubey, et al., 2016; Zhang, et al., 2018; Oki & Kizawa, 2022; Ogawa, et al., 2024). This capability facilitates a deeper and more nuanced comprehension of urban spaces. These emerging data and methods present significant opportunities to efficiently and accurately provide evidence to guide design decisions, enriching the design process with empirical data and insights.

Although imagery Big Data and AI algorithms offer substantial opportunities, significant gaps still exist. Most current research primarily highlights the potential for general urban environmental measurement using new urban data and AI algorithms (Koo, et al., 2022). However, there is a noticeable lack of systematic studies on how these imagery big data and AI algorithms can be specifically applied to evaluate walkability, in order to gather evidence that illustrates the current state of both physical and perceived walkability.

Additionally, when it comes to leveraging the large-scale evidence to inform the design of walking environments that align with people's subjective perceptions and preferences, there is a significant gap in discussing which factor is most important and how it impacts people’s responses. Regression analysis is instrumental in quantifying and exploring the relationships between streetscape factors and walking preferences and perceptions. This statistical approach offers invaluable insights, enabling urban planners and designers to comprehend how different design elements influence pedestrians' preferences and perceptions in urban spaces. Numerous studies have investigated the relationship between street-level factors and walking preferences or perceptions, primarily employing linear regression models (Borst, et al., 2009; Shatu,

et al., 2019; Basu & Sevtsuk, 2022; Guzman, et al., 2022; Nagata, et al., 2020). However, these studies frequently overlook the complex and multidimensional nature of the relationships between streetscape factors and pedestrian experiences. In real-world contexts, these relationships often exhibit non-linear characteristics and are influenced by various environmental factors. Understanding evidence solely from a linear regression perspective may lead to naive and overly simplistic conclusions, which can negatively impact informed design decisions. Additionally, there is a lack of discussion on how different types of streets might produce varying effects on walkability (Boarnet, et al., 2006; Cain, et al., 2012). This oversight highlights the need for more sophisticated analytical methods capable of capturing and modeling the intricate dynamics of how pedestrians interact with their surroundings.

Recently, advancements in AI generative techniques, particularly those leveraging Stable Diffusion models, have emerged as highly promising tools for automatically and efficiently translating public perceptions and preferences into spatial design representations (Mishra, et al., 2023). Compared to traditional methods based on regression and other statistical analyses, the Stable Diffusion method offers a more direct, visible, and intuitive approach for designers and stakeholders to understand and visualize new street design scenarios, significantly enhancing the design decision-making process (Kapsalis, 2024). Despite their potential, discussions on how to leverage Stable Diffusion to enhance walking perceptions and preferences of urban spaces, particularly in street design, are still in their infancy. There is a notable lack of discussion on establishing a reliable workflow for this integration. Developing such a workflow could pave the way for more effective and responsive urban street design, tailored to the nuanced preferences and perceptions of the public.

1.2 Research Questions

This dissertation focuses on walkability and spans multiple academic domains. The core domain is urban design, investigating how to optimize street environments through design to enhance the walking experience. The aim of this dissertation is to bridge existing gaps by answering the following questions:

Question 1: How can urban big data and AI algorithms be used to systematically and automatically capture both physical and perceived walkability features at street-level?

Question 2: How can nonlinear statistical methods enhance our understanding of the impact of physical attributes on perceived walkability at street-level, and how can this knowledge guide better street design?

Question 3: How can AI generative methods achieve automated street design and help enhance perceived walkability at street-level?

1.3 Conceptual Framework

Figure 1.1 presents a framework that showcases the interrelationships among the core concepts and dimensions discussed in this dissertation. In this conceptual framework, we focus on the walkability of streets, considering both the physical and perceived aspects of walkability.

Physical walkability encompasses factors that emphasize streetscape elements, particularly those related to visual aspects. These include factors that reflect the structural features of the street, such as street-to-building ratio and sidewalk-to-roadway ratio, as well as detailed elements like greenery, buildings, benches, street lights, street stores, and crosswalks. Perceived walkability, on the other hand, refers to the subjective evaluation of how walkable an area feels to individuals (Kang, et al., 2023). In this dissertation, we aim to understand perceived walkability from two perspectives: perceptual preference and behavioral preference.

For perceptual preference perspective, we utilized Alfonzo's hierarchy of walking needs model to establish the evaluation dimensions (Alfonzo, 2005). Alfonzo's model (2005) proposes a hierarchy organizing five levels of perceptual needs that influence the walking decision-making process: feasibility, accessibility, safety, comfort, and interestingness or pleasurability. Of these, safety, comfort, and interestingness are directly related to street-level walking environments (Adkins, et al., 2012; Alfonzo, 2005; Koo, et al., 2022). Thus, we focus on these three dimensions to represent perceived walkability. Chapter 4 of this dissertation delves into this perspective in detail.

When exploring from the behavioral preference perspective, we analyze it through the lens of street functionality and its ability to influence people's preferences in two types of walking behaviors: walking preference and lingering preference. On one hand, the concept of "walkability" primarily refers to how well streets are perceived to be designed to facilitate pedestrian traffic, thereby attracting people to use them as pathways (Frackelton, et al., 2013). This aspect fundamentally represents the streets' perceived walkability. Recently, voices have advocated that an ideal streetscape

should not only support walking but also act as a center for social activities, enhancing a sense of place and promoting social interactions (Jones & Boujenko, 2009; Jones, et al., 2008; Matthews, et al., 2004). This principle, often referred to by contemporary urban planners as "sticky streets," describes areas where pedestrians are naturally encouraged to linger and engage more actively in vibrant public life (Zapata & Honey-Rosés, 2022). Therefore, this study concentrates on two dimensions—preferable for walking and preferable for lingering —to capture the public's subjective preferences concerning walking environments. Compared to perceived safety, comfort, and interestingness, the behavioral preference's perspective provides more direct insights into walkability. Chapter 5 of this dissertation is based on this perspective to understand perceived walkability and conduct research accordingly.

However, these two perspectives are not independent; they are interrelated. Existing research suggests that safe, comfortable, and interesting streets can enhance people's preference to walk and linger (Lizárraga, et al., 2022). Chapter 6 of this dissertation explores perceived walkability from both perspectives and conducts research based on it.

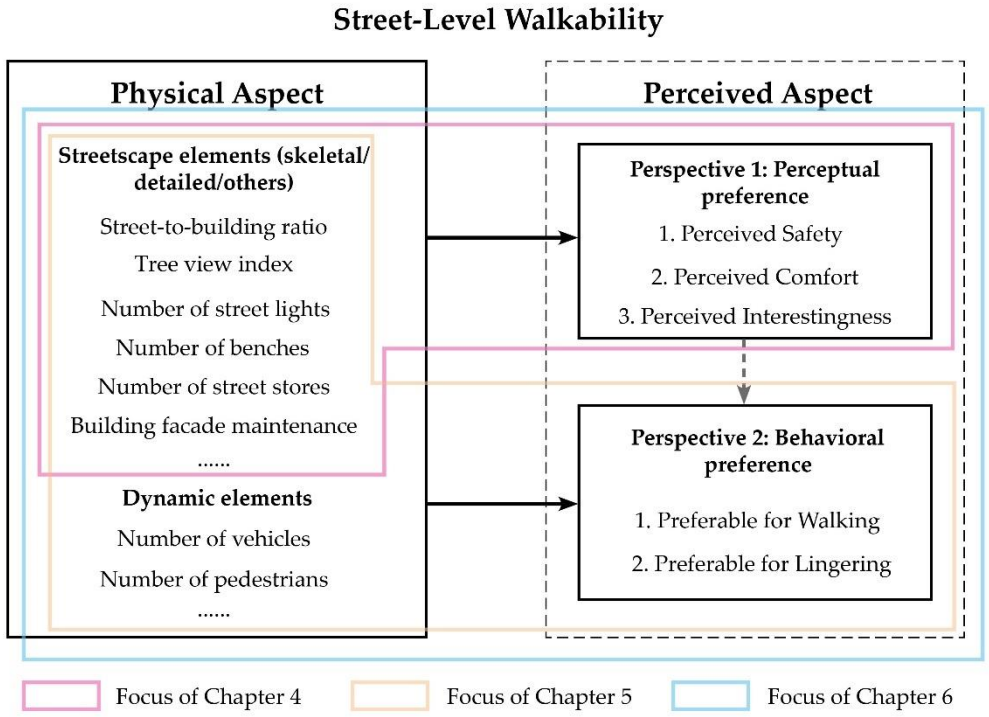


Figure 1.1 *The Conceptual framework*

1.4 Dissertation Organization

The structure of this dissertation is illustrated in Figure 1.2. Chapter 1 provides an overview of the background, research questions, and conceptual framework of the study. Chapter 2 reviews existing research and literature in the field, establishing the academic context for the dissertation. Chapter 3 summarizes the data acquisition processes and analysis methods used in the dissertation. The individual studies are presented in Chapters 4 to 6.

In Chapter 4, we delve into the exploration of street-level factors with the potential for automated measurement, focusing on both physical walkability and perceived walkability. We propose automated measurement methods using urban big data and CV techniques. By applying these proposed measurement factors and methods, we can gain insights into the physical conditions of street-level walkability and understand how people subjectively respond to them.

In Chapter 5, we further investigate how the physical attributes of street-level walkability, as automatically revealed, influence people's perceived walkability, which in this chapter refers to walking preferences. We also explore how to obtain deeper insights and evidence that can guide design practices. Building on this foundation, we employ non-linear regression analysis to examine the relationship between the streetscape and walking preferences. We then discussed how our findings can be leveraged to inform perceived walkability in street design practices.

Considering that evidence from non-linear statistics still lacks intuitiveness and specific guidance for design in particular scenarios, Chapter 6 explores the emerging Stable Diffusion method and proposes a workflow to support automatic street design. This workflow aims to enhance perceived walkability throughout the design process.

Chapter 7 offers a comprehensive discussion and conclusion, summarizing the key findings and exploring the significance and contributions of the dissertation.

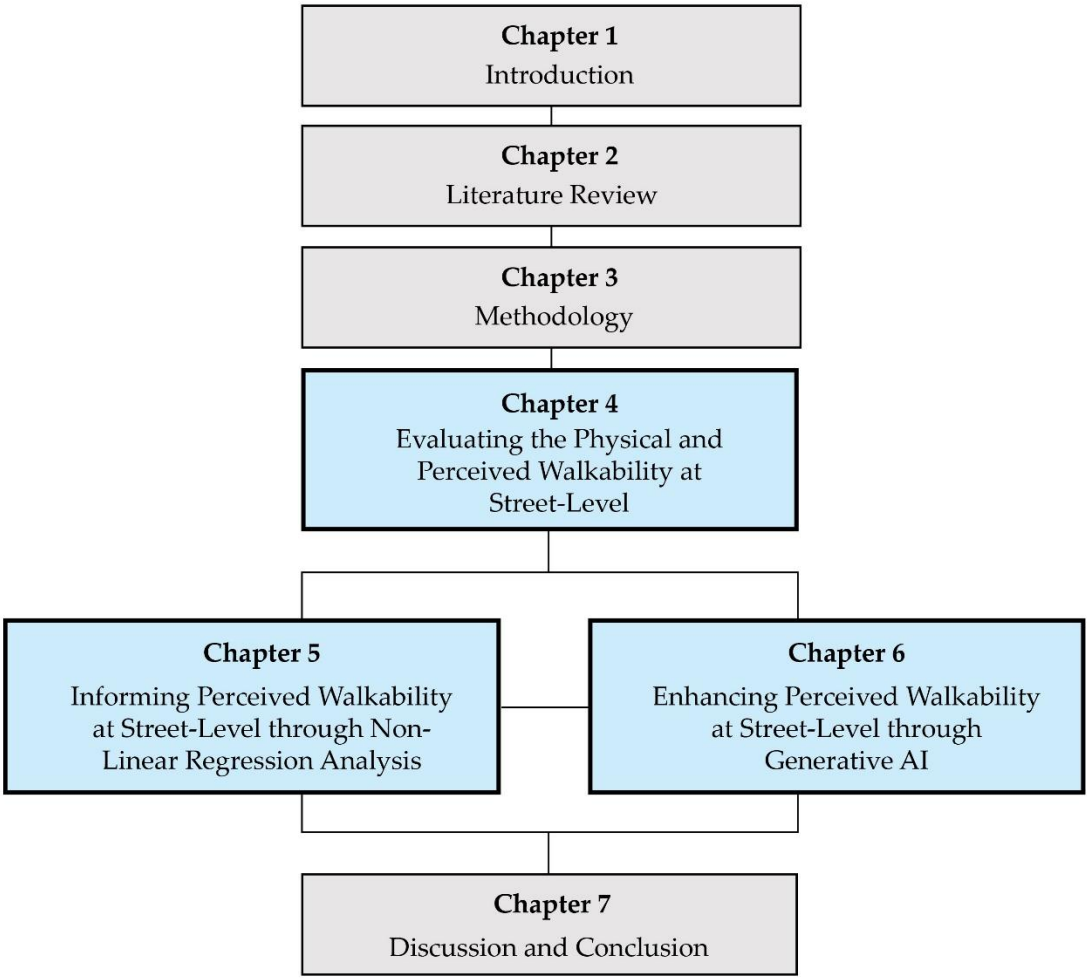


Figure 1.2 *The outline of dissertation*

CHAPTER 2: LITERATURE REVIEW

2.1 Selecting and Measuring the Street-Level Walkability

2.1.1 Selecting street-level factors

Many studies have focused on identifying the factors that affect walkability and individual preferences for walking-friendly environments. Fundamental factors often include macro-scale factors, such as urban form characteristics (i.e., density, diversity, and design) (Cervero & Kockelman, 1997), as well as destination accessibility and proximity to transit systems (Cervero, et al., 2009). However, the focus on immediate street-level factors is growing due to their cost efficiency and simpler modifications compared to macroscale factors (Cain, et al., 2012). These factors, including street width, building height, architectural style, greenery, pedestrian-focused design, street furnishings, and other fine-grained elements, enhance the aesthetic and functional qualities of streetscapes (Handy, et al., 2002). These factors are pivotal in shaping key urban design characteristics, including imageability, sense of enclosure, human scale, and transparency (Cain, et al., 2012). Several assessment tools were designed to measure the walkability and health characteristics of local environments based on the existing research findings. These tools include the Irvine Minnesota Inventory (IMI) (Boarnet, et al., 2006), Microscale Audit of Pedestrian Streetscapes (MAPS) (Cain, et al., 2012), Analytic Audit Tool and Checklist Audit Tool (Brownson, et al., 2004), Healthy Aging Network (Belza, et al., 2017), Walking Suitability Assessment Form (WSAF) (Emery, et al., 2003), PIN3 Neighborhood Audit Instrument (Evenson, et al., 2009), Neighborhood Sidewalk Assessment Tool (NSAT) (Aghaabbasi, et al., 2017), and a set of indices specifically focused on high-density Asian contexts (Ng, et al., 2016). However, there remains a lack of discussion regarding which factors and variables are suitable for automated measurement in the current data-driven context.

In addition to studies that investigate the physical factors of urban walking environments, several research efforts have focused on the environments people perceive as conducive to walking (Wang, et al., 2012; Mateo-Babiano, 2016; Humpel, et al., 2004; Vallejo-Borda, et al., 2020). Alfonzo's (2005) model outlines a hierarchy of five perceptual needs shaping the walking decision-making process: feasibility, perceived accessibility, safety, comfort, and interestingness or pleasurability. The aspects of safety, comfort, and interestingness are especially vital because they have a direct influence on street-level walking behaviors and environments (Adkins, et al., 2012;

Alfonzo, 2005; Koo, et al., 2022). Other researchers, like Mateo-Babiano (Mateo-Babiano, 2016) highlighted mobility, safety, ease, accessibility, and pleasure as key components satisfying pedestrians. Furthermore, some researchers have analyzed walkability through the lens of walking preferences. In this context, "walkability" refers to the extent to which streets are perceived as being well-designed to support pedestrian movement, enticing individuals to choose these pathways (Frackelton, et al., 2013). This perception is a core element of perceived walkability. There is a growing movement toward designing streetscapes that not only promote walking but also serve as centers for social engagement, thereby enhancing community ties and facilitating social interactions (Jones & Boujenko, 2009; Jones, et al., 2008; Matthews, et al., 2004). Modern urban planners often call these areas "sticky streets," denoting places where pedestrians are more inclined to linger and engage in vibrant public life (Zapata & Honey-Rosés, 2022).

Although there is an understanding of the concept and dimensions of perceived walkability, there is still a lack of discussion on which dimensions should be adopted for automated measurement.

2.1.2 Measuring physical walkability at street-level

Recent methodologies are increasingly shifting from traditional techniques like surveys and site visits for assessing walkability, toward more technologically driven approaches such as semantic segmentation of street view images. Zhou et al. (2019) utilized Baidu Street View (BSV) images to develop visual walkability indices that encompass psychological greenery, visual crowdedness, outdoor enclosure, and visual pavement quality. Similarly, Li et al. (2020) introduced the Walkability on Urban Street (WoUS) index, incorporating seven indicators: walk score, pedestrian flow density, noise, light, greenery, enclosure, and relative walking width, with the latter three evaluated using segmented street view images. Meanwhile, Shammas and Al Shammas (2019), Yin (2017), and Limgomonvilas and Nimanong (2018) applied GIS technology and geodata to measure urban walkability. Yin and Wang (2016) applied Artificial Neural Network and Support Vector Machine to segment the sky pixels and studied the relationship of the visual enclosure with pedestrian numbers and Walk Scores. Zhou et al. (2019), Li et al. (2020) and Koo et al. (2022) used semantic segmentation to segment streetscapes such as trees, obstacles, sky, sidewalk pavements, roads, buildings, and fences and used them to represent characters of micro walkability. Shortly after, Koo et al. (2022) proposed applying instance segmentation with GIS to measure the presence status of micro streetscapes using the MAPS-min index.

The advancement of Street View images (SVIs) and CV technology significantly expands the potential for integrating diverse objects from SVIs with GIS data to evaluate the physical walking environment. However, reliance on a single method, such as semantic segmentation, still presents limitations and inadequacies when quantifying walkability indicators.

2.1.3 Measuring perceived walkability at street-level

For an extended period, methods relying on quantitative measures, like stated-preference surveys and behavior mapping, have been employed in collecting public perceptions. Some comprehension of the connection between spatial perception and the associated design elements has been attained (Ewing & Handy, 2009; Gehl, et al., 2006). Nevertheless, these investigations are frequently carried out on manually gathered, limited datasets utilizing labor-intensive and time-consuming methods.

With the advancement of online crowdsourcing methods and CV technologies, an increasing number of studies are dedicated to automatically or semi-automatically assessing people's visual perceptions and preferences. These investigations encompass various aspects including aesthetics (Joshi, et al., 2011), memorability (Isola, et al., 2011), interest (Dhar, et al., 2011), and vitality (Deza & Parikh, 2015). This emerging research, leveraging urban imagery to gain insights into cities, has been facilitated by new data sources from commercial providers like Google Street View (GSV) and photo-sharing platforms. Ordonez et al. (2014) and Dubey et al. (2016) delved into GSV data from multiple cities, presenting extensive crowdsourcing datasets (Place Pulse) featuring pairwise image comparisons. They also proposed a neural network architecture for predicting human-labeled comparisons. Zhang et al. (2018) improved the perception prediction model and applied it to large-scale predicted urban senses using Tencent Street Views in Beijing. Oki and Kizawa (2022) proposed 22 dimensions to characterize urban perceptions and calculated them utilizing GSVs. Ogawa et al. (2024) further quantified urban perceptions in Tokyo using Zenrin street views.

However, existing perception prediction methods based on deep learning and street view data often focuses on general perceptions and lacks attention to specific aspects of walkability.

2.2 Understanding the Relationship between Street-Level Factors and Perceived Walkability

Studies have comprehensively evaluated the impact of diverse streetscape factors on pedestrians and explored the interplay between these factors. For instance, Gallimore et al. (2011) examined the link between walk-friendly routes to educational institutions and the frequency of walking among students in neighborhood settings. They identified certain IMI factors (Boarnet, et al., 2006), particularly those of street-level factors, that were correlated with increased instances of students walking to school. Numerous studies have investigated how streetscapes affect pedestrian perceptions and experiences (Ewing & Handy, 2009; Harvey, et al., 2015; Asgarzadeh, et al., 2012; Asgarzadeh, et al., 2014). Built environment characteristics at the street-level shape feelings of safety, comfort, and interest, thereby modifying pedestrian behavior (Ewing & Handy, 2009; Adkins, et al., 2012). Borst et al. (2009) proposed that the visual appeal of streets for pedestrians is influenced by three primary elements: cleanliness, aesthetic appeal, and the presence of pedestrian activities. Harvey et al. (2015) noted that streets bordered by buildings and greenery are typically perceived as safer than open and bare ones. Asgarzadeh et al. (2014) revealed a preference for low-rise buildings over high-rise ones among pedestrians. Similarly, Agrawal et al. (2008) focused on individuals commuting to railway stations and observed that, while safety and visual appeal were influential, the most critical factor was route directness. Li et al. (2015) suggested that the presence of green spaces could mitigate crime and boost pedestrians' perceived security. Rodrigue et al. (2022) highlighted the various determinants of perceived walkability in the built environment depending on the purpose of the trip.

However, these studies have failed to fully consider the intricate underlying interrelations among the built environment, walking behavior, and perceptions. To address this gap, research has increasingly employed machine learning methods to unearth complex, non-linear associations between these variables. For instance, Yin et al. (2023) explored the non-linear dynamics of various walking objectives. Tao et al. (2020) used non-linear approaches to examine the impact of built environments on active commuting and walking patterns among older adults. Similarly, Cheng et al. (2020) investigated the non-linear influences on walking duration among older adults. Furthermore, Wu et al. (2023) analyzed the non-linear interactions between streetscape attributes and the propensity for school walking in Hong Kong. Nevertheless, despite the growing body of work, studies focusing on the non-linear relationships between

streetscape factors and their effects on walking preferences, particularly those considering variations in street morphologies and structures, remain relatively scarce.

2.3 Using Generative AI to Inform Design

2.3.1 Development of automatic methods for supporting design progress

Automated design visualization involves using technology and software to create visual representations of urban designs automatically. This crucial process allows for the previewing of how urban structures will look before they are built. Historically, the initiation of urban design depended on manual sketches. However, the development of computer technology has brought about tools such as computer-aided design software and Building Information Modeling techniques. These technological advancements have simplified and automated the complex design process, significantly improving the overall workflow of design (Weber, et al., 2022; Abrishami, et al., 2020; Yu & Choi, 2023).

Deep Learning has made significant strides and has become indispensable across a wide range of fields, spanning speech recognition, natural language processing, visual object recognition, and detection (LeCun, et al., 2015; Song, et al., 2020). Notably, the emergence of Generative Adversarial Networks (GANs) in 2014 revolutionized image generation research. Research indicates that supervised GAN algorithms, such as DC-GAN, pix2pixGAN, Urban-GAN, CycleGAN and StyleGAN, excel in acquiring knowledge about urban morphological layouts (Quan, 2022; Fedorova, 2021; Wu & Biljecki, 2022; Huang, et al., 2022; Wu, et al., 2022) and authentic street-level environmental characteristics (Wijnands, et al., 2019; Wu, et al., 2022; Noyman & Larson, 2020; Kim, et al., 2022; Knuutila, 2023; Yamanaka & Oki, 2022). Nevertheless, GAN models are not without their shortcomings. They demand substantial computational resources, generate samples with limited diversity, struggle with capturing long-term dependencies and ensuring global consistency, and feature intricate architectures and training procedures.

2.3.2 Stable Diffusion in urban design and analysis

In recent years, the development of Stable Diffusion models has been a significant milestone in various fields, including arts, physics, mathematics, and spatial design. Stable Diffusion is a generation technique based on Latent Diffusion Models (LDMs) (Rombach, et al., 2022). It can generate better outcomes for image generation than the GAN model, such as unconditional image synthesis, image restoration (inpainting), super-resolution, text-to-image and image-to-image generation. Random Gaussian

noise can be gradually denoised after training. As a result, Stable Diffusion models have emerged as powerful tools, offering valuable insights into a broad spectrum of phenomena and facilitating decision-making across diverse fields. While applications in urban design and planning are still relatively limited, notable examples include Mishra et al. (2023), who proposed a methodology for minimal intervention in current road design under human supervision to mitigate accident risk, and Ma and Zheng (2023) who introduced method for generating building facades based on the Stable Diffusion model. Kapsalis (2024) discussed the fundamental application methods of Stable Diffusion in urban design, as well as the potential of plugins like inpainting.

The Stable Diffusion method is emerging as a promising approach to urban design. However, its application in street design remains relatively limited, and there is scarce research investigating its use as evidence based on perception or preference.

CHAPTER 3: METHODOLOGY

3.1 Data Collection

This dissertation utilized street view image data, GIS data, and crowdsourcing survey data as the main sources of information.

3.1.1 Street View Imagery

In recent years, with the development of CV and automated analysis technologies, street view imagery has gradually become an effective form of big data for analyzing urban characteristics at the street-level.

Street view data capture a panoramic representation of real-world environments, typically collected mainly using specialized vehicles equipped with 360-degree panoramic cameras. This data encompasses a wide range of urban elements, including streets, buildings, landscapes, traffic signs, and both artificial and natural geographic features. Prominent providers of street view data, such as GSV, BSV, and Mapillary, offer extensive archives of global locations. In this dissertation, GSV was employed to gather streetscape images from Japan, while BSV was used for collecting similar data in Hong Kong. Both platforms deliver immersive visual experiences and exhibit considerable overlap in the types of features they document, as detailed in Table 3.1. Considering that our research also utilized Zenrin images for survey collection, we have included this image dataset in our comparisons as well.

Table 3.1 *Comparison between GSVs, BSVs and Zenrin Street Views.*

	Google Street View (GSV)	**Baidu Street View (BSV)**	**Zenrin Street View**
Coverage Area	Worldwide, including major cities and rural areas.	Primarily covers cities and some rural areas in China.	Primarily covers cities in Japan.
Launch Year	2007	2013	--
Image Feature	360-degree panoramic views.	360-degree panoramic views.	360-degree panoramic views.
Camera Height	Around 2.05m	Around 2.3m	Between 2 – 2.2m
Image data density	--	--	2.5m

Platform Integration	Integrated with Google Maps and Google Earth.	Integrated with Baidu Maps, tailored for the Chinese market.	Integrated with Zenrin Map for the Japanese market
Developer API	Offers extensive APIs for integrating street view data.	Provides APIs, but with fewer features compared to Google.	--
Privacy Protection	Automatically blurs faces and license plates, with user reporting tools.	Similar privacy protection with automatic blurring of faces and plates.	The images have not yet undergone privacy protection processing.
Data Update Frequency	Regular updates.	Less frequent updates.	Very less frequent updates.
Additional Features	Historical street view allows viewing changes over time.	Provides real-scene navigation and street view tour features.	--
Application in this dissertation	Chapter 5 and 6	Chapter 4	Chapter 4, 5 and 6
Main Usage in the Dissertation	Evaluation, Statistics and Image Generation	Evaluation	Questionnaire
Reference	(Google Maps Platform, n.d.)	(Baidu Maps, 2022)	(Zenrin, 2022)

3.1.2 Street View collection

For both GSV and BSV images, we employed similar strategies and identical parameters for data collection. Initially, we gathered street network data for the target cities. For Hong Kong, we used publicly available road network data from Hong Kong Transport Department. We set a 20-meter interval for capturing SVIs and their corresponding coordinates in Kowloon West District, using QGIS version 3.22.3. This interval aligns with Gehl's concept of the human scale (around 25 meter) in outdoor urban spaces (Gehl, 1987), ensuring a detailed and human-centric perspective. By utilizing the Baidu API, we obtained approximately 14,000 panoramic street view images during the shooting period from 2017 to 2018. For GSV, we based our image capture points on the Digital Road Map (DRM) and set the interval at 30 meters, subsequently capturing approximately 50,000 panoramic images at selected coordinates in Setagaya (shooting period from 2021 to 2022).

3.1.3 GIS data collection

In this dissertation, we incorporated GIS data as supplementary information to enhance our measurements. Table 3.2 details the GIS data used in this dissertation, including the data characteristics and sources.

Table 3.2 *the differences between GSVs and BSVs*

GIS Data	Description	Source
Hong Kong		
Hong Kong Road Network	Line data in SHP format, containing classification attributes related to roads.	Hong Kong Transport Department (https://data.gov.hk/en/)
Hong Kong POI	Point data in SHP format, including classifications for various commercial, residential, office, and government public service locations.	Baidu Open Map (https://www.openstreetmap.org/) (https://data.gov.hk/en/)
Hong Kong Road Slope	5-meter resolution	DEM (https://data.gov.hk/en/)
Tokyo		
Tokyo Road Network	Line data in SHP format, including classification attributes related to roads.	DRM (https://www.drm.jp/english/)
Tokyo Building Outline	Zmap Town II, line data in SHP format	Zenrin (https://www.zenrin.co.jp/english/index.html)
Tokyo Building Height	3D data model in FBX format	PLATEAU by MLIT (https://www.mlit.go.jp/plateau/)

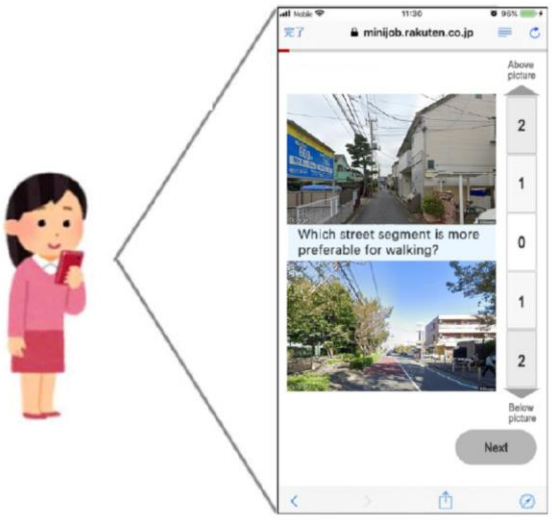
3.1.4 Crowdsourcing survey based on pairwise image comparison

We utilized two crowdsourcing surveys in this dissertation (Table 3.3 and Figure 3.1), identical in structure, gathered through the same online crowdsourcing method based on a mobile platform. The first dataset focused on 22 dimensions of environmental perception, including the three key perceptual needs identified in our research: safety, comfort, and interestingness. This crowdsourcing survey was conducted in December 2021 and garnered 400,000 responses (Ogawa, et al., 2022). The second crowdsourcing dataset captured people's behavioral preferences across 2

dimensions: the preference to walk and the preference to linger. The second survey was conducted in February 2023 and received 180,000 responses (Huang, et al., 2024). Both datasets used street view images in Tokyo as the survey data. Both datasets involved showing image pairs to participants and asking them questions such as, " Which street looks safer?" or " Which street is preferable for walking?" to gather their visual attitudes towards street view image pairs (Figure 2a). During the process, participants were asked to compare the images using a five-point scale, ranging from "Upper image is much better than the bottom image," "Upper image is somewhat better than bottom image," "Both are the same," "Bottom image is somewhat better than the upper image," to "Bottom image is much better than the upper image."

Table 3.3 *The comparison between two crowdsourcing surveys using in our study*

	Crowdsourcing Survey 1	**Crowdsourcing Survey 2**
Survey Focus Dimensions	Open, friendly, vibrant, comfortable, calm, safe, interesting, bright, traditional, liveable and so on.	Preferable to walk in street segments; preferable to walk at street intersections; preferable to linger in street segments; preferable to linger at street intersections and so on.
Survey Image Features	Front/back view perspective (not distinguishing between street segments and intersections)	Front/back view perspective (distinguishing between street segments and intersections)
Number of Survey Images	1000 street view images	1200 street view images (1000 belong to street segments, 200 belong to street intersections)
Number of Survey Image Pairs	40,000 pairs	10,000 pairs for street segments; 2000 pairs for street intersections
Survey Conducted Period	December 2021	February 2023
Survey Conducted Country	Japan	Japan
The total number of respondents in the survey, and the total number of responses.	Number of respondents: 38,525 people; Total Responses: Around 400,000 （The number of questions each respondent can answer is not limited）.	Number of respondents: 18,000; Total Responses: around 120,000（ The number of questions each respondent can answer is not limited）.
Chapters Utilized in This Dissertation	Chapter 4 and 6	Chapter 5
Reference	(Ogawa, et al., 2022)	(Huang, et al., 2024)



Crowdsourcing survey 1 - Perceputal preference

	Dimension
D1	Perceived safety
D2	Perceived comfort
D3	Perceived interestingness

Crowdsourcing survey 2 - Behavioral preference

	Dimension (Segment)
D4	Preferable to walk
D5	Preferable to linger

	Dimension (Intersection)
D6	Preferable to walk

*Collect attitude data on a mobile phone platform

Figure 3.1 *Interface and question dimensions of the crowdsourcing survey.*

3.2 Research Method

This dissertation involved employing CV techniques for automated parsing of image semantics, predicting perceptual and behavioral preferences, and generating images. Additionally, the machine learning regression method was utilized to explore the complex non-linear relationships of physical and perceived walkability. Furthermore, the CRITIC method was applied for comprehensive walkability assessment. Additionally, various GIS tools were used for the analysis of POI and street network data of different types. Table 3.4 records the specific analysis methods used for achieving each objective in each chapter.

Table 3.4 *The summary of the analysis methods in this dissertation*

Chapter	Research Objective	Research Method	Method type
Chapter 4	Objective 1a	Expert Consult and literature review	Qualitative
	Objective 1b	Semantic Segmentation	CV
		YOLOv5	CV
		Maintenance Level Prediction Method	CV
		Sidewalk Width Estimation Method	CV
		Topological Analysis and Other GIS Method	GIS

		CRITIC method	Traditional statistic
	Objective 1c	Perception Prediction Method	CV
Chapter 5	Objective 2a Objective 2b	Panoptic Segmentation	CV
		Perception Prediction Method	CV
		XGBoost	ML statistic
	Objective 2c	Interpretative analysis	Qualitative
Chapter 6	Objective 3a	Perception Prediction Method	CV
		Stable Diffusion	CV

CHAPTER 4: EVALUATING THE PHYSICAL AND PERCEIVED WALKABILITY AT STREET-LEVEL

This chapter is based on:

Huang, L., Oki, T., Muto, S., Kim, H., Ogawa, Y. and Sekimoto, Y., 2023, June. Automatic Evaluation of Street-Level Walkability Based on Computer Vision Techniques and Urban Big Data: A Case Study of Kowloon West, Hong Kong. In *International Conference on Computers in Urban Planning and Urban Management* (pp. 231-259). Cham: Springer Nature Switzerland.

4.1. Introduction

Street-level walkability factors, whether related to physical aspects or perceived qualities, are crucial (Cain, et al., 2012; Boarnet, et al., 2006). However, for a long time, the large-scale measurement and evaluation of these factors have been a significant challenge due to the substantial human and time costs involved. Recent breakthroughs in urban big data and AI technologies, especially in CV techniques, have created numerous opportunities for rapid, large-scale evaluations of street-level pedestrian environments.

While several street-level walkability indexes exist (Boarnet, et al., 2006; Cain, et al., 2012; Brownson, et al., 2004; Belza, et al., 2017; Emery, et al., 2003; Evenson, et al., 2009; Aghaabbasi, et al., 2017), and AI techniques have opened new avenues for automatically evaluating city environments, the field of automatic walkability evaluation is still in its early stages. One key challenge is that the majority of existing walkability indexes, which encompass a rich array of factors and indicators for assessing micro-walkability, are designed for manual evaluation and are not readily adaptable for automated studies. Consequently, most automated evaluations at street-level tend to rely on a limited set of factors for which automation has been proven feasible. Furthermore, while semantic segmentation, the most popular technique for automatic measurement at this level, can calculate the pixel ratio of streetscape elements, this singular approach may not effectively represent the variation characteristics of all relevant factors. In addition, some streetscape factors traditionally thought to influence pedestrian behavior—such as sidewalk width, maintenance level,

and pedestrian flow—lack appropriate automated quantification methods. Regarding the quantification of perceived walkability factors, the integration of crowdsourcing methods with deep convolutional neural networks (DCNNs) has made it possible to predict visual perceptions and preferences automatically (Ordonez & Berg, 2014; Dubey, et al., 2016; Zhang, et al., 2018; Oki & Kizawa, 2022; Ogawa, et al., 2022). Nevertheless, there is a lack of clarity regarding which aspects should be considered factors related to perceived walkability, a topic that remains relatively underexplored.

Given the gaps in the existing literature and the aim to answer the research question: How can urban big data and AI algorithms be used to systematically and automatically capture physical and perceived walkability features? The following research objectives need to be achieved:

First, to select appropriate physical and perceived walkability factors that have the potential to be automatically measured (1a);

Second, to propose measurement methods to measure physical walkability factors at street-level (1b);

Third, to propose measurement methods to measure perceived walkability factors at street-level (1c);

4.2 Methodology

The analysis framework of this chapter is depicted in Figure 4.1. The process initiates with the selection of walkability factors, both physical and perceived, based on input from an expert panel. Subsequently, we used Kowloon West in Hong Kong as a case study site, categorizing selected physical factors and developing a suitable measurement method for each category. Finally, we proposed methods for measuring the walkability factors related to perceived walkability.

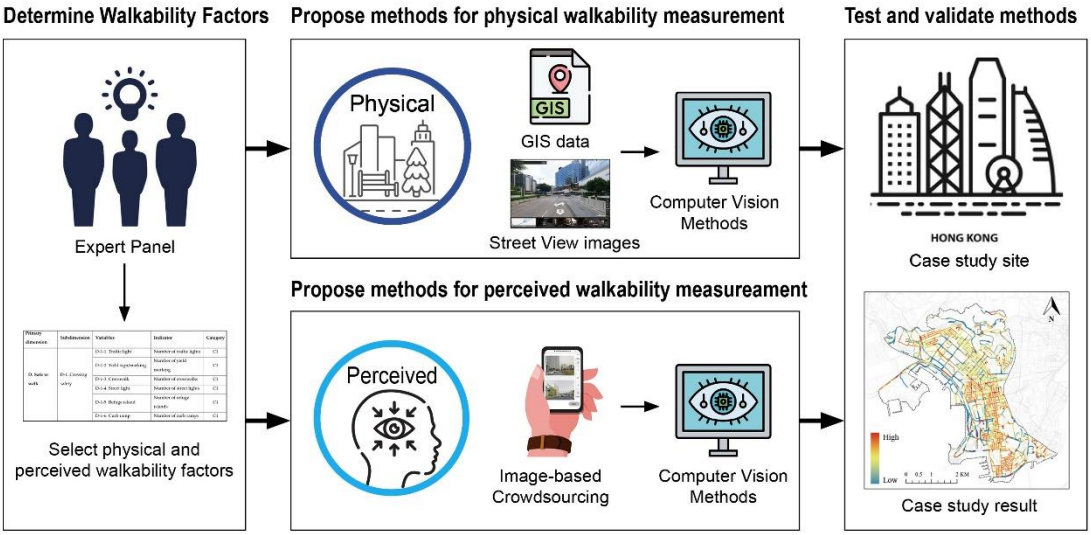


Figure. 4.1 *The Analysis framework*

4.2.1 Selecting walkability factors for automatic evaluation

Study site and scope

The case study site, Kowloon West, occupies the western part of the Kowloon Peninsula in Hong Kong (Figure 4.2). It features a coastline to the west and south, with extensive land reclamation in the waterfront area. The region's primary urban arteries trace the coastal line, while the topography becomes markedly steep to the north and northeast, where the terrain abuts mountains. The central area is characterized by high-density building and street layouts.

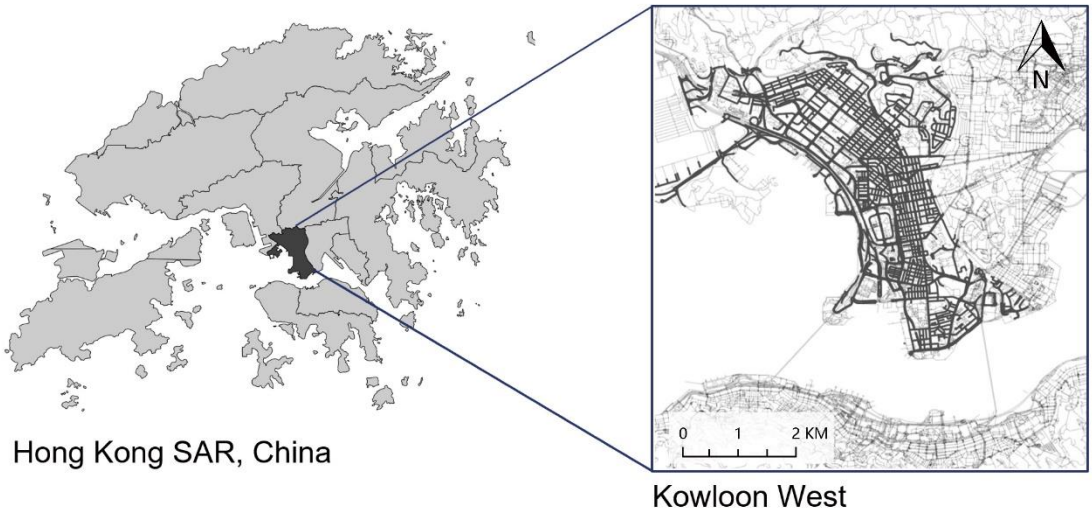


Figure 4.2 *Location of case study area*

Despite Hong Kong's possession of one of the world's most sophisticated multidimensional street networks, the most acute pedestrian-vehicle conflicts occur on ground-level streets that are shared by both pedestrians and vehicles. These streets are identified as critical for targeted walkability enhancements and were thus the primary focus of this chapter.

To delineate the scope of this chapter and the subjects of investigation, we followed the established methodologies to divide the streets into two main components: street segments and intersections (Cain, et al., 2012; Boarnet, et al., 2006). This division considers their unique morphological features and roles in facilitating pedestrian perception and movement. Segments refer to the linear portions of streets, encapsulating the vibrancy of street life, whereas intersections refer to the pivotal points of traffic flow and direction change.

Select factors related to physical walkability

A small expert team composed of individuals conversant with CV or GIS technology was organized. First, experts employed existing indexes, primarily focusing on two widely recognized indexes: IMI (Boarnet, et al., 2006) and MAPS (Cain, et al., 2012). These indexes were used to select factors that directly influence people's walking behavior and experiences. Considering the significant variations in street-level characteristics and walking traffic patterns across different countries and cities, constructing a detailed index system that is applicable universally is challenging. To address this, we also incorporated the unique characteristics of Hong Kong streetscape factors by referencing the existing Hong Kong walkability index, CEx WALKScore (Ng, et al., 2016). This comprehensive approach enabled us to identify factors possessing street-level characteristics that could be quantified through CV or GIS techniques, resulting in the identification of 105 factors in total. Next, they conducted a screening to eliminate factors with duplicate interpretations, ultimately resulting in 36 physical walkability factors belonging to street segments and 10 for street intersections (Tables 4.1 and 4.2).

In the process of deciding measurement indicators for factors, we carefully evaluated and chose indicators capable of capturing street-level variations in the factors. To address the past reliance on single indicators such as the pixel ratio for characterizing features in streetscape images, we proposed using diverse measurement indicators to consider the variety in streetscape elements. For streetscape factors with continuous or non-countable attributes, such as greenery and sidewalk pavement, we computed pixel view indices, which are typically defined as the ratio of

the pixels of a specific factor to all pixels within a street view image. For discontinuous and countable streetscape factors, such as benches and street lights, we suggest specifying exact counts, which provides clear explanations and practical guidance for urban design purposes.

In some cases, direct calculation of factors may be challenging. To overcome this difficulty, proxies were employed. One such example is the factor "eyes on the street" proposed by Jacobs (1961) as a crucial factor impacting street safety. However, it is challenging to directly measure the flow of people on the street. Therefore, we utilized "betweenness", a spatial syntax indicator, as a proxy. Betweenness is quantified as the frequency of traversal of a street segment by the "shortest" path between any two other street segments within a specified analysis radius. This calculation reflects the potential for high throughput on that particular street segment. In addition to the natural surveillance provided by pedestrians on the streets, Jacobs (1961) also highlights the natural surveillance provided by owners and customers of street-side stores. To capture this, this chapter employed the "length ratio of storefronts" as a proxy indicator for the factor. For the variable "Shelter from and exposure to sunlight" was characterized by using the level of sky openness as a proxy to represent the shading opportunities provided by the street. For the factor "Average crossing distance", the average road grade attribute of OpenStreetMap (OSM) road network was chosen as the proxy. The score assigned is inversely proportional to the grade of the road intersecting the intersection node.

Table 4.1 *Selected dimensions, factors and indicators for physical walkability (segment)*

Primary dimension	**Subdimension**	**Factors**	**Indicators**	**Category**
A. Safe to walk	A-1. Accessibility	A-1-1. Sidewalk width	Width of the sidewalk	C3
		A-1-2. Sidewalk steepness	Degree of slope	C3
		A-1-3. Curb ramp	Number of curb ramps	C1
		A-1-4. Connect to overpass	Aggregation of the exit/entrance points	C5
		A-1-5. Connect to underground footway	Aggregation of the exit/entrance points	C5

Primary dimension	Subdimension	Factors	Indicators	Category
	A-2. Maintenance	A-2-1. Sidewalk pavement maintenance	Sidewalk pavement maintenance score	C4
		A-2-2. Building façade maintenance	Building facade maintenance score	C4
	A-3. Protection against crime	A-3-1. Eyes on the streets (shop owners)	Length ratio of the street-side business	C3
		A-3-2. Eyes on the streets (pedestrians)	Betweenness level of the street	C6
		A-3-3. CCTV camera	Number of CCTV cameras	C1
		A-3-4. Street light	Number of street lights	C1
	A-4. Protection from vehicles	A-4-1. Sidewalk buffer	Length ratio of the fence	C3
		A-4-2. Painted crosswalk	Number of painted crosswalks	C1
		A-4-3. Traffic light	Number of traffic lights	C1
		A-4-4. Street light	Number of street lights	C1
B. Comfortable to walk	B-1. Protection from unpleasant weather	B-1-1. Shelter from and exposure to rain	Number of storefront awnings, and bus stop shelters	C1
		B-1-2. Exposure to sunlight	Proportion of pixels of the sky	C2
	B-2. Wind environment and ventilation	B-2-1. Street parallel with the prevailing wind	Street parallel with the prevailing wind	C6
		B-2-2. Building density	Floor area ratio of buildings on both sides of the streets	C6

Primary dimension	Subdimension	Factors	Indicators	Category
		B-2-3. Street overall width	Road grade of street network	C6
	B-3. Natural scenery	B-3-1. Greenery	Proportion of pixels of Greenery	C2
		B-3-2. Mountain scenery	Proportion of pixels of mountain scenery	C2
		B-3-3. Water scenery	Proportion of pixels of water scenery	C2
	B-4. Public realm amenities	B-4-1. Bench	Number of benches	C1
		B-4-2. Mailbox	Number of mailboxes	C1
		B-4-3. Trash-can	Number of post trash-can	C1
		B-4-4. Phone booth	Number of phone booths	C1
		B-4-5. Public toilet	Aggregation of public toilets	C5
C. Interesting to walk	C-1. Destination	C-1-1. Park, open space, and tourism attraction	Aggregation of parks, open spaces, and tourism attractions	C5
		C-1-2. Public service facilities	Aggregation of public service facilities	C5
		C-1-3. Shopping or catering facilities	Aggregation of shopping or catering facilities	C5
	C-2. Legibility & orientation	C-2-1. Wayfinding system	Number of wayfinding systems	C1

Primary dimension	Subdimension	Factors	Indicators	Category
		C-2-2. Banner	Number of banners	C1
	C-3. Active transportation	C-3-1. Bike lane	Proportion of pixels of bike lane	C2
		C-3-2. Bus stop	Aggregation of bus stops	C5
		C-3-3. Rail transit station	Aggregation of Rail transit station exit/entrance points	C5

Table 4.2. *Selected dimensions, factors and indicators for physical walkability (intersection)*

Primary dimension	Subdimension	Factors	Indicator	Category
D. Safe to walk	D-1. Crossing safety	D-1-1. Traffic light	Number of traffic lights	C1
		D-1-2. Yield sign/marking	Number of yield marking	C1
		D-1-3. Crosswalk	Number of crosswalks	C1
		D-1-4. Street light	Number of street lights	C1
		D-1-5. Refuge island	Number of refuge islands	C1
		D-1-6. Curb ramp	Number of curb ramps	C1
		D-1-7. Connect to overpass	Aggregation of the exit/entrance points	C5
		D-1-8. Connect to underground footway	Aggregation of the exit/entrance points	C5
	D-2. Crossing efficiency	D-2-1. Average crossing distance	Mean road grade	C6
		D-2-2. Number of legs	Total count of segments that intersect at a particular crossing point	C6

Aggregate individual physical factors

Considering the excessive number of individual physical walkability factors, we aggregated the factors into three primary dimensions for more intuitive pattern understanding and observation: safe to walk, comfortable to walk, and interesting to walk. These dimensions also reference the walkability evaluation structure of Hong Kong's CEx WALKScore (Ng, et al., 2016). The allocation of factors to specific dimensions was deliberated upon and confirmed by the expert panel, drawing upon existing evaluative frameworks.

In addition, we also introduced an intermediate subdimension level between the measurement factors and primary dimensions to capture more nuanced aspects of evaluation. Specifically, for street segments, we identified 11 objective subdimensions, which encompass a wide range of criteria: accessibility, maintenance, protection against crime, protection from vehicles, protection from unpleasant weather, wind environment and ventilation, natural scenery, amenities in the public realm, destinations, legibility and orientation, and facilities for active transportation. For intersections, we delineated 2 subdimensions: crossing safety and crossing efficiency.

Select factors related to perceived walkability

For factors to measure perceived walkability, we followed the conceptual framework outlined in Section 1.3 of Chapter 1, selecting factors from a perception perspective to represent perceived walkability. Consequently, three factors were chosen for street segments: perceived safety, perceived comfort, and perceived interestingness (Table 4.3). As for street intersections, pedestrians typically express how well a particular arrangement accommodates their travel based on safety (Petritsch, et al., 2005). Thus, perceived safety is the only perception factor when evaluating intersection walkability (Table 4.4).

Table 4.3 *Selected dimensions, factors and indicators for perceived walkability (segment)*

Primary dimension	Factors	Indicators	Category
E. Environmental perception (segment)	E-1. Perceived safety	Perception score	C7
	E-2. Perceived comfort	Perception score	C7
	E-3. Perceived interest	Perception score	C7

Table 4.4 *Selected dimensions, factors and indicators for perceived walkability (intersection)*

Primary dimension	Factor	Indicator	Category
F. Environmental perception (intersection)	F-1. Perceived safety	Perception score	C7

4.2.2 Developing methods for measuring street-level walkability

Collecting BSVs and GIS data

We obtained street view images (BSVs, shooting period from 2017 to 2018) from Baidu Maps (Baidu Maps, 2022), which is a leading big data vendor in China. Different BSV collection procedures were used to acquire views of street segment and intersection spaces. For street segment spaces, we collected around 14,000 panoramic images (4096*2048 pixels), with sample points generated at 20m intervals along the street network. According to Gehl (1987), distances measuring around 25m may facilitate social encounters and fit the scale of 'place' in public open spaces.

When collecting BSV for intersections, utilizing coordinates for all network junction nodes may not yield the desired results, as it could result in images of slip lanes rather than the view of the central region of the intersection. It should be noted that the imagery retrieved through the API may not exclusively depict ground-level views, but may also include views from elevated viaducts or underground tunnels.

For the first problem, we identified whether network junction nodes were in the central intersection area by calculating the included angle between the segments converging at those nodes. By examining the nodes' characteristics, we found that the majority of "suitable nodes" within our study site were connected to street segments with angles exceeding 30 degrees. Consequently, we implemented a threshold value and created a script to eliminate nodes where the angle between segments was less than 30 degrees.

For the second problem, we collected extensive street views at each intersection node. Specifically, we created 10-meter buffer radii using the cleaned intersection nodes and then randomly generated several BSV collecting points within these buffers. Using these newly created collection points, we gathered street views. Finally, we employed an image classification model, trained on two classes of images (ground-

level view or non-ground-level view), to determine the presence of ground-level views in the collected images.

The GIS data used in this study, including the street network, Points of Interest (POIs), Areas of Interest (AOIs), and Digital Elevation Model (DEM), were acquired from OpenStreetMap and the Hong Kong government's open data website (DADA.GOV.HK, 2022).

Pre-processing BSVs and GIS data

In alignment with the diverse image analysis tasks outlined in this chapter, different preprocessing strategies were employed on the acquired BSVs. This process led to the creation of five distinct image formats (Table 4.5), each meticulously tailored to support the corresponding analytical tasks discussed later in the chapter.

Table 4.5 *Pre-processed image data formats*

Data format	**Morphological section**	**Convert to perspective**	**Heading direction**	**Field of view (FOV)**	**Pitch angle**	**Size**
Image format a	Segment	No	/	/	/	4096*1400
Image format b	Intersection	No	/	/	/	4096*1400
Image format c	Segment	Yes	360/180°	90°	0°	1024*1024
Image format d	Segment	Yes	270/90°	90°	0°	1024*750
Image format e	Intersection	Yes	360/180 /270/90°	90°	0°	1024*750

Propose methods to measure physical walkability factors

Using a single method for all selected physical factors may not accurately capture the full range of their variations. Therefore, we categorized physical factors based on their features and proposed appropriate measurement methods for each category, resulting in the identification of six categories of factors. The first category (C1) includes the discontinuous and countable streetscape factors and indicators; The second category (C2) includes the continuous or non-countable streetscape factors and indicators; The third category (C3) includes the geometric characteristics of streetscapes; The fourth category (C4) includes the level of maintenance on streetscapes. The fifth category (C5) includes the aggregation of specific types of

destinations. The sixth category (C6) includes topological characteristics, and other kinds of factors.

To measure the six categories of factors and indicators, we developed six corresponding measurement methods.

Measure the discontinuous and countable streetscapes of physical walkability

To measure factors representing the number of discrete, countable streetscape elements, we employed the widely utilized object detection algorithm, YOLOv5 (Jocher, 2020). We used perspective views captured from the street side (image format c) as input data for measuring street segment sections, and cropped panoramic images (image format b) for intersection section measurements (Figure 4.3a).

Despite YOLOv5's extensive library of pre-trained weights, they are mostly trained on the COCO dataset (Lin, et al., 2014), which includes a limited number of classes relevant to our selected streetscape factors. Therefore, we selected 4000 street-side views as training images, and annotated them with a total of 13 new classes, including wayfinding systems, street lights, trash cans, mailboxes, phone booths, street stores awnings, bus stop shelters, traffic lights, benches, curb ramp, CCTV cameras, crosswalks, and banners. Similarly, 800 intersection panoramic images were selected and were annotated with 6 new classes, comprising of yield marking, curb ramps, traffic lights, street lights, refuge islands, and crosswalks. After the annotation process was completed, we used the annotated datasets to train two new models using fine-tuning and pre-trained weights (Figure 4.3b).

The trained YOLOv5 model 1 can be utilized to detect 13 classes of streetscape elements in segment sections, with a mean average precision (mAP) of 0.98 and 0.818 at a threshold of 0.5 and 0.95 respectively (precision=0.962, recall= 0.959). The trained YOLOv5 model 2 can be used to detect 6 classes of elements at street intersections, with a mAP of 0.986 and 0.808 at a threshold of 0.5 and 0.95 respectively (precision=0.979, recall= 0.974) (Figure 4.3c). Both models performed well on our image dataset, except for some errors in inferring very small elements, such as CCTV cameras (Figure 4.3d).

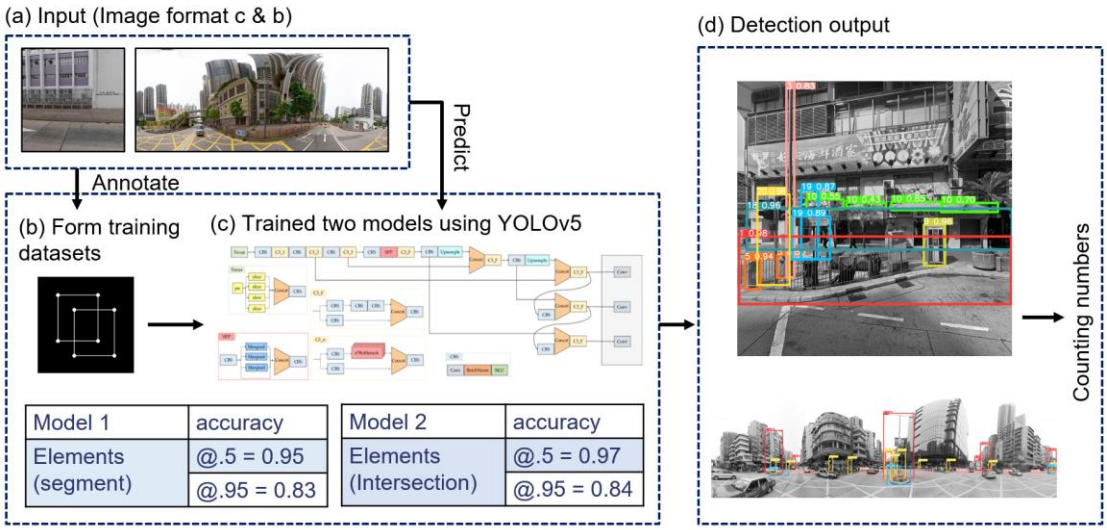


Figure 4.3 *Pipeline for measuring the discontinuous and countable streetscapes. (a) Input format, (b) forming training dataset, (c) training two YOLOv5 models, (d) Detection output.*

Measuring the continuous or non-countable streetscapes of physical walkability

In order to measure factors that signify the pixel ratio of non-countable streetscape elements, we utilized processed image dataset format a as the input (Figure 4.4a).

We employed the pre-trained Mask2Former model (Cheng, et al., 2021) with a ResNet50 backbone for semantic segmentation (Figure 4.4b). The Mask2Former introduces a novel framework for semantic segmentation by incorporating masked attention, which enhances the extraction of localized features by constraining cross-attention to within predicted mask regions. We chose Mask2Former for this section because of its extensive library of pre-trained models (Cheng, et al., 2022), which includes versions trained on the Mapillary Vistas v1 dataset (Neuhold, et al., 2017). These models are particularly adept at quantifying the pixel ratio of varied background landscapes such as greenery, mountainscapes, waterscapes, and sky (Figure 4.4c).

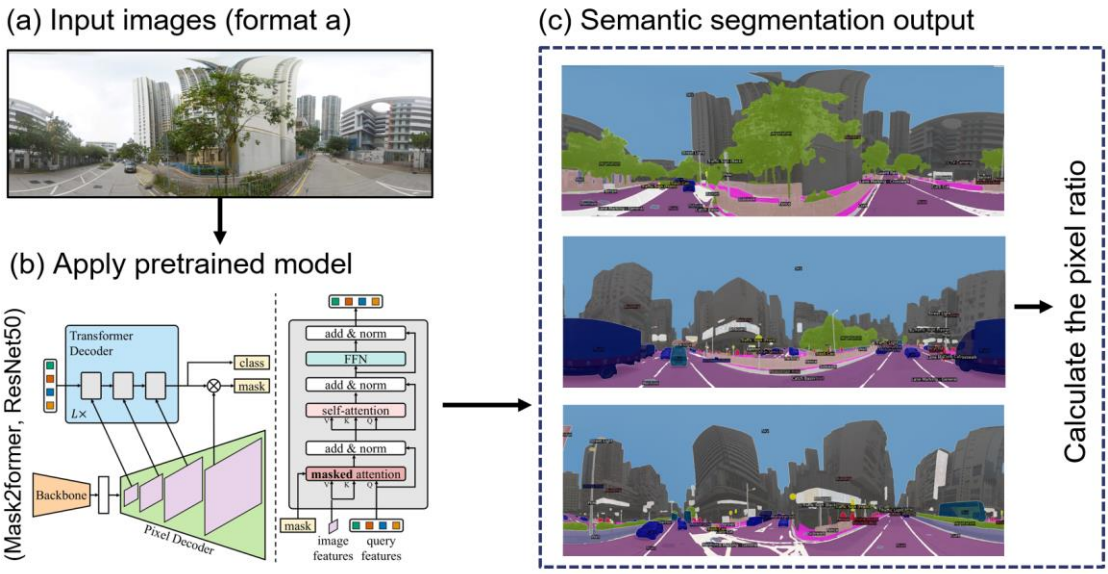


Figure 4.4 *Pipeline for measuring the continuous or non-countable streetscapes. (a) Input format, (b) Model structure, (c) Semantic segmentation output.*

Measure the geometric character of streetscapes of physical walkability

We proposed a methodological pipeline for quantifying the geometric characteristics of sidewalk width based on the principles of monocular measurement and the principle of similar triangles. This approach utilizes the relationship between the pixel sizes of BSVs and their corresponding distances in the real world to estimate the depth of the target subject. In order to acquire information on the location of sidewalk pixels in images, we employed YOLOv5 and trained a model for detecting the range of the sidewalk pavement area based on an annotated dataset (Figure 4.5b). The model achieved a mAP of 0.99 at a threshold of 0.5 and 0.91 at a threshold of 0.95, with a precision and recall of 0.972 and 0.973, respectively. Subsequently, we used image dataset c (Figure 4.5a) as the input data and implemented the estimation of the sidewalk width through the following steps. First, Equation (4.1) was utilized to estimate the distance d from the camera's foot point to the edge of the road. In this equation, h represents the height of the vehicle camera (h is about 2.3m), 512 represents half of the height of the images in pixels (since sidewalks are typically only visible in the lower half of the image), and n represents the number of vertical pixels of the road area in the image, which can be calculated using the parameters provided by the YOLO bounding box. These parameters include the ratio of the height (h) of the bounding box and the ratio of the distance from the centroid of the bounding box to the top edge of the image (y). Through the application of Equation (4.2), the pixel distance of n can be determined. In the second step, Equation -(4.3) was utilized to

estimate the distance d from the camera to the inside edge of the sidewalk. The value of n' can be calculated using Equation (4.4). Finally, Equation (4.5) was employed to measure the sidewalk width (w) by calculating the difference between d' and d (Figure 4.5c).

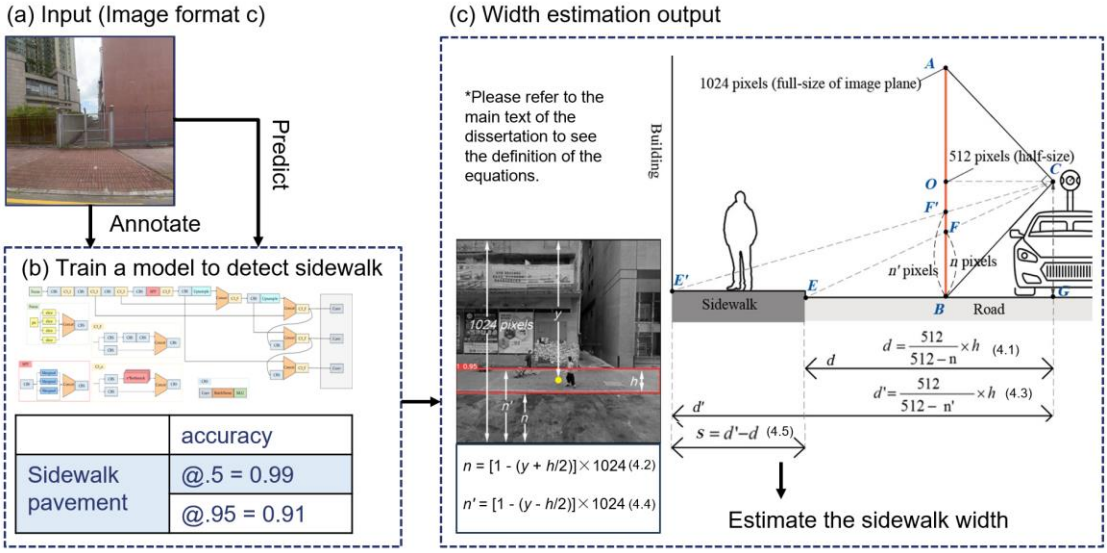


Figure 4.5 *Pipeline for measuring the geometric characteristics of streetscapes. (a) Input format, (b) Sidewalk detection model training using YOLOv5, (c) Width estimation output.*

The other two measured geometric indicators are the length ratio of storefront and fence.

Measure the maintenance of streetscapes of physical walkability

To assess the factors related to the street maintenance, we developed a pipeline that combines convolutional neural networks (CNN) regression and YOLOv5 algorithms. This pipeline was structured into several phases. Initially, we developed scoring criteria (Figure 4.6b) for assessing the maintenance levels, which involved manual annotation of a training dataset using image data format c (Figure 4.6a). These criteria were derived from existing literature and specific maintenance specifications for building façades and sidewalk pavements, as detailed in Tables 4.6 and 4.7. For the training data, we selected a sample of 1,000 images each of building facades and sidewalk pavements. These images were manually scored by multiple volunteers on a 5-point scale—where 1 indicates poor maintenance and 5 denotes excellent maintenance—based on criteria tailored for both building facade and sidewalk pavement (Figure 4.6c). The final label for each image was determined by averaging the scores across the three evaluation criteria.

Table 4.6 *Evaluation criteria of building façade maintenance*

No.	Criteria	Description	Scoring scale
B1	**Aesthetical**	Efflorescence; localized stains; uniform dirt; color change; runoff; graffiti; biological contamination; flatness deficiency; staining	1-5 (lowest - highest)
B2	**Adhesion loss and other facade defects**	Cracking; detachments; peeling; erosion; blistering	
B3	**Physical disorder**	Air conditioner external units are installed without any orders; Chaotic wiring arrangement; Stickers	

Table 4.7 *Evaluation criteria of sidewalk pavement maintenance*

No.	Criteria	Description	Scoring scale
S1	**Surface defects**	Cracking; depressions; raveling; uneven surface	1-5 (lowest - highest)
S2	**Encroachment and obstruction**	Overgrown plants encroaching on the sidewalk; Obvious piles of rubbish or litter on the pavement; presence of fixed obstructions such as tree pools, lamp poles or fire hydrants that have seriously obstruct walking activities; presence of temporary obstruction, such as temporary seating areas, garbage, and parking vehicles	
S3	**Curb defects**	Cracking; detachments; erosion	

The annotated images formed the basis of our training dataset, with the scores derived from manual assessments serving as labels. We trained a CNN regression model to predict maintenance levels using ConvNeXt (Liu, et al., 2022) as the backbone architecture. This model demonstrated robust performance, achieving R-square values of 0.937 for building façades and 0.613 for sidewalk maintenance predictions (Figure 4.6d). The effectiveness of the automatic scoring system was confirmed in real-world applications, closely matching our initial projections, as depicted in Figure 4.6.

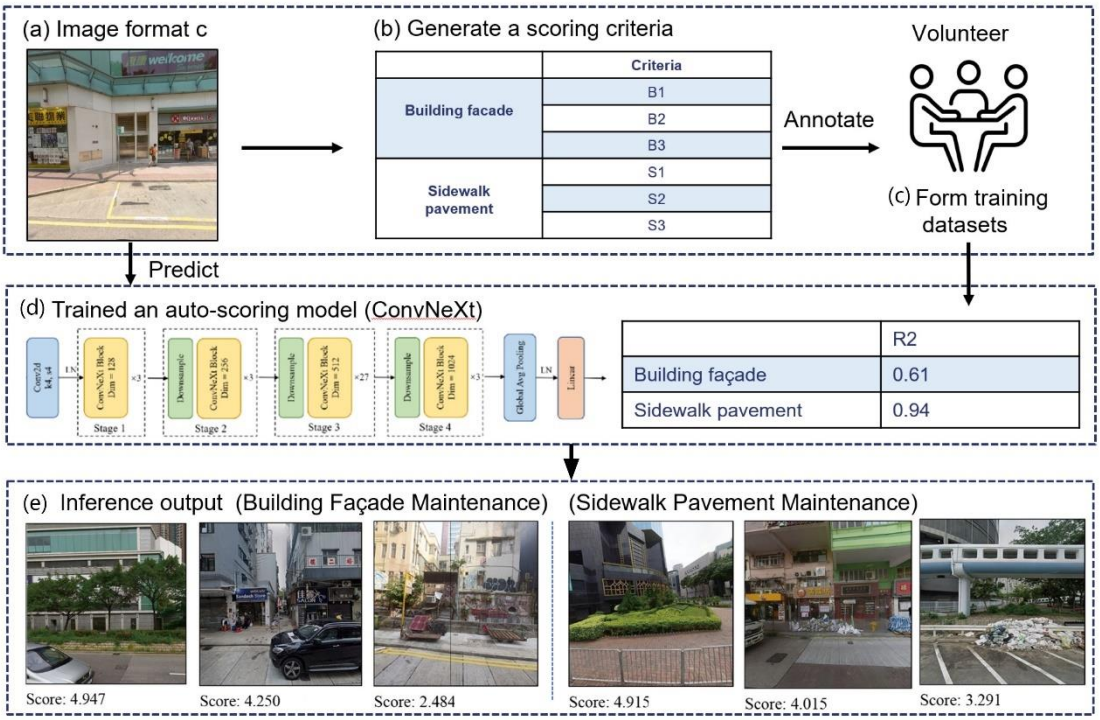


Figure 4.6 *Pipeline for measuring the level of maintenance on streetscapes. (a)Input format, (b) maintenance evaluation criteria, (c) forming training dataset, (d)training the regression models, (e) Inferring output.*

It is important to acknowledge that not all street view images used in this analysis consistently feature sidewalk pavements or building façades. To mitigate this, we employed a trained sidewalk detection model alongside a trained model for building façade detection using YOLOv5. These models helped confirm the presence of sidewalk pavements or building façades in the images. When these features were detected, the images were processed by the scoring model for automatic evaluation. In cases where neither sidewalks nor building façades were present, the images were assigned a score based on the average of previously obtained automatic scoring results, ensuring comprehensive coverage during the mapping process (Figure 4.6e).

Measure the aggregation-related, topology-related and other factors using GIS data of physical walkability

The application of CV techniques significantly enhances the study of street-level built environments by providing detailed visual insights. However, GIS methods offer distinct advantages in quantifying specific variables. For instance, in this chapter, we utilized GIS data and methods to calculate the average number of POI per meter within buffer zones along each street segment. Additionally, for variables involving connectivity to overpasses, we extracted suspension points from the network of

overpasses—specifically, the exits and entrances—before proceeding with data aggregation.

For analyzing topological factors such as betweenness centrality, we employed the Spatial Design Network Analysis (sDNA) software (Cooper & Chiaradia, 2020), which is highly regarded for spatial network analysis within GIS and CAD environments. The analysis involved pre-processing the road network data before inputting it into sDNA for calculations. Additionally, we utilized GIS tools to assess other factors, such as crossing exposure and intersection complexity. For instance, we calculated the average road grade and the number of interacting segments at each intersection to better understand the complexity and safety of pedestrian movements.

We also gauged the performance of street ventilation by calculating the angular difference between the direction of street segments and the prevailing wind direction in Kowloon West. Smaller angular differences are indicative of more effective ventilation, reflecting a more pedestrian-friendly environment.

Integrating individual factors of physical walkability into comprehensive dimensions

Following the calculation of all physical walkability factors and indictors, we used the CRITIC method, introduced by Diakoulaki et al. (1995) to integrate individual measurement results into a comprehensive evaluation. The CRITIC method is an objective weighting approach designed to determine the weights of factors based on their variability and the degree of conflict among them.

The procedure involves four key steps, starting with the identification of the positive and negative attributes of each factor. Subsequently, positive and negative normalization of these attributes is performed, using either Equation (4.6) or Equation (4.7). This normalization ensures that each indicator is appropriately scaled, facilitating effective aggregation in later phases of analysis. This systematic approach allows for the balanced integration of diverse evaluation variables, thereby enhancing the accuracy and reliability of the overall assessment results,

$$x'_{ij} = \frac{x_{ij} - \min(x_j)}{\max(x_j) - \min(x_j)} \tag{4.6}$$

$$x'_{ij} = \frac{\max(x_j) - x_{ij}}{\max(x_j) - \min(x_j)} \tag{4.7}$$

where x'_{ij} is the normalized value, x_{ij} is the original value, min (x_j), and max (x_j) are the minimum and maximum values for criterion j, respectively.

The subsequent step is to compute the variable volatility through the application of Equation (4.8),

$$S_j = \sqrt{\frac{\sum_{i=1}^{m}\left(x_{ij} - \bar{x}_j\right)^2}{n-1}} \tag{4.8}$$

where S_j represents the volatility or standard deviation of the j *-th* indicator, and x_j represents the mean value of the data associated with each indicator.

The third step in the process is to determine the degree of conflict between indicators, which is calculated utilizing Equation (4.9). In this equation, r_{ij} presents the correlation coefficient between the number of i indicators and the number of j indicators.

$$A_j = \sum_{i=1}^{n}\left(1 - r_{ij}\right) \tag{4.9}$$

Once the volatility and conflict have been quantified, the amount of information can be determined using Equation (4.10). C_j represents the relative importance of the number of j evaluation index within the overall evaluation index system, with a larger value of C_j indicating a greater weight should be assigned to that index.

$$C_j = S_j \times A_j \tag{4.10}$$

Hence, the objective weight W_j of the number of j indicators can be calculated using Equation (4.11).

$$W_j = \frac{C_j}{\sum_{j=1}^{n} C_j} \tag{4.11}$$

We employed this method to first aggregate individual variables under their corresponding sub-dimensions, and then further aggregated them under the three primary dimensions (safe to walk, comfortable to walk and interesting to walk).

Propose methods to measure perceived walkability factors

To assess individuals' attitudes towards walking perceptions, our study employed a deep learning technique that integrates crowdsourcing datasets with DCNN for predicting walking perceptions. Our methodology is rooted in the empirical basis of prior studies (Ogawa, et al., 2022) and consists of two main phases: (1) Obtaining

crowdsourcing data on people's walking perceptions through a crowdsourcing survey, and (2) training the inference model.

For obtaining crowdsourcing data, we utilized a crowdsourced dataset derived from a mobile platform that focused on 22 dimensions of environmental perception relevant to urban settings (Ogawa, et al., 2022). This dataset included three critical perceptual dimensions related to perceived walkability identified in our research: safety, comfort, and interestingness. Therefore, we only utilized data from these three dimensions. This dataset was developed using 1000 street view images to construct image comparison pairs and to conduct an online crowdsourcing survey. The survey received approximately 400,000 responses. While the survey did not specifically target the Kowloon district of Hong Kong, it focused on cities in Asia with similar economic scales and high-density environments, lending potential applicability to similar contexts.

The methodology involved presenting participants with pairs of street images and asking evaluative questions such as "Which street looks safer?" This approach helped us gather nuanced insights into their visual attitudes towards varying street views (Figure 4.7a). Participants rated their preferences using a five-point scale, which included options ranging from "Upper image is much better than the bottom image" to "Bottom image is much better than the upper image," with intermediary gradations for lesser preferences and equivalence. This scale facilitated a detailed comparative analysis of pedestrian perceptions across different urban environments.

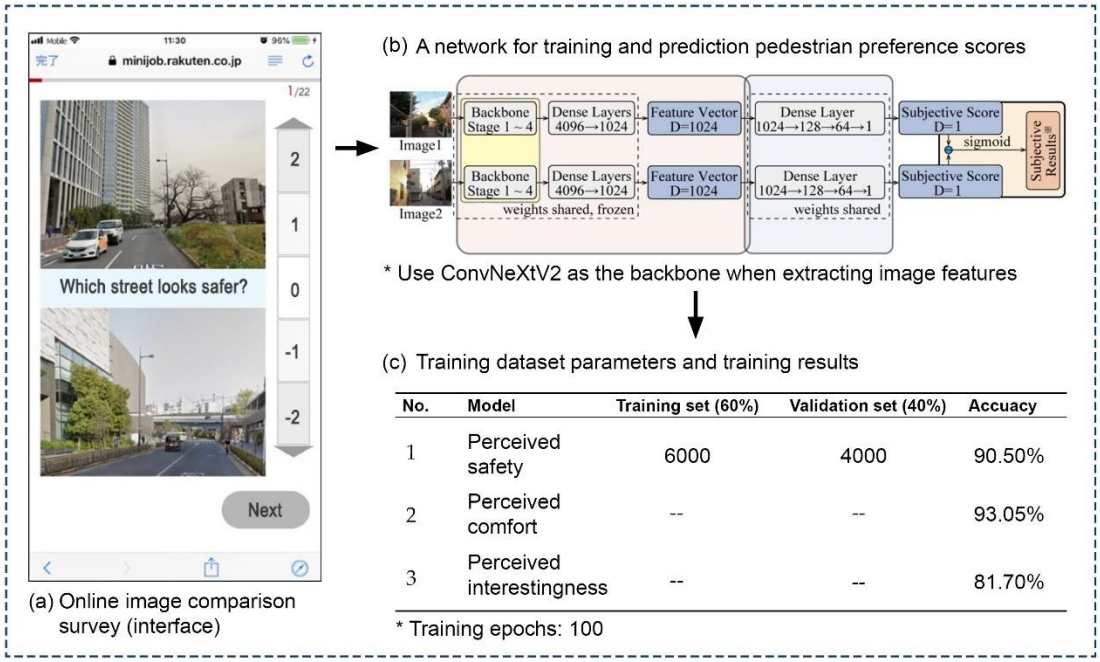


Figure 4.7 *Pipeline for measuring how people perceive the environment. (a) Image comparison survey interface (translated to English, original in Japanese); (b) a network to infer walking preference at street segments and intersections; (c) training dataset parameters and results.*

To train our perception prediction model, we begin by preprocessing the survey responses from image comparisons and converting them into binary classification labels. We then employed the ConvNeXtV2 network (Woo, et al., 2023) to extract features from these images. These extracted features are input into a Deep Neural Network (DNN) to produce two predicted scores. The difference between these scores is computed to facilitate binary classification and further model training. To normalize this difference within a range of 0 to 1, we apply a sigmoid function. This transformation helps to minimize the loss function by aligning the predicted values with the pre-processed labels derived from an online image survey. The detailed architecture of this model is illustrated in Figure 4.7b.

Accuracy was adopted as the evaluation metric for these 3 prediction models, reflecting the model's reliability in classifying the validation results. As illustrated in Figure 4.7, for the three walking perception dimensions (safety, comfort, and interestingness), the model achieved prediction accuracies of 90.50%, 93.05%, and 81.70%, respectively (Figure 4.7c).

4.3 Experimental results

4.3.1 Physical walkability measurement in Kowloon West

Using the proposed measurement methods, we calculated individual factors for physical walkability. Subsequently, utilizing the CRITIC method, we synthesized the 11 sub-dimensions that encapsulate the concept of walkability in the context of Kowloon West. As shown in Figure 4.8, an obvious spatial pattern emerges on maps a, c, e, i, j, and k. High-value evaluation results were mainly clustered in central high-density blocks. This is because high-density environments continuously attract people and commercial resources. Driven by commercial interests, various facilities will automatically cluster in these areas. Increased economic values also stimulates the renewal of street facilities. Meanwhile, high-density urban forms provide more shelter from rain and reduce heat radiation, which greatly reduces the potential that pedestrians will be affected by bad weather.

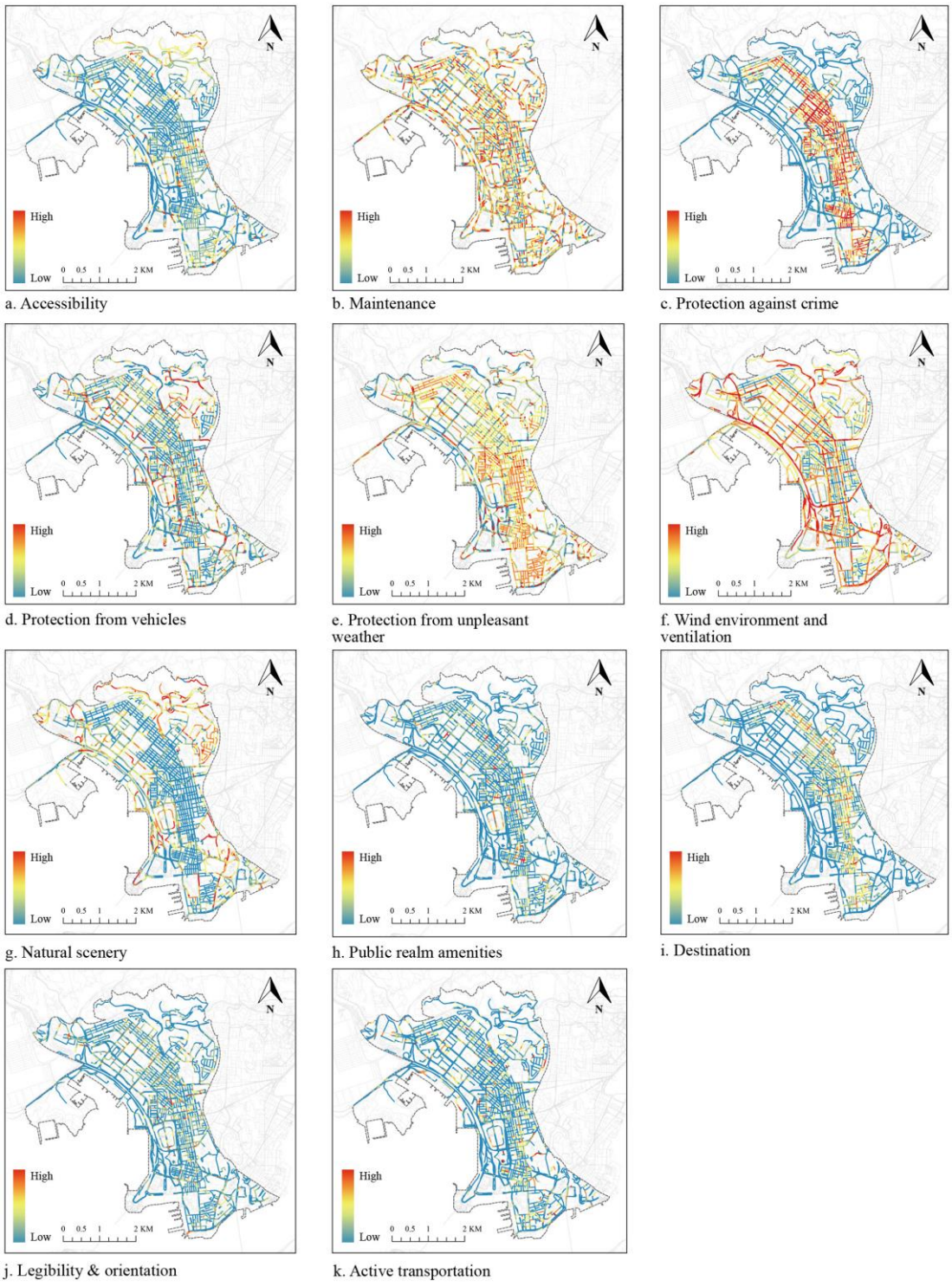
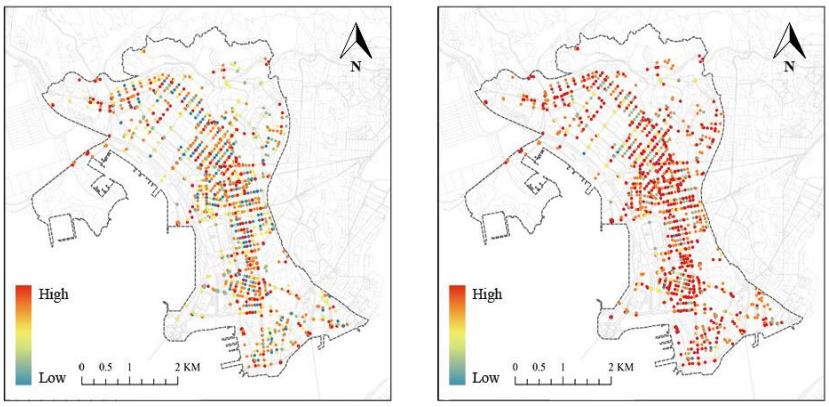


Figure 4.8 *Subdimensions of physical walkability in Kowloon West (segment)*

We found opposite spatial distribution trends for street segments in maps f and g, the latter of which reflects the uneven spatial distribution of natural landscape in Kowloon West. High-density neighborhoods in the central area (e.g., Yau Ma Tei, Mong Kok, and Sham Shui Po) suffer from a severe lack of greenery and other natural landscapes. Further, ventilation problems in high-density blocks in Kowloon West

must be taken seriously. As shown on map f, more segments with poor ventilation exist in high-density neighborhoods, which undoubtedly affects public health and summer walking experiences. As shown on maps k and j, respectively, 'Active transportation' and 'Legibility & orientation' are relatively evenly distributed throughout the study area, indicating that public transportation resources and wayfinding services are equitably accessible. While no clear spatial distribution pattern emerges on map b 'maintenance', it is roughly evident that segments with more low-maintenance features are mostly distributed in areas adjacent to main roads and neighborhoods that were built earlier. In regard to the intersection space, as depicted in Figure 4.9l and 4.9m, the high-value points of map l, which pertains to crossing safety, are primarily situated on middle and high-grade streets. This suggests that the central region and intersections connected to higher-grade roads possess relatively superior pedestrian facilities. Map m, while displaying similar trends to map l, does not exhibit notably higher values for intersection crossing efficiency in the central region as compared to the surrounding areas.



l. Crossing safety m. Crossing efficiency

Figure 4.9 *Subdimensions of physical walkability in Kowloon West (intersection)*

To reflect walkability characteristics more comprehensively, we further aggregated the 11 subdimensions into four primary dimensions. Figure 4.10 (n-q) shows the objective walkability results for Kowloon West at the primary dimensions, with visualizations of the spatial characteristics for safe to walk (n), comfortable to walk (o), and interesting to walk (p) in segment sections, as well as safe to walk (q) at intersections.

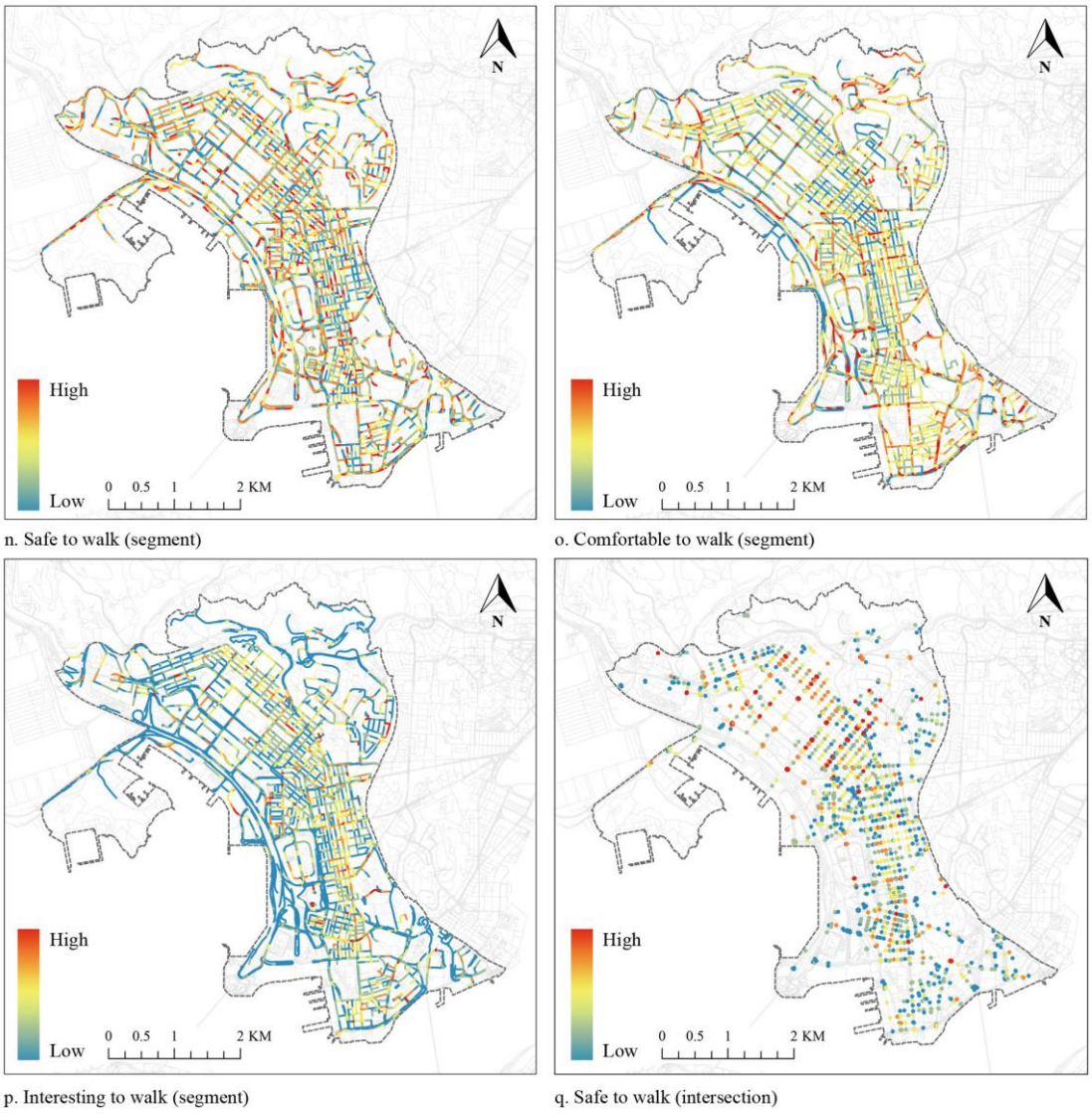


Figure 4.10 *Primary dimensions of physical walkability in Kowloon West (segment and intersection)*

4.3.2 Perceived walkability measurement in Kowloon West

We utilized a trained perception prediction model to infer how pedestrians perceived the environments of street segments and intersections in Kowloon West. Figure 4.11 illustrates the quantification results. The perceived safety (Figure 4.11r) and perceived comfort (Figure 4.11s) exhibit similar spatial distribution patterns in Kowloon West, with lower values in the high-density central areas and relatively higher values in the surrounding areas. In contrast, the perceived interest (Figure 4.11t) shows the opposite trend. These perceived measurement results align closely with our basic understanding of the site, thereby validating the effectiveness of our method.

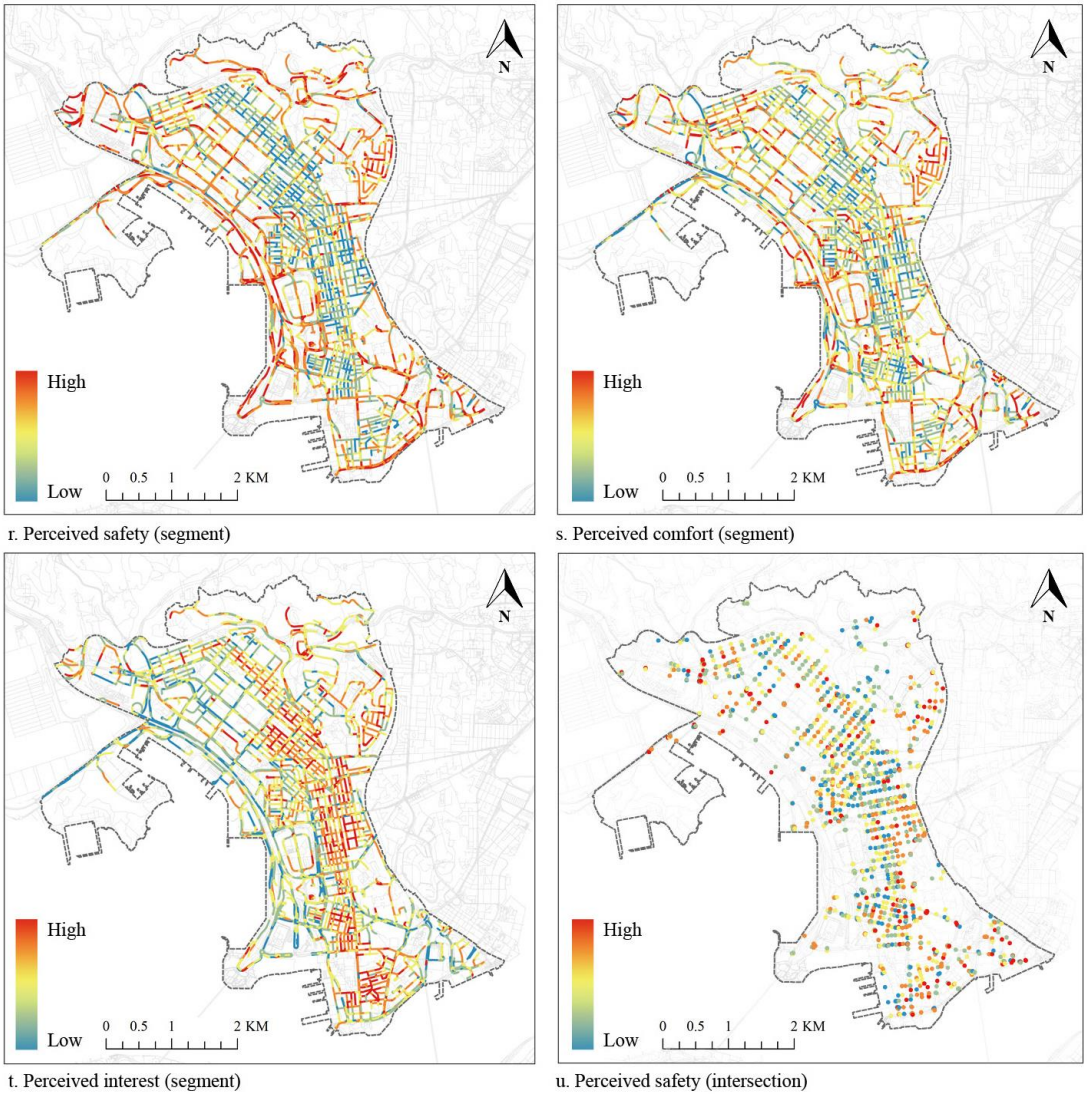


Figure 4.11 *Dimensions of perceived walkability in Kowloon West (segment and intersection)*

4.5 Discussion

4.5.1 Produce a low-cost, fast, and reliable walkability evaluation

Applying urban big data, especially street view imagery, can dramatically reduce the cost of conducting a street-level walkability evaluation. The GSV API service charges $7 USD per 1,000 images (Google Maps Platform, n.d.), while the BSV API service only charges ~$4.5 USD per 1,000 images. For the Kowloon West, the total street view acquisition cost is approximately $92 USD if using the BSV service. Because POI, street network data, and other geographical data are typically obtained for free or at minimal costs, they are considered negligible expenses. However, if the virtual audit approach is taken, assuming an average of 15 to 20 min for a proficient surveyor to audit each street segment/intersection, then it will take between 91,275 and 121,700

min (approximately1,521 to 2,028 hour) to complete the audit tasks. If surveyors are paid the standard hourly wage in Hong Kong ($26.75 USD) (Salaryexplorer, 2023), then the total cost for a virtual walkability audit of Kowloon West could reach $40,687 to $54,249 USD. If an in-field survey is adopted, then extra costs should be attached. Regarding evaluation speed and efficiency, virtual auditing and on-site surveys are substantially inferior to automatic evaluations based on CV and GIS. In terms of reliability, in-field surveys and virtual audits sometimes suffer from inconsistency when multiple surveyors assess streetscapes. However, the automatic walkability evaluation does not have this problem since the machine can fully follow the same instructions and criteria, thus producing more reliable results. This new type of data source and method undoubtedly advances the practical feasibility and cost-effectiveness of evaluation in the field of walkability.

4.5.2 Have high applicability and generalization potential

Our selected factors and measurement methods also have strong generalization potential due to the extensive coverage and consistent standard of applied urban big data. More than 140 nations/regions will have partial or complete coverage via street view images by 2022 (Google Maps Platform, n.d.). POI, street networks, and other urban big data have become widely available due to more developed electronic maps and matured crowdsourcing technologies. The proposed street-level walkability factors and measurement methods hold immense promise for transfer and application to other cities in this context. Nonetheless, there are still differences in the data sources and quality between cities/countries. Thus, any data gathering approach must be tailored to guarantee a seamless transfer.

4.5.3 Capture the characteristics of changes in various factors

Previous automatic evaluations of walkability have typically employed a single method to measure changes across all factors and indicators. While such approaches are easy to implement, they also restrict the capacity to measure a variety of walkability factors. Further, the use of only one technique precludes fine-grained measurements. For instance, using semantic segmentation to calculate the pixel ratio for countable factors such as pedestrians, vehicles, and urban amenities does not effectively reflect trends in these factors and offers limited utility for guiding design decisions. Additionally, while a single or simple method may suffice for automating the measurement of physical walkability, it falls short in measuring perceived walkability. As such, this chapter combined numerous CV techniques and GIS methods to capture as many detailed walkability characteristics as feasible.

4.6 Conclusion

This chapter explores the physical and perceived walkability factors at the street level with the potential for automated measurement, as well as the related automated measurement methods. Using Hong Kong as a case study, we have validated the practicality and reliability of the proposed index and its measurement methods. By employing them, urban planners, designers and policymakers can gain evidence and insights into the physical conditions of walkability and how these conditions are visually perceived by people. This information can significantly assist in guiding future street design and renovation practices.

This chapter still has some limitations. In terms of selecting data source, this study's reliance on street view imagery from vehicular perspectives to simulate human visual perception may introduce bias due to differences in the street structure, viewpoint elevation, and vehicular movement patterns. This focus may introduce some bias, a limitation that persists in subsequent Chapters 5 and 6. To mitigate this limitation and more accurately reflect the pedestrian experience, subsequent research should strive to include imagery data collected from pedestrian viewpoints. Besides, we utilized a perception prediction model, based on image data as input, to score images' perceived walkability. However, our proposed method primarily evaluated the visual aspect of walkability, neglecting other sensory perceptions like auditory and olfactory experiences. This focus may introduce some bias, a limitation that persists in subsequent Chapters 5 and 6. Future research should aim to incorporate a more comprehensive range of sensory perceptions to provide a fuller understanding of environmental impacts on walkability. In terms of measurement method application, this chapter proposes using different approaches to optimally calculate various factors. However, employing multiple methods might conversely affect practical efficiency. Future studies could explore simplifying the methodologies while ensuring the capability to effectively capture variations in different factors. For instance, using panoptic segmentation, which can handle both segment and counting tasks, could replace the combination of semantic segmentation and the YOLO algorithm. Finally, we selected physical street-level walkability factors based on existing experience. However, it is uncertain whether these selected indicators are necessarily suitable for representing the walkability of a specific city. Future research should validate these selected indicators.

CHAPTER 5: INFORMING PERCEIVED WALKABILITY AT STREET-LEVEL THROUGH NON-LINEAR REGRESSION ANALYSIS

This chapter is based on:

Huang, L., Oki, T., Muto, S. and Ogawa, Y., 2024. Unveiling the Non-Linear Influence of Eye-Level Streetscape Factors on Walking Preference: Evidence from Tokyo. *ISPRS International Journal of Geo-Information*, *13*(4), p.131.

5.1. Introduction

As human-centric approaches gain traction, the importance of perceived walkability has come to the forefront. It is clear that focusing solely on physical features, without taking into account the public's environmental experiences and feedback, falls short of truly improving the walking experience. By incorporating people's experience and needs, urban planning can more effectively encourage healthy and sustainable walking habits, fostering environments that are both walkable and enjoyable for pedestrians (Sung, et al., 2015; Alfonzo, 2005; Sarkar, et al., 2015). This trend necessitates a thorough understanding of how and to what extent people perceive and prefer various built environment factors at street-level (Choi, et al., 2016; Huang, et al., 2023). This chapter examines perceived walkability from the perspective of walking preferences and explores its relationship with streetscape features that characterize street-level physical walkability.

Numerous studies have explored the relationship between streetscape factors and walking preferences and behaviors in a street-level context, primarily using linear analysis models (Ewing & Handy, 2009; Harvey, et al., 2015; Asgarzadeh, et al., 2012; Asgarzadeh, et al., 2014). However, these studies have overlooked the complex and multi-dimensional nature of this relationship, which frequently exhibits non-linear characteristics in real-world contexts. Nevertheless, comprehensive research on the non-linear impact of streetscape factors on walking preferences is scarce. This limitation constrains our understanding of how the street-level built environment influences pedestrian perceptions and subsequent design guidance. A holistic approach that considers these complex relationships is essential for creating walkable, enjoyable, and safe urban spaces. Furthermore, streets can be classified into two

distinct sections based on their morphological and functional characteristics: street segments and intersections. These classifications have already been incorporated into existing walkability evaluation indexes (Cain, et al., 2012; Boarnet, et al., 2006). However, the distinct effects of streetscape factors on walking preferences within these two sections remain inadequately explored. Moreover, streets of different structural categories exhibit unique configurations of elements, and their influence on walking preferences has also not been thoroughly discussed (Nagata, et al., 2020; Harvey, et al., 2015).

This chapter delves into the non-linear impact of physical walkability factors at street-level on perceived walkability, with a specific focus on how streetscape factors influence walking preferences in a non-linear manner. In quantifying the variables, this research focused on the visual aspects of discussing and measuring physical streetscape elements and walking preferences. Considering that vision is the primary way pedestrians interact with their environment, understanding walkability in the visual aspect is highly valuable.

We posed several research objectives: To determine if streetscape factors exhibit non-linear associations with walking preferences (2a). Furthermore, to assess the variability in how streetscape factors influence walking preferences across different categories of street segments and intersections (2b). Finally, to explore how findings from non-linear regression can be utilized to inform perceived walkability (2c). To achieve these objectives, we utilized street-view big data, CV techniques, and a machine-learning regression algorithm. Our research provides valuable insights that could influence decision-making in the design of urban streetscapes to promote pedestrian perceptions.

5.2. Materials and Methods

5.2.1. Case study site

Tokyo is a high-density urban area with a substantial reliance on rail-based public transportation, enhancing pedestrian experiences and perceptions is a pivotal component of the city's spatial development strategy. This chapter focused on Setagaya Ward, a highly populous area situated on the southwestern edge of Tokyo (Figure 5.1). Setagaya is well served by extensive and efficient rail networks (Figure 5.2a). These networks guarantee swift transit of trains to major hubs, such as Shinjuku, Shibuya, and other districts of Tokyo. The ward also has diverse land-use patterns, integrating residential, commercial, and industrial zones (Figure 5.2b). This varied

urban landscape, combined with its high population density and superior transportation facilities, makes Setagaya an exemplary model for studying walkability in urban environments (Setagaya City, 2023).

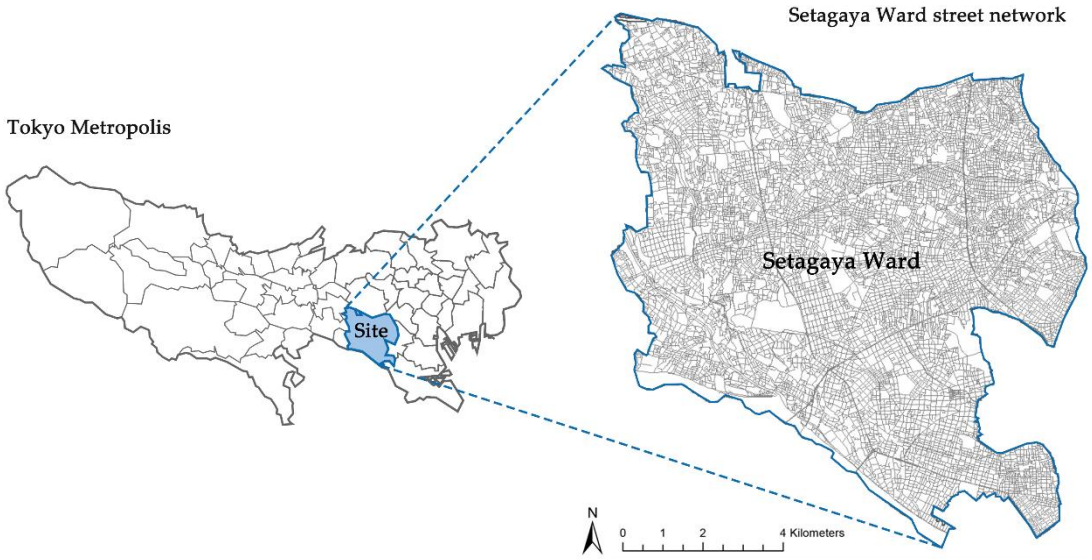
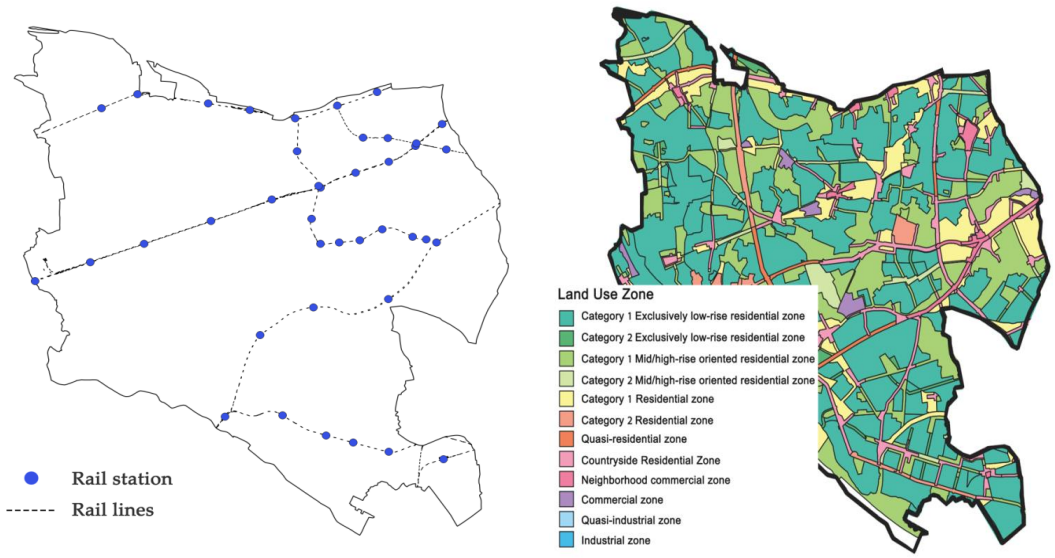


Figure 5.1 *Case study site.*



(a) (b)

Figure 5.2 *(a) Rail transit network and stations; (b) land use zones.*

5.2.2. Study scope

To delineate the scope of our study and subjects of investigation, we followed the established methodology of Chapter 4 to divide the streets into two main components: street segments and intersections.

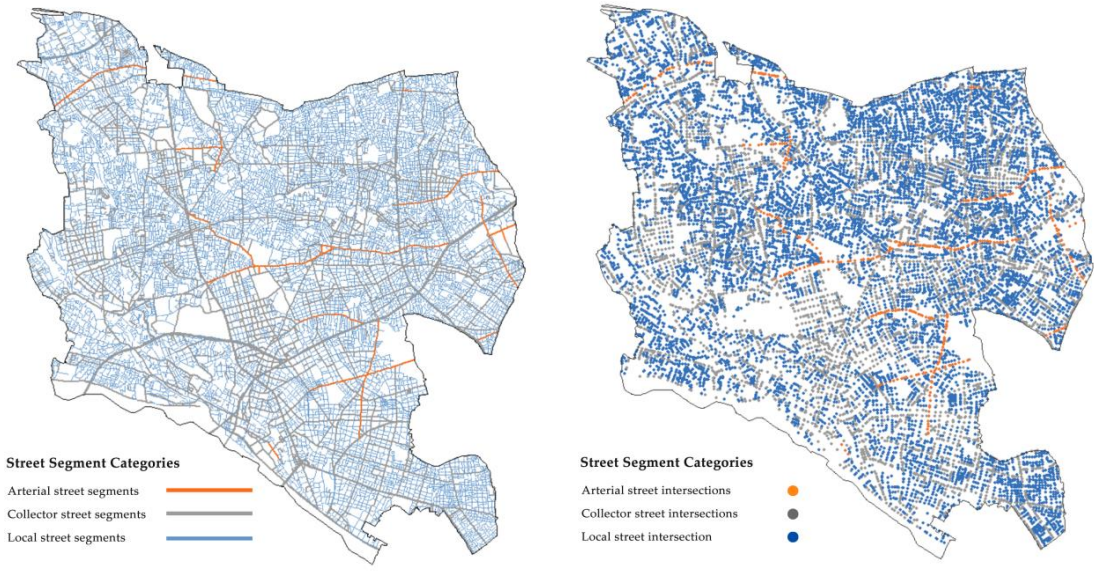
To gain a more nuanced understanding of whether differences in the street structure affect the relationship between streetscapes and walking preferences, we further classified street segments and intersections into distinct categories. We drew upon the Road Structure Ordinance (Japan Road Association, 2015) and research conducted by Neighborhood Street Research Group (Neighborhood Street Research Group, 1989), Kato and Kanki (2019), and Nagata et al. (2020) in Japanese contexts and utilized street width classes as the criteria for the classification. Through the utilization of width classification information in street network centerlines of the DRM (DRM, 2023), we discerned three categories of streets within Tokyo's Setagaya Ward accessible to pedestrians. These categories are delineated by width thresholds: streets with widths equal to or exceeding 13 m, between 5.5 and less than 13 m, and between 3 and less than 5.5 m. Each category corresponds to a specific type of street segment: arterial, collector, and local (Nagata, et al., 2020; Kato & Kanki, 2019). For street intersections, we employed a similar classification approach: intersections where arterial streets intersect were defined as arterial street intersections, those where collector streets intersect were termed collector street intersections, and those where local streets intersect were designated as local street intersections. Table 5.1 records detailed descriptions for each classification.

Table 5.1 *The categorization of street segments and intersections based on their width classes.*

Category	Width	Definition	Example
Street Segments			
Arterial street segment	W ≧ 13 m	Arterial streets constitute the basic framework of national road transportation as public roads, with the majority including bicycle lanes and pedestrian sidewalks.	
Collector street segment	5.5 m ≦ W < 13 m	Collector streets constitute the major roads forming the basic network of the road transportation system, with most of these roads including bicycle lanes and pedestrian sidewalks.	
Local street segment	3 m ≦ W < 5.5 m	Local streets serve mainly local traffic with short trip lengths. Streets of this class are usually formed in a disorderly manner adjacent to residential areas. There are few streets with bicycle and pedestrian sidewalks along the roadside.	
Street Intersection			
Arterial street intersection	The widest leg ≧ 13 m	The intersection is crossed by an arterial street and serves as a crucial node for the urban backbone street network. Such intersections typically have well-developed crossing facilities and traffic signals.	
Collector street intersection	The widest leg falls within the range 5.5 m ≦ W < 13 m	The intersection is crossed by a collector street and serves as a vital node for the neighborhood backbone road network. These intersections typically have well-developed crossing facilities and traffic signals in most cases.	

Local street intersection	The widest leg falls within the range 3 m ≦ W < 5.5 m	The intersection is crossed by a local street and forms part of the branching structure of the neighborhood street network. These intersections sometimes have crossing facilities and traffic signals.	

In Figure 5.3, the various categories of street segments (Figure 5.3a) and intersections (Figure 5.3b) within the Setagaya ward were mapped.



(a) (b)

Figure 5.3 *(a) Mapping different categories of street segments and (b) street intersections.*

5.2.3. Analysis framework

Figure 5.4 illustrates the study framework. We collected walking preference data using an online crowdsourced survey. These data were instrumental in training the deep learning models, which were then used to predict walking preferences throughout the street segments and intersections in Setagaya Ward. Subsequently, we identified and quantified the factors related to streetscapes in both segments and intersections by employing panoptic segmentation (Kirillov, et al., 2019) combined with Geographic Information System (GIS) methods. Finally, we utilized an Extreme Gradient Boosting (XGBoost) (Chen & Guestrin, 2016) machine learning model to analyze the complex influences of streetscape factors on walking preferences, with street segments and intersections as distinct categories.

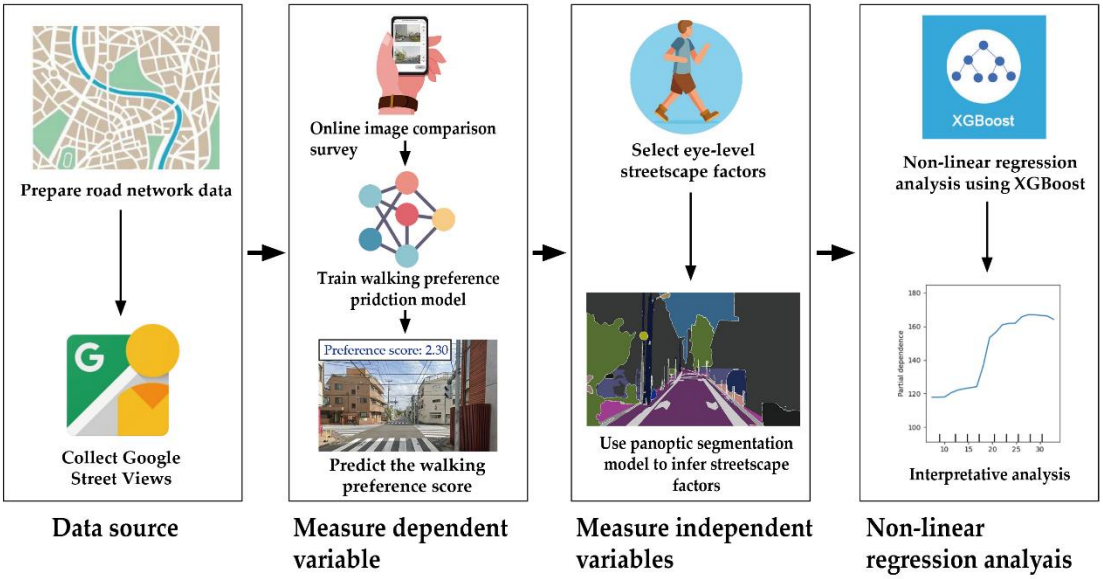


Figure 5.4 *Analysis framework.*

5.2.4. Data preparation

We used GSVs as the data source. We captured approximately 50,000 panoramic images at selected coordinates, which included around 46,000 images representing street segments and about 4,000 focusing on street intersections. These panoramas were then converted into front-view images to simulate an street-level perspective. In this chapter, to gain a detailed understanding of the relationship between streetscapes and perceived walkability, individual street view images were utilized as the units of analysis.

5.2.5. Dependent variables: walking preference scores (street-level perceived walkability)

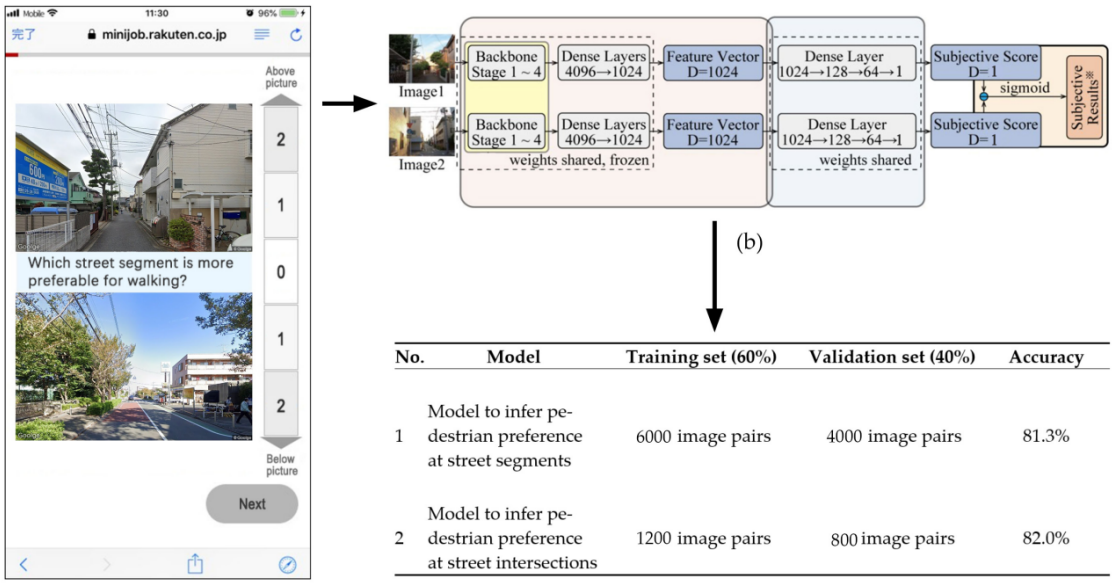
For quantifying walking preferences, this chapter adopts the method established in Section 4.2.2 of Chapter 4, which blends pairwise image comparison data with DCNN to predict walking preference scores. Regarding the dataset, we conducted a new crowdsourcing survey to collect information on walking preferences. Unlike the survey used in Chapter 4, this survey specifically focuses on the preference perspective and differentiates between street segments and street intersections for targeted information collection. Additionally, this crowdsourcing survey also gathered data on people's lingering preferences, which were not utilized in this chapter but were instead applied in Chapter 6. The process for quantifying walking preferences follows the methodology outlined in Chapter 4, involving: (1) preparing image survey data, (2)

collecting preference opinions through a pairwise image comparison survey, and (3) training the inference model to predict preferences.

Image survey data preparation

To develop a model capable of predicting walking preference scores, compiling a training dataset that includes image survey data and the corresponding preference labels from individuals is essential. To prepare our image survey data, we meticulously chose 1000 images that reflect the streetscape characteristics of street segments and 200 images that illustrate the streetscape features of intersections from our collected street-view image dataset. During the selection process, we considered the image's representativeness to ensure its ability to reflect diverse street hierarchies and intersection scenarios. Subsequently, we randomly paired these selected survey images, creating pairs in which each image was matched with the other images ten times. This process generated 10,000 image pairs for street segments and 2000 for street intersections.

We then adapted the image comparison sets to an interactive interface optimized for use on mobile devices (Figure 5.5a). The respondents were presented with questions designed for a comparative analysis of street segments and intersections, such as: "Which street segment is preferable for walking?" and "Which street intersection is preferable for walking?" Then, comparative levels were measured using a five-point scale. Although the survey also included questions related to lingering preferences, these were not utilized in this chapter but in Chapter 6.



No.	Model	Training set (60%)	Validation set (40%)	Accuracy
1	Model to infer pedestrian preference at street segments	6000 image pairs	4000 image pairs	81.3%
2	Model to infer pedestrian preference at street intersections	1200 image pairs	800 image pairs	82.0%

(a) (c)

Figure 5.5 *(a) Image comparison survey interface (translated to English, original in Japanese); (b) a network to infer walking preference at street segments and intersections; (c) training dataset parameters and results.*

Walking preference surveying

In February 2023, a survey involving pairwise image comparisons was conducted among volunteer participants of diverse ages and genders living in Japan, facilitated by a professional survey company. Approximately 18,000 anonymous respondents were recruited. Each comparison pair received responses from 10 different respondents. The highest number of responses from a single individual was 120, whereas the lowest number was one. In total, we collected over 120,000 responses with an average of 6.6 responses per respondent. The survey successfully achieved a balanced representation of gender and age among the respondents, ensuring that the results accurately reflected the broader population.

Prediction model training

We developed two distinct prediction models using the training data, which consisted of image pairs of street segments and intersections from the image surveys, along with the corresponding comparative preference scores gathered as labels. This approach builds on the DCNN prediction model (Figure 5.5b) proposed in Chapter 4. The models achieved accuracy rates of 81.3% for street segments and 82.0% for street intersections (Figure 5.5c). These accuracy rates represent the proportion of correctly predicted samples out of the total number of samples tested. During the model

training phase, we opted not to create separate models for different genders and age groups due to concerns that subdividing the training data could diminish the accuracy of the predictive models.

Prediction model application

We utilized the trained models to predict the GSV scores across the street segments and intersections in Setagaya Ward. They were then normalized within their respective categories using the z-score method for subsequent regression analysis. Figure 5.6 illustrates the spatial mapping of preference scores for street segments of different categories, arranged in descending order from warm to cold based on color. Figure 5.6a,c,e depicts street scenes with high predicted preference scores based on arterial street segments, collector street segments, and local street segments, respectively. Meanwhile, Figure 5.6b,d,f illustrate examples with lower predicted scores based on the aforementioned three categories of segments.

Figure 5.7 shows the preference-mapping pattern for street intersections. Figure 5.7a,c,e show the intersection scenes that received relatively high scores based on arterial street intersections, collector street intersections, and local street intersections, whereas Figure 5.7b,d,f show the intersection scenes that received relatively low scores on the aforementioned three categories of intersections.

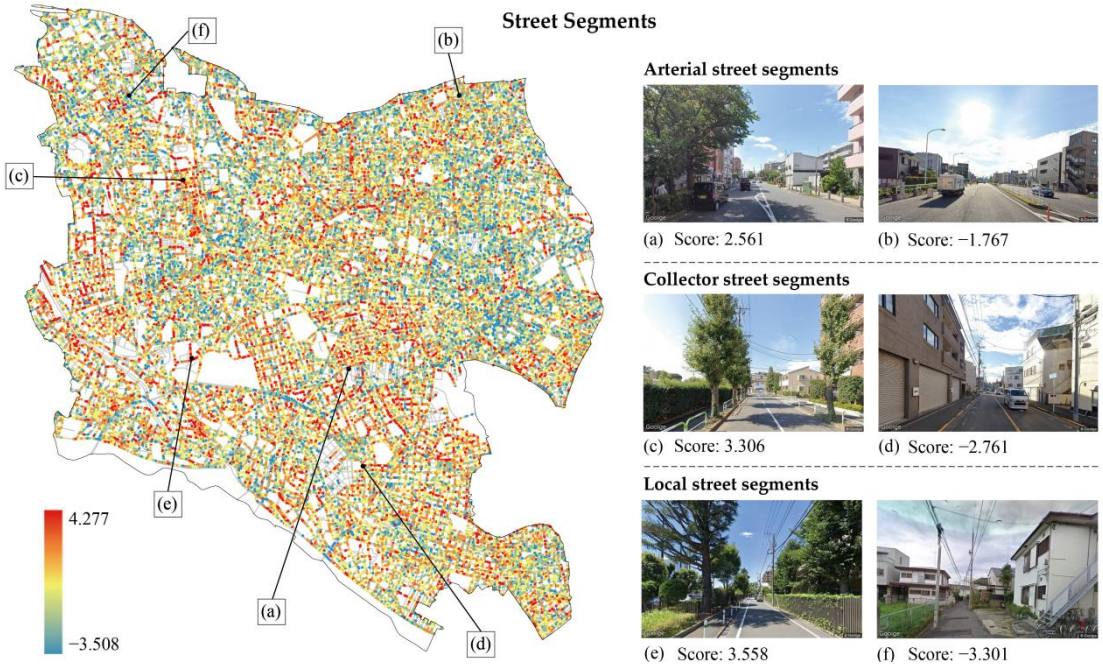


Figure 5.6 *Mapping preference scores for street segments. (a,b) Examples of preference predictions for arterial street segments, (c,d) collector street segments, and (e,f) local street segments.*

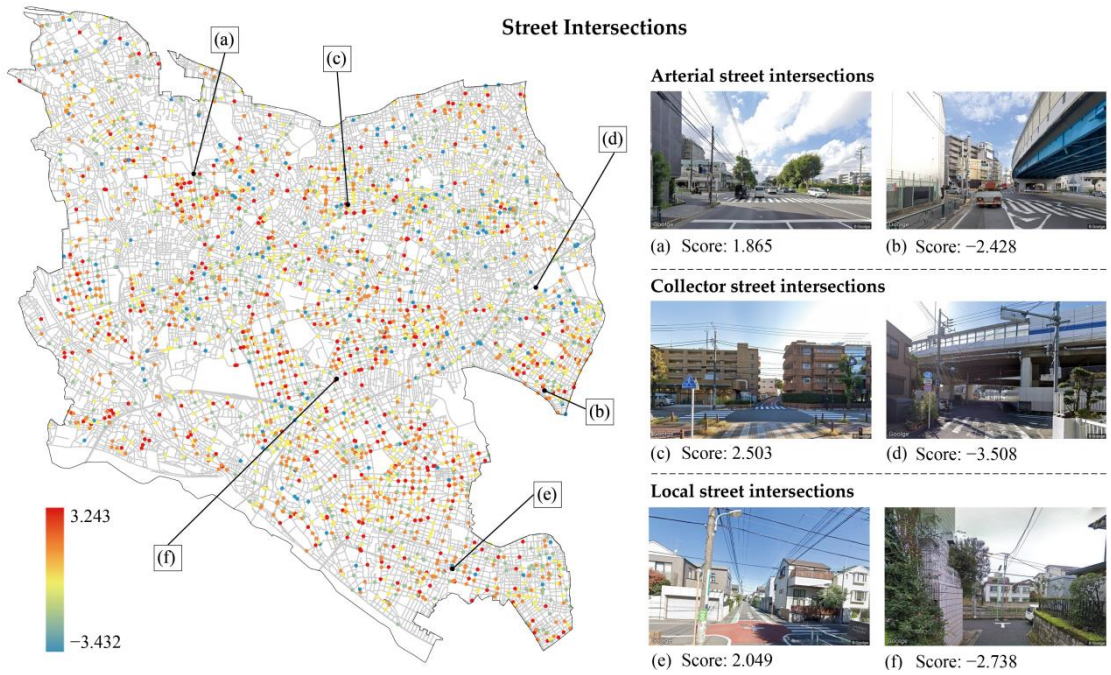


Figure 5.7 *Mapping preference scores for street intersections. (a,b) Examples of preference predictions for arterial street intersections, (c,d) collector street intersections, and (e,f) local street intersections.*

5.2.6. Independent Variables: Streetscape Factors (street-level physical walkability)

Selection of streetscape factors

We used the selected walkability factors from Chapter 4 as a reference, adjusting the factor list to meet the specific needs of this chapter's research purpose. Additionally, we considered the unique conditions of Japanese streets to select the independent variables for our analysis.

Firstly, considering that this chapter aims to explain how streetscape factors influence visually perceived walkability, we excluded those factors not directly related to visual aspects, such as destination-related factors and wind environment and ventilation, from Chapter 4. Additionally, this chapter adopts a skeletal-detail framework to distinguish between skeletal and detailed factors, aiming to understand the impact differences between skeletal and detailed elements of streets on walking preferences. Furthermore, we adjusted the measurement method for traffic flow from using the Betweenness level of the street in Chapter 4 to directly measuring the visual presence of vehicles, cyclists, and pedestrians in the images. Finally, we supplemented and removed some factors from Chapter 4 based on the characteristics of Japanese streetscapes. For example, we removed factors such as phone booths and water

scenery, as they are rarely present in the study area. Correspondingly, factors like the number of stop-lines, the number of utility poles and others were added.

Ultimately, we identified 22 factors for street segments and 18 factors for street intersections (Tables 5.2 and 5.3).

Table 5.2 *Streetscape factors of street segments.*

No	Factors	Description	Data Source	Mean			Std		
				Arterial	Collector	Local	Arterial	Collector	Local
	Skeletal streetscape								
1	Street-to-building ratio	Ratio of the street view index to the building view index	GSV	0.894	0.712	0.387	0.544	0.609	0.379
2	Sidewalk-to-roadway ratio	Ratio of the sidewalk view index to the roadway view index	GSV	0.252	0.285	0.234	0.150	0.239	0.272
	Detailed streetscape								
3	Elevated viaduct view index	Proportion of pixels of the elevated viaduct category.	GSV	0.005	0.013	0.002	0.030	0.059	0.022
4	Wall view index	Proportion of pixels of the wall category.	GSV	0.010	0.026	0.041	0.017	0.036	0.041
5	Fence view index	Proportion of pixels of the fence category.	GSV	0.023	0.031	0.038	0.019	0.038	0.044
6	Sidewalk view index	Proportion of pixels of the sidewalk category.	GSV	0.038	0.030	0.012	0.020	0.024	0.019
7	Roadway view index	Proportion of pixels of the roadway category.	GSV	0.145	0.118	0.181	0.025	0.033	0.026
8	Tree view index	Proportion of pixels of the tree category.	GSV	0.176	0.098	0.088	0.116	0.113	0.108

9	Shrub view index	Proportion of pixels of the shrub category.	GSV	0.024	0.037	0.053	0.026	0.045	0.062
10	Grass view index	Proportion of pixels of the grass category.	GSV	0.001	0.003	0.001	0.005	0.010	0.008
11	Number of street stores	Number of street stores detected.	GSV	0.718	0.652	0.155	1.025	1.018	0.656
12	Number of utility poles	Number of utility poles detected.	GSV	3.015	3.536	3.323	1.879	2.135	1.748
13	Number of street lights	Number of street lights detected.	GSV	0.989	0.883	0.597	0.817	0.894	0.699
14	Bike lane view index	Proportion of pixels of the bike lane category.	GSV	0.001	0.001	0.001	0.004	0.004	0.002
15	Number of benches	Number of benches detected.	GSV	0.010	0.003	0.003	0.100	0.058	0.065
16	Number of trash-cans	Number of trash-cans detected.	GSV	0.020	0.032	0.054	0.150	0.198	0.254
17	Number of awnings	Number of awnings detected.	GSV	0.029	0.018	0.007	0.169	0.156	0.089
18	Number of mailboxes	Number of mailboxes.	GSV	0.002	0.009	0.025	0.050	0.099	0.160
19	Number of banners	Number of banners detected.	GSV	0.125	0.097	0.045	0.394	0.374	0.258
20	Number of riders	Number of riders detected.	GSV	0.263	0.162	0.075	0.553	0.435	0.291
21	Number of vehicles	Number of vehicles detected.	GSV	4.119	2.911	1.748	2.066	2.203	1.594
22	Number of pedestrians	Number of pedestrians detected.	GSV	0.898	0.606	0.367	1.136	1.010	0.758

Table 5.3 *Streetscape factors of street intersections.*

No	Factors	Description	Data Source	Mean			Std		
				Arterial	Collector	Local	Arterial	Collector	Local
	Skeletal streetscape								
1	Average crossing distance	Mean length that pedestrians need to cover to traverse the intersection from one side to the other.	DRM	3.342	2.133	1	1.215	0.931	0
2	Number of legs	Total count of segments that intersect at a particular crossing point.	DRM	3.709	3.677	3.488	0.550	0.528	0.513
	Detailed streetscape								
3	Elevated viaduct view index	Proportion of pixels of the elevated viaduct category.	GSV	0.040	0.013	0.003	0.106	0.058	0.028
4	Corner space view index	Proportion of pixels of the corner space category.	GSV	0.033	0.033	0.020	0.018	0.021	0.017
5	Fence view index	Proportion of pixels of the fence category.	GSV	0.015	0.021	0.030	0.017	0.025	0.033
6	Crosswalk view index	Proportion of pixels of the crosswalk category.	GSV	0.033	0.023	0.003	0.027	0.021	0.010
7	Curb ramp view index	Proportion of pixels of the curb ramp category.	GSV	0.002	0.002	0.001	0.001	0.002	0.001
8	Number of vehicle traffic lights	Number of vehicle traffic lights detected.	GSV	1.118	0.709	0.042	0.891	0.922	0.270
9	Number of pedestrian traffic lights	Number of pedestrian traffic lights detected.	GSV	0.610	0.428	0.018	0.780	0.775	0.164
10	Bike lane view index	Proportion of pixels of the bike lane category.	GSV	0.001	0.001	0.001	0.003	0.004	0.005
11	Number of stop lines	Number of stop lines detected.	GSV	0.348	0.334	0.305	0.496	0.507	0.493
12	Number of street lights	Number of street lights detected.	GSV	0.877	0.751	0.674	0.711	0.770	0.727

13	Tree view index	Proportion of pixels of the tree category.	GSV	0.069	0.071	0.088	0.074	0.084	0.102
14	Shrub view index	Proportion of pixels of the shrub category.	GSV	0.012	0.025	0.048	0.020	0.033	0.052
15	Grass view index	Proportion of pixels of the grass category.	GSV	0.001	0.001	0.001	0.003	0.006	0.007
16	Number of riders	Number of riders detected.	GSV	0.465	0.269	0.105	0.707	0.548	0.345
17	Number of vehicles	Number of vehicles detected.	GSV	3.903	2.585	1.620	1.848	1.942	1.471
18	Number of pedestrians	Number of pedestrians detected.	GSV	1.069	0.751	0.458	1.213	1.092	0.842

Quantification of streetscape factors

This chapter employs deep-learning-based CV algorithms to analyze GSV images and quantify both non-countable and countable streetscape variables. For streetscape factors with continuous or non-countable attributes, such as elevated viaducts, walls, trees, and sidewalks, we computed pixel view indices. The pixel view index of a streetscape factor is commonly defined as the ratio of its pixels to all pixels within a GSV (Qiu, et al., 2023; Wu, et al., 2023), as described in Equation (5.1).

$$V_{obj} = \frac{\sum_{i=1}^{n} Pixel_{obj}}{\sum_{i=1}^{n} Pixel_{total}}, obj \in \{tree, crosswalk, sidewalk, etc\} \quad (5.1)$$

where V_{obj} is the view index, $\sum_{i=1}^{n} Pixel_{obj}$is the number of pixels of the streetscape factor, and $\sum_{i=1}^{n} Pixel_{total}$ is the total number of pixels. This ratio reflects the proportion of streetscape variables in a pedestrian's street-level view.

Skeletal streetscape factors, street-to-building ratios, and sidewalk-to-roadway ratios were calculated based on pixel statistics. The street-to-building ratio was calculated as the ratio of the sum of pixels of vertical elements, such as buildings and walls, to the sum of pixels of sidewalks, roadway, people, and vehicles on them. This method, as shown in Equation (5.2), has been outlined in previous studies (Qiu, et al., 2023; Wu, et al., 2023).

$$V_{sb} = \frac{\sum_{i=1}^{n} Pixel_{sidewalk} + \sum_{i=1}^{n} Pixel_{roadway} + \sum_{i=1}^{n} Pixel_{person} + \sum_{i=1}^{n} Pixel_{vehicle}}{\sum_{i=1}^{n} Pixel_{building} + \sum_{i=1}^{n} Pixel_{wall}} \quad (5.2)$$

where V_{sb} is the visual ratio of street-related pixels to building-related pixels; and $\sum_{i=1}^{n} Pixel_{sidewalk}$, $\sum_{i=1}^{n} Pixel_{roadway}$, $\sum_{i=1}^{n} Pixel_{person}$, $\sum_{i=1}^{n} Pixel_{vehicle}$, $\sum_{i=1}^{n} Pixel_{building}$, and $\sum_{i=1}^{n} Pixel_{wall}$ are the total pixels of the sidewalk, roadway, person, vehicle, building, and wall categories, respectively.

The sidewalk-to-roadway ratio was calculated by comparing the sum of the pixels of the sidewalk and pedestrians with the sum of the pixels of the roadway and vehicles, as shown in Equation (5.3).

$$V_{sr} = \frac{\sum_{i=1}^{n} Pixel_{sidewalk} + \sum_{i=1}^{n} Pixel_{person}}{\sum_{i=1}^{n} Pixel_{roadway} + \sum_{i=1}^{n} Pixel_{vehicle}} \quad (5.3)$$

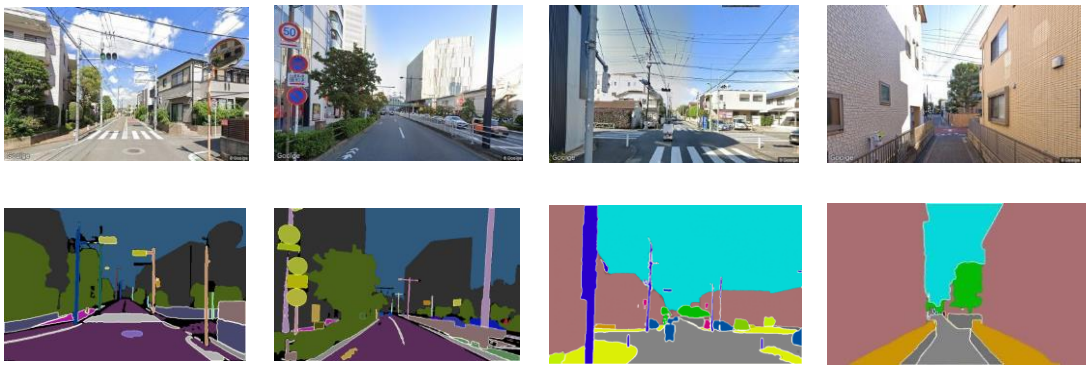
where V_{sr} is the visual ratio of sidewalk-related pixels to roadway-related pixels; and $\sum_{i=1}^{n} Pixel_{sidewalk}$, $\sum_{i=1}^{n} Pixel_{roadway}$, $\sum_{i=1}^{n} Pixel_{person}$, and $\sum_{i=1}^{n} Pixel_{vehicle}$ are the total pixels of sidewalk, roadway, person, and vehicle categories, respectively.

Although pixel statistics are effective for computing certain streetscape factors, they are less suited for quantifying discontinuous and countable streetscape factors, such as benches and street lights. Therefore, specifying the precise numbers offers clear explanations and practical guidance for design purposes (Huang, et al., 2023).

In Chapter 4, we employed semantic segmentation and YOLOv5 to quantify non-countable and countable factors. However, these methods presented complexities and lacked integration. To more efficiently and simultaneously extract both pixel- and quantity-based features from images, this chapter introduces the use of a panoptic segmentation algorithm as a superior alternative. Unlike semantic segmentation, panoptic segmentation not only provides pixel counts for each category but also identifies and counts individual instances within these countable categories. This dual capability enables a comprehensive and detailed image analysis.

We trained a model capable of inferring streetscape variables from images based on the Mask2Former framework (Cheng, et al., 2022) using a ResNet50 backbone (He, et al., 2016). The training dataset used was Mapillary Vistas v2.0 (Neuhold, et al., 2017), which encompassed a comprehensive set of 124 feature categories (116 categories for panoptic segmentation). This training dataset was carefully selected due to its ability to effectively encapsulate most of the selected variables, which was central to the focus of this chapter. Moreover, we used a pre-trained panoptic segmentation model, also based on the Mask2Former structure but trained on the ADE20K dataset (Zhou, et al., 2017), to extract pixel values for indicators, including the view index of trees, shrubs, and grasses, and number-counting information, such as awnings.

Figure 5.8 shows the results of the panoptic segmentation applied to the images. These images were segmented based on models trained using both the Mapillary Vistas v2.0 (Figure 5.8a,b) and the ADE20K dataset (Figure 5.8c,d).



(a) (b) (c) (d)

Figure 5.8 *(a) Example of panoptic segmentation at a street intersection using the Mapillary Vista v2.0 dataset; (b) panoptic segmentation at a street segment using the Mapillary Vista v2.0 dataset; (c) panoptic segmentation at a street segment using the ADE20K dataset; (d) panoptic segmentation at a street intersection using the ADE20K dataset.*

In addition to using GSV data, we utilized DRM data as a supplementary resource to quantify variables, such as leg numbers and the average crossing distance. To calculate the leg numbers, which represent the total number of segments intersecting at a specific crossing point, we counted the number of segments converging at each intersection point. For calculating the average crossing distance, which represents the difficulty people face when crossing streets in intersection areas, we measured the average number of drive lanes of the segments intersecting at the intersections. To facilitate this calculation, we retrieved the lane count classification from the DRM, which included four levels within the study area. We then assigned scores ranging from 1 to 4 to represent the gradation from the lowest to the highest level of the drive lane classification and calculated the average score to characterize the average crossing distance.

5.2.7. XGBoost regression analysis

For the regression analysis, we utilized the XGBoost machine learning algorithm (Chen & Guestrin, 2016) to explore the influence of streetscape factors on walking preferences.

XGBoost is a software library designed to be highly efficient in terms of computational speed and model performance by providing parallel-tree boosting. It optimizes Gradient Boosting Decision Tree (GBDT) training and dramatically

improves the computational speed and model performance by conducting distributed computing. It attempts to accurately predict target variables by combining a set of simpler and more fragile model estimates. While this model performs better than its counterparts, it may encounter challenges, such as overfitting or the misjudgement of component weights, when exposed to a limited number of samples or certain model-specific factors.

XGBoost modeling provided two key insights. First, the relative importance of each predictor was identified, totaling 100%, using the mean-decreasing impurity. The Mean Decrease Gini method, based on the Gini impurity index (Atkinson, 1970), quantifies the effect of each independent variable on the dependent variable. Second, the model generated partial dependence plots (PDPs). These plots illustrate the relationship between a given predictor and the predicted walking preferences, considering interactions with other predictors.

5.3. Results

5.3.1. Results of XGBoost model training

This chapter was conducted using Python within the PyCharm Integrated Development Environment, version 2022.0.2. We employed the XGBRegressor module from the XGBoost library for our regression analysis, aiming to train separate models to explore the non-linear impacts of streetscape factors on walking preferences across various categories of street segments and intersections. To ensure the performance of the models, we employed GridSearchCV from scikit-learn to adjust the parameters. Table 5.4 presents an overview of the optimal parameter configurations for all models. We used the pseudo-R-squared as the assessment metric to evaluate the training performance of the models. In total, six models were developed. Among them, three models focused on street segments—specifically arterial street segments, collector street segments, and local street segments—with R-squared values of 0.38, 0.41, and 0.31, respectively. The other three models targeted intersections connected to arterial streets, collector streets, and local streets, achieving R-squared values of 0.35, 0.39, and 0.29, respectively.

Table 5.4 *XGBoost model training parameters and results.*

	Arterial (Segment)	Collector (Segment)	Local (Segment)	Arterial (Intersection)	Collector (Intersection)	Local (Intersection)
Gamma	0	0.1	0.1	0.1	0	0.1
Learning_rate	0.01	0.1	0.1	0.01	0.01	0.01
Max_depth	4	3	5	4	4	4
Min_child_ weight	3	1	3	5	3	5
n_estimators	200	300	300	300	300	300
Pseudo R2	0.38	0.41	0.31	0.35	0.39	0.29

5.3.2. Results for relative importance

Figure 5.9 delineates the relative importance of streetscape factors in predicting walking preferences across street segments. As depicted in Figure 5.9, skeletal and detailed streetscape factors based on arterial street segments, collector street segments, and local street segments explain 12.042% and 87.958%, 19.333% and 80.667%, and 26.273% and 73.727% of the variance, respectively.

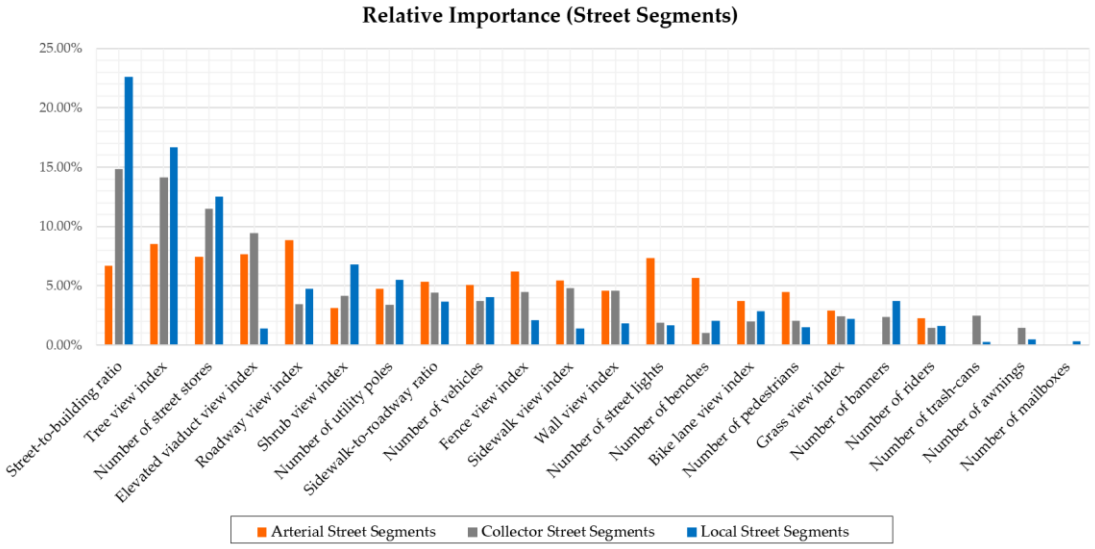


Figure 5.9 *Relative importance of streetscapes on walking preference (street segments).*

Specifically, for arterial street segments, the roadway view index was a key factor, with a significant impact of 8.847%, followed by the tree view index (8.529%) and elevated viaduct view index (7.657%). Other significant variables included the number of street stores (7.428%) and number of street lights (7.325%). In terms of collector streets, the most influential variables impacting walking preferences were the street-

to-building ratio (14.850%), tree view index (14.154%), number of street stores (11.505%), elevated viaduct view index (9.416%), and sidewalk view index (4.822%). For local streets, in addition to the street-to-building ratio, tree view index, and number of street stores, the shrub view index and number of utility poles are also key influencing factors. Conversely, the presence of trash cans, mailboxes, and awnings exerted a lesser effect on walking preferences across nearly all categories of segments, indicating their lower predictive power for walking preferences.

In the analysis of street intersections (Figure 5.10), we observed that the proportions of skeletal factors and detail factors at intersections linked to arterial streets, collector streets, and local streets were found to be 8.791% and 91.209%, 5.035% and 94.965%, and 13.64% and 86.36%, respectively.

For arterial street intersections, the elevated viaduct view index was the most influential factor, accounting for an importance of 16.046%. This was closely followed by the tree view index at 15.483% and crosswalk view index at 6.884%. These three factors also stand out as key determinants for collector street intersections. Meanwhile, at local street intersections, the predominant factors included the number of intersection legs (13.638%), the grass view index (9.383%), and the crosswalk view index (8.397%). In contrast, across all categories of intersections, factors, such as the number of stop lines, riders, and pedestrian traffic lights, had a considerably lower impact on walking preferences.

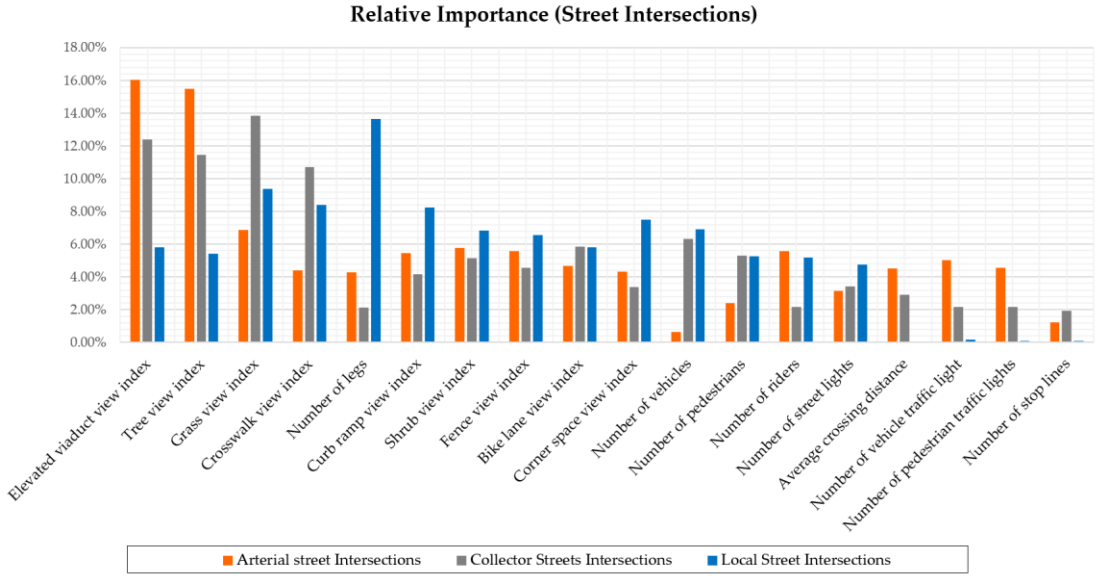


Figure 5.10 *Relative importance of streetscapes on the walking preference (street intersections).*

5.3.3. PDP results

The PDP depicts a fine-grained analysis of the relationship between independent and dependent variables. For each plot, the X-axis shows the distribution and variation of the independent variable, and the Y-axis represents the level of walking preference.

Skeletal streetscapes in segment sections

Figure 5.11 illustrates the non-linear effects of skeletal streetscape factors on walking preferences within street segments. The plot revealed an inverted U-shaped pattern for the street-to-building ratio based on arterial street segments (Figure 5.11a). When the ratio ranged from 0.5 to approximately 1.3, a positive correlation with the walking preference was observed. However, when the ratio exceeded 1.3, the positive impact began to plateau, indicating a diminishing influence on the walking preference. Furthermore, when the ratio exceeded 3, walking preferences were adversely affected. This observed pattern can be compared with studies examining the distance-to-height (D/H) ratio (Ashihara, 1983), which contrasts the street width with building height. This non-linear trend is not unique to arterial street segments but is also observed in other segment categories, suggesting that certain skeletal factors have a universal impact on walking preferences across different categories of street segments.

Another skeletal streetscape factor, the sidewalk-to-roadway ratio (Figure 5.11b), displayed a threshold effect as well, particularly on collector streets and arterial streets. The beneficial effects on walking preferences increase until the ratio reaches around 0.5 for both arterial and collector street segments, after which the marginal benefits significantly decline. This indicates that there is an optimal proportion of the sidewalk width to roadway width based on arterial and collector street segments that encourages pedestrian preferences, beyond which the added value decreases. However, variations in the sidewalk-to-roadway ratio have a less pronounced impact on walking preferences for local streets.

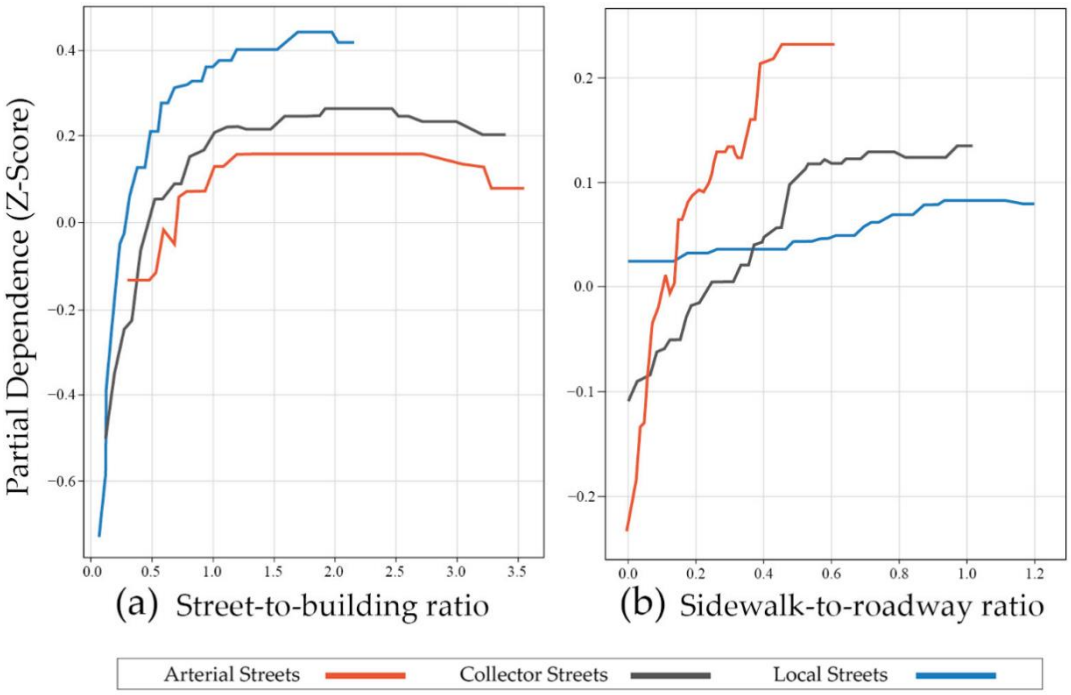


Figure 5.11 *(a,b) PDP of skeletal factors affecting walking preferences for street segments.*

Detailed streetscapes in street segments

Figure 5.12 shows the influence of detailed streetscape factors on walking preferences across different categories of street segments. Initially, several factors reveal consistent patterns that either enhance or discourage walking in specific street categories. For instance, the increase in street stores (Figure 5.12a) on both arterial and collector streets demonstrated a predominantly positive effect, indicating that when this factor is more visible, it makes walking more attractive. Likewise, the visibility of the roadway (Figure 5.12b) has a reliably positive effect on the walking preference on local street segments, despite displaying a different trend in the other segment categories. On the other hand, factors, like the fence (Figure 5.12c), wall (Figure 5.12d), and elevated viaduct view index (Figure 5.12e), tend to discourage walking across almost all segment categories, indicating an overall negative impact on the desire to walk.

Additionally, we observed clear threshold effects for some detailed streetscapes. Factors, like the tree view index (Figure 5.12f) and sidewalk view index (Figure 5.12g), on arterial street segments and collector street segments show an inverted L pattern. Specifically, enhancing the visibility of trees to around 0.32 on arterial, 0.31 on collector, and 0.35 on local street segments has a positive effect on people's preference to walk. However, after reaching these visibility levels, the additional advantage of more tree

visibility for walking preferences markedly diminishes (Figure 5.12f). For the sidewalk view index (Figure 5.12g), on arterial and collector street segments, visibility within the range of 0.06 to 0.07 and from 0 to 0.06 positively influences walking preferences. Beyond these ranges, any increase in visibility does not have a significant impact. However, the trend in variations in the sidewalk view on local street segments shows a different trend (Figure 5.12g), presenting a left-slanted L pattern in its impact on walking preferences. This left-slanted L pattern is also observed in factors, such as the number of vehicles (Figure 5.12h) across all segment categories and the number of utility poles (Figure 5.12i) on arterial street segments.

Moreover, an inverted U-shaped pattern can be identified in certain detailed streetscape factors, such as the shrub view index (Figure 5.12j) based on collector and local street segments and the number of pedestrians (Figure 5.12k) based on collector street segments. Regarding the shrub view index (Figure 5.12j), changes in visibility within the ranges of 0–0.15 for collector street segments and 0–0.2 for local street segments positively influence walking preferences. However, surpassing these tipping points leads to a negative impact on walking preferences. The number of pedestrians (Figure 5.12k) on collector street segments has a positive influence on walking preferences up to a count of eight, after which it starts to have a negative effect, highlighting a balance point beyond which more pedestrians become a deterrent rather than an encouragement to walk on collector streets. Other factors that exhibit a similar pattern include the roadway view index (Figure 5.12b) based on arterial and collector street segments and the number of street stores (Figure 5.12a) based on local street segments.

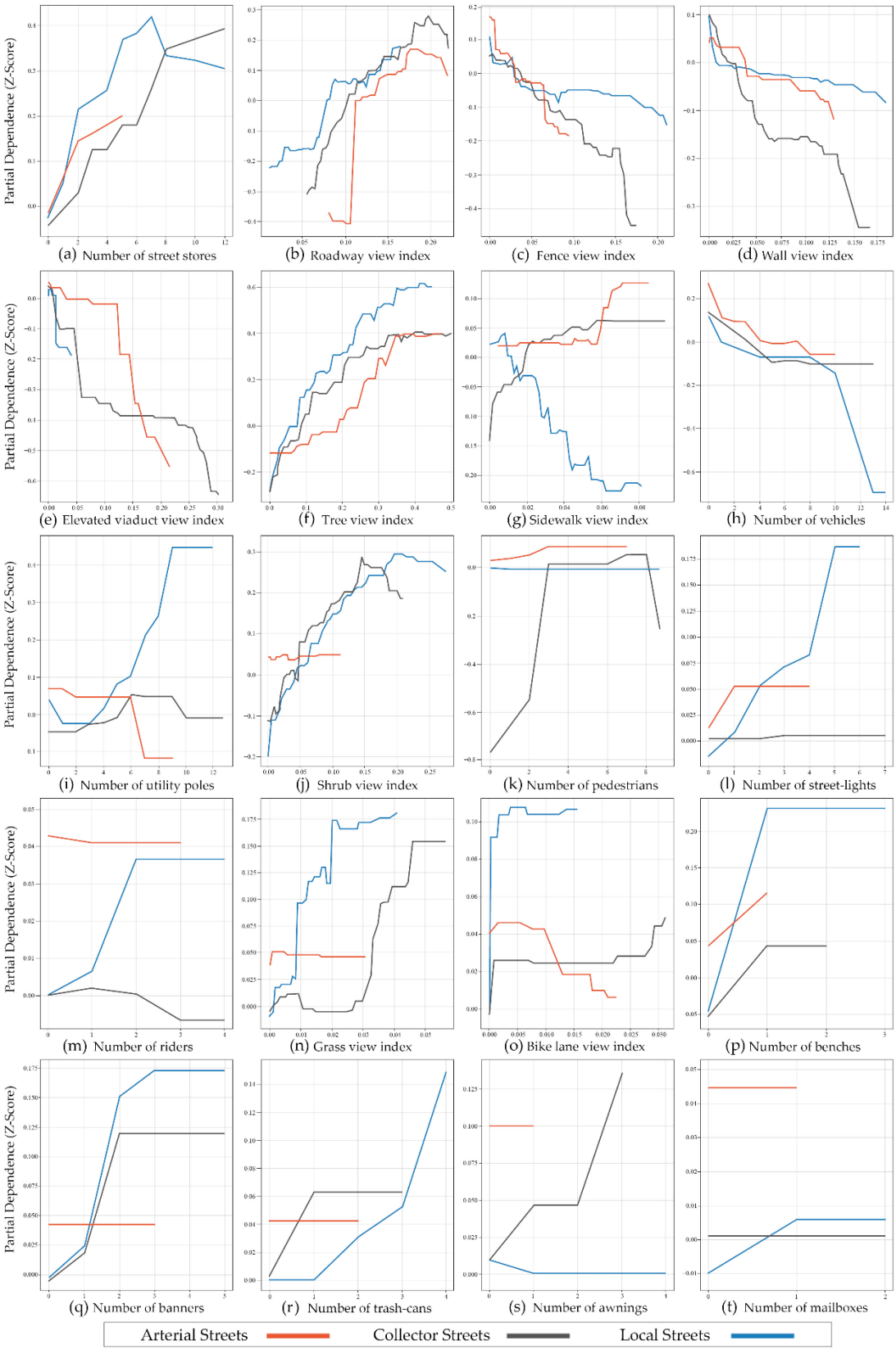


Figure 5.12 *(a–t) PDP of detailed factors affecting walking preferences for street segments.*

Skeletal streetscapes in street intersections

Figure 5.13 depicts the impact of skeletal streetscape factors on walking preferences at street intersections of varying categories. For arterial street intersections, the average crossing distance (Figure 5.13a) showed a generally decreasing preference for walking, indicating that larger average crossing distances are associated with lower walking preference. Conversely, for collector street intersections, a greater crossing distance tends to enhance walking preferences, up to a score of 2.7. However, for local street intersections, this variable did not vary because the values at this category of intersections were consistent. Therefore, it did not have an impact on walking preferences. Regarding the number of intersection legs (Figure 5.13b), a notable inverted-L pattern was observed based on local street intersections. Intersections with four legs have been observed to positively influence walking preferences. However, an increase beyond this number does not contribute to further enhancements in walking preferences. A different trend is observable for both arterial and collector street intersections, where the presence of four legs also acts as a tipping point. Beyond this level, an increase in the number of legs is found to have a negative impact on walking preferences, a trend that is particularly pronounced for arterial street intersections.

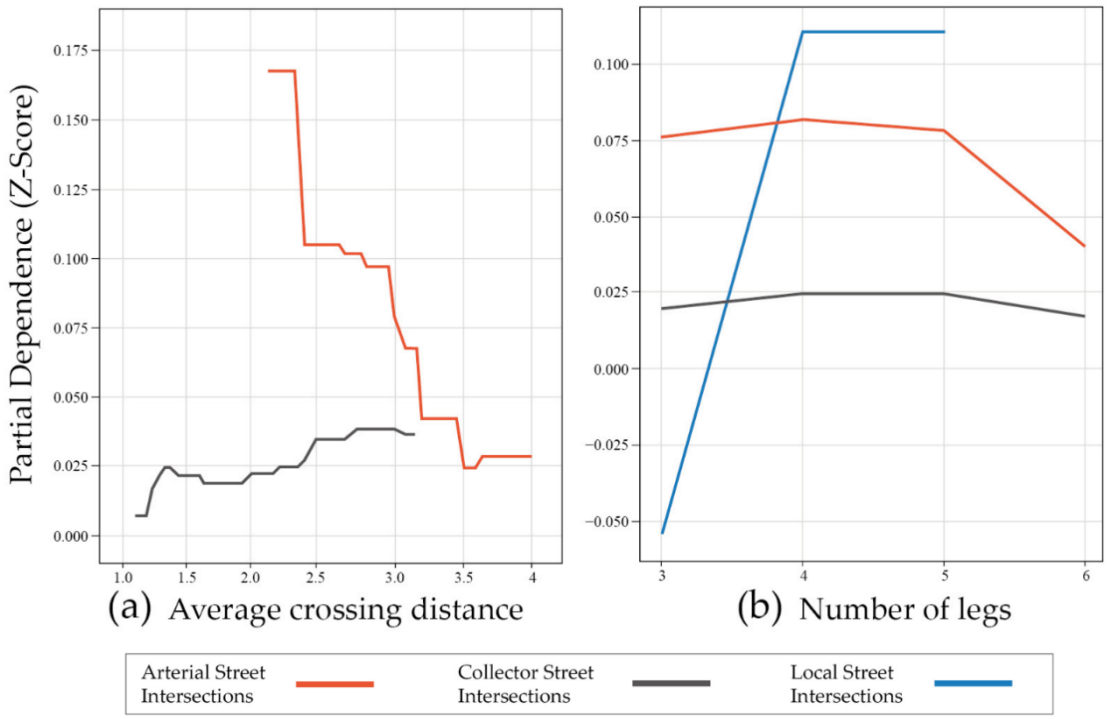


Figure 5.13 *(a,b) PDP of skeletal factors affecting walking preferences for street intersections.*

Detailed streetscapes in street intersections

In the detailed streetscape analysis of street intersections, factors, such as the crosswalk view index (Figure 5.14a), across intersections of all categories exhibited an overall positive correlation with walking preferences, reflecting that the number of marked crossings was positively correlated with perceived safety levels (Blečić, et al., 2016; Wang & Akar, 2018). Conversely, factors, like the number of vehicles (Figure 5.14b), at intersections interacting at collector streets exhibited a contrasting trend, showing that a higher presence of vehicles negatively impacts walking preferences.

Factors, such as the curb ramp view index (Figure 5.14c), the number of street lights (Figure 5.14d), and the corner space view index (Figure 5.14e), exhibited inverted L-shaped patterns in relation to walking preferences. Specifically, the curb ramp view index exhibited a positive correlation within specific ranges: from 0 to 0.002 at arterial and collector street intersections, and from 0 to 0.001 at local street intersections; however, beyond these thresholds, the marginal effect was significantly diminished. This trend can be understood as the strategic positioning of curb-ramp visibility at intersection corners aiding the transition from the sidewalk to street level (Blečić, et al., 2016) and offered accessible routes for people with physical disabilities, as well as for those using shopping carts or strollers. However, an excessive visibility of the ramp facility did not bring additional benefits. Additionally, a left-slanted L pattern can also be identified in some detailed factors, such as the elevated viaduct view index (Figure 5.14f), which exhibits a decreasing trend within specific ranges and then leveling off across all categories of intersections. This trend diverges from the pattern observed along street segments, where the elevated viaduct view index (Figure 5.12e) tends to exhibit a nearly negative correlation with walking preferences on street segments.

Regarding factors, such as the fence view index (Figure 5.14g), an inverted U-pattern can be detected. Specifically, the observed trend indicated a significant positive correlation with walking preferences within specific ranges: from 0.04 to 0.07 for arterial street intersections, and from 0 to 0.04 for collector street intersections. However, when the value exceeded the tipping points, the positive influence began to diminish and a negative correlation emerged, suggesting nuanced interplay between perceived safety and the desire for unrestricted movement around intersections. While a moderate number of safe barriers improves perceived safety and positively influences walking preferences, an excess can lead to negative perceptions. This adverse effect is likely due to the feelings of restriction or confinement that the barrier creates, particularly at local street intersections. In addition, this trend was markedly different from that observed for street segments (Figure 5.12c). Other factors exhibiting

an inverted U-pattern include the tree view index (Figure 5.14h) and shrub view index (Figure 5.14i) across all categories of intersections.

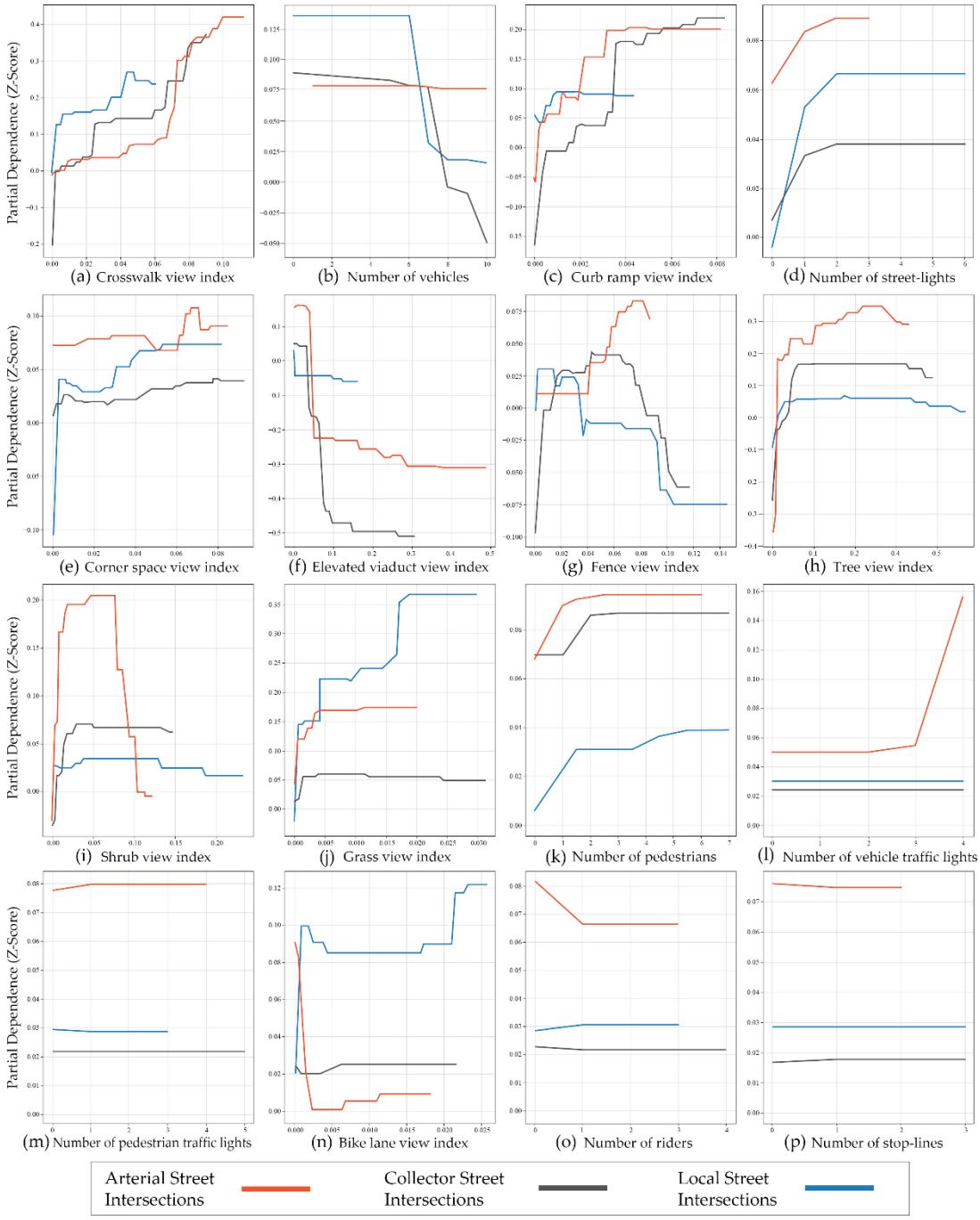


Figure 5.14 *(a–p) PDP of detailed factors affecting walking preferences for street intersections.*

5.4. Discussion and Implementation

The results of this chapter are important for guiding design practices aimed at creating pedestrian-friendly street environments. First, pinpointing critical factors is

vital for deciding which streetscape factors to prioritize for design intervention. By acknowledging these key variables, urban planners can optimize resource utilization, minimize waste, and significantly improve the overall efficiency of city development.

Existing studies have highlighted numerous key street-level factors in street segment sections that can influence walking preferences, such as the street-to-building ratio (Koo, et al., 2022), number of street stores (Chiang, et al., 2017; Herrmann-Lunecke, et al., 2021), roadway view index (Wu, et al., 2023), and tree view index (Ki & Lee, 2021). Similarly, for street intersections, factors, such as the crosswalk view index (Wang & Akar, 2018; Blečić, et al., 2016), elevated viaduct view index (Wang & Akar, 2018), and number of legs index (Wang & Akar, 2018; Blečić, et al., 2016), have been identified as key variables affecting pedestrian behavior and preferences. Our study underscores the significance of these factors but goes further by incorporating the heterogeneity of the street structure into our analysis. By categorizing streets into three categories for analysis, we have unearthed deeper insights. Our investigation reveals that while there are common effects of streetscape factors on walking preferences across different segment and intersection categories, significant differences also exist. For instance, within street segments, factors, such as the roadway, tree, and elevated viaduct view index, are crucial in influencing walking preferences on arterial streets (Figure 5.9). However, for collector and local streets, the key variables that predict walking preferences shift to the street-to-building ratio, tree view index, and the number of street stores. By highlighting these key factors, tailored to various street categories in practical design initiatives, an opportunity arises to improve pedestrian experiences and perceptions. This is achieved through the implementation of more precise and context-sensitive design strategies.

Second, our study employed PDPs to uncover potential non-linear trends across various streetscape factors, unveiling diverse patterns, such as upward, downward, and threshold effects. By observing these patterns, planners and designers can discern the optimal range for design interventions. For street segments, some streetscape factors showing an upward trend in relation to walking preferences, like street stores on arterial and collector street segments (Figure 5.12a), indicate that future urban renovations and design projects could enhance walkability by promoting ground-level small businesses along arterial and collector streets. For local street segments, integrating an appropriate number of stores can similarly contribute to increasing pedestrian activity and enhancing interest in walking. Conversely, the presence of elevated viaducts constantly undermines walking preferences, affecting both street segments (Figure 5.12e) and intersections (Figure 5.14f) negatively. Strategies aimed

at mitigating their impact—such as burying these structures or improving their permeability—should be considered to enhance walkability. Acknowledging threshold effects is vital for urban planners and designers, as it allows them to precisely adjust the number of specific streetscape factors and establish the most suitable extent for interventions. In this chapter, the findings suggest that enhancing tree visibility on streets of all categories up to a certain threshold can optimally boost the walking preference (Figure 5.12f). However, beyond this level, further improvements in tree visibility cease to yield additional benefits. These insights are consistent with established research (Yang, et al., 2021). These findings not only demonstrate the superiority of non-linear approaches but also provide designers with the necessary tools to precisely determine the optimal scale and intensity of design interventions.

Third, our research uncovers that while shared trends exist in the non-linear influence of streetscape factors on walking preferences across different categories of street segments and intersections, such as the effects of the wall view index (Figure 5.12d), elevated viaduct view index (Figure 5.12e), and tree view index (Figure 5.12f) on street segments, and the impacts of the crosswalk view index (Figure 5.14a), shrub view index (Figure 5.14i), and the number of street-lights (Figure 5.14d) at intersections, notable differences are also evident. For example, the sidewalk view index along street segments (Figure 5.12g) shows an initial increase followed by a plateauing trend based on arterial and collector streets, indicating an inverted L-shaped pattern. However, on local streets, sidewalk visibility tends to negatively correlate with walking preferences. This phenomenon can be attributed to the fact that in Japan, local streets often lack sidewalks and are subject to regulatory limits on traffic flow and speed, leading residents to become accustomed to freely walking on any part of the road (Akasaki, 1989; Kato & Kanki, 2019). Consequently, the introduction of sidewalks on local streets might be perceived as an impediment to the freedom of walking. Conversely, on arterial and collector streets, the high volume of vehicular traffic prompts a desire for dedicated and adequately wide pedestrian sidewalks, aligning with existing research findings (Herrmann-Lunecke, et al., 2021). Identifying these heterogeneous trends aids designers in more accurately and contextually applying non-linear statistical results in their work.

Last, our study revealed that the same factors can elicit different perceptual responses in street segments and intersections, as evidenced by factors, such as the fence view index (Figures 5.12c and 5.14g). This discovery underscores the significance of discerning and conducting a comprehensive examination of how various

streetscapes affect walking preferences within different morphological street sections when formulating street design strategies.

5.5. Conclusion

This chapter aimed to unravel the intricate and non-linear relationships between streetscape factors and individual walking preferences. We used the XGBoost regression method and importance plots to identify the relative importance of streetscape factors across different street categories and sections that influence walking preferences. Furthermore, by employing PDP, this chapter illustrated the non-linear trends of streetscape factors in street segments and intersections of various categories, aiding the identification of optimal design intervention thresholds for promoting walking preference. These insights offer planners and designers the opportunity to craft evidence-based and context-sensitive intervention strategies for different categories of street segments and intersections at the initial stages of design.

This chapter had several limitations. First, this study focused on walking preferences and desires. However, the role of streets as places is also critically important. Later research should consider both the function of streets as pathways and as places, as well as the impact of the streetscape on each of these roles. Second, our analysis explored the heterogeneity among street segments and intersections, along with different structural categories. However, it fell short of thoroughly investigating the effects of land use differences and other more detailed contextual factors on streetscape variations and their impact on walking preferences. Future research should aim to integrate these contextual variances. In addition, this chapter unearthed complex relationships between streetscape factors and walking preferences using non-linear machine learning methods. However, the conclusions drawn from this approach predominantly reflect statistical trends and are not suited for providing specific guidance in particular design scenarios. Future research should employ other methods that can offer direct guidance for individual cases. Finally, the evidence and guidance provided by this method are suitable for the initial stage of street design, but there is a lack of discussion on how to obtain continuous evidence during the project operation by post-occupancy evaluation to allow urban designers and planners to optimize design schemes based on feedback. This limitation is also present in Chapter 6. Subsequent research could focus on the entire process of evidence-based design.

CHAPTER 6: ENHANCING PERCEIVED WALKABILITY AT STREET-LEVEL THROUGH GENERATIVE AI

This chapter is based on:

Huang, L., Oki, T. Enhancing People's Walking Preferences in Street Design Through Generative Artificial Intelligence and Crowdsourcing Surveys: The Case of Tokyo [to be submitted]

6.1. Introduction

Streets designed to prioritize pedestrians are increasingly prevalent in urban areas around the world. This trend towards "walkable streets" has not only increased the number of pedestrians but has also enhanced overall resident satisfaction, and in some cases, even boosted retail sales for businesses located along these streets. At the same time, the growing emphasis on a human-centric perspective demands that planners and designers consider people's perceived walkability as a critical factor in the design, renovation, and redevelopment of streets (Choi, et al., 2016). In Chapter 5, we obtained non-linear statistical evidence by analyzing the relationships between streetscape factors and walking preferences, identifying key factors and detailed patterns of their influence. However, we recognized the inherent limitations of this statistical method, namely that the evidence lacks intuitive representation and is not suitable for individual case design guidance. Therefore, in this chapter, we explore the application of a cutting-edge AI-generative method to address this gap.

AI generative techniques, particularly through the use of Stable Diffusion models, have emerged as an exceptionally promising avenue for automatically and efficiently translating public attitudes into spatial design representations (Mishra, et al., 2023). This provides a visible and intuitive understanding for both designers and stakeholders, facilitating design decision-making processes (Kapsalis, 2024). Nonetheless, the integration of crowdsourcing survey data with Stable Diffusion to improve environmental perceptions and preferences of urban spaces remains in its nascent stage, especially in the field of street design, with a conspicuous absence of discussion on a reliable workflow. Therefore, this chapter aims to address the aforementioned gaps by exploring a workflow to help automatically generate street

designs that improve people's perceived walkability using crowdsourcing surveys and Stable Diffusion models (3a). In this chapter, the representation of perceived walkability comprehensively considers both perceptual and behavioral preference perspectives.

6.2. Methodology

6.2.1 Analysis framework

Figure 6.1 provides an analysis framework of the technical processes of this chapter. The initial step involved quantifying the people's preferences for walking using crowdsourcing datasets. Based on the quantified results, sample images demonstrating exceptional performance in walking preferences were selected as the Stable Diffusion's training data. Following this, the Low-Rank Adaptation (LoRA) technique (Hu, et al., 2021) was employed to fine-tune the Stable Diffusion model based on the chosen data. Finally, the Stable Diffusion model and trained LoRA model, in conjunction with ControlNet and 3D model, were used to automatically generate street design scenes. The generated scenes were subsequently evaluated to confirm that they effectively enhance people's preferences for walking.

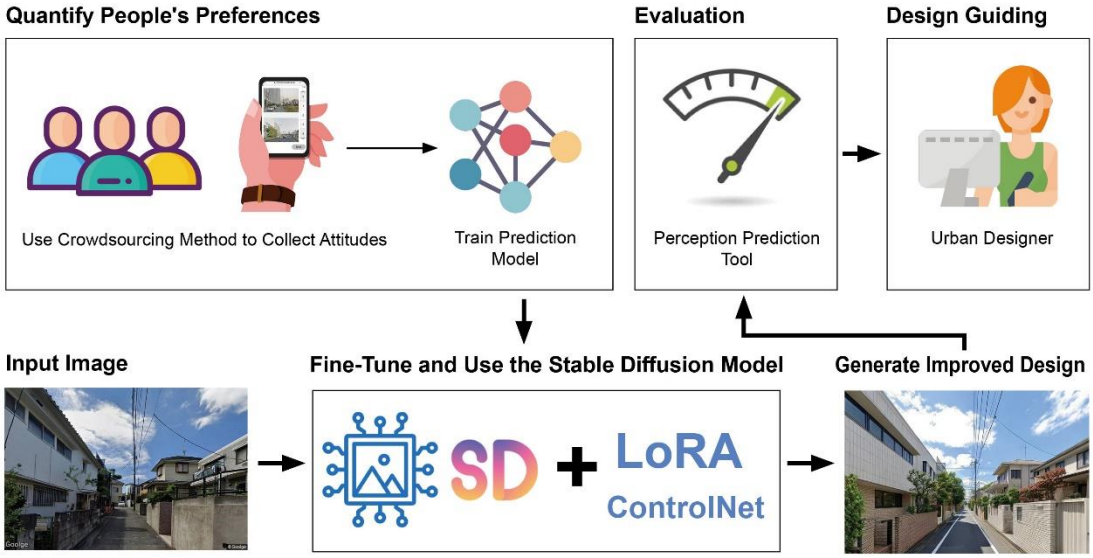


Figure 6.1 *Analysis framework.*

6.2.2 Case study area and target

Data was collected, and the feasibility of the proposed workflow was demonstrated using the same case study area as in Chapter 5, which is Setagaya Ward in Tokyo. This

chapter specifically examines street segment sections to explore methods for the automated improvement of urban scenes.

6.2.3 Street view data collection

We utilized GSV images as the primary data source for quantifying preferences for walking, as well as for generating images with Stable Diffusion. In terms of data acquisition, we continue to use the data obtained in Chapter 5.

6.2.4 Identify the key dimensions of preferences for walking

In understanding perceived walkability at street-level, this chapter focuses simultaneously on both perceptual and behavioral preference perspectives, guided by the conceptual framework established in Chapter 1. This framework requires us to quantify three specific dimensions of perceptual preferences (safety, comfort, and interestingness) and the defined dimension of behavioral preferences. However, given the limitations of solely exploring walking preferences in representing the behavioral preferences for walking (As mentioned in Section 5.5 of Chapter 5), we have expanded preference dimension to include lingering preference as an additional measurement dimension.

6.2.5 Quantify people's perceptual and behavioral preferences for walking

Obtain people's perceptual and behavioral preferences for walking

This chapter utilized two prepared crowdsourcing datasets, specifically the first crowdsourcing survey dataset (Ogawa, et al., 2022) used in Chapter 4 and the second dataset collected in Chapter 5. From the first dataset, we obtained training data concerning the three perceptual dimensions of walkability: safety, comfort, and interestingness. Subsequently, from the second dataset, we extracted training data on two dimensions related to behavioral preferences for walking: the preference to walk and the preference to linger (Huang, et al., 2024). For specific details and the structure of these two survey datasets, please refer to the respective descriptions in Chapters 4 and 5.

Train perceptual and behavioral preference prediction models

After preparing the training data, we trained 5 perception prediction models to forecast individuals' attitudes toward the aforementioned dimensions of walking perceptual preferences (safety, comfort and interesting) and behavioral preferences (preference to walk and preference to linger). For details on the structural features and

specifics of these models, you can refer to Section 4.2.2 of Chapter 4. Regarding the training accuracy of the five prediction models, as detailed in Table 6.1, the models achieved notable accuracies in the three perceptual dimensions: 90.0% for safety, 93.0% for comfort, and 81.7% for interestingness. For the two dimensions related to behavioral preferences—preference to walk and preference to linger—the models recorded training accuracies of 81.3% and 83.3%, respectively.

Table 6.1 *Evaluation accuracy of 5 preference prediction models*

	Perceptual preference			Behavioral preference	
Dimension	Safety	Comfort	Interestingness	Preference to walk	Preference to linger
Accuracy	90.0%	93.0%	81.7%	81.3%	83.3%

Prediction model application

We utilized the trained perceptual and behavioral preference prediction models to predict the GSV scores across street scenes in Setagaya Ward. These scores were subsequently normalized within their respective categories using the max-min normalization method. Figures 6.2 illustrate examples of predicted results for each perceptual preference dimension, ranked from high to low while Figure 6.3 shows examples of predicted results for each behavioral preference dimension.

Figure 6.2 *Example results of predicted perception levels.*

Figure 6.3 *Example results of predicted behavioral preference levels.*

6.2.6 Walkable street scene generation

Stable Diffusion image-to-image model

Stable Diffusion is capable of performing various image generation tasks, including Text-to-Image, Image-to-Image, image editing, and image inpainting. This chapter's study primarily focuses on the Image-to-Image task. It is similar to Text-to-Image but, in addition to a prompt, an initial image should be used as a starting point and leverages its diffusion-denoising mechanism to generate an image.

The image-to-image model consists of key components: Text-Encoder, VAE-Encoder, U-Net, and VAE-Decoder (Figure 6.4). Firstly, text information is encoded alongside the original image, generating Latent Features using the VAE-Encoder. Secondly, these features are input into the U-Net-based image optimization module within the Stable Diffusion model. Finally, the optimized Latent Features from the image optimization module undergo iterative refinement and are fed into the VAE Decoder to reconstruct them into pixel-level images (Rombach, et al., 2022).

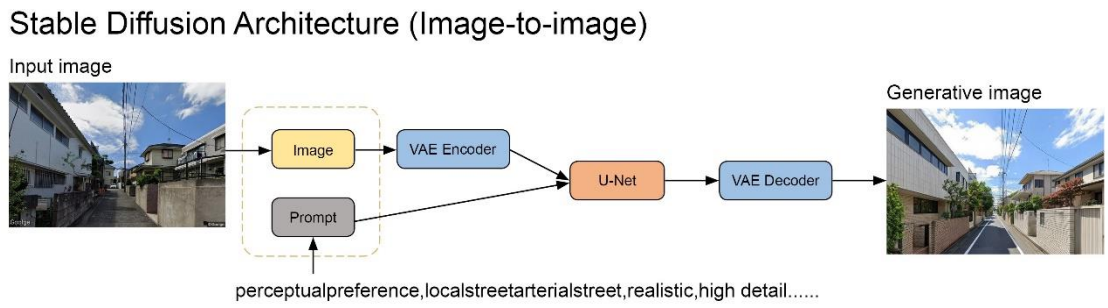


Figure 6.4 *Architecture of Stable Diffusion image-to-image model.*

LoRA

To ensure that Stable Diffusion generates content aligned with our goal of enhancing people's perceptual and behavioral preferences regarding walking, it is essential to fine-tune the model. However, the Stable Diffusion model's enormous parameter count means that direct training or fine-tuning of Stable Diffusion can consume a significant amount of time and computational power.

To address this challenge, we utilized a technique known as LoRA to fine-tune the Stable Diffusion model. LoRA is a potent and efficient pre-trained model designed for the fine-tuning of large models. It enables the training of smaller models with fine-tuning quality comparable to full models, even with a smaller dataset, thereby significantly enhancing our research efficiency. This approach allows us to align the output of Stable Diffusion with the desired perceptions and preferences while optimizing computational resources (Hu, et al., 2021).

Dataset for LoRA training

This chapter prepared two LoRA training datasets from the perspectives of perceptual and behavioral preferences, respectively. The process consisted of two steps: (1) screening and pre-processing image data, and (2) preparing corresponding text tags.

First, we used inferred perceptual and behavioral preference scores in Section 3.5.3 to filter out street view images with good walking perceptions or high behavioral preferences as training images. For the perceptual preference perspective, we identified images with high perception scores across all three selected dimensions: safety, comfort, and interest, to represent the perceptions walkable streets should have. Subsequently, we curated 200 images whose combined scores across the three dimensions exceeded the 99th percentile within the perceptual preference prediction dataset. Similarly, for the behavioral preference perspective, we selected another 200 images whose combined scores across the two dimensions exceeded the 99th percentile within the behavioral preference prediction dataset. After completing the image selection, all selected images in the two datasets were resized to a consistent resolution of 768 x 512 pixels to meet the requirements of model training (Table 6.2). During this process, the Upscaler tool from Stable Diffusion was applied to enhance the resolution of the images.

Table 6.2 *Feature and parameter of training image dataset.*

	Dataset 1 (Perceptual preference)	**Dataset 2 (Behavioral preference)**
Image Data Feature	The images received high scores across dimensions of safety, comfort, and interestingness	The images received high scores in both walking preference and lingering preference dimensions.
Number of images	200	200
Image Resolution	768*512	768*512
Trigger words	perceptualpreference, arterialstreet/collectorstreet/localstreet	behavioralpreference, arterialstreet/collectorstreet/localstreet
Other tags	outdoors, building, sidewalk, motor vehicle, road, sky, power lines, utility pole,	outdoors, building, sidewalk, motor vehicle, road, sky, power lines, utility pole,

Next, we tagged the images selected for the two datasets. We differentiated the tagging into normal tags and trigger words. For normal tags, we first used the WD14 tagger, an image tag interrogation web UI, for initial tagging. This tool can infer tags based on the features of the input images. We used ViTv2 as the backbone network for the tagger and set the weight threshold and the minimum tag fraction in batch and interrogations to 0.35 and 0.05, respectively. Following the initial generation of tagging information for the images, we opted to preserve only fundamental element tags like 'building', 'road', 'sky', and 'sidewalk'. Tags pertaining to detailed street elements such as 'bench', 'shop', 'lamppost', and 'trash can', as well as descriptors of pavement and color styles, were omitted. This decision was made to empower LoRA to discern features not explicitly tagged, leveraging differential learning processes (Table 6.2).

For trigger words, it is best if they are words that do not exist in the dictionary. For Dataset 1, the first trigger word we assigned is "perceptualpreference," while for Dataset 2, the first trigger word is "behavioralpreference. Additionally, our tagging works also account for variations in street levels. This consideration arises from the significant differences in street configurations and associated amenities, a phenomenon particularly pronounced in Japan. We adopted the standard outlined Chapter 5, utilizing the width classification parameters from the DRM to categorize streets into three levels. These levels correspond to arterial streets, collector streets, and local streets, with widths equal to or exceeding 13 m, between 5.5 and less than 13 m, and between 3 and less than 5.5 m, respectively. Subsequently, we assigned the corresponding trigger word to images belonging to each street level: "arterialstreet" for arterial streets, "collectorstreet" for collector streets, and "localstreet" for local streets.

This classification system enhances the precision of our training process, enabling LoRA to better understand the distinguishing features of streets during training.

Street scene generation with ControlNet

The Stable Diffusion model for image-to-image generation exhibits a significant degree of unpredictability when no control constraints are added. To address this, we incorporate generative conditions by introducing the ControlNet model (Zhang, et al., 2023). ControlNet offers several control models to enhance the generation process, each suited to different applicable domains. In this chapter, we tested five distinct ControlNet models—Canny, MLSD, Lineart, Depth, and Segmentation (Figure 6.5)—to control the generated images. We then conducted comparisons among these models to ascertain the most suitable one for our task.

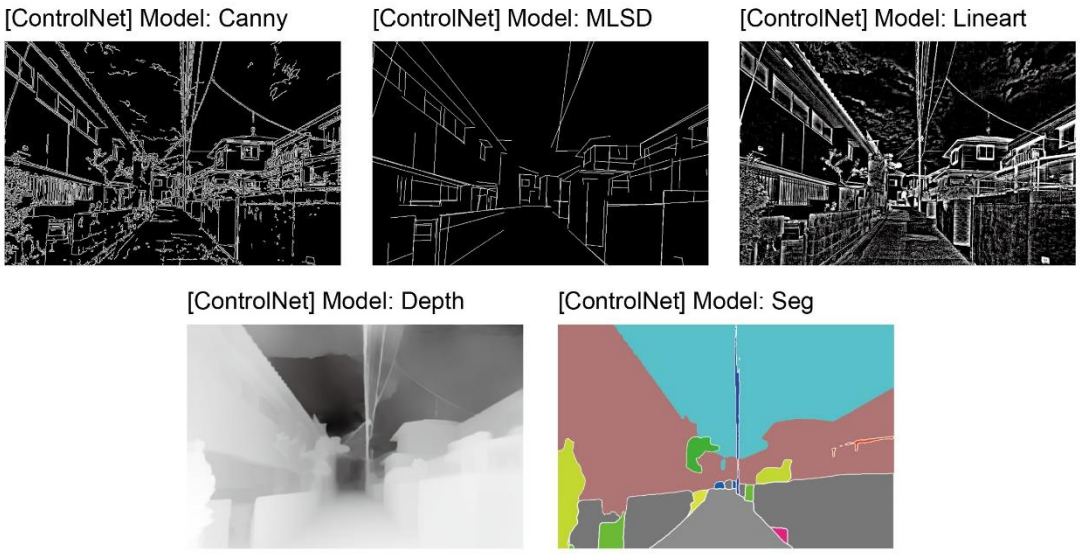


Figure 6.5 *five different ControlNet models.*

Street scene generation with ControlNet and 3D model

In some scenarios, adherence to specific planning regulations governing the design dimensions of urban elements, such as sidewalk width, building height, and drive lane number, is crucial to achieve particular development objectives. These objectives may encompass attaining an optimal balance between pedestrian and vehicular traffic, ensuring clear evacuation and rescue routes during natural disasters, and fostering equitable natural lighting on both sides of buildings. Compliance with these regulations necessitates minimizing randomness in the generation process and achieving precise control over these dimensions. However, ControlNet alone lacks the

capability to accurately control the generation output in terms of spatial dimensions. Therefore, in cases where rigorous control over the spatial dimensions of specific street elements is required, an additional 3D model is introduced in conjunction with ControlNet. This additional model effectively manages changes to the dimensions of urban elements during the generation process.

To elaborate, the process commences with the construction of a 3D model to accurately represent the before-design street scene. Following this, modifications are made to the model space based on specified dimensions, representing the after-design street scene. Next, we converted the 3D model into four variations of ControlNet masks tailored to different ControlNet models, including MLSD, Depth, and Segmentation. We then conducted comprehensive comparisons among these models and masks to ascertain the most suitable option for the task.

6.3. Results and Discussions

6.3.1 Results of LoRA model training

We utilized the two datasets prepared and the SD-Trainer LoRA script (Akegarasu, 2024) to train two LoRA models. The Model 1 aimed to automatically enhance the street walkability of street scenes from a perceptual preference perspective, while the Model 2 aimed to automate the enhancement of street scenes from a behavioral preference perspective. The base model selected was the v1-5-pruned base model. Successfully, we trained the models on an NVIDIA GeForce RTX 3090 graphics card with 32GB of memory. Regarding the key parameters during training, we set them as shown in Table 6.3. Figure 6.6 illustrates the loss values of the training processes of the two LoRA models. Overall, all two models demonstrated a consistent decreasing trend in loss values with the selected parameters.

Table 6.3 *LoRA training parameters of LoRA model 1 and 2*

Parameters	
Base model	V1-5-pruned
Batch size	2
Train_epoches	20
Repeat	20
Learning rate	1e-4
Optimizer_type	AdamW8bit

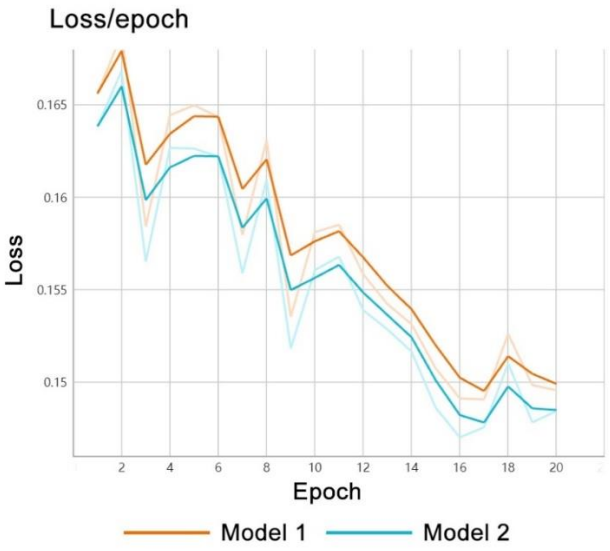


Figure 6.6 *Training process of LoRAs*

6.3.2 Generation results of street scenes

In the following sections, we conducted experiments to generate street scene images using only the LoRA model, different ControlNet models, and a combination of ControlNet and 3D reference. Additionally, we evaluated the generation results to ascertain whether the proposed workflow could help enhance people's perceptual and behavioral preferences regarding street walkability.

For model loading and image-to-image generation tasks, this paper uses Stable Diffusion Web-UI (Figure 6.7) (AUTOMATIC1111, 2024). Stable Diffusion Web-UI is a browser-based user interface designed for generating and editing images using the Stable Diffusion model. It simplifies the process of interacting with this powerful AI tool, making it accessible and user-friendly for a wide range of users.

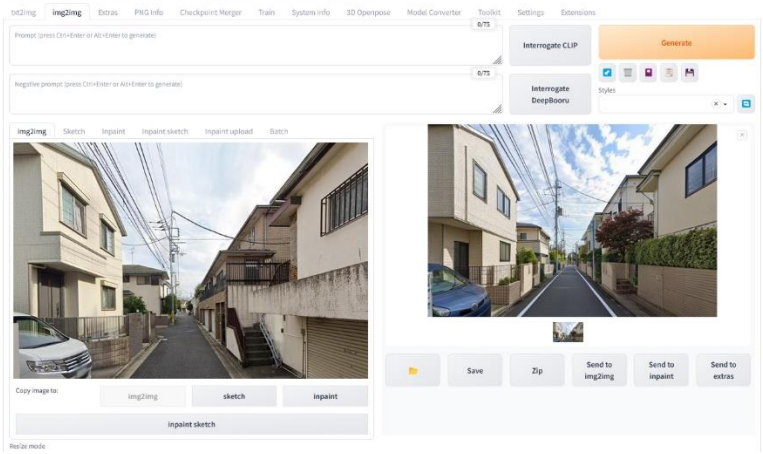


Figure 6.7 *Stable Diffusion Web-UI interface.*

Generation results via different LoRAs

In this section, we conducted an experiment to evaluate the image generation reliability of two trained LoRA models across different types of street scenes while also assessing the optimal choice of LoRA weights. We adjusted the LoRA weights to 0, 0.6, 0.8, and 1 to observe their impact on the image generation results. Throughout the generation process, we consistently employed DPM++ SDE Karras as the sampler, with sampling steps set to 30, the CFG scale set to 7, and the Denoising value set to 0.6. Additionally, we ensured that each model utilized the same random seed to maintain consistency during the generation process. Figure 6.8 illustrates the comparative results.

By observing the generation results, we found that compared to the case without LoRA influence (weight = 0), the introduction of LoRA indeed transformed the generated images. When the LoRA weight is at 0.8, the generation effects are relatively stable while also providing detailed outputs. However, excessively high LoRA weights may introduce instability in the generated outputs. In this section's experiment, at a value of 1, some generated images start to display distorted lines and shapes, accompanied by issues related to color and texture instability.

Figure 6.8 *Generation Results via different LoRAs and weights.*

Generation results via different ControlNets

Five commonly applied ControlNet models—Canny, MLSD, Lineart, Depth, and Segmentation—were utilized to test their control effects on image generation, using model 1 as an example. We set the control weights of different ControlNet models to 0.4, 0.6, 0.8, and 1 to examine various weight values' effects on output and determine the optimal weight range for our task.

Figure 6.9 depicts the generation effects of a selected image under different ControlNet models and weights while keeping other parameters and random seed constant. Upon comparing the generated images, we observe that images generated

based on the Depth Map model exhibit relatively accurate spatial relationships, particularly when the weight is set within the range of 0.6 to 0.8.

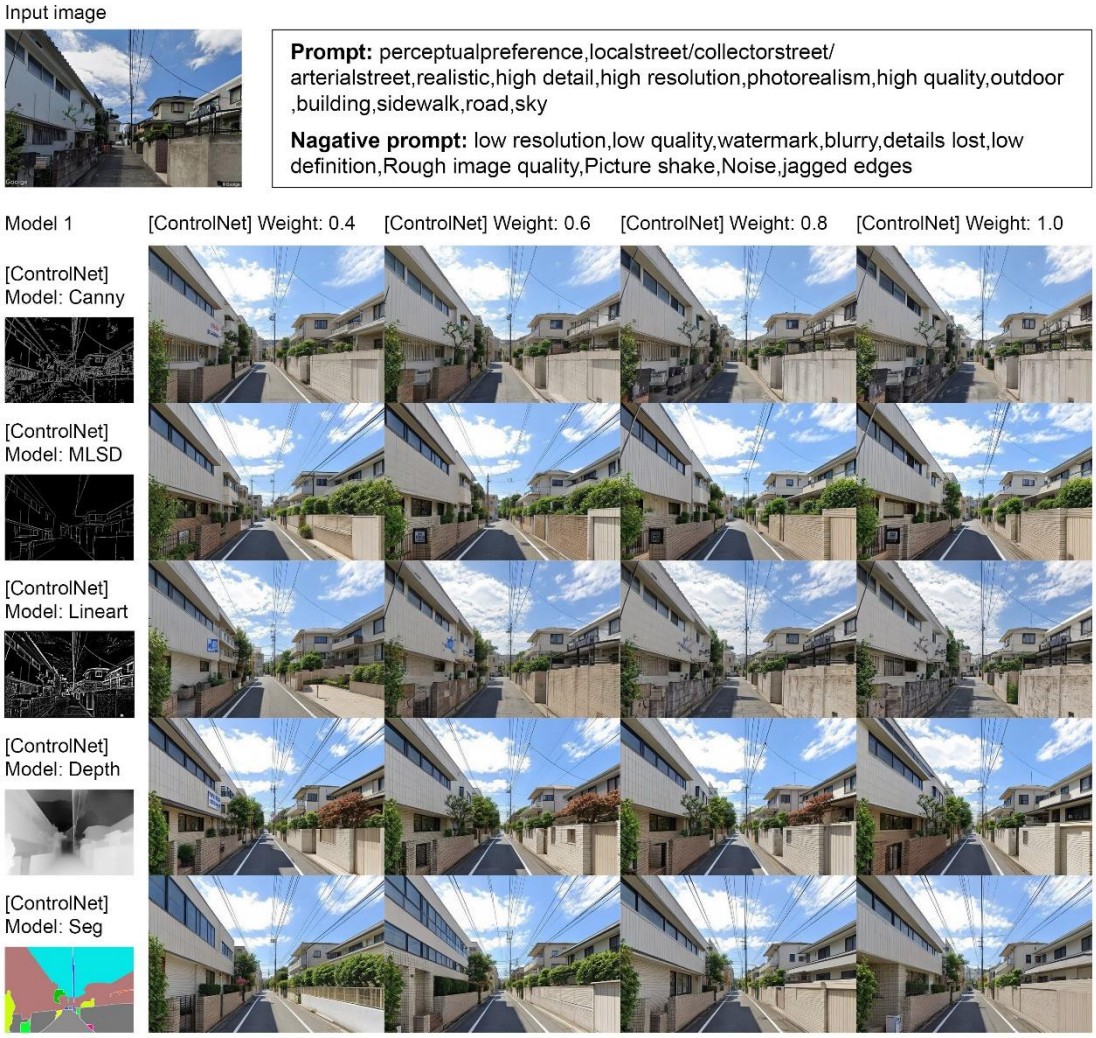


Figure 6.9 *Generation Results via different ControlNet models and weights.*

Generation results via ControlNets and 3D models

This section experimented with the output effects of precisely controlling specific dimensions using different ControlNet models paired with mask files created from 3D models. The experiments were conducted on the planned road redevelopment project in the Setagaya ward of Tokyo, identified as Auxiliary Route No. 54. This road stretches from Kamisoshigaya 4-chome to Kamisoshigaya 5-chome in Setagaya. The project, scheduled to span from 2023 to 2032, aims to expand the street from its current width of approximately 8 meters, which includes a single sidewalk, to 15 meters with two traffic lanes and 3-meter-wide sidewalks on both sides. This redevelopment project

seeks to reduce local traffic congestion, enhance the disaster resilience of the community, and create safe, comfortable pedestrian pathways (TMG, 2024).

We utilized SketchUp 2022 as the modeling platform, with the contour data and building heigh information sourced respectively from Zenrin twon II (ZENRIN, 2024) and PLATEAU 3D building dataset (MLIT, 2022).

Figure 6.10 showcases the designed street scenes generated using LoRA Model 1, as an example, in combination with three types of ControlNet masks derived from 3D models. It is observable that, the use of the segmentation model for control results in generated street scenes with higher accuracy in dimension control. On the other hand, the use of MLSD and Depth for control exhibits greater randomness, which is evident in the inability to accurately represent the dimensional proportion between sidewalks and roadways in most cases. This suggests that using segmentation control mask output from 3D models to control the generation output is more suitable for generation tasks that require high precision in the spatial dimensional relationships.

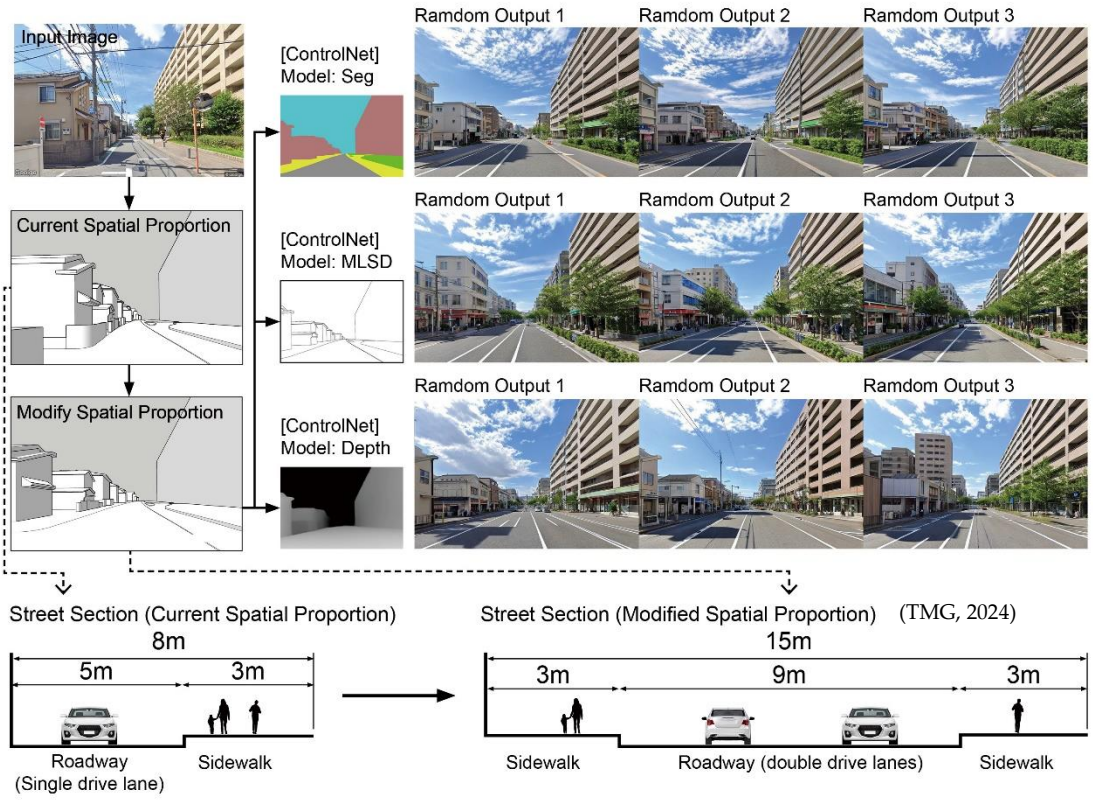


Figure 6.10 *Generation Results via ControlNets and 3D Models.*

6.3.3 Evaluation on the generative street scenes

To determine whether the generated images can enhance public perceptions and preferences and by how much, we evaluated them by comparing the scores of the images before and after generation using our workflow. The evaluation tool used was our trained perceptual and behavioral preference prediction models.

Figure 6.11 considers 3 different types of street scenes (Arterial, Collator and Local Streets) using LoRA Model 1 and the Depth ControlNet model. We observed that all generated street scenes exhibited improvements in all perception dimensions. Among them, Figure 6.11a showed respective improvements in safety, comfort, and interest by 0.213, 0.138, and 0.187. Meanwhile, Figures 6.11b and 6.11c showed improvements in the dimensions of safety, comfort, and interest by 0.026 and 0.100, 0.184 and 0.205, 0.211 and 0.126 respectively. This indicates that LoRA Model 1 can effectively learn from the perceptual dimension evidence and automate the improvement design of street scenes.

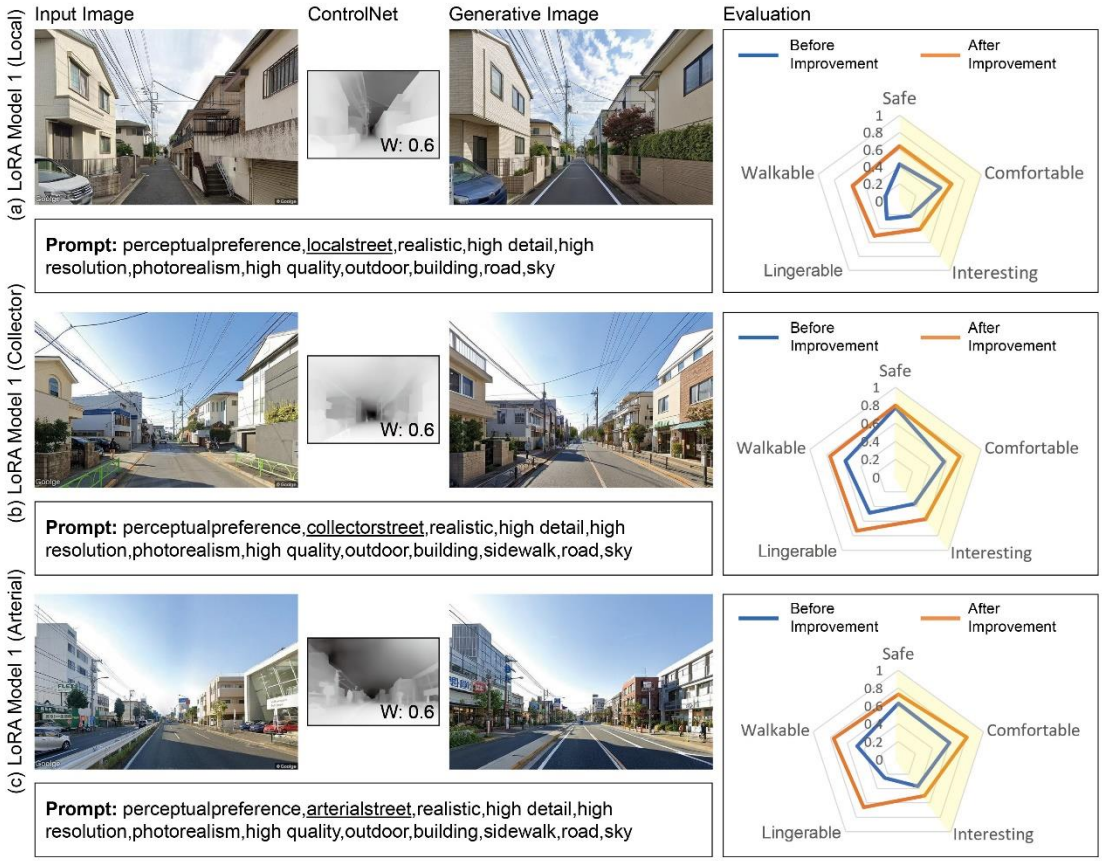


Figure 6.11 *Evaluation of generative images using LoRA model 1.*

The improvement in scores across multiple dimensions in the generated scenes can be explained by comparing the original and the automatically generated street scenes.

Several key improvements can be observed: compared to the original images, the generated images enhance the street interface, particularly the first-floor interface's permeability, by increasing commercial storefronts (Figure 6.11b and 6.11c). This can be interpreted as more permeable building interfaces bringing about safer street environments while simultaneously creating vibrant zones (Carmelino & Hanazato, 2019). Simultaneously, this aligns with the findings of Huang et al. (2024), who highlight the critical role of ground-level storefronts in shaping street perception and behavioral preferences. Additionally, more green elements appear in the generated street scenes, used to soften the edge between buildings and roads, creating a more comfortable transitional space (Figure 6.11a, 6. 11b, and 6. 11c). This is considered a key space for stimulating social potential and vitality in previous research (Thwaites, et al., 2020). Furthermore, the materials of some street-facing building facades have also been updated to warmer tones (Figure 6.11a and 6.11b), which previous studies have shown that diverse architectural colors can impact people's environmental perception and behavioral preferences (Fan, et al., 2023). Furthermore, wider sidewalks have been extended (Figure 6.10c), a strategy frequently adopted in Tokyo's street redevelopment and renewal design, aimed at balancing the interplay between vehicular and pedestrian traffic. Overall, the improvement shown in the improved street senses generated by using the trained LoRA combined with Stable Diffusion and ControlNet is generally interpretable and aligned with practical design strategies.

Figure 6.12 evaluates three distinct levels of enhanced street scenes generated by LoRA Model 2 and the Depth ControlNet model. All of them show significant improvements across walking and lingering preference dimensions. Specifically, Figure 6.12a displayed improvements in the preference to walk and linger by 0.175 and 0.067, respectively. Meanwhile, Figures 6.12b and 6.12c exhibited enhancements in these two dimensions, with gains of 0.201 and 0.388, 0.341 and 0.315 respectively. These results affirm that LoRA Model 2 is also effective in enhancing the behavioral preferences for walking.

Figure 6.12 *Evaluation of generative images using LoRA model 2.*

Several significant transformations are evident in Figure 6.12. A transitional edge dominated by greenery has been established, serving to bridge the gap between buildings, which function as private spaces, and the roads, which constitute public space. Additionally, a setback strategy for buildings has been applied by LoRA, and the setbacks have been transformed into sidewalk pavement, providing more walking space (Figure 6.12b). Moreover, the LoRA Model 2 has notably enhanced the permeability of first-floor building interfaces.

From a comprehensive observation of the evaluation results, we find that utilizing images trained on datasets featuring high levels across all three perceptual preference dimensions enhances not only the perception levels themselves but also boosts behavioral preference scores. Conversely, training LoRA on datasets featuring high behavioral preference levels across two dimensions also results in improvements across all three perceptual preference dimensions. This suggests a correlation between visual walking perceptual and behavioral preferences, indicating that heightened perceptions of safety, comfort, and interest enhance people's willingness to walk and linger.

6.4. Conclusion

This chapter explored a workflow to help generate high-fidelity and renewal street scenes that enhance people's walking perceptions and preferences using crowdsourcing surveys and Stable Diffusion models. It provides valuable inspiration and options for urban planners and designers in the early stages of street design and redevelopment. Our study includes the following contributions:

Firstly, our proposed workflow aims to enhance perceptual and behavioral preference for walking specifically tailored to certain street design scenarios, offering a more intuitive and visually-oriented approach. Unlike traditional statistical methods, including non-linear ones commonly used in many evidence-based design practices, our method provides targeted insights rather than general trends. Moreover, traditional statistical outputs, typically numerical or tabular, may lack the immediate clarity and intuitiveness provided by our visual approach.

Secondly, the workflow we proposed advances street design and redevelopment towards a human-centric and evidence-based approach. On one hand, unlike traditional street design that mainly focuses on the physical built environment or actual human behaivors, our method emphasizes a subjective perspective and focuses on people's visual walking preferences. This approach makes the human-centric effort more intuitively reflected in the street design and redevelopment. On the other hand, the method promotes an evidence-based design approach. By utilizing neural networks, we associate previously hard-to-quantify subjective attitudes with corresponding street scenes. We use these quantified subjective attitudes as evidence to train AI generation networks, enabling them to automatically improve streetscapes based on this evidence.

Thirdly, the workflow we proposed facilitated the formation of a new model of public participation and co-design, allowing the public to become decision-makers in the street design and redevelopment process. By collecting crowdsourcing attitudes on a large scale and combining them with DCNN models, we were able to predict people's perceptual and behavioral preferences for walking. With this understanding of large-scale public perceptions and preferences, we could further filter images that exhibit favorable perception and preference characteristics. Subsequently, leveraging trained LoRA models, we could furnish early-stage design proposals aimed at enhancing safety, comfort, interest, walking preference, and lingering preference. These generated street scenes, reflecting the collective perceptions and preferences of the public, can, in turn, further facilitate the co-design process by lowering the barriers

to the public during the visualization of spatial design proposals, making it possible to integrate public opinions seamlessly into the design process.

Fourthly, the workflow we proposed enhances the efficiency of street design and drives the role transition of urban planners from pure designers to urban coordinators. By utilizing trained Stable Diffusion models, designers can quickly generate design schemes in the early stages without spending too much time on modeling and image rendering. When urban designers are freed from repetitive design workflows, they can devote more time to analysing and coordinating urban policies, local history, culture, and social characteristics, allowing these deeper aspects to be reflected in the design.

Overall, this workflow has shown enormous potential when applied to the renovation of street design and redevelopment practices in Japan and globally.

However, this study also has a limitation. Since our method relies on public perceptions and preferences as training data for the generative model, the variations in the generated images do not extend beyond the collective preferences of the public, limiting the potential for more diverse and impressive possibilities. Future research should optimize the methodology to ensure that the generated results reflect public preferences while also allowing for greater creative freedom.

CHAPTER 7: DISCUSSION AND CONCLUSION

7.1 Summary and Findings

In Chapter 4, we aim to address the question: How can urban big data and AI algorithms be used to systematically and automatically capture both physical and perceived walkability features at street-level? Firstly, we use Hong Kong as a case study to select suitable variables for automated measurement based on existing street-level factors, achieving our proposed research objective 1a. For quantifying physical aspect of walkability, we used deep CV techniques and Street View imagery, complemented by GIS data and analysis methods, to propose quantification methods for each type of variable. This allows us to automatically analyze and quantify these features. Specifically, we distinguished between quantifiable and non-quantifiable streetscape variables and utilized semantic segmentation and YOLOv5 for their measurement. We also explored the use of CNN regression models to measure maintenance quality and leveraged YOLO and monocular measurement methods to estimate sidewalk width and other geometric variables. To quantify the perceived aspect of walkability, we investigated the feasibility of applying deep DCNN alongside image crowdsourcing surveys. This innovative approach enabled us to accurately capture and measure perceived walkability, demonstrating the capability of advanced AI techniques to understand and evaluate these subjective experiences effectively. We applied all proposed methods in a case study focusing on Kowloon West, Hong Kong, testing their practical applicability and robustness, achieving our proposed research objectives 1b and 1c.

Through this research, we confirmed that the integration of Street View big data and advanced CV algorithms has significantly enhanced our ability and efficiency to capture both the physical and perceived characteristics of walkability, thereby providing valuable evidence. Traditional methods of urban data collection, such as on-site surveys and questionnaires, often suffer from limitations in both detail and coverage. In contrast, emerging data sources like street view imagery, combined with CV methods, offer new opportunities to efficiently assess fine-grained street environments. However, we also highlight our concerns that despite the opportunities brought by street view big data combined with CV algorithms, they may also introduce redundant information and unverified second-hand experiences. Effectively and accurately filtering and parsing this evidence becomes crucial.

Chapter 5 builds upon Chapter 4, addressing its concerns by delving into the use of a non-linear regression method to explore key influencing factors on visual perceived

walkability from the perspective of walking preferences and their impact trends, and answering the question: How can nonlinear statistical methods enhance our understanding of the impact of physical walkability on perceived walkability at street-level, and how can this knowledge inform perceived walkability in street design? This chapter's research considers street-level factors as dependent variables and walking preferences as independent variables, using Setagaya Ward in Tokyo as a case study to implement the measurements. We then employed XGBoost as a regression algorithm to explore the non-linear relationships between streetscapes and walking preferences.

Our findings reveal that the significance of specific streetscape factors varies across different grades of street segments and intersections. The XGBoost regression analysis uncovers diverse patterns in these relationships, including upward trends, downward trends, and threshold effects, demonstrating the complex interplay between urban design elements and walking preferences. This accomplishment fulfills our research objective 2a. Moreover, our study highlights the heterogeneous nature of structural characteristics across different street segments and intersections. These results enable us to more precisely identify critical influencing factors and pinpoint the optimal range for design interventions for each variable—capabilities that traditional linear methods cannot match. Consequently, we achieve research objectives 2b and 2c. However, our research also revealed a notable limitation of both linear and non-linear regression statistics: they can only provide general trends or impact thresholds based on mathematical and numerical analyses, which are not suitable for offering specific and visually intuitive guidance for enhancing perceived walkability in particular designs or individual cases.

In Chapter 6, we developed a street scene generation workflow and trained two LoRA models designed to automatically create immersive and revitalized street environments that enhance people's perceived walkability. The Stable Diffusion-based workflow addresses the limitations of the previously mentioned statistical methods by enabling the creation of tailored, optimized, and visually intuitive design schemes for specific street scenes, thereby providing valuable, actionable insights. Moreover, unlike traditional methods that present statistical trends through charts and data, Stable Diffusion can provide photorealistic previews of expected outcomes and generate a wide range of alternative design options. This capability is especially beneficial in the early design stages. Our approach also considers the need to control specific spatial proportions in design and introduces three-dimensional models to achieve effective control. To validate the optimal output parameters of our method,

we conducted experiments on LoRA outputs, ControlNet outputs, and 3D model control outputs. To demonstrate the effectiveness of our proposed workflow, we utilized a trained perception and preference prediction model for evaluation, confirming that our proposed workflow and models can optimize street scenes while improving perceptual and behavioral preferences. The above work achieves our stated research objective 3a.

7.2 Research Contributions

This dissertation makes contributions in several areas. From a methodological perspective, our research introduces innovative techniques for measuring and evaluating urban walkability at a street-level, allowing micro-scale walkability features to be presented to people in a systematic, automated, and large-scale manner. Additionally, by employing non-linear regression approaches, we offer a more accurate and realistic framework for informing perceived walkability design. Furthermore, we explore the integration of emerging AI generative technologies, specifically stable diffusion, with crowdsourced data to develop automated workflows for perceived walkability design. These efforts advance the integration of evidence-based design principles into walkable street design at the methodological level.

Practically, our research offers actionable insights and tools for various stakeholders involved in street-level walkability design, including policymakers and city managers, urban planners and designers, and concerned citizens. For policymakers and city managers, they can leverage the proposed data-driven methods to quickly and thoroughly understand the current state of urban spaces at a fine-grained scale. This understanding enables them to formulate policies or design guidance that promote sustainable and healthy urban environments. Moreover, an accurate comprehension of the overall characteristics of urban spaces helps them better coordinate different departments and stakeholders for effective collaboration. Urban planners and designers, as direct beneficiaries of these methods, can utilize our findings and tools to promote evidence-based design and more effectively create walkable and livable streets, enhancing the overall quality of urban life. The proposed methods are particularly useful in the early stages of design, providing valuable insights for planners and designers. For concerned citizens, their perceptual needs and preferences can be considered in the design process using the methods proposed in this dissertation, without the complexity of traditional public participation processes. This human-centric approach ensures that the public's attitudes, collected through

crowdsourcing methods, are integrated into urban design, making it more attuned to their needs and expectations.

ACKNOWLEDGEMENT

I am profoundly grateful to my supervisor, Dr. OKI Takuya, for the opportunity to undertake this study and for their steadfast support throughout its duration. Dr. Oki's dedication, passion, and visionary insight have been incredibly inspiring. Under their expert guidance, I have sharpened my abilities to conduct comprehensive research and articulate my findings with precision and significance. It has been a true honor to learn from and work under Dr. Oki's tutelage.

I also wish to express my sincere appreciation to the dissertation reviewers—Prof. FUJII Haruyuki, Prof. OSARAGI Toshihiro, Prof. SAIO Naoko, and Prof. MATSUOKA Masashi. Their meticulous critiques and valuable insights have greatly enriched the quality of my work. Your detailed evaluations and constructive feedback were crucial.

My deep gratitude extends to the Tokyo Tech SPRING (JPMJSP2106) and KAKENHI (JP22K04490) funds for their financial support, which was instrumental in allowing me to conduct my research with the necessary scope and depth.

Furthermore, I must acknowledge the unwavering personal support from my family, whose love and encouragement have been my bedrock throughout this doctoral study journey. I owe you an immense debt of gratitude.

In addition, special thanks are due to Zenrin and JoRAS for their provision of essential street view and map data, along with other critical resources facilitated by CSIS. These contributions were fundamental to the execution and success of my research. Furthermore, I am deeply grateful to all the respondents of the questionnaire. Your participation and the insights you provided were vital for ensuring the validity and richness of my findings.

Each of you has uniquely contributed to my academic journey, and for this, I am eternally thankful.

REFERENCE

Abrishami, S., Goulding, J. and Rahimian, F., 2021. Generative BIM workspace for AEC conceptual design automation: Prototype development. *Engineering, Construction and Architectural Management*, *28*(2), pp.482-509.

Adkins, A., Dill, J., Luhr, G. and Neal, M., 2012. Unpacking walkability: Testing the influence of urban design features on perceptions of walking environment attractiveness. *Journal of urban design*, *17*(4), pp.499-510.

Aghaabbasi, M., Moeinaddini, M., Shah, M.Z. and Asadi-Shekari, Z., 2017. A new assessment model to evaluate the microscale sidewalk design factors at the neighbourhood level. *Journal of Transport & Health*, *5*, pp.97-112.

Akasaki, K., 1989. *The Concept of 'Mid-Streets' in Sprawl Urban Areas.* s.l.:Kashima Publishing Co..

Akegarasu 2024. *Akegarasu/lora-scripts*. [online] GitHub. Available at: https://github.com/Akegarasu/lora-scripts.

Al Shammas, T. and Escobar, F., 2019. Comfort and time-based walkability index design: a GIS-based proposal. *International journal of environmental research and public health*, *16*(16), p.2850.

Alfonzo, M.A., 2005. To walk or not to walk? The hierarchy of walking needs. *Environment and behavior*, *37*(6), pp.808-836.

Asgarzadeh, M., Koga, T., Hirate, K., Farvid, M. and Lusk, A., 2014. Investigating oppressiveness and spaciousness in relation to building, trees, sky and ground surface: A study in Tokyo. *Landscape and Urban planning*, *131*, pp.36-41.

Asgarzadeh, M., Lusk, A., Koga, T. and Hirate, K., 2012. Measuring oppressiveness of streetscapes. *Landscape and Urban Planning*, *107*(1), pp.1-11.

Yoshinobu Ashihara (1983). *The aesthetic townscape*. Cambridge, Mass.: Mit Press.

Atkinson, A.B., 1970. On the measurement of inequality. *Journal of economic theory*, *2*(3), pp.244-263.

AUTOMATIC1111 2022. *Stable Diffusion web UI*. [online] GitHub. Available at: https://github.com/AUTOMATIC1111/stable-diffusion-webui.

Baidu Maps. 2022. *Baidu Maps*. [online] Available at: https://map.baidu.com.

Basu, R. and Sevtsuk, A., 2022. How do street attributes affect willingness-to-walk? City-wide pedestrian route choice analysis using big data from Boston and San Francisco. *Transportation research part A: policy and practice*, *163*, pp.1-19.

Belza, B., Altpeter, M., Smith, M.L. and Ory, M.G., 2017. The Healthy Aging Research Network: modeling collaboration for community impact. *American journal of preventive medicine*, *52*(3), pp.S228-S232.

Blečić, I., Cecchini, A., Canu, D., Cappai, A., Congiu, T. and Fancello, G., 2016. Evaluating the effect of urban intersections on walkability. In *Computational Science and Its Applications--ICCSA 2016: 16th International Conference, Beijing, China, July 4-7, 2016, Proceedings, Part IV 16* (pp. 138-149). Springer International Publishing.

Boarnet, M.G., Day, K., Alfonzo, M., Forsyth, A. and Oakes, M., 2006. The Irvine–Minnesota inventory to measure built environments: reliability tests. *American journal of preventive medicine*, *30*(2), pp.153-159.

Borst, H.C., de Vries, S.I., Graham, J.M., van Dongen, J.E., Bakker, I. and Miedema, H.M., 2009. Influence of environmental street characteristics on walking route choice of elderly people. *Journal of Environmental Psychology*, *29*(4), pp.477-484.

Brownson, R.C., Hoehner, C.M., Brennan, L.K., Cook, R.A., Elliott, M.B. and McMullen, K.M., 2004. Reliability of 2 instruments for auditing the environment for physical activity. *Journal of Physical Activity and Health*, *1*(3), pp.191-208.

Cain, K., Rachel, M., Millstein, A., Carrie, M. and Geremia, M. 2012. *Microscale Audit of Pedestrian Streetscapes (MAPS): Data Collection & Scoring Manual*. [online] Available at:

https://www.drjimsallis.com/_files/ugd/a56315_852577b5ff554ec891cb8586fb75022c.pdf [Accessed 11 Apr. 2024].

Carmelino, G. and Hanazato, T., 2019. The built environment of Japanese shopping streets as visual information on pedestrian vibrancy. *Frontiers of Architectural Research*, *8*(2), pp.261-273.

Cavill, N., Kahlmeier, S., Rutter, H., Racioppi, F. and Oja, P., 2007. Economic assessment of transport infrastructure and policies. Methodological guidance on the economic appraisal of health effects related to walking and cycling.

Cervero, R. and Kockelman, K., 1997. Travel demand and the 3Ds: Density, diversity, and design. *Transportation research part D: Transport and environment*, *2*(3), pp.199-219.

Cervero, R., Sarmiento, O.L., Jacoby, E., Gomez, L.F. and Neiman, A., 2009. Influences of built environments on walking and cycling: lessons from Bogotá. *International journal of sustainable transportation*, *3*(4), pp.203-226.

Chan, J.A., Bosma, H., Drosinou, C., Timmermans, E.J., Savelberg, H., Schaper, N., Schram, M.T., Stehouwer, C.D., Lakerveld, J. and Koster, A., 2023. Association of perceived and objective neighborhood walkability with accelerometer-measured physical activity and sedentary time in the Maastricht Study. *Scandinavian Journal of Medicine & Science in Sports*, *33*(11), pp.2313-2322.

Cheng, B., Misra, I., Schwing, A.G., Kirillov, A. and Girdhar, R., 2022. Masked-attention mask transformer for universal image segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 1290-1299).

Cheng, B., Schwing, A. and Kirillov, A., 2021. Per-pixel classification is not all you need for semantic segmentation. *Advances in neural information processing systems*, *34*, pp.17864-17875.

Cheng, L., De Vos, J., Zhao, P., Yang, M. and Witlox, F., 2020. Examining non-linear built environment effects on elderly's walking: A random forest approach. *Transportation research part D: transport and environment*, *88*, p.102552.

Chen, T. and Guestrin, C., 2016, August. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).

Chiang, Y.C., Sullivan, W. and Larsen, L., 2017. Measuring neighborhood walkable environments: A comparison of three approaches. *International journal of environmental research and public health*, *14*(6), p.593.

Choi, J., Kim, S., Min, D., Lee, D. and Kim, S., 2016. Human-centered designs, characteristics of urban streets, and pedestrian perceptions. *Journal of Advanced Transportation*, *50*(1), pp.120-137.

Cooper, C.H. and Chiaradia, A.J., 2020. sDNA: 3-d spatial network analysis for GIS, CAD, Command Line & Python. *SoftwareX*, *12*, p.100525.

Data.gov.hk. 2022. *DATA.GOV.HK*. [online] Available at: https://data.gov.hk/en/.

De Nadai, M., Vieriu, R.L., Zen, G., Dragicevic, S., Naik, N., Caraviello, M., Hidalgo, C.A., Sebe, N. and Lepri, B., 2016, October. Are safer looking neighborhoods more lively? A multimodal investigation into urban life. In *Proceedings of the 24th ACM international conference on Multimedia* (pp. 1127-1135).

Deza, A. and Parikh, D., 2015. Understanding image virality. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1818-1826).

Dhar, S., Ordonez, V. and Berg, T.L., 2011, June. High level describable attributes for predicting aesthetics and interestingness. In *CVPR 2011* (pp. 1657-1664). IEEE.

Diakoulaki, D., Mavrotas, G. and Papayannakis, L., 1995. Determining objective weights in multiple criteria problems: The critic method. *Computers & Operations Research*, *22*(7), pp.763-770.

DRM, 2023. *Digital Road Map*. [online] Available at: https://www.drm.jp/english/ [Accessed 23 Jul. 2024].

Dubey, A., Naik, N., Parikh, D., Raskar, R. and Hidalgo, C.A., 2016. Deep learning the city: Quantifying urban perception at a global scale. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14* (pp. 196-212). Springer International Publishing.

Emery, J., Crump, C. and Bors, P., 2003. Reliability and validity of two instruments designed to assess the walking and bicycling suitability of sidewalks and roads. *American Journal of Health Promotion*, *18*(1), pp.38-46.

Evenson, K.R., Sotres-Alvarez, D., Herring, A.H., Messer, L., Laraia, B.A. and Rodríguez, D.A., 2009. Assessing urban and rural neighborhood characteristics using audit and GIS data: derivation and reliability of constructs. *International Journal of Behavioral Nutrition and Physical Activity*, *6*, pp.1-16.

Ewing, R. and Handy, S., 2009. Measuring the unmeasurable: Urban design qualities related to walkability. *Journal of Urban design*, *14*(1), pp.65-84.

Fan, Z.M., Zhu, B.W., Xiong, L., Huang, S.W. and Tzeng, G.H., 2023. Urban design strategies fostering creative workers' sense of identity in creative and cultural districts in East Asia: An integrated knowledge-driven approach. *Cities*, *137*, p.104269.

Fedorova, S., 2021. GANs for urban design. *arXiv preprint arXiv:2105.01727*.

Frackelton, A., Grossman, A., Palinginis, E., Castrillon, F., Elango, V. and Guensler, R., 2013. Measuring walkability: Development of an automated sidewalk quality assessment tool. *Suburban Sustainability*, *1*(1), p.4.

Gallimore, J.M., Brown, B.B. and Werner, C.M., 2011. Walking routes to school in new urban and suburban neighborhoods: An environmental walkability analysis of blocks and routes. *Journal of environmental psychology*, *31*(2), pp.184-191.

Gehl, J. (1987). *Life between Buildings: Using Public Space*. Washington, DC: Island Press.

Gehl, J., Kaefer, L.J. and Reigstad, S., 2006. Close encounters with buildings. *Urban design international*, *11*, pp.29-47.

Google Maps Platform (n.d.). *Street View Static API*. [online] Google for Developers. Available at: https://developers.google.com/maps/documentation/streetview/overview?hl=zh-cn [Accessed 23 Jul. 2024].

Guzman, L.A., Arellana, J. and Castro, W.F., 2022. Desirable streets for pedestrians: Using a street-level index to assess walkability. *Transportation research part D: transport and environment*, *111*, p.103462.

Handy, S.L., Boarnet, M.G., Ewing, R. and Killingsworth, R.E., 2002. How the built environment affects physical activity: views from urban planning. *American journal of preventive medicine*, *23*(2), pp.64-73.

Harvey, C., Aultman-Hall, L., Hurley, S.E. and Troy, A., 2015. Effects of skeletal streetscape design on perceived safety. *Landscape and Urban Planning*, *142*, pp.18-28.

He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

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

Herrmann-Lunecke, M.G., Mora, R. and Vejares, P., 2021. Perception of the built environment and walking in pericentral neighbourhoods in Santiago, Chile. *Travel behaviour and society*, *23*, pp.192-206.

Huang, C., Zhang, G., Yao, J., Wang, X., Calautit, J.K., Zhao, C., An, N. and Peng, X., 2022. Accelerated environmental performance-driven urban design with generative adversarial network. *Building and Environment*, *224*, p.109575.

Huang, L., Oki, T., Muto, S., Kim, H., Ogawa, Y. and Sekimoto, Y., 2023, June. Automatic Evaluation of Street-Level Walkability Based on Computer Vision Techniques and Urban Big Data: A Case Study of Kowloon West, Hong Kong. In *International*

Conference on Computers in Urban Planning and Urban Management (pp. 231-259). Cham: Springer Nature Switzerland.

Huang, L., Oki, T., Muto, S. & Ogawa, Y., (2023). Capturing Walking-Related perceptions and willingness within Tokyo's station areas: leveraging Crowd-Sourced Methods and AI approach. *GISA & IAG'i 2023,* Volume https://confit.atlas.jp/guide/event-img/gisa2023/E1-01/public/pdf?type=in.

Huang, L., Oki, T., Muto, S. and Ogawa, Y., 2024. Unveiling the Non-Linear Influence of Eye-Level Streetscape Factors on Walking Preference: Evidence from Tokyo. *ISPRS International Journal of Geo-Information*, *13*(4), p.131.

Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L. and Chen, W., 2021. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*.

Hu, F.B., Sigal, R.J., Rich-Edwards, J.W., Colditz, G.A., Solomon, C.G., Willett, W.C., Speizer, F.E. and Manson, J.E., 1999. Walking compared with vigorous physical activity and risk of type 2 diabetes in women: a prospective study. *Jama*, *282*(15), pp.1433-1439.

Hu, F.B., Stampfer, M.J., Solomon, C., Liu, S., Colditz, G.A., Speizer, F.E., Willett, W.C. and Manson, J.E. (2001). Physical Activity and Risk for Cardiovascular Events in Diabetic Women. *Annals of Internal Medicine*, 134(2), p.96. doi:https://doi.org/10.7326/0003-4819-134-2-200101160-00009.

Hu, G., Qiao, Q., Silventoinen, K., Eriksson, J.G., Jousilahti, P., Lindström, J., Valle, T.T., Nissinen, A. and Tuomilehto, J., 2003. Occupational, commuting, and leisure-time physical activity in relation to risk for Type 2 diabetes in middle-aged Finnish men and women. *Diabetologia*, *46*, pp.322-329.

Humpel, N., Owen, N., Leslie, E., Marshall, A.L., Bauman, A.E. and Sallis, J.F., 2004. Associations of location and perceived environmental attributes with walking in neighborhoods. *American Journal of Health Promotion*, *18*(3), pp.239-242.

Isola, P., Xiao, J., Torralba, A. and Oliva, A., 2011, June. What makes an image memorable?. In *CVPR 2011* (pp. 145-152). IEEE.

Jacobs, J. (1961). *The Death and Life of Great American Cities*. New York: Random House.

Japan Road Association, (2015). *Explanation and Application of the Road Structure Ordinance.* Revised Edition ed. s.l.:Japan Road Association.

Jocher, G. (2020). *ultralytics/yolov5*. [online] GitHub. Available at: https://github.com/ultralytics/yolov5.

Jones, P. and Boujenko, N., 2009. 'Link'and'Place': A new approach to street planning and design. *Road & transport research: A journal of Australian and New Zealand research and practice*, *18*(4), pp.38-48.

Jones, P., Marshall, S. and Boujenko, N., 2008. Creating more people-friendly urban streets through 'link and place'street planning and design. *IATSS research*, *32*(1), pp.14-25.

Joshi, D., Datta, R., Fedorovskaya, E., Luong, Q.T., Wang, J.Z., Li, J. and Luo, J., 2011. Aesthetics and emotions in images. *IEEE Signal Processing Magazine*, *28*(5), pp.94-115.

Jun, H.J. and Hur, M., 2015. The relationship between walkability and neighborhood social environment: The importance of physical and perceived walkability. *Applied Geography*, *62*, pp.115-124.

Kang, Y., Kim, J., Park, J. and Lee, J., 2023. Assessment of perceived and physical walkability using street view images and deep learning technology. *ISPRS International Journal of Geo-Information*, *12*(5), p.186.

Kapsalis, T., 2024. UrbanGenAI: Reconstructing Urban Landscapes using Panoptic Segmentation and Diffusion Models. *arXiv preprint arXiv:2401.14379*.

Kato, H. and Kanki, K., 2019. Effectiveness of Walkability Indicator from the perspective of Subjective Evaluation on Streets in Sprawl Urban Areas. *Journal of the City Planning Institute of Japan*, *54*(1).

Ki, D. and Lee, S., 2021. Analyzing the effects of Green View Index of neighborhood streets on walking time using Google Street View and deep learning. *Landscape and Urban Planning*, *205*, p.103920.

Kim, D., Guida, G. and García del Castillo y López, J.L., 2022. PlacemakingAI: Participatory Urban Design with Generative Adversarial Networks. 485–494.

Kingma, D.P. and Welling, M., 2013. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.

Kirillov, A., He, K., Girshick, R., Rother, C. and Dollár, P., 2019. Panoptic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 9404-9413).

Koo, B.W., Guhathakurta, S. and Botchwey, N., 2022. Development and validation of automated microscale walkability audit method. *Health & Place*, *73*, p.102733.

Koo, B.W., Guhathakurta, S. and Botchwey, N., 2022. How are neighborhood and street-level walkability factors associated with walking behaviors? A big data approach using street view images. *Environment and Behavior*, *54*(1), pp.211-241.

LeCun, Y., Bengio, Y. and Hinton, G., 2015. Deep learning. *nature*, *521*(7553), pp.436-444.

LIMGOMONVILAS, T. and NIMANONG, R., 2018. Multi-criteria analysis and network analysis for walkability score in Amphoe Muang, Nonthaburi, Thailand. *Journal of Advanced Research in Social Sciences and Humanities*, *3*(4), pp.125-135.

Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P. and Zitnick, C.L., 2014. Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13* (pp. 740-755). Springer International Publishing.

Liu, Z., Mao, H., Wu, C.Y., Feichtenhofer, C., Darrell, T. and Xie, S., 2022. A convnet for the 2020s. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 11976-11986).

Li, X., Zhang, C. and Li, W., 2015. Does the visibility of greenery increase perceived safety in urban areas? Evidence from the place pulse 1.0 dataset. *ISPRS International Journal of Geo-Information*, *4*(3), pp.1166-1183.

Li, Y., Yabuki, N., Fukuda, T. and Zhang, J., 2020, September. A big data evaluation of urban street walkability using deep learning and environmental sensors-a case study around Osaka University Suita campus. In *proceedings of the 38th eCAADe conference, TU Berlin, Berlin, Germany* (pp. 319-328).

Lizárraga, C., Martín-Blanco, C., Castillo-Pérez, I. and Chica-Olmo, J., 2022. Do university students' security perceptions influence their walking preferences and their walking activity? A case study of Granada (Spain). *Sustainability*, *14*(3), p.1880.

Ma, H. and Zheng, H., 2023, July. Text Semantics to Image Generation: A method of building facades design base on Stable Diffusion model. In *The International Conference on Computational Design and Robotic Fabrication* (pp. 24-34). Singapore: Springer Nature Singapore.

Mateo-Babiano, I., 2016. Pedestrian's needs matter: Examining Manila's walking environment. *Transport Policy*, *45*, pp.107-115.

Matthews, H., Limb, M. and Taylor, M., 2004. The 'street as thirdspace'. In *Children's geographies* (pp. 54-68). Routledge.

Meng, S.A., Yang, Y. and Lewis, R., 2023. Subjective Versus Objective: Divergency Between Subjective Walkability and Walk Score During the COVID-19 Pandemic. *Transportation Research Record*, p.03611981231165023.

Mishra, S., Mishra, M., Kim, T. and Har, D., 2023. Road Redesign Technique Achieving Enhanced Road Safety by Inpainting with a Diffusion Model. *arXiv preprint arXiv:2302.07440*.

MLIT, 2021. *Comfortable and inviting to walk: Ground-level design - Elements and points learned from examples.*, s.l.: MLIT.

Plateau. 2022. *Open Data | 3D 都市モデルオープンデータ | PLATEAU [プラトー]*. [online] Available at: https://www.mlit.go.jp/plateau/open-data/ [Accessed 23 Mar. 2024].

Moura, F., Cambra, P. and Gonçalves, A.B., 2017. Measuring walkability for distinct pedestrian groups with a participatory assessment method: A case study in Lisbon. *Landscape and Urban Planning*, *157*, pp.282-296.

Nagata, S., Nakaya, T., Hanibuchi, T., Amagasa, S., Kikuchi, H. and Inoue, S., 2020. Objective scoring of streetscape walkability related to leisure walking: Statistical modeling approach with semantic segmentation of Google Street View images. *Health & Place*, *66*, p.102428.

Neighborhood Street Research Group, 1989. *The Creation of Roads for Coexistence Between People and Cars: Considerations for Neighborhood Street Planning.* Chuo-ku, Tokyo, Japan: KAJIMA INSTITUTE PUBLISHING CO., LTD.

Neuhold, G., Ollmann, T., Rota Bulo, S. and Kontschieder, P., 2017. The mapillary vistas dataset for semantic understanding of street scenes. In *Proceedings of the IEEE international conference on computer vision* (pp. 4990-4999).

Ng, S., Lai, C., Liao, P., Lao, M., Lau, W., Govada, S. and Spruijt, W., 2016. Measuring and improving walkability in Hong Kong. *Civic Exchange and UDP International*.

Noyman, A. and Larson, K., 2020, May. A deep image of the city: Generative urban-design visualization. In *Proceedings of the 11th annual symposium on simulation for architecture and urban design* (pp. 1-8).

Ogawa, Y. et al., 2022. Development of a Model for Evaluating Subjective Impressions of Streetscapes Using Omnidirectional Street Image Big Data. *31st Annual Conference of Geographical Information Systems Association, proceedings.*

Ogawa, Y., Oki, T., Zhao, C., Sekimoto, Y. and Shimizu, C., 2024. Evaluating the subjective perceptions of streetscapes using street-view images. *Landscape and Urban Planning*, *247*, p.105073.

Oki, T. and Kizawa, S., 2022. Model for evaluating impression of streets in residential areas based on image big data and a large questionnaire survey. *J Archit Planning*, *87*(800), pp.2102-2113.

Ordonez, V. and Berg, T.L., 2014. Learning high-level judgments of urban perception. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part VI 13* (pp. 494-510). Springer International Publishing.

Petritsch, T.A., Landis, B.W., McLeod, P.S., Huang, H.F., Challa, S. and Guttenplan, M., 2005. Level-of-service model for pedestrians at signalized intersections. *Transportation research record*, *1939*(1), pp.54-62.

Qiu, W., Li, W., Liu, X., Zhang, Z., Li, X. and Huang, X., 2023. Subjective and objective measures of streetscape perceptions: Relationships with property value in Shanghai. *Cities*, *132*, p.104037.

Quan, S.J., 2022. Urban-GAN: An artificial intelligence-aided computation system for plural urban design. *Environment and Planning B: Urban Analytics and City Science*, *49*(9), pp.2500-2515.

Rodrigue, L., Manaugh, K., El-Geneidy, A., Daley, J., Wasfi, R., Ravensbergen, L. and Butler, G., 2022. Factors influencing subjective walkability. *Journal of Transport and Land Use*, *15*(1), pp.709-727.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P. and Ommer, B., 2022. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 10684-10695).

Salaryexplorer. 2023. *Average Salary in Hong Kong 2021 - The Complete Guide*. [online] Available at: http://www.salaryexplorer.com/salary-survey.php?loc=97&loctype=1 [Accessed 2 Feb. 2023].

Sarkar, C., Webster, C., Pryor, M., Tang, D., Melbourne, S., Zhang, X. and Jianzheng, L., 2015. Exploring associations between urban green, street design and walking: Results from the Greater London boroughs. *Landscape and Urban Planning*, *143*, pp.112-125.

Setagaya City 2023. *Setagaya City*. [online] 世田谷区ホームページ. Available at: https://www.city.setagaya.lg.jp/.

Shatu, F., Yigitcanlar, T. and Bunker, J., 2019. Shortest path distance vs. least directional change: Empirical testing of space syntax and geographic theories concerning pedestrian route choice behaviour. *Journal of Transport Geography*, *74*, pp.37-52.

Song, J., Lee, J.K., Choi, J. and Kim, I., 2020. Deep learning-based extraction of predicate-argument structure (PAS) in building design rule sentences. *Journal of Computational Design and Engineering*, *7*(5), pp.563-576.

Sung, H., Go, D., Choi, C.G., Cheon, S. and Park, S., 2015. Effects of street-level physical environment and zoning on walking activity in Seoul, Korea. *Land Use Policy*, *49*, pp.152-160.

Tao, T., Wang, J. and Cao, X., 2020. Exploring the non-linear associations between spatial attributes and walking distance to transit. *Journal of Transport Geography*, *82*, p.102560.

Thwaites, K., Simpson, J. and Simkins, I., 2020. Transitional edges: A conceptual framework for socio-spatial understanding of urban street edges. *Urban Design International*, *25*, pp.295-309.

TMG 2023. *Tokyo Sustainability Action 2023 ＃FutureTokyo*. [online] Available at: https://www.metro.tokyo.lg.jp/english/about/sustainable/documents/tokyo_sustainability_action2023.pdf.

TMG 2024. *東京都公式ホームページ*. [online] www.metro.tokyo.lg.jp. Available at: https://www.metro.tokyo.lg.jp/index.html.

Vallejo-Borda, J.A., Cantillo, V. and Rodriguez-Valencia, A., 2020. A perception-based cognitive map of the pedestrian perceived quality of service on urban sidewalks. *Transportation research part F: traffic psychology and behaviour*, *73*, pp.107-118.

Wang, K. and Akar, G., 2018. Street intersection characteristics and their impacts on perceived bicycling safety. *Transportation research record*, *2672*(46), pp.41-54.

Wang, W., Li, P., Wang, W. and Namgung, M., 2012. Exploring determinants of pedestrians' satisfaction with sidewalk environments: Case study in Korea. *Journal of Urban Planning and Development*, *138*(2), pp.166-172.

Weber, R.E., Mueller, C. and Reinhart, C., 2022. Automated floorplan generation in architectural design: A review of methods and applications. *Automation in Construction*, *140*, p.104385.

Irvin, K.A.T.J.A., 2008. How far, by which route and why? A spatial analysis of pedestrian preference. *Journal of urban design*, *13*(1), pp.81-98.

Wijnands, J.S., Nice, K.A., Thompson, J., Zhao, H. and Stevenson, M., 2019. Streetscape augmentation using generative adversarial networks: Insights related to health and wellbeing. *Sustainable Cities and Society*, *49*, p.101602.

Williams, P.T. and Thompson, P.D., 2013. Walking versus running for hypertension, cholesterol, and diabetes mellitus risk reduction. *Arteriosclerosis, thrombosis, and vascular biology*, *33*(5), pp.1085-1091.

Woo, S., Debnath, S., Hu, R., Chen, X., Liu, Z., Kweon, I.S. and Xie, S., 2023. Convnext v2: Co-designing and scaling convnets with masked autoencoders. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 16133-16142).

Wu, A.N. and Biljecki, F., 2022. GANmapper: geographical data translation. *International Journal of Geographical Information Science*, *36*(7), pp.1394-1422.

Wu, A.N., Stouffs, R. and Biljecki, F., 2022. Generative Adversarial Networks in the built environment: A comprehensive review of the application of GANs across data types and scales. *Building and Environment*, *223*, p.109477.

Wu, F., Li, W. and Qiu, W., 2023. Examining non-linear relationship between streetscape features and propensity of walking to school in Hong Kong using machine learning techniques. *Journal of transport geography*, *113*, p.103698.

Yamanaka, R. & Oki, T., 2022. Development of a simulation method capable of integratively predicting changes in street landscape images and impression evaluation values. *2022 Annual Conference of the Artificial Intelligence Society of Japan (36th edition),* p. Kyoto.

Yang, L., Ao, Y., Ke, J., Lu, Y. and Liang, Y., 2021. To walk or not to walk? Examining non-linear effects of streetscape greenery on walking propensity of older adults. *Journal of transport geography*, *94*, p.103099.

Yan, Y. and Ryu, Y., 2021. Exploring Google Street View with deep learning for crop type mapping. *ISPRS Journal of Photogrammetry and Remote Sensing*, *171*, pp.278-296.

Yin, C., Cao, J., Sun, B. and Liu, J., 2023. Exploring built environment correlates of walking for different purposes: Evidence for substitution. *Journal of transport geography*, *106*, p.103505.

Yin, L., 2017. Street level urban design qualities for walkability: Combining 2D and 3D GIS measures. *Computers, Environment and Urban Systems*, *64*, pp.288-296.

Yin, L., 2017. Street level urban design qualities for walkability: Combining 2D and 3D GIS measures. *Computers, Environment and Urban Systems*, *64*, pp.288-296.

Yu, E. and Choi, J., 2023. Development of building information modeling-based automation assessment process for universal design of public buildings. *Journal of Computational Design and Engineering*, *10*(2), pp.641-654.

Zapata, O. and Honey-Rosés, J., 2022. The behavioral response to increased pedestrian and staying activity in public space: A field experiment. *Environment and Behavior*, *54*(1), pp.36-57.

Zeng, L., Lu, J., Li, W. and Li, Y., 2018. A fast approach for large-scale Sky View Factor estimation using street view images. *Building and Environment*, *135*, pp.74-84.

ZENRIN. 2023. *MaaS*. [online] Available at: https://www.zenrin.co.jp/product/category/iot/maas/index.html [Accessed 23 Jul. 2024].

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

Zhang, L., Rao, A. and Agrawala, M., 2023. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 3836-3847).

Zhou, B., Zhao, H., Puig, X., Fidler, S., Barriuso, A. and Torralba, A., 2017. Scene parsing through ade20k dataset. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 633-641).

Zhou, H., He, S., Cai, Y., Wang, M. and Su, S., 2019. Social inequalities in neighborhood visual walkability: Using street view imagery and deep learning technologies to facilitate healthy city planning. *Sustainable cities and society*, *50*, p.101605.