Leveraging New Forms of Data for Human-Centric Urban Analytics

Yu Wang

SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

SCHOOL OF GEOGRAPHICAL & EARTH SCIENCES

COLLEGE OF SCIENCE & ENGINEERING

University of Glasgow

SEPTEMBER 2025

To my parents,
and my wife

Abstract

Urban environments are becoming increasingly diverse, dynamic, and complex ecosystems. This transformation is driven by global urbanisation, rising population densities, and technological advancements. Understanding urban environments is becoming both essential and challenging due to their rapid evolution and inherent dynamism. Traditional methods often struggle to meet such demand. Therefore, there is a growing need to leverage new forms of urban data that can capture contextual and behavioural insights at a higher frequency to facilitate timely and responsive decision-making within urban environments.

In recent decades, the widespread use of smart devices, data-sharing platforms, and improvements in computational hardware have facilitated the continuous generation of new forms of urban data. Advances in artificial intelligence and deep learning models, along with such data, have opened up significant opportunities for developing cost-effective and scalable applications that enhance our understanding of urban environments from new perspectives. This thesis makes three complementary contributions to the field of human-centric urban analytics.

First, it presents an innovative approach for building height estimation by leveraging ubiquitous mobile signals. The proposed approach provides a cost-effective, globally accessible, and efficient solution for creating 3D maps, which requires no dedicated equipment beyond consumer-level mobile phones. Second, the thesis contributes human-centric research to support humans in cities. It demonstrates the effectiveness of Ultra-Wideband (UWB) signals for fine-grained human activity recognition, further emphasising its cost-effective, non-intrusive, and reliable potentials. Additionally, the thesis analyses urban pedestrian disorientation in complex city environments based on survey data from the Greater London Area. The research identifies and quantifies factors leading to disorientation, employing expert-led Analytical Hierarchy Process (AHP) and data-driven regression methods. Both studies offer insights for designing more inclusive and navigable urban spaces. Third, this thesis addresses an often-overlooked issue in the development of deep learning models for positioning: temporal bias in visual urban datasets. Using cross-view geo-localisation (CVGL) as a case study, this thesis evaluates the performance of two state-of-the-art CVGL models on an original benchmark dataset and a custom dataset spatially aligned with the benchmark. Our findings reveal significant degradation in model performance due to temporal variations between two datasets. Semantic segmentation and SHapley Additive exPlanations (SHAP) explainability framework are used to further illustrate how temporal visual changes affect model reliability.

In summary, this thesis presents an effective framework for urban analytics that prioritises human-centric approaches. It covers the collection and creation of custom datasets in new formats, as well as the development of novel sensing methods at various scales. The research offers valuable insights aimed at improving positioning and navigation services, while also expanding humans' understanding of urban environments. Importantly, it identifies hidden biases in urban data applications. Overall, this work demonstrates how emerging data sources and analytical techniques can improve our ability to model, understand, and design smarter and more inclusive urban environments.

Contents

University of Glasgow
College of Science & Engineering
Statement of Originality

Name: Yu Wang
Registration Number: xxxxxxx

I certify that the thesis presented here for examination for a PhD degree of the University of Glasgow is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it) and that the thesis has not been edited by a third party beyond what is permitted by the University's PGR Code of Practice.

The copyright of this thesis rests with the author. No quotation from it is permitted without full acknowledgement.

I declare that the thesis does not include work forming part of a thesis presented successfully for another degree, unless explicitly identified and as noted below.

I declare that this thesis has been produced in accordance with the University of Glasgow's Code of Good Practice in Research.

I acknowledge that if any issues are raised regarding good research practice based on review of the thesis, the examination may be postponed pending the outcome of any investigation of the issues.

Signature: Yu Wang
Date: 15 October 2025

Abbreviations

- AE - Absolute Error
- AHP - Analytic Hierarchy Process
- AI - Artificial Intelligence
- BIM - Building Information Modeling
- CDF - Cumulative Distribution Function
- CIR - Channel Impulse Response
- CNN - Convolutional Neural Network
- CSI - Channel State Information
- CVGL - Cross-View GeoLocalisation
- CVUSA - Cross-View USA dataset
- DBSCAN - Density-Based Spatial Clustering of Applications with Noise
- DL - Deep Learning
- FCN - Fully Convolutional Neural Network
- GL - Getting Lost
- GNSS - Global Navigation Satellite Systems
- GPU - Graphics Processing Unit
- GRU - Gated Recurrent Unit
- HAR - Human Activity Recognition
- IoT - Internet of Things
- IQR - Interquartile Range
- kNN - k-Nearest Neighbours
- LEO - Low Earth Orbit
- LiDAR - Light Detection and Ranging
- LOD - Level of Detail
- LOS - Line of Sight
- LSTM - Long-short Term Memory
- LSTM-FCN - Long-short Term Memory – Fully Convolutional Network
- ML - Machine Learning

- `MLP` - Multilayer Perceptron
- `mmWave` - Millimetre Wave
- `MSE` - Mean Squared Error
- `NBC` - Naive Bayes Classifier
- `NLOS` - Non Line Of Sight
- `NMSE` - Normalised Mean Squared Error
- `OLS` - Ordinary Least Squares
- `OS` - Ordnance Survey
- `OSM` - OpenStreetMap
- `OSRM` - Open Source Routing Machine
- `PDFM` - Population Dynamics Foundation Model
- `POI` - Point of Interest
- `RF` - Random Forest
- `RNN` - Recurrent Neural Network
- `RR` - Ridge Regression
- `RSS` - Received Signal Strength
- `SAFA` - Spatial-Aware Feature Aggregation
- `SAR` - Synthetic Aperture Radar
- `SHAP` - SHapley Additive exPlanations
- `SVI` - Street View Images
- `SVM` - Support Vector Machine
- `Tx` - Transmitter
- `Rx` - Receiver
- `UAV` - Unmanned Aerial Vehicle
- `UWB` - Ultra-Wideband
- `VGG` - Visual Geometry Group (VGG16/VGG19 backbone)
- `VIF` - Variance Inflation Factor

Acknowledgements

I want to thank my wonderful supervisor, Ana Basiri. I would not have reached this point without her endless support, constructive advice, and insightful discussions throughout my PhD journey. Thank you for your guidance, patience, and generous sharing of knowledge, both academic and personal, during this time. I am especially grateful for your constant encouragement, which enabled me to explore and seek answers to interesting research questions freely. I am equally thankful to my co-supervisor, Mingshu Wang, who has always been generous with his invaluable advice and insightful conversations whenever I needed guidance or support in my research. I would also like to thank my colleagues, Petrus Gerrits, Dominick Sutton, Joseph Shingleton, and Guy Solomon. I truly appreciate the constructive and supportive conversations we have shared over these years, which have made this journey more meaningful.

My heartfelt gratitude also goes to my wife's parents, whose encouragement a few years ago inspired me to take this significant step in my life. Their constant support, kindness, and belief in me have sustained me throughout this journey, without which I would not have reached this point. To my beloved wife, Wanying: from the toughest times during the pandemic, when we were separated at opposite ends of the world, to the day we finally stand here side by side, you have been my best friend, my mentor, and my soulmate. Every step of this journey has been memorable and joyful because of you. To my mum and dad: you may think your greatest pride is me, but in truth, my greatest pride has always been you.

Finally, I would like to acknowledge my own life journey—with its struggles, mistakes, missed opportunities, regrets, frustrations, and lessons I wish I had learned earlier, as well as the efforts, brave attempts, achievements, and milestones—every step counts, and they have shaped me into who I am.

List of Publications

- **Wang, Y.**, & Basiri, A. (2025). Bit to brick: from cellular mobile signals to 3D city map creation. *Big Earth Data*, 1–25. *DOI: 10.1080/20964471.2025.2561319.*
- **Wang, Y.**, & Basiri, A. (2024). Advancing human activity recognition using ultra-wideband channel impulse response snapshots. In *2024 International Conference on Activity and Behavior Computing (ABC)* (pp. 1-10). IEEE. *DOI: 10.1109.ABC61795.2024.10651886.*
- **Wang, Y.**, Basiri, A., Gerrits, P., Solomon, G., Woelk, S., & Fidel Pereira, M., Why do pedestrians get lost? A case study of personal, situational, and environmental factors in Greater London. *Journal of Location Based Services.* (accepted).
- **Wang, Y.**, & Basiri, A., From static to dynamic: evaluating model robustness with historical street view images in cross-view geolocalisation. *IEEE Access* (under revision).

Chapter 1

Introduction

Nowadays, urban areas are expanding rapidly at an unprecedented pace (Angel 2023), evolving from simple administrative boundaries into diverse, highly dynamic, and constantly changing ecosystems where people live, move, and interact (Iossifova et al. 2017; Brenner and Schmid 2015; Wegener et al. 1986). This dynamism introduces frequent changes in land use, mobility demand, and the built environment, thereby creating an urgent need for rich, low-cost, and readily available data to capture these changes and support timely decision-making for infrastructure maintenance, transportation optimisation, and emergency responses (Angel 2023; Iossifova et al. 2017; Brenner and Schmid 2015; Wegener et al. 1986; Glaeser et al. 2018; Biljecki et al. 2015; Kwan and Lee 2005). However, traditional data acquisition methods, such as large-scale surveys or fine-grained sampling from extensive data corpora, face challenges in this scenario (Bradley et al. 2021; Meng 2018). They struggle to address contemporary urban challenges due to the necessary trade-off between data representativeness and acquisition efficiency, the lack of timely reflection of situations through the data (United Nations Economic Commission for Europe, Conference of European Statisticians Bureau 2024), or questionable data quality when sampling tiny groups from holistic populations (Arribas-Bel 2014; Rao and Molina 2015). In particular, multi-year survey cycles and delayed publication can make it difficult to reflect rapid redevelopment or short-term variations in cities, limiting their usefulness for responsive urban services (United Nations Economic Commission for Europe, Conference of European Statisticians Bureau 2024). The emerging use of digital technology in day-to-day activities, however, presents new opportunities for urban data collection, improving both scale and granularity. With the high penetration of end-user devices and changing lifestyles that involve voluntarily sharing data collected by such devices (Zaman et al. 2024; Rehan 2023), the new crowd-sourcing platforms relying on volunteer contributions (Heipke

2010; Huang et al. 2024), and the increasing availability of commercial data (Kim and Jang 2023), new types of data are becoming more common, widespread, and accessible. Consequently, new forms of data are generated within urban areas, including satellite and street-level images (Yu and Fang 2023; Cinnamon and Jahiu 2021), wireless signals (Minoli and Occhiogrosso 2018; Kapoor et al. 2017), and mobility traces (Kafi et al. 2013). These valuable data sources provide an opportunity to observe urban environments with continuous or near-real-time updates, rather than periodic snapshots from survey-based workflows, enabling more responsive analysis of dynamic urban systems (United Nations Economic Commission for Europe, Conference of European Statisticians Bureau 2024; Senousi et al. 2021).

From a spatial perspective, urban environments encompass both small-scale and large-scale features, ranging from individual buildings and street segments to neighbourhood- and city-level structures, which poses challenges for integrated sensing and modelling across scales. Human activities (Hasan et al. 2013; Ji et al. 2016; Fu et al. 2020), buildings (Alsafery et al. 2023), transportation networks (Silva et al. 2018), and public spaces (Lau et al. 2017) can now be sensed, mapped, and studied using many different methods. Taken together, the modern concept of “urban” not only includes its traditional definition as a collection of physical locations, but also a digital platform in which heterogeneous streams (e.g., imagery and wireless signals) can be integrated to model both the built environment and human activity (Yu and Fang 2023; Liu et al. 2019).

Along with the emergence of urban data proliferation, numerous urban applications are being developed to enhance citizens’ lives. A significant portion of them rely on location-based techniques, supporting applications such as urban mobility, transportation, and driving (Huang et al. 2021). Navigation is the most widely used and frequently required urban service in modern city life, with evidence showing that Google Maps is used by 2 billion users per month (Miriam Daniel, Google Maps 2025), highlighting the societal scale and practical importance of reliable positioning and navigation services in everyday urban contexts. This scale implies that even modest degradations in positioning reliability in dense urban settings would affect large numbers of users. Technologies such as the Global Navigation Satellite System (GNSS) have become prevalent in urban navigation due to their availability, accessibility, and affordability via mass-market devices, playing a vital role in navigation services (Zangenehnejad and Gao 2021). However, several vulnerabilities are associated with GNSS signals, compromising not only location accuracy but also diminishing their reliability. Its weak signals are increasingly the subject of interference, spoofing, or jamming (Zidan et al.

2020; Papadimitratos and Jovanovic 2008; Psiaki and Humphreys 2016). Additionally, within the urban environment, multiple static or dynamic obstacles lead to multipath transmission, further compromising the reliability of false distance measurements and resulting in questionable positional accuracy and navigation reliability (Budiyono 2012; Groves et al. 2012). The rapid growth of urban populations has led to an increase in buildings, vehicles, and other physical structures, resulting in highly cluttered environments (Schläpfer et al. 2015; De Bellefon et al. 2021). Consequently, in such dense urban areas, GNSS signal availability and quality are often degraded due to signal blockage, non-line-of-sight (NLOS) reception, multipath propagation, and related environmental challenges (Zidan et al. 2020). Typical consumer-grade GNSS receivers achieve horizontal errors below 10 m for about 95% of observations under open-sky conditions, while performance degrades to tens of metres in urban environments due to signal blockage and multipath (U.S. Department of Defense 2020; Groves 2013). Existing research tackling these limitations generally follows two approaches: either improving GNSS accuracy using methods such as shadow matching supported by detailed 3D urban models (Groves 2011a; Adjrad and Groves 2016; Bai et al. 2022), or developing alternative approaches that are equally free, accessible, and globally available like GNSS but also reliable under urban conditions (Han et al. 2016), options included millimetre Wave (mmWave) wireless signals (Saleh et al. 2022), vision information (Berton et al. 2022; Liu et al. 2024a), and multi-sensor integration (Wang 2023). The first approach relies on precise 3D maps (Groves et al. 2012; Wen and Hsu 2022; Ng and Hsu 2021; Zhang et al. 2018b), while the second strand has seen increasing attention in recent years, particularly in vision-based geolocalization using computer vision techniques (Durgam et al. 2024).

Although diverse forms of data are continuously generated by smart devices and (Chambers and Evans 2020; Salih et al. 2025), sensors (Lau et al. 2017; Bok et al. 2014), and commercial organisations (Google Maps 2025; Apple Maps 2025; Bing Maps 2025), and are widely shared through various online platforms (Mapillary 2025; KartaView Contributors 2025), their full potential to support and advance urban applications has yet to be fully realised, particularly in terms of robust integration across modalities, scalable deployment, and reliable performance under spatial and temporal data biases (Bibri 2019; Korada 2021; Koutra and Ioakimidis 2022; Dyer et al. 2024).

The effort to integrate these diverse data sources into current urban applications can be viewed from two complementary perspectives. On the one hand, innovative methodologies for integrating these data offer new opportunities to develop efficient and effective urban solutions. On the other hand, it is also crucial to examine the scalability and reliability of such methodologies, as well as the spatial-temporal biases embedded in the generated data. Specifically, “bias” is used in the thesis as a term for systematic deviations arising from data generation, sampling, processing, and modelling assumptions, which may lead to non-random deviations or imbalances across space, time, or contexts. In geographic data science, these issues are further involved with spatial heterogeneity and dependence, which complicate model outcomes in terms of both interpretation and generalisation. This thesis therefore treats bias as an empirical question and evaluates it through concrete case studies, including temporal bias in benchmark imagery datasets for cross-view geolocalisation (Zhao et al. 2025; Deuser et al. 2025).

Human-centric urban analytics refers to analysis and interpretation of urban data, moving beyond purely technological or infrastructure-centric analytics towards insights that are directly relevant to citizens’ interactions with their environment (Resch and Szell 2019). Building on this background, the central objective of this thesis is to advance human-centric urban analytics by developing and evaluating methods grounded in human-generated data and oriented towards improving the efficiency and reliability of human-centred urban services. Across its chapters, the thesis demonstrates how infrastructure- or environment-focused analyses can contribute to human-centric outcomes when they draw on data arising from human activity and are explicitly designed to support positioning, navigation, and wayfinding for urban residents.

The thesis encompasses three primary research strands.

- Research Strand 1 develops a data-driven approach to large-scale 3D mapping by leveraging opportunistic cellular mobile signals for building height estimation, with the aim of providing a cost-effective and scalable means of improving GNSS positioning and navigation reliability in dense urban environments.
- Research Strand 2 adopts a human-centred perspective to investigate how wireless sensing can recognise human physical activities through Ultra-Wideband signals and explain pedestrian disorientation by analysing the combined effects of personal, environmental, and situational factors on urban navigation.
- Research Strand 3 examines the impact of temporal bias in benchmark visual datasets on deep learning-based positioning models, using cross-view geolocalisation as a case study to assess how dataset design influences model performance, robustness, and reliability.

Taken together, by framing urban areas as dynamic laboratories of human–environment interaction, where heterogeneous data streams coexist and enable urban sensing, my thesis advances methodological tools and conceptual insights to improve positioning and navigation systems, ultimately supporting everyday mobility and the citizens' experience in urban environments.

1.1 Background

Urban analytics has emerged as an interdisciplinary research domain covering geography, urban planning, geographic information science, and computational social science, with a broader aim of understanding, modelling, and supporting urban systems through data-driven inquiry (Batty 2013; Kitchin 2014; Townsend 2015). In this thesis, cities are considered not only as physical systems that can be measured and optimised, but also as socio-technical environments shaped by human behaviour, governance structures, and uneven access to resources (Batty et al. 2012; Luusua et al. 2023). As a result, urban analytics encompasses a wide range of methodological traditions, including qualitative inquiry, statistical modelling, spatial analysis, and computational approaches, each providing distinct but complementary insights into urban phenomena (Kitchin 2014; Shi 2021). Our cities are denser than ever and are expanding at a faster rate (Angel 2023; Frolking et al. 2024). This introduces additional challenges related to urban sustainability, inequality, and resilience, which increasingly manifest through the spatial configuration, accessibility, and reliability of urban infrastructure and services (Michalina et al. 2021; Batty et al. 2012; Xu et al. 2025; Zeng et al. 2022; Batty 2016). In particular, due to the dynamic nature of urban areas, high-frequency changes in cities require timely responses for prompt decision-making, which traditional methods, such as surveys, could hardly satisfy (Senousi et al. 2021; Radwin 2009). Therefore, to cope with the rising demand for understanding urban environments, there is a need to leverage new forms of data that can help better capture the context and insights of urban environments, as well as advanced methodologies to process and analyse these increasingly complex datasets (Pan et al. 2016; United Nations 2024). At the same time, alongside rapid urban expansion, the availability of new data generated from user devices is also growing swiftly (Li et al. 2022c). These gadgets are becoming new types of sensing devices that can describe urban environments in novel ways by enabling large-scale, fine-grained, and frequently updated observations that are difficult to achieve through conventional survey-based data collection alone. Moreover, thanks to the proliferation of crowdsourcing platforms (Crooks et al. 2015), social media (Martí et al. 2019), and commercial organisations (Mapillary 2025; Bing Maps 2025; Google Maps 2025; Apple Maps 2025), such data is increasingly being made publicly accessible through various information propagation channels, with increasing availability and accessibility. Across the urban analytics and geographic data science literature, there is no single standard for how urban data should be categorised or evaluated, with perspectives ranging from technological capability and spatial coverage to social meaning, governance, and ethical implications (Kitchin 2014; Goodchild and Li 2012; Batty 2020). In this thesis, the term *new forms of data* is used in a broad sense to encompass both emerging data modalities collected through non-traditional means and novel uses of established data sources enabled by advances in digital collection and processing. This includes data types that are relatively new to urban studies, such as cellular mobile signals, ultra-wideband channel impulse responses, and large-scale street-view imagery, as well as more conventional sources, such as surveys, when collected via crowdsourcing platforms rather than through traditional sampling frameworks.

Recent technology advances have enabled the emergence of new data formats that continuously fill gaps in urban data (Manley and Dennett 2019). These include street view images (SVI) (Google Maps 2025; Biljecki and Ito 2021), remote-sensing imagery (Toth and Jóźków 2016; Dong et al. 2024), wireless and opportunistic signals (Liu et al. 2019; Han et al. 2016), data streams from Internet of Things (IoT) sensors (Salih et al. 2025), and so on. Together, these diverse sources are forming a new ecosystem for urban studies, moving beyond the limitations of traditional approaches and providing a versatile description of urban areas (Shi 2021). Each data modality offers unique perspectives, for example, SVIs can be used for the study of human perception (Dubey et al. 2016) and urban infrastructure (Li et al. 2022d) remote-sensing imagery is widely adopted for the study of urban land use (Yin et al. 2021), urban green spaces evaluation (Shahtahmassebi et al. 2021), and building detection and height estimation (Li et al. 2022a) mobile data can be used for the analysis of human mobility (Hu et al. 2021) and urban travel behaviours (Yang et al. 2023b). These multi-modal data sources can provide integrated, more comprehensive descriptions of urban elements and processes, which is essential for understanding the complexity and dynamics of contemporary cities (Lau et al. 2019). Beyond their technical characteristics, these data modalities have been framed in the literature as measurements, proxies, or representations of urban elements, raising ongoing debates about what aspects of urban life can be meaningfully captured, abstracted, or inferred from data-driven observations (Kitchin 2014; Shi 2021). The intrinsic complexity and heterogeneity of urban systems further underscore the need for new data forms. Cities are inherently multi-scale, involving interactions among diverse actors (residents, visitors, vehicles, and infrastructure) and processes that unfold across different spatial and temporal scales (Batty 2013; Kitchin 2014). Leveraging these new forms of data streams is therefore crucial for advancing urban analysis and for addressing challenges that remain intractable with conventional single-source approaches (Zheng et al. 2014). However, although the size and richness of these datasets continue to grow, their inherent biases are still insufficiently examined and often overlooked in the development of urban applications (Kamar et al. 2015; Hecht and Stephens 2014; Zhao et al. 2025).

Despite the complexity and inherent biases in new forms of urban data, their extensive spatial coverage, high update frequency, and relatively low cost of collection have made them an attractive resource for urban research (Martí et al. 2019; Biljecki and Ito 2021; Goodchild and Li 2012). Consequently, there is growing interest in exploring to what extent these data can be effectively and responsibly used to advance urban analytics and support data-informed decision-making (Grekousis 2019). For instance, multi-modal AI approaches can potentially enhance several applications and services. They include localisation in environments where GNSS signals are unreliable (Zhang et al. 2021), or creating automated mapping of rapidly changing urban features (Neupane et al. 2021), and provide fine-grained analysis of urban perception and behavioural patterns at scale (Dubey et al. 2016; Belhadi et al. 2021). These innovations have effectively processed urban data with increasing complexity and derived corresponding solutions from it. At the same time, rapid advancements in computational resources, such as Graphics Processing Unit (GPU)-accelerated processing, have made it feasible to handle the increasing data volumes and enable the effective integration of new forms of urban data (Reuther et al. 2021). Unlike traditional methods limited by static, single-source data, AI approaches enable the leveraging of diverse and rapidly growing urban data streams to build comprehensive, adaptable representations of urban environments. Deep learning models (LeCun et al. 2015), such as Convolutional Neural Network (CNN) (Fukushima 1980; LeCun et al. 1989) and Transformers (Vaswani et al. 2017), have shown remarkable success in extracting urban features, enabling robust visual geolocalisation, and recognising complex human activities from heterogeneous sources (Durgam et al. 2024; Alzubaidi et al. 2021; Grekousis 2019; Lin et al. 2022). In the context of human-centric urban analytics, neural network models are widely used to capture and interpret heterogeneous urban data. This thesis therefore focuses on the systematic evaluation of established model families, with detailed architectural and implementation choices discussed in the relevant chapters alongside their empirical applications. Moreover, AI-driven models are inherently scalable and adaptive, capable of learning from new data formats and transferring knowledge across different urban contexts (Wei et al. 2016; Wang et al. 2018a). The ongoing convergence of AI and multi-modal urban data is enabling more dynamic, precise, and actionable urban analytics, representing a fundamental shift towards truly data-driven city science and governance (Yi et al. 2025; Dao 2022; Feng et al. 2025). However, the increased reliance on data-driven models also introduces new challenges, including the risk of bias, issues with interpretability, and the need to ensure fair and responsible integration of AI insights into urban policy and design (Batty 2020; Luusua et al. 2023).

In parallel with the shift toward new data formats, the landscape of information propagation, data creation and data maintenance has also evolved (Goodchild 2007; Kitchin 2014). Historically, 3D mapping was carried out and maintained primarily by national mapping agencies (Lawrence 2004). In recent decades, commercial companies have taken a leading role, with platforms such as Google Maps (Google Maps 2025), Bing Maps (Bing Maps 2025), and Apple Maps (Apple Maps 2025) becoming integral to daily urban life. In addition to these commercial initiatives, crowdsourcing platforms like OpenStreetMap (OSM) (OpenStreetMap contributors 2017) have emerged as influential sources of open and collaborative mapping, relying on volunteers to create and maintain up-to-date 3D maps using a mix of official, open, and user-generated data. Similar trends are observed in other domains; for example, Mapillary (Mapillary 2025) and KartaView (KartaView Contributors 2025) now offer open, crowdsourced SVI as a complement to traditional providers like Google (Google Maps 2025). These developments highlight the growing diversity, scale, and participatory nature of urban data ecosystems, which further emphasise the importance and challenge of multi-modal data integration for urban intelligence.

Across the literature, new forms of urban data are frequently positioned as powerful solutions to long-standing urban challenges, particularly in sensing, localisation, and decision support. However, despite their increasing availability and technical sophistication, a fundamental gap remains insufficiently addressed. Specifically, there exists a disconnect between the *assumed problem-solving capacity* of new forms of urban data and their *empirically demonstrated ability* to support reliable, human-centred urban services under dynamic real-world urban conditions. The prevailing assumption that richer, larger, or more granular data will naturally translate into improved urban services is rarely examined holistically. This gap does not arise from a single technical limitation, but from the lack of integrated evaluation that connects data, models, and human outcomes. As a result, improvements are often demonstrated at the level of sensing, prediction, or classification, while their implications for service reliability, human experience, and robustness over time remain unclear. This thesis addresses this overarching gap by systematically examining how new forms of urban data perform when they are explicitly tasked with supporting positioning, navigation, and wayfinding for urban residents, and by evaluating the conditions under which their promised benefits hold or break down.

1.2 Motivations and Objectives

Given the scale and dynamism of cities, it is vital to explore effective, scalable, and cost-efficient approaches for sensing and modelling urban environments (Boyle et al. 2013). A wide range of urban information streams, such as satellite imagery, street-level views, and wireless signals, are now routinely generated (Yu and Fang 2023; Biljecki and Ito 2021). Moreover, the digitalisation of everyday life, particularly through ubiquitous mobile technologies, provides unprecedented opportunities for multi-modal urban sensing and data integration (Ghahramani et al. 2020). Nevertheless, alongside these opportunities, cities face enduring and evolving challenges. The United Nations has highlighted pressing urban issues, including housing shortages, resilience, inequality, waste management, and governance (United Nations 2024). In addition, urban environments are increasingly shaped by rapid population growth, financial and resource inequalities, and the rising complexity of socio-technical systems (Zhang 2016). As cities expand and adapt, these existing challenges often intensify, while new problems emerge through technological progress and shifting urban dynamics (Batty 2018; United Nations 2022). Consequently, addressing these intertwined challenges requires innovative approaches that move beyond traditional methods. For instance, many services and applications rely on precise location information. they are the foundation for location-based services (Usman et al. 2018), such as emergency rescue (Kim and Choi 2025; Tyagi et al. 2024), route navigation (Galbrun et al. 2016; Steiniger et al. 2006), and autonomous driving (Yurtsever et al. 2020). Currently, positioning and navigation are predominantly dependent on GNSS. These systems rely on relatively weak radio-frequency signals, typically around -160 dBW, transmitted from medium Earth orbit satellites located approximately 20,000–25,000 km above the Earth's surface. Each satellite continuously broadcasts a precisely time-stamped ranging code and ephemeris data, maintained through atomic clocks and control segments. A GNSS receiver determines its position by measuring the signal travel time from at least four satellites and then applying trilateration (a form of geometric triangulation) to compute its three-dimensional coordinates and clock offset (Enge 1994; Hofmann-Wellenhof et al. 2008; Kaplan and Hegarty 2017). At a conceptual level, GNSS can be understood as an inference system that estimates position from indirect, noisy, and incomplete observations. The receiver does not observe location directly, but infers it by combining multiple imperfect range measurements under simplifying assumptions about signal propagation, clock synchronisation, and satellite geometry. In this sense, the accuracy of the final output is strongly conditioned by data quality or environmental context (Langley et al. 2017). Despite their broad adoption, persistent difficulties remain in achieving accurate localisation and navigation within complex urban environments

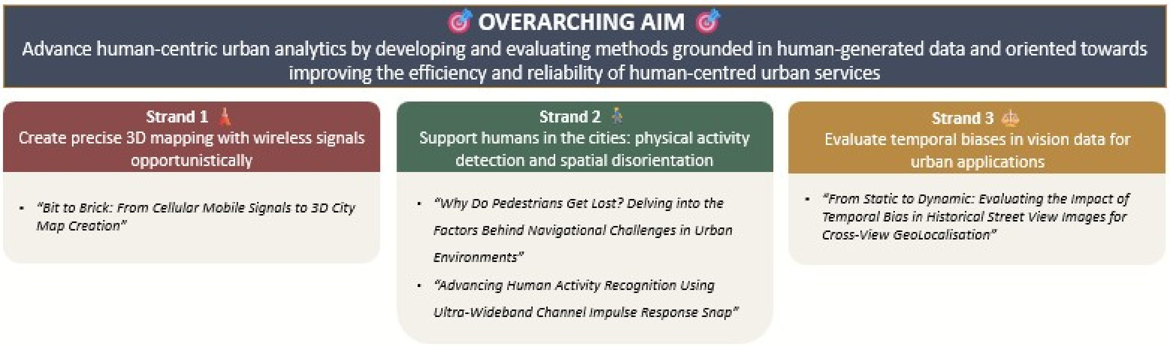
(Schön et al. 2022). Because GNSS-based localisation can often be challenging in the built environment—particularly within the so-called urban canyons where signals are frequently obstructed, reflected, or diffracted by buildings and other structures—users often encounter noise, signal blockage, and inaccurate position estimates (Zidan et al. 2020). While advanced algorithms and sensor optimisations have been proposed to mitigate multipath and signal blockage, the integration of up-to-date and precise 3D maps has emerged as a critical complementary solution (Groves and Adjrad 2019; Wang et al. 2013b). Such maps are not only useful for modelling signal propagation and interactions with urban structures, but also form the foundation of 3D map-aided GNSS positioning techniques, which have been shown to improve accuracy in dense urban canyons markedly. For example, shadow matching and related methods use detailed 3D building models to predict line-of-sight and non-line-of-sight satellite visibility, thereby reducing multipath errors. Empirical evaluations have demonstrated significant benefits Groves 2011a; Wang et al. 2013a, underscoring why high-quality, up-to-date 3D maps are critical enablers of reliable localisation in modern cities.

Despite the significance of precise and up-to-date 3D maps, building and updating comprehensive 3D maps remains a long-term goal due to several factors (Balado et al. 2025). First, creating and maintaining 3D mapping pose challenges, particularly in low-income countries. For example, in the United Kingdom, the creation and maintenance of a national 3D map has been estimated at around £75 million, illustrating the considerable costs of scaling 3D mapping initiatives (Wong 2018). Moreover, 3D mapping reveals marked disparities in coverage and quality between developed and developing regions. For example, even in many European countries, where technical capacity is higher, recent research shows that more than half still have less than 50% completeness in 3D building data (Milojevic-Dupont et al. 2023). The gap is likely to be even more severe in Global South countries, where rapid urbanisation synchronises with limited financial resources (Myers 2021; Biljecki 2020). Accordingly, a consistent effort is required to explore scalable methods to enhance 3D map coverage. Last but not least, urban environments are dynamic; redevelopment, demolition, and transformation continuously change the form of cities, thus requiring efforts for regular mapping updates to support cooperation and accurately represent urban areas (Zhao et al. 2020). Therefore, it is essential to explore scalable, cost-effective, and globally accessible alternatives for 3D mapping that extend beyond the use of dedicated equipment.

In addition, supporting humans in cities through human-centric research forms a second primary motivation of this thesis. Supporting people in their daily urban lives requires more than an understanding of the built environment, such as 3D mapping; it also depends on recognising and interpreting human behaviours at finer scales. This thesis aims to explore how sensing approaches can be extended across different scales of urban life, from large-scale representations of the built environment, such as 3D mapping, to small-scale observations of human activities through wireless sensing. At the same time, as a wide range of smart devices continuously generates new forms of data, people benefit from abundant external support for daily navigation. Yet, they still face the risk of getting lost in complex urban environments (Prandi et al. 2023; Farr et al. 2012). The growing complexity of urban areas, characterised by the diverse and dynamic flows of both people and vehicles, exacerbates the challenges of orientation and wayfinding (Vaez et al. 2016). Although a variety of navigation applications are now readily available on end-user devices (Ishikawa 2019), many users continue to experience disorientation when travelling in cities, a phenomenon commonly referred to as "getting lost". The underlying causes of getting lost are highly heterogeneous (Farr et al. 2012; Iftikhar et al. 2021). A getting lost event is the result of a complex interplay of personal, environmental and situational factors (Hölscher et al. 2012; Gath-Morad et al. 2022). Navigation ability shows substantial variation across individuals. For example, large-scale evidence demonstrates a near-linear decline in navigation performance with age from early adulthood, alongside a male advantage that varies markedly in all countries (Spiers et al. 2023). Education has been proven to be positively and causally associated with spatial navigation ability (Coutrot et al. 2025). Cultural background has been shown to influence discrepancies between self-evaluated navigational ability and actual wayfinding performance (Walkowiak et al. 2023a). Environmental factors, such as urban layout and the complexity of road networks, may either facilitate or hinder effective navigation (Coutrot et al. 2022). Ultimately, a "getting lost" event results from the interaction between these factors, making it a truly multifaceted phenomenon. Therefore, effective solutions should aim to understand and evaluate these factors in combination, assessing their relative importance and interactions rather than treating each factor in isolation. Together, these two strands of human activity recognition and urban disorientation analysis embody the human-centric perspective of this thesis, seeking to better understand how people interact and navigate within cities, and how technology can more effectively support them.

Data plays a crucial role in uncovering latent patterns that inform urban applications, particularly in the era of artificial intelligence and deep learning, where such models increasingly rely on large and diverse datasets. Data itself underpins both the need for long-term goals, such as infrastructure digitalisation, as well as high-frequency service updates (Bayat and Kawalek 2023; Shibasaki et al. 2020). New forms of urban data have provided new opportunities for timely responses for urban decision-making. The proliferation of these data has also brought new capabilities to build urban applications than before. For example, with the increasing availability of urban data, researchers and practitioners can now develop powerful multimodal foundation models for a wide range of urban analytics tasks (Mai et al. 2023). For example, recent advances have enabled the large-scale training of deep learning models, such as Google's Population Dynamics Foundation Model (PDFM) (Agarwal et al. 2024) and the trajectory-based mobility foundation model (Choudhury et al. 2024), which have the potential to facilitate downstream applications. However, the transparency, representativeness, and fairness of the datasets need to be fully explored and understood for building urban applications that rely on data-driven and AI-based methods (Wang et al. 2019b). The lack of investigation into these components, such as geographic granularity (Liu et al. 2022; Mandal et al. 2021), temporal variation (Shah and Sureja 2025), or cultural background differences (Zhou et al. 2024; Zhou et al. 2024), raises questions about the outcomes of AI models trained on these datasets. The biases or exclusion represented inherently in the datasets are the byproduct of the nature of user-generated and volunteered-based data. They remain invisible until their existence can be thoroughly examined, or in the worst case, never be recognised, thus leading to skewed results and flawed conclusions (Tommasi et al. 2017; Zhang et al. 2018a).

In this thesis, visual-based geolocalisation is used as a case study to explore the extent to which spatial-temporal biases may contribute to the outcome of deep learning models. In recent years, the availability of geo-tagged images has surged, driven by the widespread penetration of smart devices that continually capture multimodal urban data (Zhang et al. 2024) and by digitalised lifestyles in which user-generated content is routinely shared and made publicly accessible across online platforms (Gao et al. 2021). Alongside advances in deep learning, these developments have created new opportunities to train models on visual data and infer geolocation directly from image content (Mai et al. 2025). Such visual information, captured by mobile phones, social media posts, or surveillance cameras, has become especially valuable in scenarios such as disaster response, crime reporting, and other emergencies where rapid and accurate location identification supports timely decision-making (Amiruzzaman et al. 2021; Lopez-Fuentes et al. 2018; Kustu and Taskin 2023). Nevertheless, a significant challenge for visual-based applications lies in generating datasets that are both sufficiently large and representative, while minimising systematic biases (Torralba and Efros 2011). For example, in geolocalisation, benchmark datasets typically aim for spatial balance and diversity but often neglect temporal coverage (Deuser et al. 2025). As a result, models trained on such datasets may generalise poorly across time, thereby introducing unexpected temporal biases (Tommasi et al. 2017). Similarly, when models are trained to infer subjective perceptions such as safety, beauty, or liveliness, they frequently display cultural or regional biases if tested on imagery outside the scope of the training distribution (Dubey et al. 2016). Therefore, although urban applications increasingly rely on AI-driven methods, important aspects such as transparency, representativeness, and fairness of the underlying datasets remain underexplored. Insufficient attention to geographic granularity, temporal variation, or demographic coverage raises concerns about the accuracy of model outputs. Accordingly, recognising, quantifying, and mitigating such biases remain critical for the reliable adoption of these models in smart city design and data-driven urban modelling.



OVERARCHING AIM

Advance human-centric urban analytics by developing and evaluating methods grounded in human-generated data and oriented towards improving the efficiency and reliability of human-centred urban services

Strand 1

Create precise 3D mapping with wireless signals opportunistically

- *"Bit to Brick: From Cellular Mobile Signals to 3D City Map Creation"*

Strand 2

Support humans in the cities: physical activity detection and spatial disorientation

- *"Why Do Pedestrians Get Lost? Delving into the Factors Behind Navigational Challenges in Urban Environments"*
- *"Advancing Human Activity Recognition Using Ultra-Wideband Channel Impulse Response Snap"*

Strand 3

Evaluate temporal biases in vision data for urban applications

- *"From Static to Dynamic: Evaluating the Impact of Temporal Bias in Historical Street View Images for Cross-View GeoLocalisation"*

Figure 1.1: Overview of the overarching aim of the thesis and its supporting research strands.

Specifically, building on the overarching aim of this thesis, three interrelated research strands are defined to (i) advance 3D mapping through innovative methods, (ii) support urban residents via human-centred urban services, and (iii) identify and assess the impact of data bias in contemporary deep learning models for urban applications. The relationships between these strands and their collective contribution to the overarching aim are illustrated in Figure 1.1. The objetive of each research strand is detial as follows:

- Research Objective 1. The thesis aims to address the challenges of rapidly changing urban environments, where timely data are crucial to support location-based services. To facilitate 3D mapping, it seeks to investigate the feasibility of leveraging new forms of data for 3D mapping. In particular, the research aims to examine the potential of ubiquitous wireless signals as opportunistic data sources for large-scale and cost-effective 3D mapping. In highly dynamic urban settings, this approach intends to deliver spatial information, enabling more responsive and reliable location-based services.

- Research Objective 2. By focusing on humans as active agents within urban environments, my thesis aims to examine how wireless sensing can capture and interpret human physical activities, offering new opportunities for unobtrusive monitoring and improved responsiveness in urban spaces. Additionally, the thesis aims to investigate the interaction between humans and urban environments through the phenomenon of pedestrian disorientation, systematically assessing the environmental, situational, and personal factors that contribute to people getting lost in cities. By integrating these two strands, the thesis tries to offer a human-centred perspective on urban applications, highlighting both how people interact in the city and how the city's complexity influences their navigation.
- Research Objective 3. The thesis aims to reveal inherent biases in urban data and evaluate their impact on deep learning-based models for positioning, taking cross-view geolocalisation as a case study. The objective is to investigate how temporal biases present in benchmark datasets can affect the performance of pre-trained models, ultimately enhancing their robustness and fairness.

By addressing these interconnected challenges, this thesis aims to demonstrate how diverse data streams and novel methodologies can overcome current limitations of urban studies, offering richer, fairer, and more actionable insights for urban modelling and understanding. Additionally, it highlights the risk of data bias in this process, with the ultimate aim of contributing to the development of smarter and more inclusive cities.

1.3 Contributions

Firstly, the thesis demonstrates a standalone approach for 3D mapping that leverages cellular mobile signals as opportunistic data sources. This method offers a free, globally available, and easily accessible alternative to traditional mapping techniques. It enables accurate large-scale building height estimation using ubiquitous cellular signals. Secondly, the thesis contributes to human-centric research on how people interact with their urban environments through two complementary sub-strands of work. The first sub-strand extends wireless sensing from environmental mapping to fine-grained human activity recognition, demonstrating its potential for non-intrusive analysis of human behaviour in urban environments. The second sub-strand provides empirical evidence on the relative influence of personal, environmental, and situational factors on pedestrian disorientation, advancing understanding of why people get lost in complex urban environments. Thirdly, the thesis contributes to the study of data bias within concurrently benchmarked datasets that are extensively used in the era of AI, using the cross-view geolocalisation task as a case study. This contribution demonstrates that temporal bias in benchmark visual datasets can systematically affect the performance and reliability of deep learning-based geolocalisation models. Together, these contributions demonstrate how heterogeneous emerging urban data can be systematically integrated to support scalable, human-centred, and bias-aware urban analytics, with particular relevance to positioning and navigation services.

1.4 Orgnisation of Thesis

The remainder of this thesis is organised into five chapters. The literature review is presented in Chapter 2. The subsequent three chapters, Chapter 3, 4, and 5, comprise my published and under-review research papers, corresponding to the three principal strands of my research: 3D mapping, humans in cities, and temporal bias evaluation, respectively. Finally, Chapter 6 summarises the overall findings of the thesis and outlines potential directions for future research. The details of each chapter are outlined below.

- Chapter 2 provides a literature review that establishes the background for the works included in this thesis. The chapter begins with an overview of new data forms that are currently utilised in the field of intelligent urban systems, as well as technology advancements that underpin the data proliferation. Then, three key strands are reviewed in detail: the use of wireless signals as opportunistic sources for urban environment sensing in Section 2.1, reviewing works for supporting humans in cities through two streams of directions, namely recognising human physical activities in Section 2.2 and pedestrian wayfinding research in Secton 2.3, and the growing role of visual data in positioning in Section 2.4.
- Chapter 3 presents the paper “Bit to brick: from cellular mobile signals to 3D city map creation”. This chapter outlines the proposed approach for 3D mapping using cellular mobile signals, achieving building height estimation accuracy comparable to that of national mapping agencies. By leveraging both in-lab simulated data conducted with MATLAB and real-world data collected via the custom mobile app “BitToBrick”, the proposed approach has demonstrated both financial and computational efficiency. Importantly, it enables effective 3D mapping using ubiquitous consumer-level devices, without the need for specialised or expensive equipment.
- Chapter 4 includes two papers focusing on human in the cities. More specifically, Section 4.1 is the paper “Advancing human activity recognition using Ultra-Wideband channel impulse response snapshots”. The effectiveness of eleven machine learning and deep learning models for the HAR task was evaluated under three different indoor layout scenarios: line-of-sight (LOS), NLOS, and complex LOS, where distracting objects are present along the LOS. The dataset was generated by continuously recording UWB CIR data as four volunteers performed six different types of body movements, encompassing both small- and large-scale activities. This study establishes a benchmark for HAR using UWB CIR, providing methodological insights and a reference dataset for broader effectiveness eval-

uation within the research community. Additionally, Section 4.2 is the paper "Why Do Pedestrians Get Lost? A Case Study of Personal, Situational, and Environmental Factors in Greater London". Drawing on survey responses from individuals who experienced getting lost in the Greater London Area, I systematically mapped their accounts to a set of high-level factors. By applying both data-driven analysis and expert-led weighting through the Analytic Hierarchy Process (AHP), this study evaluates the relative importance of these factors and uncovers the underlying interactions that contribute to pedestrian disorientation in urban settings. The findings offer new insights into the multidimensional nature of urban wayfinding and inform the design of more inclusive and effective navigation aids.

- Chapter 5 is the paper "From Static to Dynamic: Evaluating the Impact of Temporal Bias in Historical Street View Images for Cross-View Geolocalisation", which presents the investigation into the impact of temporal bias in SVIs on the performance of deep learning models for urban applications, with a focus on cross-view geolocalisation. To enable controlled evaluation, a large dataset was constructed that is spatially aligned but temporally diverse, based on the CVUSA validation set. Two state-of-the-art deep learning models, originally developed for the Cross-View GeoLocalisation (CVGL) task and pretrained on CVUSA, were used to perform geolocalisation on custom historical SVIs that share locations with the CVUSA validation set but differ in time. To further interpret the geolocalisation outcome from this custom dataset, both semantic segmentation and model explainability analysis were employed to enhance the explainability of how feature variations contribute to the models' retrieval performance. This study establishes an initial benchmark for evaluating the temporal bias inherent in visual data sources and highlights the need to incorporate both spatial and temporal diversity to enhance the generalisability, reliability, transferability and robustness of deep learning models built on these datasets.
- Chapter 6 summarises the thesis by highlighting the main findings and proposing directions for future research.

Chapter 2

Literature Review

Recent advances in digital sensing and sharing have enabled large-scale, high-frequency urban data streams that complement conventional sources. (Roser 2023; Shin 2009; Zhang and He 2020; Shirky 2010; Darwish and Lakhtaria 2011; Aichner et al. 2021; Wazny 2017; Al Tareq et al. 2024; Elhanashi et al. 2024; Carminati et al. 2021; Zhao et al. 2024; Yang et al. 2022; Liu et al. 2019; Ramírez-Moreno et al. 2021). These developments have fundamentally altered how cities are studied by enabling data-driven analyses at spatial and temporal scales that were previously infeasible. For example, mobile phone records and social media traces have been shown to capture urban mobility patterns at hourly or sub-hourly resolution, far exceeding the temporal granularity of traditional travel surveys (Calabrese et al. 2014; Niu and Silva 2020).

As a result, recent years have seen substantial advances in both data collection and data processing, enabling a more granular and timely understanding of cities: Data now play a central role in the understanding, modelling, and mapping of urban systems, enabling quantitative representations of urban structure, dynamics, and behaviour (Long and Liu 2016; Townsend 2015; Manandhar et al. 2023; Gorjian 2025; Kandt and Batty 2021). For instance, recent studies have demonstrated how multi-source urban data can support fine-grained representations of land use, mobility flows, and functional urban regions at city-wide scales (Wang et al. 2020; Yan et al. 2024). With the growing complexity of urban environments, there has been a marked shift in research focus toward leveraging diverse and heterogeneous data streams that encompass the spatial, temporal, and behavioural aspects of urban systems (Yan et al. 2024; Liu et al. 2020; Wang et al. 2020). Meanwhile, urban environments are changing rapidly, requiring timely data that can capture their continuous and swift transformations (Deng et al.

2009; Yu and Fang 2023). With the accelerated pace and growing frequency of data collection enabled by the proliferation of sensors, mobile phones, and IoT devices, it has become increasingly feasible to harness such sources as a complementary input to traditional datasets to understand and respond to such dynamic changes (Niu and Silva 2020; Calabrese et al. 2014; Yu and Fang 2023; Ang and Seng 2016; Psyllidis 2020). These data sources emerge from two broad trajectories. On one hand, technologies initially designed for dedicated usages, such as GNSS (Langley et al. 2017), street view images (He and Li 2021; Kang et al. 2020), and remote sensing systems (Toth and Jóźków 2016; Zhang and Zhu 2023), are increasingly repurposed for urban applications. On the other hand, mass-market consumer devices, such as mobile phones (Reades et al. 2007) and IoT devices (Merenda et al. 2020; Zanella et al. 2014), generate continuous data streams as byproducts of everyday usage. Together, these streams provide an unprecedented richness and frequency of urban information, creating new opportunities for analysis, modelling, and decision-making across various urban applications (Huang et al. 2019; Lines and Basiri 2021; Li et al. 2018; Yu and Fang 2023; Salih et al. 2025). Second, advances in computation, storage, and device capabilities have transformed how urban data are generated and shared (Alahi et al. 2023; Liu et al. 2025; Rulff et al. 2024).

Accordingly, this chapter provides a literature review on the data and methodological advances that underpin the three core strands of this thesis: 3D mapping, human in cities with particular attention to activity recognition and disorientation, and visual-based geolocalisation. These domains are central to how people and technologies interact with cities. First, 3D maps are indispensable not only for urban planning, environmental monitoring, and navigation applications (Biljecki et al. 2015; Groves 2011a; Kadhim and Mourshed 2017). 3D maps are increasingly recognised as essential for a wide range of urban applications. Studies have shown that incorporating 3D building models can significantly improve GNSS positioning accuracy in dense urban environments by explicitly modelling signal blockage and multipath propagation (Groves 2011a; Wang et al. 2013a). In environmental analysis, the inclusion of 3D morphology has been demonstrated to improve the performance of noise and air pollution dispersion modelling compared to 2D representations (Stoter et al. 2008; Zhang et al. 2022a). Second, this thesis examines the human dimension of urban environments, with a particular focus on human activity recognition and pedestrian disorientation. These two sub-strands share a common aim of better supporting people in cities by examining both their physical behaviours and their navigational challenges. In 3D mapping research, wireless signals have been applied for sensing larger objects such as buildings. HAR further investigates how wireless signals can also be employed at smaller scales to capture human physical activities. By leveraging the ubiquitous wireless signals generated by smart devices, this work demonstrates the potential of non-intrusive and cost-effective approaches for human activity recognition. In parallel, positioning and navigation are essential for daily urban life while remaining challenging in dense urban areas (Vaez et al. 2016). Understanding why people get lost thus provides critical insight for designing more inclusive and accessible navigation systems (Darken and Peterson 2002a). Therefore, the study of “getting lost” explores how pedestrians interact with complex urban environments, identifying the contribution of personal, situational, and environmental factors to disorientation. Third, the proliferation of new forms of data, such as imagery from crowdsourcing initiatives (Wazny 2017; Hecht and Stephens 2014) and social media platforms (Martí et al. 2019), together with advances in infrastructure, hardware, and data processing capabilities (Psyllidis 2020), has created new opportunities for timely and responsive urban applications. At the same time, however, the biases embedded in large-scale datasets remain insufficiently addressed (Liu and He 2024). This thesis takes visual-based geolocalisation as a case study to examine how temporal bias systematically affect the outcomes of learning models, thereby highlighting both the potential and the limitations of such approaches in advancing positioning and navigation research.

The following section reviews the literature in relation to the main strands of this thesis. More specifically, Section 2.1 reviews 3D mapping techniques, with a particular focus on the data of various modalities and approaches leveraged in this field. Section 2.2 then turns to HAR, highlighting the role of sensing data in capturing urban behaviours. Section 2.3 reviews studies on urban navigation and disorientation, reflecting the human-centred challenges of interacting with complex urban spaces. Finally, Section 2.4 examines vision-based geolocalisation. Collectively, these four strands establish the conceptual and methodological foundation for the research contributions that follow.

2.1 Review of 3D mapping

A 3D map is a digital representation of the built environment, typically encompassing buildings, vegetation, and infrastructure (Herman and Řeznik 2015; Ying et al. 2023). In recent years, the role of 3D maps has expanded from visualisation to support a wide range of applications (Biljecki et al. 2015). In environmental studies, 3D maps can provide morphological information that supports noise propagation (Stoter et al. 2008; Chen et al. 2024a) and air pollution dispersion (Zhang et al. 2022a) analysis.

For location-based services, they can improve GNSS positioning accuracy in challenging urban environments by modelling signal propagation and mitigating multipath errors (Groves 2011a; Wang et al. 2013a). In emergency management, 3D maps enable more effective planning of evacuation routes and timely disaster rescue (Kwan and Lee 2005). Additionally, 3D maps are increasingly important for urban management (Bitelli et al. 2018; Skondras et al. 2022), cultural heritage documentation (Marques and Roca 2021; Li et al. 2023; Hu and Minner 2023), and sustainability assessments (Benedetti et al. 2022). More recently, 3D maps have been increasingly demanded with the rise of autonomous vehicles (Ilci and Toth 2020; Munir et al. 2019), drone navigation (Chen and Gao 2019), and immersive VR/AR technologies (Rohil and Ashok 2022), all of which require fine-grained and reliable spatial contextual information. In summary, 3D maps are becoming critical enablers in concurrent urban applications, covering accurate positioning and navigation, facilitating environmental analysis, emergency responses and location-based services. Despite their usefulness, producing and maintaining large-scale 3D maps remains a challenging task. It is seen that both the coverage and completeness of 3D maps show significant differences across countries, with notable gaps still existing even in some of the most developed regions (Biljecki 2020; Milojevic-Dupont et al. 2023; Bernard et al. 2022). Moreover, the financial cost for creating and updating accurate 3D maps remains high, posing particular challenges for less developed regions with limited resources (Wong 2018). Accordingly, it is becoming essential to explore cost-effective methods for 3D maps using available and accessible new forms of data. The emergence of advanced technologies for collecting sensor data has significantly advanced the building and maintenance of 3D maps. Over the past decades, a wide range of sensing modalities has been developed to capture the geometric and semantic characteristics of urban environments (Toschi et al. 2017; Over et al. 2010; Willenborg et al. 2017), encompassing active high-resolution instruments (Luo et al. 2024; Yan and Huang 2022), as well as passively collected signals (Lines and Basiri 2021; Basiri et al. 2023). These sensing technologies differ in terms of spatial coverage, resolution, cost, and ease of deployment, thereby shaping their suitability for city-scale modelling (Blaschke et al. 2011). This section reviews the main categories of sensing data used in 3D mapping, grouped by their acquisition modality and physical principle: active 3D scanning, image-based reconstruction, remote sensing imagery, and signal-based measurements.

3D mapping generated from Light Detection and Ranging (LiDAR) data primarily focuses on reconstructing buildings and other urban objects in the form of point clouds (Wang et al. 2018b), which are generally used for the geometrical reconstruction of urban objects, such as buildings or vegetation (Wang et al. 2019c). The typical processing pipeline for building 3D reconstruction using point clouds begins with, in general, ground filtering and semantic classification to isolate building-related points from vegetation and terrain (Morgan and Habib 2002; Liu et al. 2013; Milioto et al. 2019). Building reconstruction methods can be classified into model-driven and data-driven approaches (Wang et al. 2018b). Model-driven methods rely on approaching a list of pre-defined geometric primitives, such as planes and cuboids, to roof structures and walls (Dorninger and Pfeifer 2008). On the other hand, data-driven methods bypass the geometric assumptions. For instance, Zhang et al. (2024) introduced Point2Building, a learning-based approach that learns to reconstruct polygonal meshes from raw point clouds, even in cases of occlusion or partial data (Liu et al. 2024b). A key application of LiDAR data is the estimation of building heights. The most widely used approach involves generating a digital surface model (DSM) and a digital terrain model (DTM), then computing height as the difference between them. This method is widely adopted in urban studies and environmental analysis (Lab 2020). Alternatively, height can be estimated by calculating statistical measures (e.g., maximum or median elevation) from points within the building footprint. For example, (Park and Guldmann 2019) introduced a method for roof point classification to enhance the reliability of these metrics using machine learning models. Despite their utility, LiDAR-based methods for 3D modelling are still facing several limitations, including occlusions in dense urban areas, relatively high costs due to equipment and on-site surveys, and infrequent data updates (Wang and Menenti 2021; Li and Ibanez-Guzman 2020; Lin et al. 2025). These constraints have stimulated the adoption of data fusion techniques, where LiDAR is combined with other data formats, such as photogrammetry or OpenStreetMap data, to improve model completeness and semantic richness (Bok et al. 2014; Barranquero et al. 2023; Zhang and Lin 2017). In addition, LiDAR-based reconstructions often suffer from data incompleteness in dense urban canyons and require costly acquisition campaigns, which limits update frequency and global scalability (Wang and Menenti 2021; Li and Ibanez-Guzman 2020). These constraints motivate the exploration of alternative or complementary data sources for large-scale 3D mapping.

Remote sensing imagery, particularly from satellite and aerial platforms, provides extensive spatial coverage and frequent revisit cycles, making it a valuable source for large-scale 3D modelling (Zhu et al. 2017). Although raw 2D imagery lacks direct depth information, several indirect techniques have been developed to estimate 3D structures. Stereophotogrammetry leverages multi-angle or multi-temporal image pairs to derive elevation models through feature correspondence (Do and Nguyen 2019), while DSMs generated from stereo pairs or optical flow estimation capture surface elevation with varying degrees of accuracy (Sun and Wang 2018). Additionally, shadow-based methods infer building height from projected shadows under known sun angles, particularly in high-resolution imagery (Kadhim and Mourshed 2017; Dong et al. 2024). The accuracy of these methods depends on factors such as image resolution, viewing geometry, urban density, and scene complexity (Li et al. 2022a). Occlusion and seasonal variability often introduce uncertainties (Pan 2020; Ajayi and Ojima 2022; Yu et al. 2019). Nevertheless, the scalability, availability, and cost-effectiveness of remote sensing imagery make it suitable for preliminary 3D mapping, especially in areas with limited access to LiDAR or ground-based surveys (Qin and Liu 2022; Christodoulides et al. 2025). Recent work has also explored the integration of optical satellite data with SAR (synthetic Aperture Radar), particularly in urban canyons or cloud-prone regions (Koukiou 2024). While satellite imagery alone may not match the precision of ground-based sensing relying on LiDAR, its complementary role in multi-modal fusion frameworks continues to be crucial in building globally scalable 3D maps (Bagheri et al. 2018; Koukiou 2024).

Beyond traditional optical sensing modalities, recent research has begun to explore the use of opportunistically collected wireless signals for 3D modelling, particularly for inferring building height through signal–environment interactions such as shadowing, attenuation, and multipath effects (Esrafilian and Gesbert 2017; Peng et al. 2022; Lines and Basiri 2021; Basiri et al. 2023). These approaches leverage the interaction between radio-frequency signals and urban structures to infer spatial geometry, particularly building height, without relying on explicit visual or laser-based inputs. One common strategy involves modelling signal shadowing effects in GNSS, where the presence of buildings obstructs satellite visibility, creating detectable patterns in received signals (Lines and Basiri 2021; Basiri et al. 2023). Other techniques exploit Received Signal Strength (RSS) gradients or LOS inference using signals from terrestrial cellular towers or low Earth orbit (LEO) satellites (Esrafilian and Gesbert 2017; Peng et al. 2022). These methods enable indirect height estimation by analysing signal attenuation, multipath behaviour, or signal blockage caused by urban morphology. Signal-based data are particularly suitable for large-scale 3D mapping because they rely on already-deployed communication infrastructure and passively collected measurements, significantly reducing data acquisition cost and operational overhead compared to LiDAR or dedicated surveying efforts (Oughton and Frias 2018; Wang and Basiri 2025). The widespread presence of wireless infrastructure in urban environments, combined with the high penetration rate of user-end devices, enables scalable and passive data collection (Oughton and Frias 2018; Ofcom 2020). For example, as demonstrated in Chapter 3, building height estimation using crowdsourced cellular RSS data can achieve metre-level accuracy, relying solely on Android-based smartphones and open-access cellular infrastructure databases (Wang and Basiri 2025). This empirical result provides evidence that signal-based approaches can deliver practically meaningful geometric information at city scale. Although signal-based 3D mapping is still emerging, it offers a promising and underexplored avenue for cost-effective 3D modelling, especially in regions where conventional sensing is inaccessible or prohibitively expensive. Furthermore, the potential integration with other data modalities, such as visual data, could further enhance the completeness and reliability of hybrid modelling pipelines.

2.2 Review of Human Activity Recognition

HAR aims to recognise human physical activities and body movements using data that reflect or capture human motion patterns (Yin et al. 2024; Kumar et al. 2024; Kaseris et al. 2024). HAR plays an important role in contemporary society, with demonstrated applications in healthcare monitoring, fall detection, and activity-aware assistance systems (Kumar et al. 2024; Lentzas and Vrakas 2020). In urban contexts, HAR has been increasingly used to analyse mobility behaviours, pedestrian dynamics, and the usage of public infrastructure, offering behavioural insights that are difficult to capture through conventional survey-based approaches (Javed et al. 2021; Hu et al. 2023). In recent years, HAR has also contributed to urban analytics, for applications including mobility analytics, crowd behaviour modelling, and infrastructure usage assessment (Javed et al. 2021; Hu et al. 2023).

From a sensing perspective, the data used for HAR tasks can be broadly categorised into three families: sensor-based, vision-based, and device-free approaches (Wang and Basiri 2024). Sensor-based HAR relies on data collected from sensors that are worn or carried by users (Yin et al. 2024). These sensors are often embedded in smartphones, smartwatches, or wearable devices and include accelerometers, gyroscopes, magnetometers, and inertial measurement units (IMUs) (Ramanujam et al. 2021; Webber and Rojas 2021). The data typically captures motion dynamics or orientation changes of the user, which are then fed into supervised or unsupervised learning models for activity classification (Zhang et al. 2022b). Although sensor-based approaches provide high accuracy in controlled settings, they require user consent and continuous device attachment, which may not be feasible in many passive monitoring contexts (Wang et al. 2016). Vision-based HAR utilises visual inputs—such as images or videos—captured by cameras to detect and classify human activities. These methods benefit from rich spatial and temporal information, enabling detailed analysis of posture or gesture (Mahbub and Ahad 2022; Holte et al. 2012; Pareek and Thakkar 2021). Large-scale datasets, such as Kinetics (Kay et al. 2017) and NTU RGB+D (Shahroudy et al. 2016), have significantly advanced the training of deep vision-based HAR models. However, vision-based approaches often raise privacy concerns, necessitating additional methods for preserving privacy while performing such tasks. (Pareek and Thakkar 2021; Cui et al. 2024; Ahuja et al. 2021). In addition, vision-based HAR may also suffer from environmental constraints such as lighting conditions, occlusion, or camera placement (Pareek and Thakkar 2021; Thi et al. 2010). Device-free HAR has recently gained traction due to its non-intrusive nature and suitability for passive sensing, which leverages ambient signals in the environment, such as radio-frequency or acoustic signals, to infer human activities without requiring users to wear or carry any devices (Yang et al. 2023a; Moghaddam et al. 2025; Mohtadifar et al. 2022). Common data sources include WiFi Channel State Information (CSI), UWB CIR, and cellular RSS (Moshiri et al. 2021; Wang and Basiri 2024; Lafontaine et al. 2023; Bibbò et al. 2022; Sarsodia et al. 2022). These methods interpret variations in signal propagation, caused by human movement or presence, using signal processing and machine learning techniques. Acoustic sensing, using microphones or ultrasonic transceivers, also falls into this category, offering another passive modality for activity inference (Stuchbury-Wass et al. 2022; Hu et al. 2025). While each sensing data modality presents distinct advantages and limitations, their integration is increasingly explored to enhance robustness and coverage across diverse HAR scenarios (Zhou et al. 2023; Chung et al. 2019; Vidya and Sasikumar 2022; Webber and Rojas 2021).

These diverse data modalities form the basis of HAR, and their full potential is realised when coupled with methodological frameworks designed to extract and classify activity-related patterns. For both device-free and sensor-based approaches, human activities are generally inferred from the measurable effects they produce on recorded signals (Vrigkas et al. 2015). In device-free methods, wireless signal transmissions can be reflected, diffracted, or scattered by the human body's large-scale movements (e.g., walking) or small-scale motions (e.g., breathing). These interactions induce variations in the wireless channel characteristics, which can be mapped to specific activities (Hussain et al. 2020). Similarly, in sensor-based methods, physical activities are captured through readings from sensors carried by the subject (Chen et al. 2021). A typical example is the tracking of daily activities, such as running, walking, or step counting, which can now be derived through the integration of IMUs, accelerometers, and gyroscopes in smartphones and smartwatches. These readings are processed by algorithms that transform raw sensor data into interpretable metrics for end-users (Yin et al. 2024). In both categories, signals from wireless transmissions or body-mounted and wearable sensors are collected while participants perform activities. Due to the high volume of collected data and the subtle nature of activity-related patterns, traditional statistical approaches often fail to deliver accurate recognition performance (Yang et al. 2023a). Consequently, recent HAR research has increasingly adopted deep learning models, which have demonstrated strong capabilities in automatically learning representative features from raw signals. In typical deep learning-based HAR pipelines, the model architecture consists of a feature extractor that transforms raw input into numerical embeddings, followed by a classifier that maps each embedding to a corresponding activity label (Yang et al. 2023a). Classifiers are commonly implemented as fully connected layers, which have proven effective for activity classification (De Leonardis et al. 2018). While classifier designs are relatively straightforward, the choice of feature extractor remains a central focus of research. A number of deep learning architectures have been employed as feature extractors in device-free and sensor-based HAR tasks, including multi-layer perceptrons (MLPs) (Rustam et al. 2020; Geravesh and Rupapara 2023; Wan et al. 2020), CNNs (Münzner et al. 2017; Yang et al. 2015), recurrent neural networks (RNNs) (Pienaar and Malekian 2019; Inoue et al. 2018), and transformer-based models (Dirgová Luptáková et al. 2022; Shavit and Klein 2021). Variants and hybrid architectures that combine the strengths of multiple paradigms, such as CNN–LSTM hybrids or two-stream networks, have also been proposed to improve feature representation and recognition accuracy (Nadia et al. 2023; Xia et al. 2020; Mutegeki and Han 2020; Li et al. 2021).

Vision-based methods infer human activities from visual data captured by RGB cameras. Early approaches relied on handcrafted visual features such as Histograms of Oriented Gradients (HOG) (Dalal and Triggs 2005), space-time features (Laptev et al. 2008), and space-time interest points (Dollár et al. 2005). The extracted features are then fed into classifiers, such as support vector machines, to perform HAR tasks. In recent years, with the rise of deep learning, models such as CNNs have become dominant for extracting spatial features from images or video frames (Karpathy et al. 2014; Simonyan and Zisserman 2014). To jointly capture spatial and temporal features, 3D CNNs, such as C3D (Tran et al. 2015), process video as spatiotemporal volumes. Two-stream CNN architectures further incorporate motion information by modelling appearance and optical flow in separate streams and fusing their predictions (Feichtenhofer et al. 2016). Pose-based methods estimate skeletal keypoints from frames (Cao et al. 2017) and then apply temporal models, such as graph convolutional networks (GCNs), to model motion patterns in the skeletal domain (Yan et al. 2018). Vision-based approaches thus provide rich spatial and temporal cues for recognising human activities, enabling both fine-grained motion analysis and large-scale video-based HAR applications.

2.3 Review of Urban Navigation and Disorientation

In this thesis, the concept of getting lost has been used to describe a temporary or prolonged breakdown in wayfinding, where individuals become uncertain about their current location or intended route (Farr et al. 2012; Hunter et al. 2016a; Darken and Peterson 2002a; Gath-Morad et al. 2022). As outlined in Section 4.2, this thesis adopts a broad definition of "getting lost", encompassing both minor wayfinding errors and more severe disorientation events.

Within this broad framing, the research community has addressed urban wayfinding from multiple perspectives. One major line of work investigates individual differences in wayfinding ability, considering factors such as age, gender, educational background, cultural background, and access to navigation technologies (Davies and Pederson 2001; Farr et al. 2012; Liu et al. 2011). Complementing these studies, Vaez et al. (2016) reviews research examining how urban form, including both natural landscapes and man-made features, affects people's spatial cognition. This body of work often focuses on how environmental variables such as road layout, signage, visual accessibility, and location-specific characteristics influence the formation of "cognitive maps". These maps reflect the process through which individuals build, store, remember, and decode spatial knowledge to facilitate navigation along their itineraries (Rapoport 2013; Lynch 2023; Montello 2015; Golledge 1999). In parallel, advances in navigation and positioning technologies have transformed the tools available for acquiring spatial knowledge. The increasing integration of GPS-enabled devices, digital maps, and location-based services has been widely investigated by the research community into their performance, reliability, and accessibility in supporting wayfinding (Baldwin 2003; Gupta et al. 2020; Prandi et al. 2023). Another significant methodological direction focuses on enhancing wayfinding effectiveness and reducing the likelihood of disorientation through urban route optimisation (Huang et al. 2017; Koletsis et al. 2017; Tyagi et al. 2022).

In summary, urban navigation and disorientation result from diverse and interactive domains, including the influence of urban form, individual characteristics, route planning, and the effectiveness of technological aids (Gath-Morad et al. 2022; Hölscher et al. 2012). Consequently, the methodological approaches employed in the literature are highly heterogeneous, reflecting the multifaceted nature of the phenomenon. Understanding these interacting factors is essential for the design of inclusive navigation systems, as empirical studies have shown that wayfinding difficulties disproportionately affect elderly users, visitors, and individuals with cognitive or perceptual impairments (Farr et al. 2012; Prandi et al. 2023).

2.4 Review of Visual-based Geolocalisation

Urban positioning via visual information has rapidly emerged as a significant research direction (Durgam et al. 2024; Wilson et al. 2021). This research area is enabled by the technological advances and the widespread adoption of smart devices which have created unprecedented opportunities to employ new forms of data for urban applications (Glaeser et al. 2018; Paiva et al. 2021). Among these, images stand out as one of the most prominent and widely utilised data modalities (Zhang et al. 2024; He and Li 2021). They are continuously generated by commercial platforms (Google Maps 2025; Mapillary 2025; KartaView Contributors 2025), as well as through crowdsourcing initiatives and social media contributions (Wazny 2017; Hecht and Stephens 2014; Martí et al. 2019), and have become the backbone for constructing a variety of benchmark datasets serving different research purposes. Enabled by advances in AI and deep learning, this field leverages spatially diverse image datasets to develop and evaluate geolocation methods (Zhu et al. 2021; Workman et al. 2015; Liu and Li 2019a). However, while the volume and diversity of visual data have expanded considerably, recent studies have identified substantial spatial and temporal biases in widely used datasets, including uneven geographic coverage and inconsistent image update cycles, which can lead to systematic performance degradation when models are applied to unseen regions or time periods (Zhao et al. 2025; Yue 2025; Fan et al. 2025).

In urban environments, the demand for reliable geolocation and navigation systems is ubiquitous (Karimi et al. 2013; Delikostidis et al. 2016). These services are typically realised via user-end devices, such as smartphones or smartwatches, that receive and process signals from GNSS (Wang et al. 2022b; Hsu et al. 2015). Although GNSS provides the backbone for most location-based services and remains indispensable for global positioning, its performance can be challenged in dense urban areas where signal blockage, multipath effects, and NLOS propagation are common (Zhu et al. 2018; Zidan et al. 2020; Groves 2011a). To complement GNSS under such conditions, recent research has explored alternative or supplementary data modalities, among which vision-based geolocalisation has shown particular promise. By exploiting visual information from provided images, these methods can enhance positioning in areas where satellite signals alone may be insufficient (Yang and Soloviev 2020; Hrabar and Sukhatme 2009; Ben-Afia et al. 2014). Among these, ground-level imagery has emerged as a promising alternative with considerable potential in urban localisations (Lu et al. 2021). Visual-based geolocalisation leverages image features, including street layouts, façades, and intersections, to infer the location of a query photo, either directly through learned image-to-coordinate mappings or indirectly by retrieving the most visually similar image from a geo-referenced database (Castaldo et al. 2015; Masone and Caputo 2021; Durgam et al. 2024).

A wide variety of image sources have been explored for this purpose, including commercial street view platforms (Weyand et al. 2020), images captured from vehicles (Warburg et al. 2020), and photos taken by mobile phones or wearable cameras (Crooks and See 2022; Deuser et al. 2025). More recently, crowdsourced street view imagery has gained traction through Volunteered Geographic Information (VGI) (Flanagin and Metzger 2008; Goodchild 2007) initiatives such as Mapillary (Mapillary 2025) and KartaView (KartaView Contributors 2025), which allow users to upload geo-tagged images captured at diverse locations (Neuhold et al. 2017). These platforms provide an alternative to proprietary services, enabling more frequent and locally varied data collection (Juhász and Hochmair 2016). However, the quality and coverage of SVIs present several challenges. First, the spatial completeness of SVIs varies significantly across regions, with noticeable disparities between urban and rural areas (Yue 2025; Zhang et al. 2025). Second, not all SVIs contain distinctive or stable visual features, which limits their utility for localisation (Rodrigues and Tani 2023). Third, update frequency is inconsistent, leading to temporal mismatches between image content and the current physical environment, which is an issue especially evident in regions undergoing rapid redevelopment or post-disaster reconstruction (Kang et al. 2020; Cândido et al. 2018).

Beyond traditional SVI-to-coordinate geolocalisation, a major advancement in this domain is CVGL, which seeks to match a ground-level image with a corresponding geo-referenced aerial or satellite image (Durgam et al. 2024). This task is inherently more challenging due to the drastic viewpoint differences (Shi and Li 2022). CVGL methods typically rely on deep learning models trained to learn joint embedding spaces, where matching cross-view pairs are pulled closer together and non-matching pairs are pushed apart (Shi et al. 2019; Rodrigues and Tani 2023; Zhu et al. 2021; Zhu et al. 2022). To support such research, a number of benchmark datasets have been developed, including

CVUSA (Workman et al. 2015; Zhai et al. 2017), CVACT (Liu and Li 2019a), Vo (Vo and Hays 2016), and Vigor (Zhu et al. 2021). These datasets vary in terms of street-view field of view, aerial image resolution, and geographical coverage, and serve as the foundation for developing and evaluating CVGL algorithms.

In the research field of CVGL, computer vision and deep learning techniques are widely employed to match images captured from drastically different viewpoints (Durgam et al. 2024). A typical CVGL pipeline aims to retrieve a geo-referenced aerial image corresponding to a query ground-view image, most commonly a street-view image. The large viewpoint gap between aerial and ground perspectives introduces substantial challenges in feature alignment and similarity measurement (Zhu et al. 2021). Consequently, the primary objective of CVGL methodologies is to learn a joint embedding space in which geographically matched ground–aerial pairs are projected close together, while unmatched pairs are pushed far apart. This means the model learns a common feature representation in which images of the same place look similar to the computer, even though they appear very different to the human eye. To achieve this, many CVGL methods adopt a siamese or two-branch network architecture (Bromley et al. 1993), where each branch acts as an encoder to extract features from one modality, either ground or aerial. The feature vectors are then compared in the embedding space using a distance metric, with training guided by metric learning losses such as contrastive loss or triplet loss (Hu and Lee 2020; Cai et al. 2019; Zhang et al. 2023). Deep CVGL models, such as Cross-View Matching Network (CVM-Net) (Hu and Lee 2020), employed CNNs as a local feature extractor to learn modality-specific features and then applied NetVLAD pooling (Arandjelovic et al. 2016) as global descriptor generator. Other works have explored attention mechanisms (Zhu et al. 2022), multi-scale feature fusion (Shi et al. 2019), and multi-modality feature fusion (Rodrigues and Tani 2023) to improve cross-view alignment. Both CNN- and Transformer-based architectures have attracted significant attention in CVGL due to their effectiveness for feature aggregation and spatial reasoning. This thesis adopts representative models from both families (Shi et al. 2019; Zhu et al. 2022) in order to examine how architectural choices influence model behaviour under spatial and temporal data biases.

In addition to architectural innovations, CVGL research has explored data augmentation and training strategies to improve model robustness. Viewpoint simulation, in which ground-view images are synthetically transformed to match aerial viewpoints, has been used to reduce the domain gap between modalities (Regmi and Borji 2018). Ground-view rotation augmentation, where images are rotated during training to account for unknown camera orientations, has been applied in several works, including hard exemplar reweighting approaches (Cai et al. 2019) and spatial-aware feature aggregation networks (Shi et al. 2019). Multi-task learning has also been employed to incorporate auxiliary objectives that provide complementary spatial or semantic cues. For example, Zhai et al. (2017) predicted semantic layouts from aerial imagery as an auxiliary task to enhance cross-view matching; Shi et al. (2020) jointly optimised cross-view retrieval and height estimation to embed richer spatial context into the learned features; Rodrigues and Tani (2023) introduced both semantic segmentation and pixel-wise keyword embeddings to exclusively original images for CVGL tasks.

In summary, visual data—particularly street-level images and their aerial-view counterparts—play a central role in both direct coordinate inference from SVIs and cross-view matching approaches. The growing availability of benchmark datasets has greatly enabled and accelerated research in this field, allowing the development of vision-based urban positioning systems. On the methodological side, most CVGL approaches leveraged siamese or dual-branch network architectures to learn a joint embedding space that aligns aerial and ground-view imagery. Advances have been driven by innovations in feature extraction modules (e.g., attention mechanisms, multi-scale fusion, and transformer-based designs) as well as by training strategies such as viewpoint simulation, rotation augmentation, and auxiliary tasks like semantic segmentation. These developments have enhanced the ability to match images across significant viewpoint differences, allowing for the extraction of location information from them. However, the potential risks of biases embedded in benchmark datasets remain underexplored, posing challenges to the performance and robustness of models in real-world applications. Addressing these limitations, this thesis investigates the effect of temporal biases in benchmark datasets on deep learning models, with the aim of supporting more reliable and robust vision-based geolocalisation for urban location-based services.

Chapter 3

Bit to Brick: From Cellular Mobile Signals to 3D City Map Creation

Publication and Licence Statement

This chapter is based on the following open access publication:

Yu Wang and Ana Basiri (2025). *Bit to Brick: From Cellular Mobile Signals to 3D City Map Creation. Big Earth Data.*

Abstract

3D city maps play an essential role in various applications, including emergency services, environmental impact assessment, urban planning, and location-based service. Existing approaches for building height estimation largely rely on expensive technologies such as LiDAR or aerial photogrammetry. This paper proposes a novel approach for building height estimation leveraging ubiquitous cellular network signals. We collected cellular data using the BitToBrick app in six areas to measure buildings with different types and roof shapes. The collected received signal strength (RSS) and 2D footprint of the buildings are used to estimate building heights through unsupervised learning methods. The results show that five out of six buildings had errors under 2.39 m, with the smallest error being 0.02 m and the largest error being 4.68 m. This is comparable with the accuracy of Great Britain's national mapping agencies. Our method has several advantages: it uses existing cellular infrastructure without specialised equipment, relies on stationary antennas to reduce computational overhead, and employs a lightweight mobile app to access mobile signals for building height estimation. These benefits showcase its effectiveness and scalability for this purpose.

3.1 Introduction

3D city maps are digital representations of urban areas that illustrate elements such as buildings, vegetation, and infrastructure in a 3D format. They are increasingly vital across a range of domains (Biljecki et al. 2015; Ying et al. 2023). In environmental studies, 3D maps offer detailed morphological information that facilitates the analysis of noise propagation (Stoter et al. 2008; Chen et al. 2024a) and air pollution dispersion across urban areas (Zhang et al. 2022a). For location-based services, 3D city models enhance GNSS positioning accuracy in complex environments such as urban canyons (Groves 2011a). In the context of emergency management, such maps support the planning of evacuation routes and the development of effective rescue strategies (Kwan and Lee 2005).

Building height information is fundamental to accurate 3D city mapping, yet its availability remains highly inconsistent, even across developed regions. For example, OpenStreetMap (OSM), a widely used open-source mapping platform, contains height data for only 0.1% of buildings in Paris, with just over half having level tags (Bernard et al. 2022). Similarly, the EUBUCCO v0.1 dataset (Milojevic-Dupont et al. 2023), which includes nearly 202 million buildings across 27 EU countries and Switzerland, shows that while the average height coverage is 73%, 12 countries report single-digit percentages and over half have less than 50% completeness. These statistics underscore the fragmented and uneven documentation of building heights, even in well-resourced urban areas. The challenge is likely to be more acute in developing regions, where institutional support and consistent data collection are often lacking (Biljecki 2020). Moreover, producing and maintaining national-scale 3D maps is economically demanding—costs for creating LOD2 maps in countries like the UK can reach up to £75 million (Wong 2018)—with maintenance expenses likely even higher. These limitations point to a critical need for alternative approaches: methods that are low-cost, scalable, and suitable for crowdsourced or widely accessible data collection, offering a viable path toward broader and more equitable 3D city map generation.

Several established methods are available for estimating building heights, serving as a foundation for 3D city mapping. These methods include total station surveying, Light Detection and Ranging (LiDAR), Synthetic Aperture Radar (SAR) technologies, remote sensing imagery, learning-based models, and the use of wireless signals. (Mill et al. 2013) used terrestrial laser scanning and total station surveying to detect façade damage and produce accurate building models. In previous studies (Baltsavias 1999; Li et al. 2020; Dowman 2004), detailed 3D maps were constructed through LiDAR technology, a scanning system that measures distances from a sensor placed on an airborne platform. Additionally, airborne or spaceborne SAR serves as another technique by estimating building heights from radar imagery (Brunner et al. 2009). These methods expand data modalities for estimating building heights, but the specific equipment requirements limit their scalability for broader applications. High-resolution remote sensing imagery captured by satellites is commonly used in height estimation. Comber et al. 2012 conducted approximate estimates of the number of stories in buildings by analysing the cast shadows captured through high-resolution imagery. Incorporating both building shadows, footprints and solar data, Kadhim and Mourshed 2017 estimated building heights from very high-resolution multispectral images. Recent advances have explored shadow-based height estimation techniques using high-resolution satellite imagery and deep learning methods, which demonstrate promising results in urban settings through the use of building shadows and semantic segmentation (Yan et al. 2023; Chen et al. 2024b). While such methods benefit from detailed visual cues, they are inherently sensitive to environmental and geometric conditions—including sun angle, cloud cover, image occlusions and resolution, and seasonal variation—which can limit their robustness in diverse urban contexts. Despite these advances and the effectiveness of remote sensing for building height estimation, this method is still highly dependent on the quality of the image (Muhmad Kamarulzaman et al. 2023), which could be compromised by weather-related interference such as fog, rainfall or cloud occlusions (Wang and Li 2019). Recent research has focused on estimating building heights based on how buildings interact with environmental signal transmission. More specifically, this method typically relies on classifying the propagation conditions between the signal source and ground users into two categories: line-of-sight (LOS) and non-line-of-sight (NLOS). This approach can be implemented using Unmanned Aerial Vehicles (UAVs) (Esrafilian and Gesbert 2017), Low Earth Orbit (LEO) satellite (Peng et al. 2022), or GNSS satellites (Lines and Basiri 2021) in relation to ground users. Esrafilian and Gesbert 2017 utilised low-altitude UAV-based radio measurements to reconstruct a city map through simulation, achieving less than 0.4 normalised mean square error (NMSE). Using UAVs as signal transmitters, however, raises concerns about battery life, scalability in urban settings, and privacy issues. Lines and Basiri 2021 proposed a method for height estimation using GNSS signals from mobile phones, reducing the height estimation error to less than 5 metres. Basiri et al. 2023 then advanced this approach by utilising edge detection techniques to generate 3D city maps directly from GNSS signals, bypassing the need to classify transmission paths. Peng et al. 2022 introduced a 3D map reconstruction method using LEO communication satellites that have a larger quantity and spatial diversity compared to GNSS. In their simulation results, a height reconstruction NMSE of 0.09 was achieved. Despite the diversity and high accuracy of existing approaches, they are prohibitively expensive or require specialised equipment and infrastructure, resulting in high operational costs and restricted applicability in resource-constrained settings. High-resolution optical imagery, while widely used, suffers from sensitivity to environmental conditions, such as cloud cover and lighting variations, which can compromise data reliability. Signal-based methods are innovative but typically use movable transmitters (e.g., UAVs), requiring complex tracking.

This paper aims to address these limitations of existing methods by proposing a novel, low-cost method for estimating building heights using cellular signal propagation data. Leveraging the widespread infrastructure of mobile networks and the static nature of cellular antennas, our approach offers three distinct advantages: First, the presence of shadow zones caused by cellular signals can indirectly reveal relative building heights. Second, mobile networks cover the majority of urban and rural areas, enabling high geographic scalability (Ofcom 2024). Finally, the fixed position of transmitters simplifies the computational pipeline compared to methods reliant on moving platforms. We demonstrate the feasibility of our approach through both simulation and real-world data collected via a custom Android application in dense urban settings. The resulting accuracy, combined with minimal hardware and cost requirements, highlights the method's potential for scalable 3D city map construction and urban analytics.

3.2 Methods

This section details the proposed methodologies for building height estimation, beginning with a proof-of-concept study to estimate building heights using the cellular signal coverage map in MATLAB simulation, followed by a real-world experiment involving data collection through our custom Android application. The subsequent data processing methodologies employed for height estimation are also explained.

In both simulated and real-world experiments, the 2D building footprints are crucial for height estimation. Different strategies were employed for extracting these building 2D footprints for height estimation. In the simulation, we first performed Received Signal Strength (RSS) sampling to generate a coverage map, then upsampled it into a high spatial distribution density, similar to a crowdsourced platform where participants can ideally contribute from any available location. After applying K-means clustering (Lloyd 1982) on RSS samples, we extracted the coordinated boundaries between clusters and derived the building's footprint from them, independent of third-party building footprints. In the real-world experiment, data collected via our dedicated Android app provided a much lower sampling density than the simulation. Therefore, we utilised a third-party building footprint as an additional resource for height estimation.

It is worth noting that the proposed method does not depend on modelling RSS as a continuous function of distance or elevation. Rather, the approach leverages the significant RSS drop typically observed when signal paths are obstructed by buildings. Since the receiver locations are placed within a limited spatial area around each building, the variation in signal strength due to distance is negligible. This design allows the method to detect LOS-NLOS transitions more robustly without requiring assumptions about signal propagation models.

3.2.1 Height estimation from simulated data

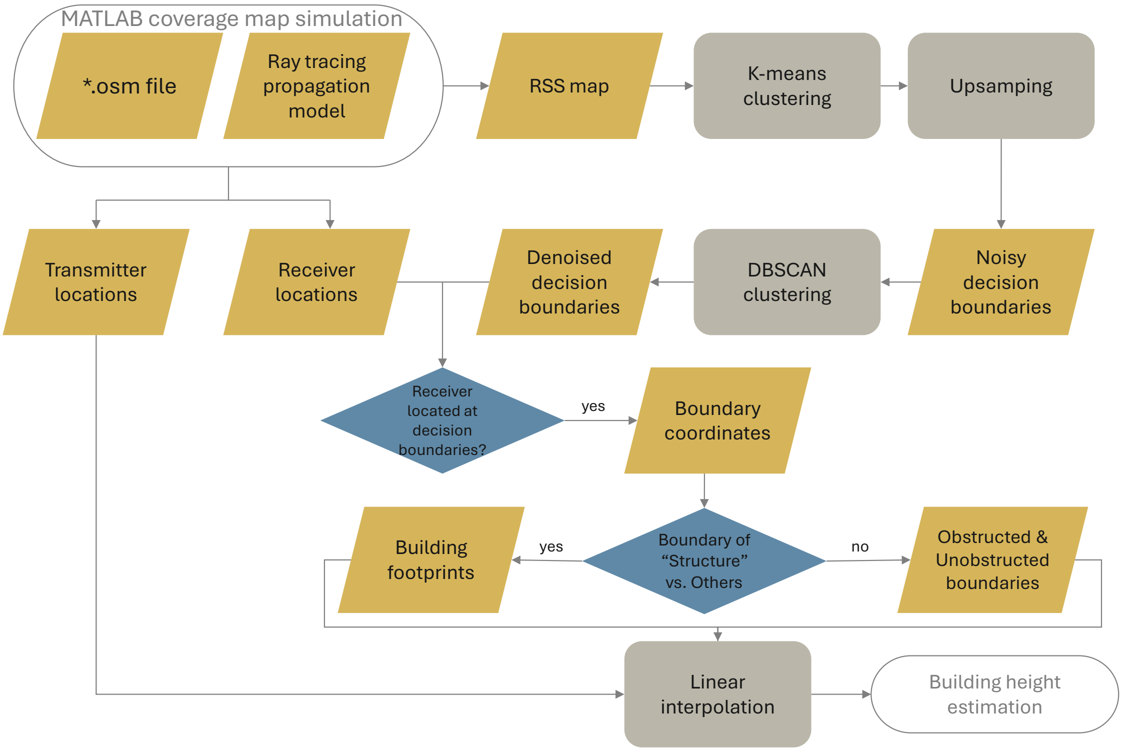

MATLAB coverage map simulation
*.osm file
Ray tracing propagation model
RSS map
K-means clustering
Upsamping
Transmitter locations
Receiver locations
Denoised decision boundaries
DBSCAN clustering
Noisy decision boundaries
Receiver located at decision boundaries?
yes
Boundary coordinates
Building footprints
yes
Boundary of "Structure" vs. Others
no
Obstructed & Unobstructed boundaries
Linear interpolation
Building height estimation

Figure 3.1: Overview of the proposed workflow for building height estimation using simulation. The pipeline starts from building footprints extracted from OpenStreetMap and a ray-tracing-based propagation model to simulate received signal strength maps. Clustering and denoising steps are applied to identify decision boundaries between obstructed and unobstructed signal regions. Receiver measurements located near these boundaries are then associated with building structures, and linear interpolation is used to estimate building height.

The proposed methodology for assessing the feasibility of height estimation through simulation consists of four phases, namely coverage map simulation, RSS clustering, boundary detection, and height estimation. Figure 3.1 provides a schematic representation of each phase.

Coverage map simulation : A cellular signal coverage map is generated using MATLAB by integrating data describing the selected urban area, which is extracted from an OSM file. The resulting RSS map provides a detailed representation of signal strength distribution across the defined area. See “MATLAB coverage map simulation” step in Figure 3.1.

RSS clustering : The RSS samples are categorised into three distinct classes using unsupervised K-means clustering: (1) LOS samples with high signal strength, (2) NLOS samples blocked by buildings, and (3) a structural cluster representing the building footprint area with no RSS data. This process, from the RSS map output generated by simulation to the identification of noisy decision boundaries, is illustrated in Figure 3.1.

Boundary detection : To refine cluster boundaries, upsampling is applied to the clustered RSS samples. Subsequently, smoothed decision boundaries are further clustered by Density-Based Spatial Clustering of Applications with Noise (DBSCAN) (Ester et al. 1996) for denoising and preserving the informative decision boundaries for height estimation. In Figure 3.1, the output of this step is shown as “Boundary Coordinates”, derived from both “Receiver locations” and “Denoised decision boundaries”.

Height estimation : The identified boundaries, combined with the fixed transmitter location and the locations of sampled receivers (Rx), are used to estimate the height of the target building via a proposed linear interpolation technique. This step is represented in Figure 3.1 by the “Linear Interpolation” block, which leads to the “Final Height Estimation” output.

The decision boundary between RSS clusters, determined by the farthest shadow cast by the target building, implicitly contains building height information. Considering the linear propagation of signal transmission, line segments joining the transmitter antenna location and each boundary-located receiver can be viewed as signal transmission pathways. Consequently, intersections between these line segments and the 2D building outline can be identified on the latitude/longitude plane, as shown in Figure 3.2. Therefore, we can deduce that for any arbitrary boundary-located receiver, the intersection point closest to it contributes to the generation of that boundary point.

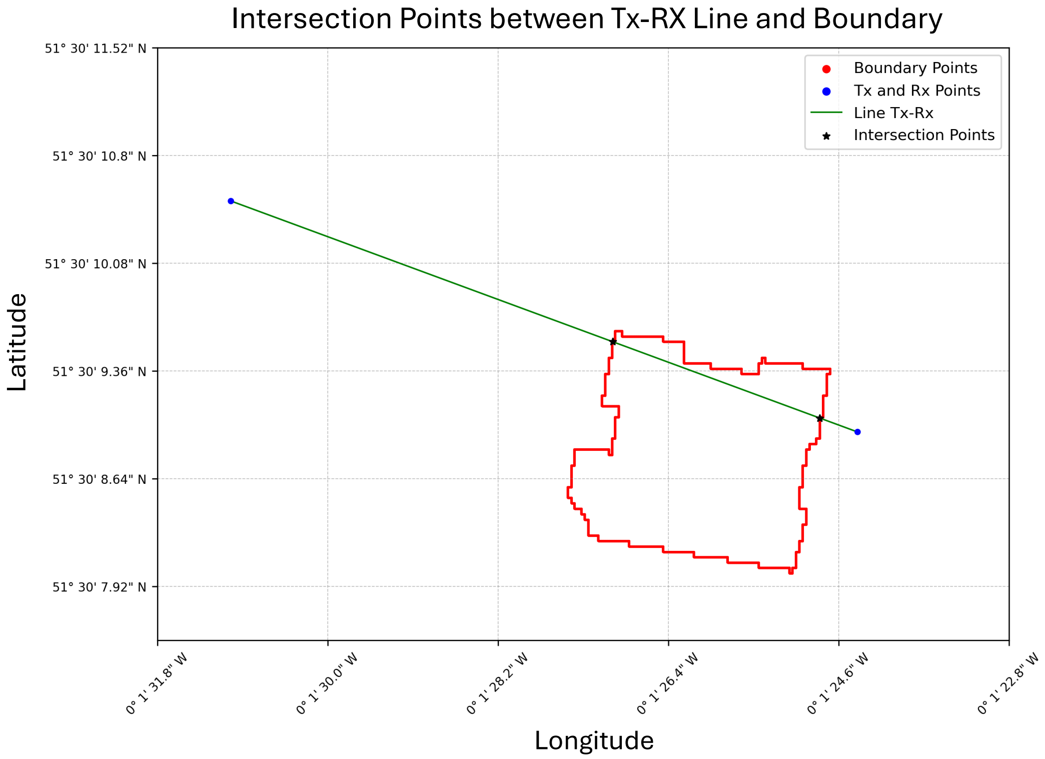


Figure 3.2: Intersection points between the line connecting Tx and Rx and the building footprint. Among all the detected intersection points, the nearest one to the Rx is considered to be the most informative for height estimation.

The process of identifying this intersection point can be formally defined by Equation 3.1:

$$p = \arg\min_{p_j, j \in J} \| p_{rx}^{(i)} - p_j^{(i)} \|_2 \tag{3.1}$$

where $p_{rx}^{(i)}$ is the i-th receiver location, J is the set incorporating all the possible intersection points with the building outline of $p_{rx}^{(i)}$, $p_j^{(i)}$ is the j-th intersection point and $\|\cdot\|$ is the norm-2 distance. Given the short distances involved in signal propagation, the Earth's curvature was ignored, thus permitting the use of Euclidean distance seems to be a reasonable approximation. It is also worth noting that this linear interpolation operates in 3D space, and remains valid even in areas with non-flat terrain, as long as the heights of the transmitter and receiver are referenced to a consistent vertical datum (e.g., above sea level).

The intersection point's height for a given $rx^{(i)}$ location, $h_{p^{(i)}}$, is estimated using linear interpolation, as illustrated in Equation 3.2:

$$h_{p^{(i)}} = h_{Tx} + (h_{Rx^{(i)}} - h_{Tx}) \cdot \frac{d_{p^{(i)}Tx}}{d_{TxRx^{(i)}}} \tag{3.2}$$

where $h_{p^{(i)}}$, h_{Tx}, and $h_{Rx^{(i)}}$ signify the heights of intersection point $p^{(i)}$, transmitter, and receiver elevations, while $d_{p^{(i)}Tx}$ and d_{TxRx} denote distances from the target to the transmitter and from the transmitter to the receiver on the latitude/longitude plane, respectively.

3.2.2 Height estimation from real-world data

Simulation provides an approach for validating the height estimation methodology and can achieve high-density RSS sampling that emulates the potential performance of future crowdsourced platforms. To verify the effectiveness of the proposed approach in real-world settings, we conducted a field experiment to collect relevant data in a dense urban area of London, United Kingdom as mapped in Figure 3.11. We performed a sequential double-level clustering operation of the collected data. In the first level of clustering, we employed DBSCAN to group the collected data points according to their distance-based proximity. This stage allowed us to organise the data into distinct clusters, each corresponding to a specific target building. After establishing these initial clusters, we conducted a second level of K-means clustering to identify two clusters within each group. This second step focused on the RSS levels, enabling us to further subdivide the data points within each building's cluster and find the decision boundaries.

Compared to the simulated data, the real-world data is more sparse. This makes it challenging to use the same upsampling technique for reconstructing building 2D footprints. Therefore, we used third-party 2D building footprints from Microsoft (Bing Maps 2025) to define the outlines of the target buildings. This data was used due to its ready availability. However, other sources such as OSM could be similarly employed, and our method is not tied to a specific footprint provider. This approach allowed us to reduce the number of K-means clusters from three in the simulation to two in the real-world setup, eliminating the need for an additional cluster to perform this task. We approximated the boundary-located receiver by leveraging the central point between two K-means cluster centres. The following equation can be used for height estimation:

$$h_p = h_{Tx} + (h_{Rx\_centre} - h_{Tx}) \cdot \frac{d_{pTx}}{d_{TxRx\_centre}} \tag{3.3}$$

In this equation, h_p represents the height of the intersection point between the signal transmission path defined by Tx and Rx_centre, and the target building's 2D outline. h_{Tx} and h_{Rx_centre} are the elevations of the transmitter and the detected centre point of clusters, respectively. d_{pTx} denotes the distance from the intersection point to the transmitter, and d_{TxRx_centre} signifies the distance from the transmitter to the centre point of the clusters.

3.3 Implementation

In this section, we present the detailed implementation for both the simulation and the real-world use case. Section 3.3.1 outlines the generation of the simulated cellular RSS map, with subsequent clustering of RSS samples and boundary detection methodologies discussed in Sections 3.3.1.1, 3.3.1.2, and 3.3.1.3, respectively. For the real-world use case, we begin with an introduction to the dedicated Android application in Section 3.3.2, followed by a detailed explanation of the data collection process in Section 3.3.2.1. The collected data is then processed using a double-level clustering approach, as described in Section 3.3.2.2, from which decision boundary points are derived to assist in height estimation.

3.3.1 Simulation

3.3.1.1 Cellular coverage map

We employed MATLAB and OpenStreetMap to simulate a cellular coverage map for a designated dense urban area. Firstly, we exported an *.osm formatted file from OpenStreetMap which includes the crowdsourced map data for the area of interest (OpenStreetMap contributors 2017). The crowdsourced data incorporates the 2D position information of streets and buildings, as well as their spatial information such as the width of streets and heights of buildings. Building heights in OSM are specified through attributes contributed by users. The height information is either directly provided by the height tag, or it can be inferred from other related attributes, such as the number of building levels (OpenStreetMap 2021; OpenStreetMap 2009). When importing the OSM map into the simulation, these attributes were used to generate a 3D model representing the building geometries based on the available data. As a result, the buildings are automatically depicted as simple, crude blocks rising from their 2D footprints registered in OSM, as shown in Figure 3.3. This generated 3D model was utilised in MATLAB to calculate signal propagation for transmission. The simulation accounted for relevant factors such as the distance between the transmitter and receivers, as well as signal reflection and diffraction caused by obstacles. In our simulation, a cellular antenna transmitting signals was assumed to be installed on a building rooftop with a known latitude, longitude, and altitude. This is a reasonable assumption as mobile operators generally own this information in their databases, though it is not publicly available. However, some crowdsourced databases, such as OpenCellID (OpenCellID 2008), could provide approximate locations of cell towers, thus making such an assumption seems to be realistic. Next, a target building within the range of the transmitter was selected for height estimation.

Since RSS reflects both path attenuation and the obstruction effects caused by buildings and other obstacles (Fan et al. 2019), we selected it as the key feature for clustering receiver locations. As a result, we uniformly sampled the RSS across both latitude and longitude around the vicinity of the target building. Figure 3.3 illustrates the simulation setup and outcome for a coverage map, including the transmitter location, the target building, and the RSS sampling area for receivers. The Rx altitude was set to 1 metre above ground level, approximating the height of a mobile phone in a trouser pocket. Due to the uneven terrain surrounding the target building, variations in elevation naturally occurred across different locations. Figure 3.4 shows the distribution of Rx elevations, with an average value of 10.47 metres and a 95th percentile value of 11.66 metres. For simplification, an estimated altitude of 11 metres was applied to all Rx locations. Two RSS coverage maps were simulated, differing primarily in their maximum diffraction settings. The first simulation assumed no diffraction, meaning only LOS or reflected signals were considered, while the second simulation accounted for both signal reflection and diffraction effects. The parameters used in these simulations are detailed in Table 3.1. Both simulations employed the Ray-tracing propagation model using the method of

Table 3.1: Cellular coverage map parameters.

No.	Propagation model	Transmitter frequency	Transmitter antenna height	Transmitter power	Radiation range	Max reflections	Max diffractions	Receiver antenna height (above ground)
A.	Ray-tracing	3.6 GHz	4 metres	3 watts	500 metres	3	0	1 metre
B.	Ray-tracing	3.6 GHz	4 metres	3 watts	500 metres	3	1	1 metre

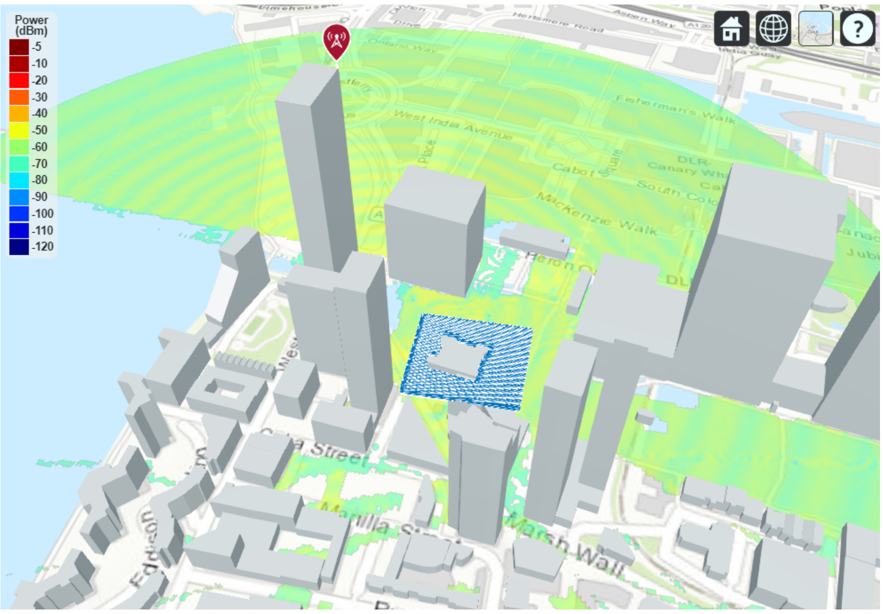


Figure 3.3: Simulation of cellular coverage utilising both MATLAB and OSM 3D models. The transceiver is positioned on a building rooftop with the RSS sampled near the target building.

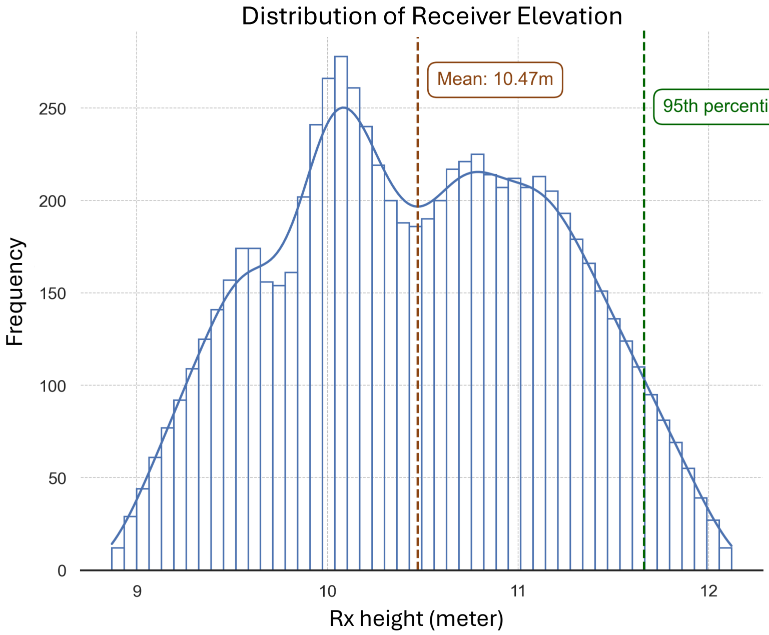
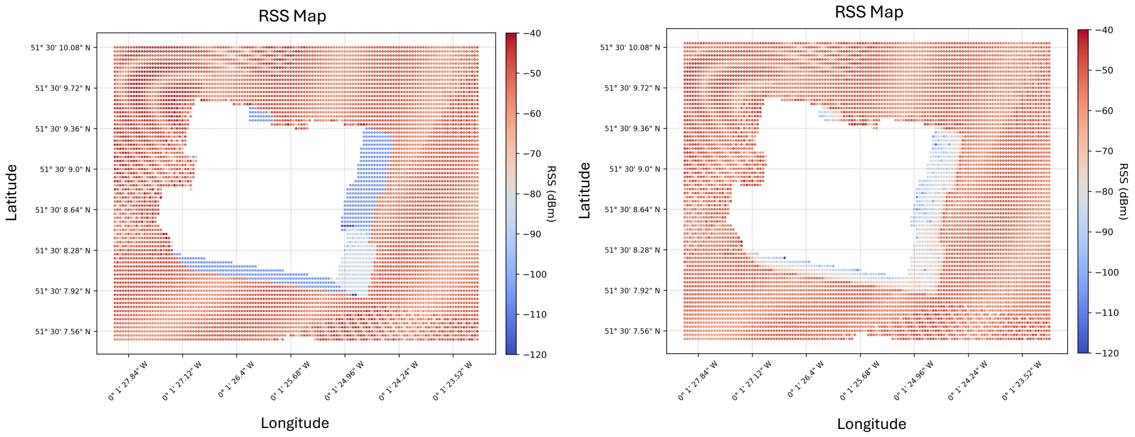


Figure 3.4: The height distribution of receivers.

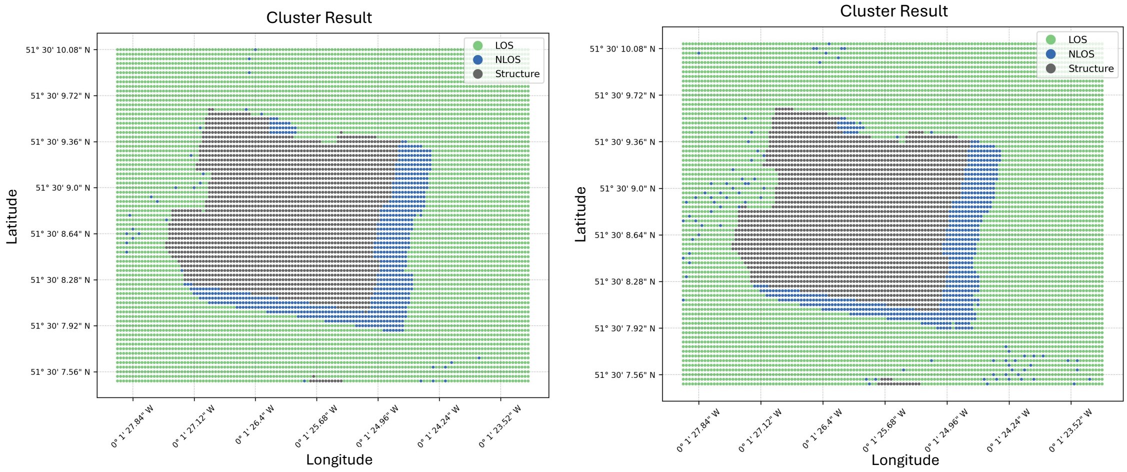
shooting and bouncing rays (Ling et al. 1989). Additionally, the simulation parameters for building and terrain materials were set to “glass” and “brick”, respectively, to closely approximate the reflection, penetration, and absorption characteristics of these materials in real-world environment.



(a) Without diffraction

(b) With diffraction

Figure 3.5: Simulated RSS results in the vicinity of the target building: (a) RSS map excluding signal diffraction. (b) RSS map considering a maximum of one signal diffraction.



(a) Simulation A

(b) Simulation B

Figure 3.6: Clustering results from two simulated RSS maps. The ray-tracing model using (a) maximum three reflections and without diffraction, and (b) maximum diffraction number set to one.

3.3.1.2 RSS clustering

Figure 3.5 reveals the RSS maps of Simulation A and B. In general, it can be observed that a receiver distanced from the target building is more likely to exhibit higher RSS levels, attributed to unhindered signal transmission by the building. Accordingly, its RSS level is only influenced by the distance to the transmitter and possibly augmented by reflected and/or diffracted signals (e.g., façade of nearby buildings). Conversely, a receiver situated within the building's shadow area exhibits a significantly reduced

RSS level. By definition, the shadow area cannot receive any direct signals; however, depending on the experimental setup, it can potentially receive reflected signals (Simulation A) or both reflected and diffracted signals (Simulation B), contributing to the overall weaker RSS levels.

The RSS observations are organised into three clusters. The first cluster consisted of data points of receivers located in LOS positions, where the signal was unobstructed by the target building or any nearby obstacles, resulting in the highest RSS levels. The second cluster represented the data points in the NLOS area where the received signal was blocked by the target building, resulting in the weakest RSS level. The third cluster "Structure" was defined to simplify the detection of the building footprint. This cluster included the area bounded by the outlines of the target building, where no data points were available. Figure 3.6 illustrates the result of K-means clustering for both of the simulations.

3.3.1.3 Boundary detection

Upsampling For boundary detection among distinct clusters, we first identified the polygons representing each cluster and extracted their bounding outlines. We employed nearest-neighbour interpolation to upsample the original receiver coordinates to refine these outlines further. Specifically, each interpolated point was assigned the cluster label of its nearest receiver, resulting in a smoother and more continuous boundary across the spatial grid. The reason for adopting the upsampling technique is twofold. First, we simulated dense sampling intervals to mimic future crowdsourcing environments, where many contributors would provide their data in a concentrated manner. Second, the results obtained through upsampling resemble the data found on a crowdsourcing platform. This means that building footprints could be derived directly from this data without the need to involve a third party. The size of the original dataset sampling grid consisted of 135×73 receiver data points across the longitude and latitude axis, respectively. Following the proposed upsampling process, we extended the quantity of these grids from 135×73 to 1000×1000. Figure 3.7 illustrates the polygons from the first simulation for all clusters identified in Section 3.3.1.2. Scattered points in the NLOS cluster represent outlier receiver locations possibly caused by multipath propagation and signal reflections, leading to weak or inconsistent received signal strength measurements away from the building shadow. These points are treated as noise and are suppressed in the subsequent boundary denoising step.

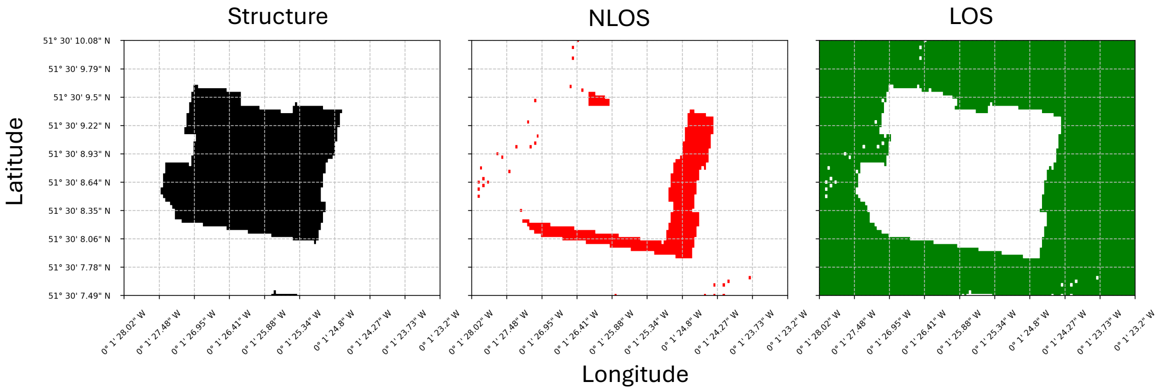


Figure 3.7: Extracted polygons corresponding to distinct clusters, demonstrated using the first simulation.

Cluster boundary extraction To estimate the height of the target building, two decision boundaries among clusters are crucial. The first boundary is between LOS and NLOS, indicating the transition from unobstructed to obstructed signals due to the building's height. The second boundary differentiates the "Structure" cluster from all other clusters, representing the derived building footprint. The boundary detection is achieved by constructing a mask grid for each cluster pair, calculating axial differences across the grid, and identifying geographic coordinates at points of cluster transitions.

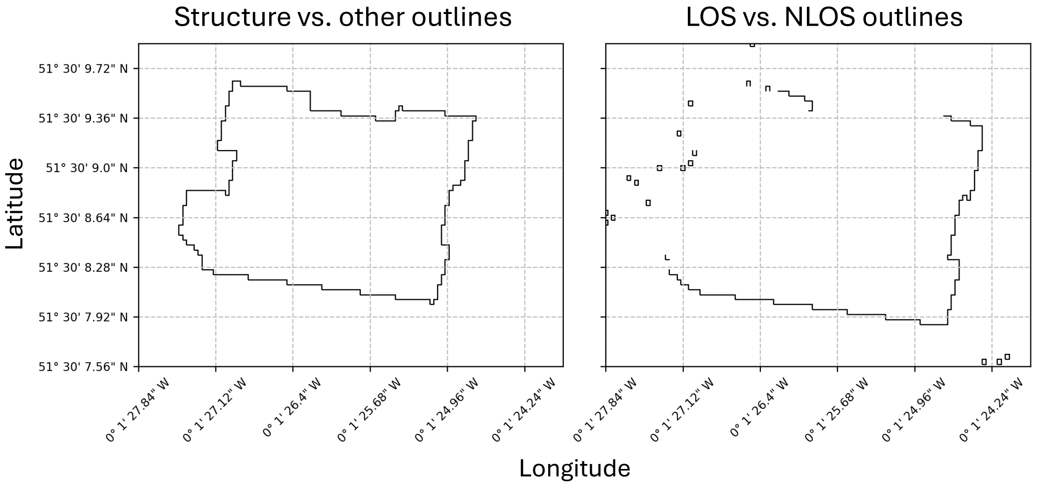


Figure 3.8: Detected boundaries delineating the target building's outline from other clusters, and distinguishing LOS from NLOS clusters.

Figure 3.8 shows the results of our boundary detection method. Successful boundary detections typically appeared as line segments at transitions between clusters. However, some detections were sporadically located in isolated regions, which provided limited value for height estimation. To address this issue, we applied the DBSCAN clustering algorithm to reduce noise in the detected boundaries. In the denoising phase, each point on the detected boundaries was treated as a sample and clustered. This method enables the grouping of closely packed points into clusters while distinguishing points

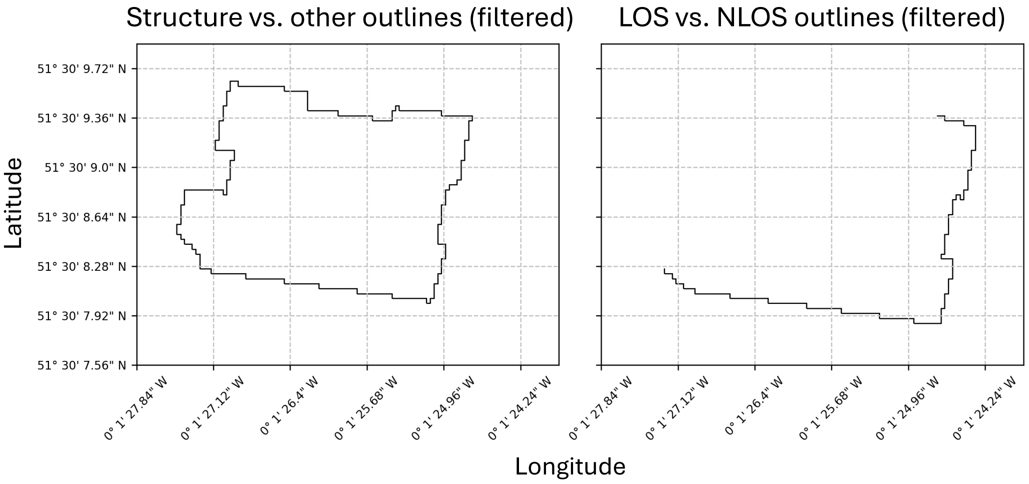


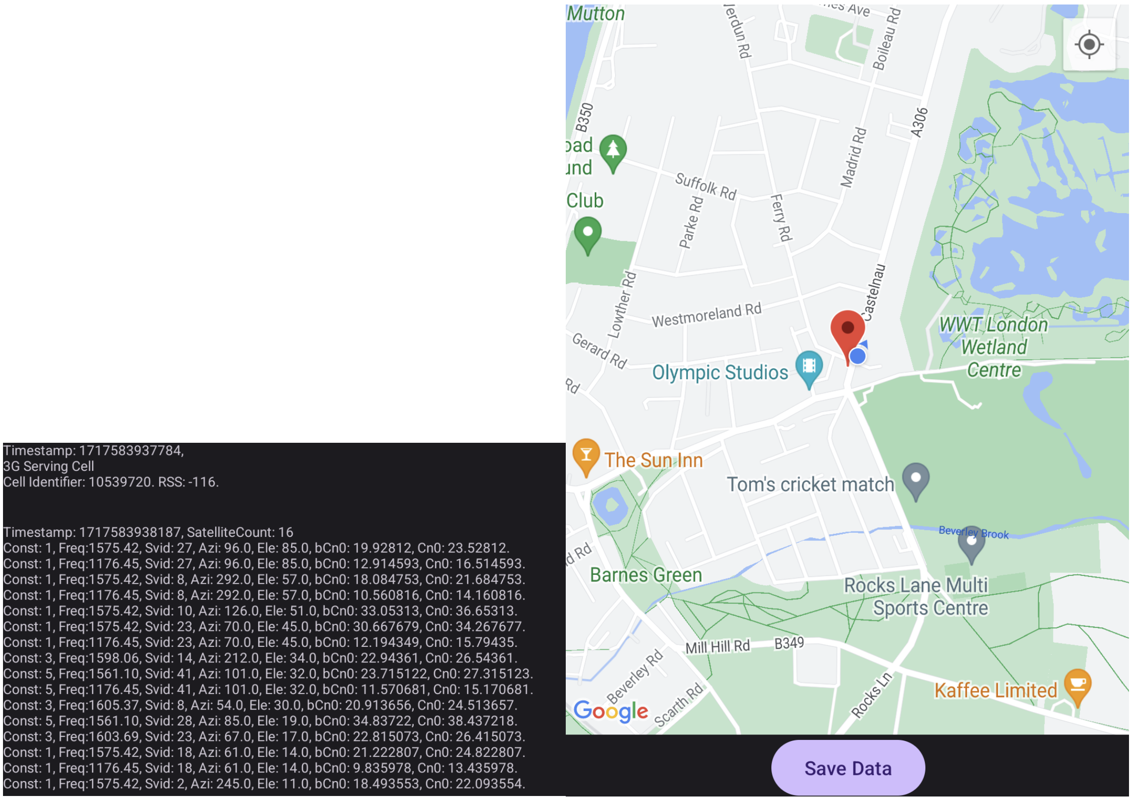
Figure 3.9: Denoised boundaries via DBSCAN, excluding incorrectly identified boundaries.

that lack sufficient nearby neighbours as noise. Figure 3.9 shows the results of boundary denoising. The DBSCAN parameters were fine-tuned using the heuristic approach described in (Ester et al. 1996). This denoising process effectively removed incorrectly clustered points that formed non-informative boundaries, preserving the line-segment-shaped points for further analysis.

3.3.2 Usage of real-world data for building height estimation

To further validate our proposed approach, we conducted an experiment in a real-world scenario, performing height estimations for six buildings in a dense urban area.

After conducting an exhaustive search in the Google Play Store, we found no existing Android mobile application that supports simultaneous recording of cellular signal-related data and the user's geolocations, which are essential in our proposed approach. Therefore, we developed a dedicated Android app, *BitToBrick*, to collect key information for height estimation, including RSS at the receivers' locations, unique identifier (CellID) of the transmitter, and both GNSS-resolved and user-reported geo-locations. We plan to make this application publicly available to the broader research community in the future. Figure 3.10 shows a screen capture of the application in operation. The application primarily collects five key pieces of information:



```
Timestamp: 1717583937784,
3G Serving Cell
Cell Identifier: 10539720. RSS: -116.


Timestamp: 1717583938187, SatelliteCount: 16
Const: 1, Freq:1575.42, Svid: 27, Azi: 96.0, Ele: 85.0, bCn0: 19.92812, Cn0: 23.52812.
Const: 1, Freq:1176.45, Svid: 27, Azi: 96.0, Ele: 85.0, bCn0: 12.914593, Cn0: 16.514593.
Const: 1, Freq:1575.42, Svid: 8, Azi: 292.0, Ele: 57.0, bCn0: 18.084753, Cn0: 21.684753.
Const: 1, Freq:1176.45, Svid: 8, Azi: 292.0, Ele: 57.0, bCn0: 10.560816, Cn0: 14.160816.
Const: 1, Freq:1575.42, Svid: 10, Azi: 126.0, Ele: 51.0, bCn0: 33.05313, Cn0: 36.65313.
Const: 1, Freq:1575.42, Svid: 23, Azi: 70.0, Ele: 45.0, bCn0: 30.667679, Cn0: 34.267677.
Const: 1, Freq:1176.45, Svid: 23, Azi: 70.0, Ele: 45.0, bCn0: 12.194349, Cn0: 15.79435.
Const: 3, Freq:1598.06, Svid: 14, Azi: 212.0, Ele: 34.0, bCn0: 22.94361, Cn0: 26.54361.
Const: 5, Freq:1561.10, Svid: 41, Azi: 101.0, Ele: 32.0, bCn0: 23.715122, Cn0: 27.315123.
Const: 5, Freq:1176.45, Svid: 41, Azi: 101.0, Ele: 32.0, bCn0: 11.570681, Cn0: 15.170681.
Const: 3, Freq:1605.37, Svid: 8, Azi: 54.0, Ele: 30.0, bCn0: 20.913656, Cn0: 24.513657.
Const: 5, Freq:1561.10, Svid: 28, Azi: 85.0, Ele: 19.0, bCn0: 34.83722, Cn0: 38.437218.
Const: 3, Freq:1603.69, Svid: 23, Azi: 67.0, Ele: 17.0, bCn0: 22.815073, Cn0: 26.415073.
Const: 1, Freq:1575.42, Svid: 18, Azi: 61.0, Ele: 14.0, bCn0: 21.222807, Cn0: 24.822807.
Const: 1, Freq:1176.45, Svid: 18, Azi: 61.0, Ele: 14.0, bCn0: 9.835978, Cn0: 13.435978.
Const: 1, Freq:1575.42, Svid: 2, Azi: 245.0, Ele: 11.0, bCn0: 18.493553, Cn0: 22.093554.
```

(a)

(b)

Figure 3.10: Illustration of Android app BitToBrick. (a) Details of cellular and GNSS signals captured, and (b) Map with GNSS-resolved location & movable pin for precise location.

- **CellID:** This is the unique serial number identifying a cellular antenna to which the mobile device is currently connected. Additionally, the online platform (Cellmapper 2025) is an open database that provides the geolocation of the cell tower associated with the specified CellID.
- **Received signal strength:** This is the signal strength received by the mobile phone from the corresponding cellular antenna, measured in dBm.
- **GNSS position:** The location calculated by the native Android GNSS components is displayed as a blue dot in Figure 3.10.
- **User-reported location:** This supplementary location-related functionality allows users to record their precise location by moving a red pin. This feature is particularly useful for accurate location calibration when GNSS-derived locations are unreliable, especially in urban areas.
- **GNSS satellites information:** In addition to cellular signals, the application is designed to report and record information about the GNSS satellites to which the mobile device is connected. While this satellite-related data is not used in our current research, it may be leveraged for height estimation in conjunction with cellular data in future studies.

To mitigate the GNSS degradation in urban, the data collection application allows users to manually adjust their reported location when the GPS-resolved position is perceived to be inaccurate. A unified user-reported location is then adopted for height estimation, enabling the method to tolerate GPS errors that are perceptible at the human scale while maintaining consistency in the estimation process. The app uses 4G cellular signals, which offer wider coverage and more stable user-cell associations than 5G, which typically operates at higher carrier frequencies with denser cell layouts, leading to shorter propagation ranges, increased sensitivity to environmental dynamics, and more frequent handovers within limited spatial extents. The "Save Data" functionality enables the recording of the above-mentioned information at the moment the button is tapped, enabling subsequent analysis. Based on our observations, the RSS and GNSS position update frequencies may vary slightly depending on network conditions but typically fall within the range of one update every 1–2 seconds. To ensure synchronisation between these data streams, we record only when both RSS and location signals appear stable and no significant fluctuations are observed. Due to restrictions imposed by the Android SDK, the application can only access and record data from the currently connected CellID, even though the device may receive signals from multiple neighbouring cells. This limitation could potentially be mitigated by using devices equipped with multiple SIM cards from different network providers, or more effectively, by incorporating crowdsourced data contributed by multiple users across different networks. Currently, the data can be saved locally on the mobile device and has to be exported for further processing. However, *BitToBrick* has the potential to evolve into a crowdsourcing platform for large-scale data collection, with the possibility of uploading data to a centralised online database.

3.3.2.1 Data collection

The data collection in the experiment comprised two parts: Firstly, the RSS and precise geolocation of each Rx location were collected using BitToBrick. Secondly, we used an open-source repository that contains building footprints from around the world (Bing Maps 2025).

RSS and location We showcase a dense urban area near London for our data collection, where multiple cellular antennas were installed on the rooftop of a high-elevation hotel building (marked as red star in Figure 3.11). Using the *BitToBrick* application, we identified the exact CellID to which our mobile device was connected. We then used the online platform Cellmapper (Cellmapper 2025) for localising the cellular tower associated with this CellID, which was confirmed to be one of the antennas on the rooftop.

Next, we selected six target buildings surrounding the transmitter antenna to estimate their heights. These buildings were chosen because their elevations were visibly lower than the rooftop antenna, and their surroundings were broad enough to encompass both LOS and NLOS scenarios, offering geometric convenience as a starting point for height estimation. Information provided by BitToBrick confirmed that these buildings and their surrounding areas were within the coverage area of the antenna installed on the hotel rooftop. Furthermore, the buildings feature various geometric types, such as flat or pitched roofs, and we did not make any previous assumptions about the presence of perpendicular walls. They serve different functions, encompassing both commercial and residential uses. Additionally, their distances from cellular antennas and the characteristics of their surrounding environments vary. For instance, some buildings are positioned on quiet streets with minimal vehicle traffic and tall vegetation that acts as a barrier against signal interference. In contrast, other buildings are located in busier areas with more dynamic environmental conditions. During the experiment, the mobile phone used for testing remained connected to the same antenna throughout the entire data collection session for all six targets. In total, we measured RSS levels at 96 locations for the six target buildings, including both LOS and NLOS areas. On average, 16 RSS locations are used to estimate the height of each building. Due to environmental dynamics, measurements at the exact location may vary slightly due to signal interference. To minimise the influence of signal noise and outliers on clustering performance, we recorded RSS measurements multiple times at each receiver location, typically collecting between 5 and 14 readings per point. This strategy helps mitigate the effects of transient fluctuations caused by multipath propagation, attenuation, or environmental interference, allowing for a more stable and representative RSS value at each location. The primary criterion for selecting these locations was to ensure coverage of both obstructed and unobstructed areas relative to the target building, allowing us to establish the decision boundary between high and low RSS clusters.

Building footprints Similar to the MATLAB simulation, the target building footprint is essential for height estimation. In real-world experiments, accurate building footprints are essential for estimating height, just as they are in simulations. However, unlike the simulated environment, which provides easily accessible and densely packed data samples, the current state of real-world data collection is significantly sparser compared to what we expect from its future evolution of crowdsourcing. This limitation poses a challenge in directly and accurately deriving building footprints from the real-world data we have gathered. Therefore, we opted to use the public dataset from Microsoft building footprints (Bing Maps 2025), a growing global repository containing 2D building outlines. These building footprints were generated using techniques including semantic segmentation with deep neural networks, and polygonization that interprets image pixels into polygons that represent buildings (Bing Maps 2025).

3.3.2.2 Double-Level clustering

The collected RSS data points are spread throughout the area, with clusters of points representing measurements for a specific target building. It is evident from Figure 3.11 that the gathered RSS data are distributed around the buildings. To distinguish the measurements and perform height estimation for each target building, we implemented a double-level clustering approach. First, we applied DBSCAN to separate individual groups of points based on density, isolating the data corresponding to each target building in the urban area. We set 30m as the minimum distance for a point to be considered within the neighbourhood of a core point. As illustrated in Figure 3.11, the data points were clustered into six distinct groups based on their proximity. The second level of clustering further divided the RSS points into two distinct groups for the identification of decision boundaries.

3.3.2.3 Decision boundary identification

After the double-level clustering described in Section 3.3.2.2, each DBSCAN-identified cluster was further divided into two distinct clusters based on RSS values. Similar to the simulated experiment, the decision boundaries between RSS clusters implicitly carry the height information of the corresponding target building. While a full boundary could, in principle, be extracted using multiple transition points, this would introduce additional processing complexity and is difficult to achieve under realistic urban data-collection constraints. In practice, measurements are often spatially uneven and limited by accessibility, resulting in sparse and discontinuous samples rather than a well-defined continuous boundary. Under these conditions, the midpoint between identified LOS and NLOS clusters provides a stable and computationally efficient approximation of the underlying signal transition. For computational convenience, we approximated that the decision boundary lies at the midpoint between the centres of these two clusters. For each DBSCAN cluster i, the midpoint (or boundary point) $(\mathrm{Lat}_\mathrm{b}, \mathrm{Lon}_\mathrm{b})$ is computed as:

$$(\mathrm{Lat}_\mathrm{b}, \mathrm{Lon}_\mathrm{b}) = \left(\frac{\mathrm{Lat}_{\mathrm{ctr}_0} + \mathrm{Lat}_{\mathrm{ctr}_1}}{2}, \frac{\mathrm{Lon}_{\mathrm{ctr}_0} + \mathrm{Lon}_{\mathrm{ctr}_1}}{2} \right) \tag{3.4}$$

Here, $\mathrm{Lat}_{\mathrm{ctr}_j}$ and $\mathrm{Lon}_{\mathrm{ctr}_j}$ are the latitude and longitude of the center of cluster j (where $j \in \{0,1\}$ represents the two K-means clusters within the i-th DBSCAN cluster), computed as:

$$\mathrm{Lat}_{\mathrm{ctr}_j} = \frac{1}{N_j} \sum_{k=1}^{N_j} \mathrm{Lat}_{k_j},$$
$$\mathrm{Lon}_{\mathrm{ctr}_j} = \frac{1}{N_j} \sum_{k=1}^{N_j} \mathrm{Lon}_{k_j}$$

where N_j is the number of points in cluster j, and $(\mathrm{Lat}_{k_j}, \mathrm{Lon}_{k_j})$ are the latitude and longitude of the k-th point in cluster j. The resulting midpoint is used as the estimated receiver located at the decision boundary for height estimation. The boundary points of each cluster are represented by flag icons in Figure 3.11.

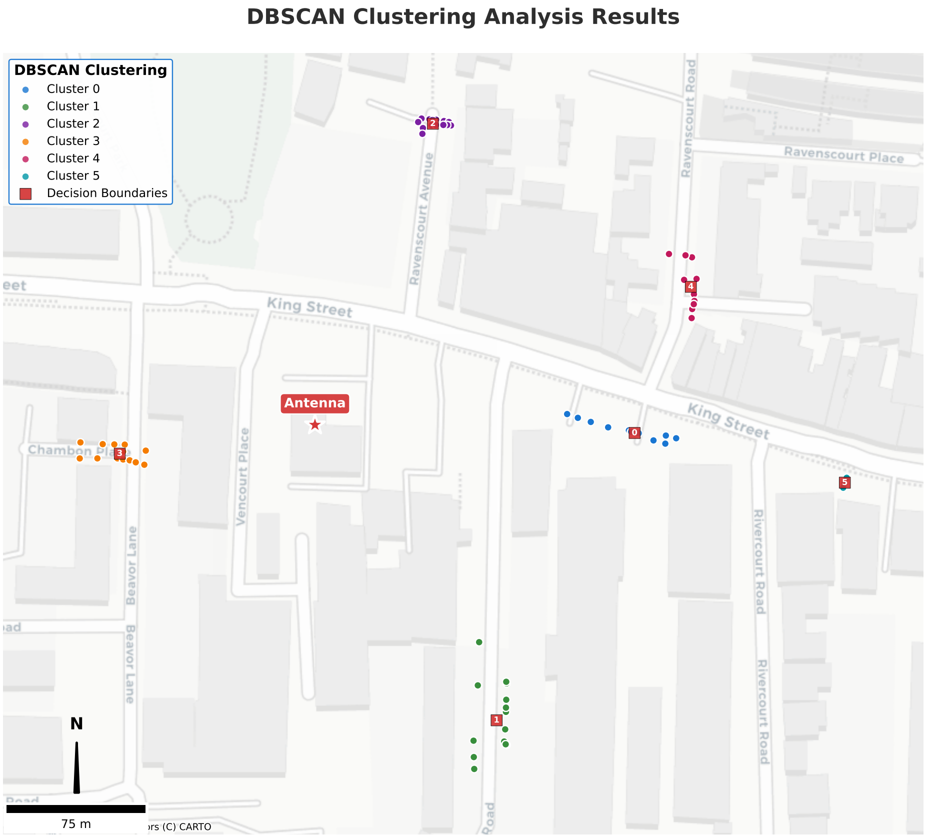


Figure 3.11: Result of double-level clustering. RSS data points of various colours indicate different DBSCAN clusters. The red square icons represent detected boundary points within each cluster, as outlined in Section 3.3.2.3. Antenna location is marked in red star.

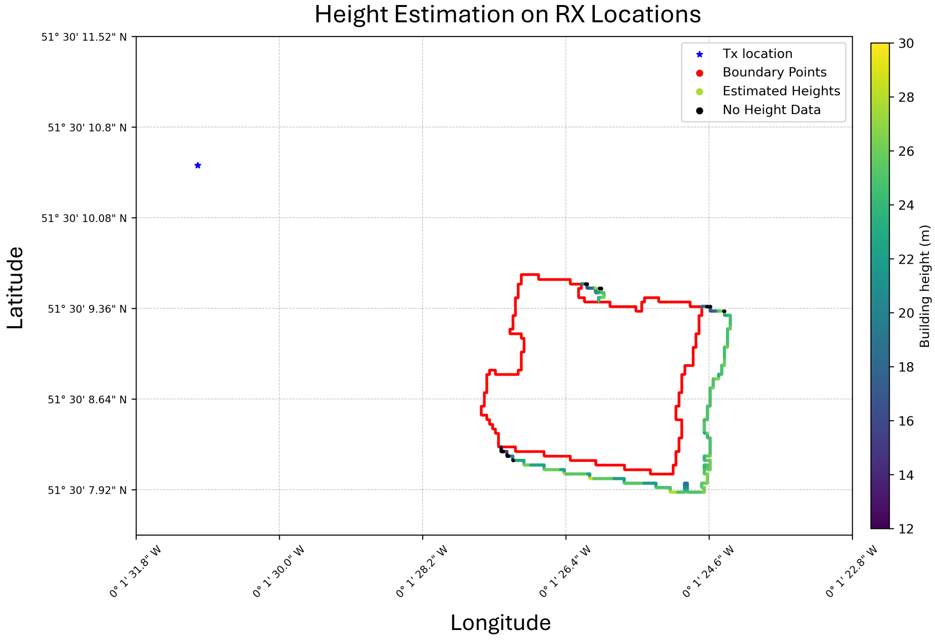
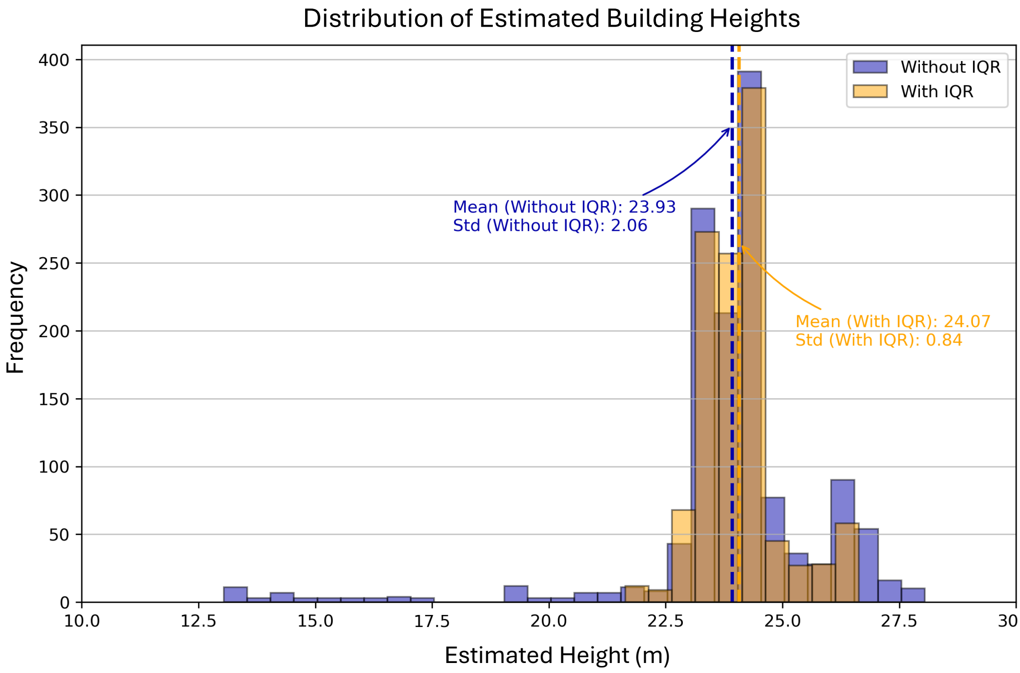
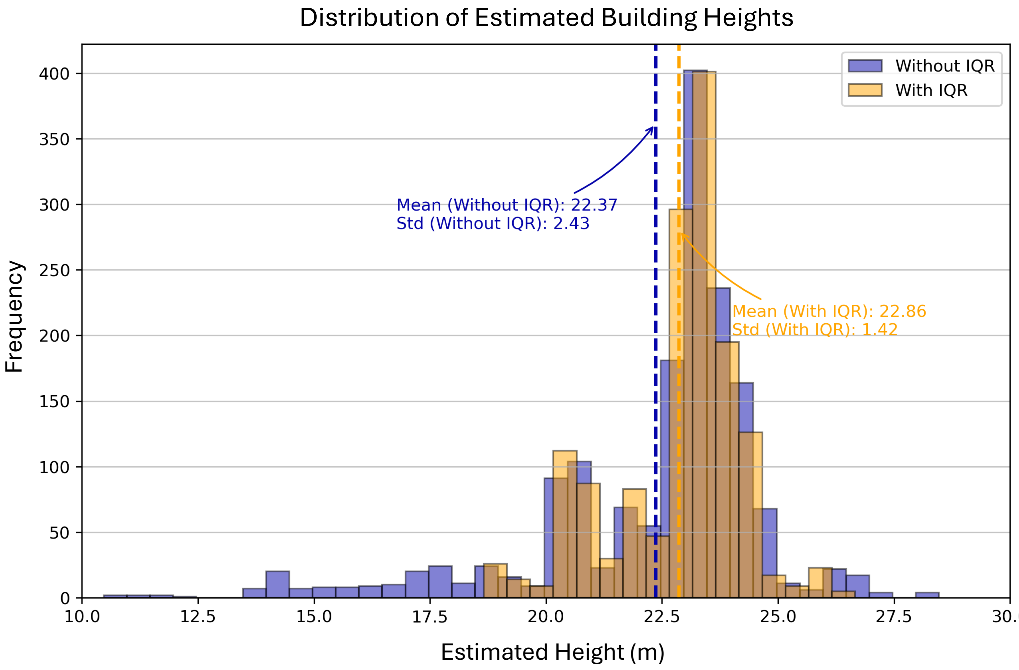


Figure 3.12: Height estimation results corresponding to each Rx positioned at shadow boundaries.



(a) Distribution of height estimation: Simulation A



(b) Distribution of height estimation: Simulation B

Figure 3.13: Distribution of height estimation for (a) RSS Simulation a, and (b) RSS Simulation b. Pre- and post-IQR processing results are depicted in light blue and light orange, respectively.

3.4 Result

3.4.1 Height estimation results on simulated data

For the simulated data, we applied the height estimation method outlined in Section 3.2.1 at each Rx location identified on the boundary. Figure 3.12 depicts the height estimation results of Simulation b at each boundary-located receiver location. The results illustrate the presence of outliers, particularly at the decision boundaries formed by verticle building edges. The distribution of the unprocessed height estimates is illustrated in Figure 3.13 using light blue histograms. The figure highlights the presence of outliers, which predominantly originate from boundary regions affected by the vertical edges of buildings. To address these outliers, the interquartile range (IQR) method (Vinutha et al. 2018) was employed to confine outlier bounds and eliminate these beyond the identified bounds. The lower and upper outlier bounds are computed as:

$$\begin{aligned}\text{Lower Outlier Bound} &= Q1 - 1.5 \times (Q3 - Q1) \\ \text{Upper Outlier Bound} &= Q3 + 1.5 \times (Q3 - Q1) \qquad (3.5)\end{aligned}$$

where $Q1$ and $Q3$ are the first and third quantiles of height estimation distribution, respectively. As a result, the majority of outliers were effectively removed from the initial height estimation results, as demonstrated in Figure 3.13 by light orange histograms.

The target building in the simulation was modelled with a height of 25 metres. The results are detailed in Table 3.2. When comparing simulations with two different setups, Simulation A, which excluded the diffraction effect, produced height estimates that were closer to the modelled height than Simulation B, which included diffraction. Additionally, incorporating the IQR to remove outliers improved the accuracy of the results for both simulation setups, bringing them closer to the ground truth.

Table 3.2: Height estimation results.

	Simulation A	Simulation B
Without IQR	23.93 ± 2.06 m	22.37 ± 2.43 m
With IQR	24.07 ± 0.84 m	22.86 ± 1.42 m

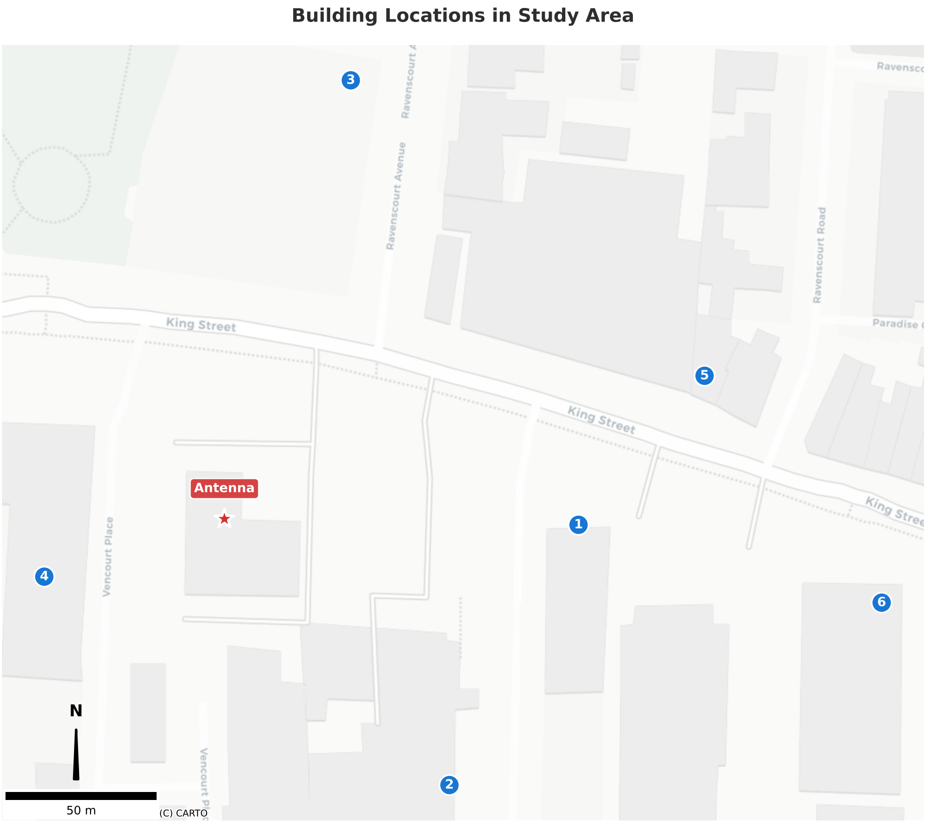


Figure 3.14: Selected target buildings for height estimation. Antenna location is marked with a red star.

3.4.2 Height estimation results on real-world data

We have identified six target buildings for height estimation, all of which are located within the coverage area of the selected cellular antenna. These target buildings encompass a variety of types, including residential, public service structures (such as schools), and commercial properties. They feature different roof types, including both flat and pitched roofs. In terms of layout, the buildings include terraced houses and semi-detached houses. Their locations are shown in Figure 3.14. Table 3.3 presents their addresses, ground truth heights, and estimated heights through the proposed methods. We utilised Absolute Errors (AE) to assess the measurement error as shown in Table 3.3. AE calculation is illustrated in Equation 3.6, note that we take the midpoint value between the highest and lowest bounds as the ground truth height for pitched roof buildings. In our real-world experiment, a single estimation was performed for each building by using the detected central points, as described in Equation 3.4. The ground truth data were retrieved from Google Earth (Google 2001) for evaluation purposes.

$$\text{AE} = |\text{Estimated Height} - \text{Ground Truth Height}| \tag{3.6}$$

The cellular antenna elevation was reported as 48 metres from Google Earth, which corresponds to h_{Tx} in Equation 3.3. Additionally, we verified the ground user's elevation, finding that the average ground elevation was approximately 4 metres above mean sea level. While collecting cellular data, a volunteer held the mobile phone at chest level. Given the volunteer's height of 1.8 metres, the phone was approximately 1.6 metres above ground level or 5.6 metres above sea level. This height was used as h_{Rx_centre} in Equation 3.3.

Among the six buildings, two had flat roofs with ground truth heights of 17 and 21 metres (No. 2 and 4). The remaining buildings had pitched roofs, with the lowest and highest roof elevations listed in the Ground Truth column of Table 3.3. For the flat-roofed building No. 2, the AE is 0.25 m, indicating a precise height estimation compared to its ground truth. Despite the simplification of the complexities of the roof shape and pitch, which can contribute to the errors in the height estimation for pitched-roofed buildings, such as No. 1, No. 5 and No. 6 the overall accuracy is still high. The AEs for these buildings were 0.02 m, 2.39 m and 0.72 m, respectively.

For building No. 4, the height was initially estimated at 12.07 metres, compared to its ground truth of 21 metres, resulting in a significant discrepancy. Upon careful investigation, we discovered that in the Microsoft Building Footprints data, the polygon representing the target building was incorrectly merged with its neighbour. This error led to an incorrect intersection point between the line segment formed by Tx and Rx and the 2D footprint. After manually correcting the intersection point to an approximated location, the revised height was 21.04 metres, yielding an AE of only 0.04 m. Similarly, for building No. 6, the intersection did not occur due to the imprecise placement of the target polygon, which was shifted on the map relative to its ground truth location. After manually approximating the intersection point, the estimated height was 14.28 metres. This is close to the actual height of 14–16 metres for a pitched roof. The least accurate estimation occurred for the pitched roofed building No. 3, with a ground truth height range of 14–17 metres, and an estimated height of 20.18 metres, resulting in an AE of 4.68 m. This inaccuracy can be attributed to two main factors. First, unlike other successfully estimated cases, the line between Tx and Rx for building No. 3 was nearly parallel to the intersecting 2D outline edge of the building rather than perpendicular, which made distinguishing RSS clusters more difficult. Secondly, during data collection, tree foliage obstructed the line of sight between the building's roof and the ground user. This likely introduced additional interference or attenuation to the signal transmission, causing the received signal to be noisier than expected.

Table 3.3: Height estimation results.

No.	Location	Ground Truth	Estimated Height	Absolute Error
1	239 King St, W6 0RF	17–20 m pr✧	18.52 m	0.02 m
2	Palingswick House241 King St, London W6 9LP	17 m fr✧	16.75 m	0.25 m
3	Main building, 16 Ravenscourt Ave, W6 0SL	14–17 m pr	20.18 m	4.68 m
4	273 King St, W6 0RF	21 m fr	12.07 m to 21.04 m†	0.04 m
5	236a King St, W6 0RF	13–15 m pr	16.39 m	2.39 m
6	233 King St, W6 0RF	14–16 m pr	14.28 m†	0.72 m

✧ pr and fr represent "pitched roof" and "flat roof", respectively.
† Manually corrected intersection point at 2D outline.

3.5 Discussion

Building height estimation is crucial for generating 3D city maps. This paper presents a novel method for estimating height using widely available cellular networks, providing a practical enhancement to existing techniques. First, our method relies on the existing cellular network infrastructure, requiring no additional equipment beyond standard mobile phones. Second, the method uses stationary cellular antennas located at predetermined positions, reducing the computational cost. Third, the RSS data used for this method is easily accessible through our proposed Android app.

Specifically, we used ground user RSS measurements from a single cellular antenna. By classifying these measurements, we extracted shadowing information from the target building, which, along with the locations of the receiver and transmitter, was used for height estimation. Our method relies solely on the absolute RSS values and groups them based on whether they are influenced by the target building. More importantly, this approach does not require any prior knowledge of modelling or calculations related to signal propagation, which ensures its relatively low computational complexity. Despite the relatively small number of samples collected for each target building, the estimation results distinguished between affected and unaffected RSS groups, demonstrating its consistent robustness throughout the experiment. The resulting estimate for a target building with an actual height of 25 m was 24.07 ± 0.84 m and 22.86 ± 1.42 m from the simulated RSS map without and with signal diffraction setup, respectively. The difference between the RSS of LOS and NLOS clusters in the second simulation is less pronounced than in the first due to signal diffraction, leading to greater deviation in the resulting estimate based on detected cluster boundaries compared to the actual modelled height.

Our prototype BitToBrick app has proven effective in collecting data for estimating building heights, indicating its potential for scalability if developed into a crowdsourcing platform. Two key factors support this potential: First, the widespread use of mobile phones in daily life makes a mobile-based crowdsourcing platform both cost-efficient and highly scalable, as it leverages existing, readily available technology. Second, crowdsourced data from users connected to different telecom networks can offer diverse and comprehensive coverage by capturing wireless signals from cellular antennas at various locations. In the real-world application, we revealed the effectiveness of our proposed method. The heights of three out of six target buildings (No. 1, 2, and 5) were accurately measured, including both flat-roofed and pitched-roofed cases. The AEs for buildings No. 1 and No. 2 were exceptionally low, at 0.02 m and 0.25 m, respectively, indicating a high level of precision. A typical residential floor-to-floor height in the UK is around 3.2 m per storey (Mayor of London 2026); therefore, a 2.39 m height error for building No. 5 corresponds to less than a typical storey height in the UK. For context, airborne LiDAR products often achieve decimetre-level vertical accuracy when evaluated against ground control, with reported discrepancies around 0.12 m (Elaksher et al. 2023). In contrast, large-scale satellite-based (SAR/optical) building-height mapping typically exhibits metre-level errors, e.g., frequency-weighted RMSE of 2.9–3.5 m for Sentinel-1/2 models (Frantz et al. 2021). Compared to these techniques, the proposed approach achieves lower estimation accuracy, but it leverages existing mobile network infrastructure and mass-market smartphones, requires no specialised equipment, and supports low-cost, scalable deployment, particularly in settings where dedicated 3D mapping resources are limited. Considering its pitched roof, the estimated height of 16.39 metres is closely aligned with the maximum roof height of 15 metres. The results also highlighted the dependence on the environmental sensitivity of cellular signals. For building No. 3, we observed that additional stationary or dynamic objects likely induced extra noise, leading to imprecise clustering decision boundaries and reduced accuracy in height estimation. In this regard, collecting additional data with both temporal and spatial diversity could help mitigate noise and interference caused by dynamic environmental factors. RSS measurements can be taken during off-peak hours with fewer distracting nearby objects, or from different sides of the target building, especially if contributed by users through antennas operated by various telecom providers. On top of this, while the buildings incorporated in this study are predominantly masonry structures constructed from brick, other materials such as glass façades, concrete, and metal components can significantly affect signal attenuation, reflection, and penetration behaviour. Nevertheless, as the proposed method identifies decision boundaries based on relative RSS clustering rather than absolute signal values, variations in material-dependent attenuation are less likely to fundamentally compromise the boundary detection process. Future work will explicitly investigate the robustness of the proposed approach across a wider range of building materials.

Environmental factors such as vegetation can introduce RSS fluctuations. Transient variations are mitigated through repeated sampling and outlier removal, while more persistent attenuation, for example from dense foliage, may affect absolute signal strength. Nevertheless, the method relies on clustering RSS observations into LOS and NLOS regimes. Even when tree-induced attenuation is present, the contrast between building-shadowed regions and relatively unobstructed regions remains discernible, allowing building-induced signal transitions to be identified. Variability arising from receiver hardware, antenna sensitivity, and weather conditions is treated as part of the overall measurement noise and addressed through aggregation across repeated observations rather than device-specific calibration. Similarly, factors such as distance to the serving cell and antenna orientation influence baseline path loss but do not systematically alter the presence of LOS/NLOS transitions. By focusing on relative RSS patterns and clustering behaviour rather than absolute signal levels, the method demonstrates robustness to these sources of variability.

Our proposed method has certain limitations that need to be addressed. We found that precise 2D outlines are crucial for accurate height estimation using our proposed methods. In the cases of buildings No. 4 and 6, the estimations were not as expected due to the spatial deviation of the 2D building outlines from their real-world positions. The intersection points were either incorrectly calculated or overlooked, which led to less accurate height estimations. The current approach relies on third-party building footprints, which can introduce additional error into the height estimation. In a multimodal setting, visual data could be used to refine building footprints.

Although there are quality issues with the third-party building footprints, which is a dataset that is still developing and gradually maturing, our results show that we can achieve comparable outcomes without relying on more accurate government datasets, such as those from the Ordnance Survey. The current evaluation is based on a representative compact urban area, and future work will extend this framework through crowdsourced data acquisition to ensure broader applicability and scalability across diverse urban forms. However, deploying the proposed approach in a crowdsourcing setting also introduces challenges related to data heterogeneity, uneven spatial and temporal coverage, device and platform variability, and privacy protection. We will focus on a robust back-end design that ensures scalability and effective data management. Additionally, we will prioritise data security, protect user privacy, and comply with relevant regulations concerning crowdsourced data. These challenges will need to be addressed as the approach is extended to crowdsourcing platforms. In practice, inaccurate transmitter elevation may also influence the estimated building height. Future work will consider incorporating uncertainty quantification or error propagation analysis to better accommodate this source of variance. In addition, the current formulation estimates a single representative height per building and therefore does not resolve intra-building height variation, such as pitched or stepped roof structures, which would require boundary observations from multiple viewing directions. Future work could consider a data fusion approach from street-view or satellite imagery, which could help distinguish roof structures or constrain plausible height ranges, while cellular measurements provide independent evidence for the vertical dimension. The geometric layout may prevent ground users from receiving signals from both obstructed and unobstructed clusters. Furthermore, we will integrate data in other types collected by the app with cellular measurements, specifically GNSS signals, to provide signals from higher-elevation satellites.

In general, increasing the number of observations improves robustness to noise and supports clearer separation between LOS and NLOS signal clusters, leading to more stable height estimates, as illustrated by the idealised simulation results. While, in theory, a minimal set of well-positioned observations from distinct propagation regimes could be sufficient for height estimation, such conditions are rarely achievable in real-world settings due to limited prior knowledge of signal propagation and environmental variability. Practical deployment therefore relies on aggregating multiple samples to ensure reliable clustering and robust estimation. The relationship between sampling density and estimation stability also suggests a potential avenue for future work, where the number and spatial distribution of samples could be examined as an independent research question.

3.6 Conclusion

This paper presents a novel approach for estimating building heights using ubiquitous cellular signal data collected from mass-market mobile phones, addressing the need for scalable, cost-effective, and efficient methods to create 3D city maps essential for various applications. Results from six pilot areas, encompassing diverse urban scenarios with varying roof types and shapes, demonstrate absolute height estimation errors ranging from 0.02 m to 4.68 m. Five of the six buildings exhibited errors under 2.39 m, with the smallest error being 0.02 m and the largest error, equivalent to approximately one floor, remaining within acceptable limits. Future work will aim to improve the mobile application through crowdsourced data collection, while explicitly accounting for device heterogeneity. Furthermore, incorporating additional mobile signal types, particularly GNSS data, will improve localisation accuracy and leverage signals from higher-elevation transmitters.

Chapter 4

Humans in Cities

4.1 Advancing Human Activity Recognition Using Ultra-Wideband Channel Impulse Response Snapshots

Publication and Permission Statement

This chapter reproduces the accepted version of the following IEEE publication:

Yu Wang and Ana Basiri (2024). *Advancing Human Activity Recognition Using Ultra-Wideband Channel Impulse Response Snapshots.* Proceedings of the 2024 International Conference on Activity and Behavior Computing (ABC).

Abstract

Human Activity Recognition (HAR) has a significant contribution to a wide range of applications, including security surveillance and healthcare. It provides a non-intrusive, cost-effective solution building upon wireless signal technologies. Despite the recent growing availability of Ultra-Wideband (UWB) in the mass market, its potential in HAR is still under-researched compared with other wireless technologies. In this paper, we use UWB's high-resolution Channel Impulse Response (CIR) for human activity classification tasks. A diverse UWB CIR dataset in real-world indoor settings is collected to evaluate the performance in recognition of six independent activities. Our novel approach treats each CIR as a unique snapshot, capturing specific activities and creating a dynamic activity representation through concatenated feature vectors. A thorough grid search is conducted to explore optimal parameters in constructing these vectors. In applying eleven different learning models for classification analysis on the dataset, it is generally observed that deep learning methods yielded enhanced classification accuracy compared to traditional machine learning techniques. Our research demonstrates the potential of UWB in HAR and provides insights into effective feature vector creation and model performance. The code and dataset are publicly accessible at `https://bit.ly/422WJEB`

4.1.1 Introduction

Human Activity Recognition (HAR) plays an important role in many applications, ranging from smart homes (Bouchabou et al. 2021), crowd behaviour surveillance (Sun et al. 2020) to remote healthcare monitoring and assisted living (Taylor et al. 2020; Yoshida et al. 2022). In modern office environments, prolonged sedentary behaviour has become an increasing health concern, particularly among desk-based workers (Wu et al. 2023). The detection and monitoring of extended sitting periods represents an important application of HAR to provide timely intervention for decreasing health risks (Mekruksavanich et al. 2018). HAR refers to inferring human physical behaviour from various sensor data through different approaches. The common approaches for HAR can be categorised into sensor-based, vision-based, and device-free approaches (Dang et al. 2020; Hussain et al. 2020; Yu et al. 2016). The sensor-based methods recognise the activities from the data collected by sensors that are embedded in smartphones or other wearable devices carried by the user (Straczkiewicz et al. 2021; Zhang et al. 2022b). This means the user is expected to carry devices all the time. Also, the cooperation of the users is necessary for the success of HAR (Wang et al. 2016). This may not be possible for some HAR application scenarios, such as fall detection in care homes, disaster rescue, and outliers detection in security surveillance. Another widely used group of approaches for HAR is vision-based approaches (Beddiar et al. 2020; Dang et al. 2020). However, the privacy concerns associated with visual data and the relatively high cost of implementation associated with cameras can introduce significant challenges to the adaptation of vision-based HAR (Beddiar et al. 2020; Demrozi et al. 2020; Adachi et al. 2021). In addition, the achievable accuracy may vary according to the occlusion and lighting conditions thus making this approach potentially unreliable (Dang et al. 2020). On the other hand, Device-free HAR infers the users' activities from radio-frequency signal patterns (Hussain et al. 2020). In this regard, the wireless signals for HAR may provide more flexibility and consistent accuracy while not requiring users to carry any hardware or cooperate (Xue et al. 2020).

The pervasive presence of radio-frequency signals, such as WiFi and Ultra-Wideband (UWB), in areas frequently occupied by humans makes these technologies promising assets for device-free HAR, with minimal associated costs in implementation and deployment.

In WiFi signals, the Channel State Information (CSI) and Received Signal Strength Indicator (RSSI) are commonly used to infer the users' activities while interacting with the environment. For example, Wang et al. 2017a developed a system using WiFi devices to estimate human activities, where an accuracy of 96% was reported. In (Wang et al. 2022a), a system for identifying intruders from their gait feature patterns of CSI was proposed. In addition to WiFi CSI, Hsieh et al. 2020 used the WiFi RSSI to distinguish between moving and stationary activities.

While WiFi can provide acceptable activity classification accuracies, extracting CSI and RSSI from off-the-shelf WiFi routers usually requires dedicated tools, introducing additional complexity. The growing availability of commercial UWB devices, such as the EVK1000 by Decawave (now part of Qorvo) (Qorvo n.d.), enhances the feasibility of implementing UWB wireless systems in diverse applications. Compared to WiFi systems, it provides straightforward methods to access relevant information. This increased accessibility opens up new routes for HAR using UWB technology, indicating a growing potential for its application in this field.

UWB is characterized by the transmission of extremely narrow pulses over a wide bandwidth exceeding 500 MHz (Commission et al. 2002). Its application in ranging and indoor positioning is well-noted, especially attributed to its robust resistance to interference from narrow-band signals, high resolution in the time domain, and resilience to multipath effects in dense environments (Zafari et al. 2019). Although UWB technology has achieved notable success in positioning and navigation with centimetre-level accuracy (Sung et al. 2023; Malajner et al. 2015) and increasing integration into consumer electronics like iPhones and AirTags, its potential in HAR has not been fully leveraged compared to more established technologies such as WiFi (Bocus et al. 2021). Given its broader bandwidth in the frequency domain and higher resolution in the time domain (Molisch 2009) as well as its growing availability, UWB stands out as a particularly promising alternative for undertaking HAR.

Human interaction with the environment alters the propagation of the signals, which can be revealed by the patterns in Channel Impulse Response (CIR) data. Based on the inverse relationship between time and bandwidth, the CIR of UWB systems has a very high resolution in the time domain due to its wide bandwidth, which allows for detecting propagation variations more accurately (Molisch 2009; Li et al. 2022b).

In this paper, we employ fine-grained UWB CIR data to recognise the users' activities performed in three indoor environments representing different levels of complexity. The UWB's CIR data are recorded while six types of human activities are performed. They include waving hands, walking, sitting down, and standing up, i.e., the four dynamic activities, as well as remaining seated and standing still, i.e., the two stationary activities. The background CIR for the layout without human interaction, i.e. the empty room of each indoor scenario, is also recorded and included in the dataset as an additional class. Then we evaluate the performance of eleven supervised learning models in terms of their classification accuracy for different activities.

In this paper, we make the following contributions:

- We introduce an innovative approach to pre-process CIR data, akin to video creation, for generating feature vectors from CIR snapshots.
- Our study systematically evaluates various learning models across different indoor scenarios by varying the length of CIR and the number of CIR snapshots. These findings are crucial for enhancing the energy efficiency of real-world, real-time HAR systems.

- A dataset comprising UWB CIR recordings of six human activities performed by volunteers, as well as background CIR recordings in three distinct indoor layouts is collected. This dataset and the associated code developed for our research are now publicly accessible.
- By feeding the generated feature vectors into our proposed learning models, we transfer the HAR task into a sequence classification problem, reducing the computational complexity for HAR tasks.
- Model Evaluation: We extensively evaluate eleven traditional and deep learning models to evaluate their achievable accuracies for activity classification. This evaluation provides us with insight into the performance of various models for HAR using CIR data.

The rest of this paper is structured as follows: Section 4.1.2 provides an overview of the literature on using UWB CIR in HAR tasks. The paper's methodology is outlined in Section 4.1.3, including UWB CIR introduction, signal pre-processing, and concatenation for feature vector generation, and concludes with employed classification models. Section 4.1.4 provides details of the experiment setup employed in this study. Section 4.1.5 presents the performance of the models in HAR, and Section 4.1.6 provides our conclusion and discussion. And finally, we have Section 4.1.7 outlining future work.

4.1.2 Related Work

The advantages of UWB, such as wide bandwidth and high temporal resolution, can provide more accurate ranging capabilities and less interference with narrow-band signals, making UWB a promising technology for positioning (Mazhar et al. 2017).

However, compared with extensive research on using UWB for positioning purposes, its application on HAR is still emerging. Gorji et al. 2021; Peng and Guo 2020; Du et al. 2020 exploited UWB radar and machine learning-based solutions to classify different human activities. In (Cheng et al. 2020), a range measurement-based system was developed to localise and recognise human activities in an indoor environment simultaneously. The accuracy of the one-against-all experiment was reported at 80.2%, 68.8%, and 77.4% in support vector machine, hidden Markov model and artificial neural network, respectively. Sharma et al. 2019 extracted CIR data from Off-the-shelf UWB devices to build a device-free HAR system with four simple machine-learning algorithms and achieved an overall accuracy of 95.6% for the line-of-sight scenario. However, this work included only four stationary human activities. Bocus et al. 2021 compared the performance of HAR by WiFi CSI and UWB CIR under the same environment layout through six classification algorithms to classify five activities. They concluded that using UWB for HAR can provide a better classification performance than WiFi and requires less signal preprocessing. However, both works of (Sharma et al. 2019; Bocus et al. 2021) did not provide detail in signal pre-processing for feature generation, and they conducted their experiment solely under the line of sight scenario. This limits the applicability in different environment settings and reduces the robustness of the proposed systems.

4.1.3 Methodology

In this section, we depict our proposed methodologies, starting with a brief introduction of UWB CIR in Section 4.1.3.1, followed by a representation of signal pre-processing techniques and the details of applied learning models for activity classifications. The sequential framework of our system design is illustrated in Figure 4.1. Initially, a data collection phase was undertaken, during which CIR samples were recorded as subjects performed predefined activities, creating our raw CIR dataset. The dataset was then annotated with corresponding activity labels for further processing, as elaborated in Section 4.1.3.2. After pre-processing, which includes phases of noise filtering and concatenation of CIR snapshots following data labelling, the dataset was split into training and test sets. Subsequently, we trained eleven machine learning models, encompassing both traditional machine learning and deep learning algorithms, as detailed in Section 4.1.3.3. These models were then applied to the test set for activity label prediction. The final phase was the evaluation of each model's performance, with a focus on classification accuracy.

4.1.3.1 Channel Impulse Response

Within a multipath wireless channel, the signal can propagate from transmitter to receiver through different paths due to the scattering, diffraction, and reflection caused by the objects or humans present in the environment (Molisch 2009). Each multipath component (MPC) is a replica of the signal with a certain level of attenuation and delay, and the received signal can be described as the sum of all MPCs. Consequently,

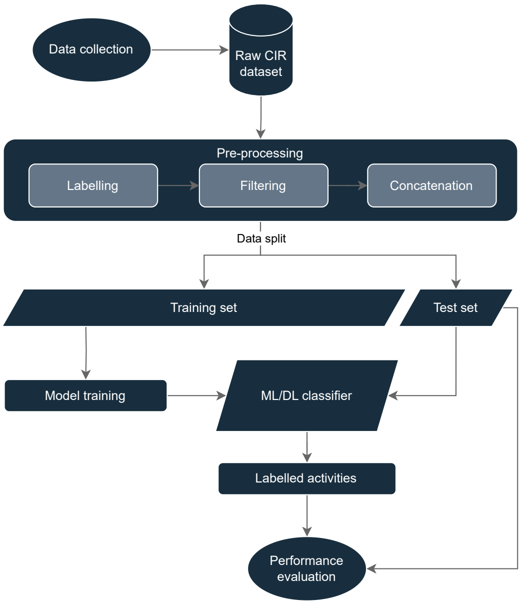


Figure 4.1: The illustration of system design

the channel impulse response can be represented as:

$$h(t) = \sum_{i=1}^{N} \alpha_i \delta(t - \tau_i) \tag{4.1}$$

where $\delta(.)$ refers to the Dirac function, N is the total number of MPCs within the multipath channel, α_i and τ_i are the gain and delay of ith MPC, respectively. As the CIR is used to depict the wireless channel, its pattern varies accordingly when the layout changes or the objects move within the surroundings. For example, Figure 4.2 compares two denoised background CIRs collected from both line-of-sight (LOS) and non-line-of-sight (NLOS) scenarios where no activity was performed. Only the first 150 samples starting from the first path are used for this comparison. By looking into the profile of both curves, the CIR of the LOS condition has fewer MPCs than NLOS, which can be observed from its clearer first path and less significant tails than NLOS. This observation confirms that the CIR can reflect environmental change. Intuitively, as human activities performed within the environment also change the propagation conditions, they can be interpreted from pattern variations of the CIR.

The hardware we used in this work is EVK1000. It contains two 802.15.4a standard evaluation boards equipped with a Decawave DW1000 chip. When activated, they continuously exchange messages and provide calculated ranging measurements through time-of-flight and two-way-ranging protocol. From the perfect periodic autocorrelation property of the preamble code of the UWB system, the receiver can determine the exact CIR (Qorvo 2025), which can be extracted from the EVK1000 for further processing. In

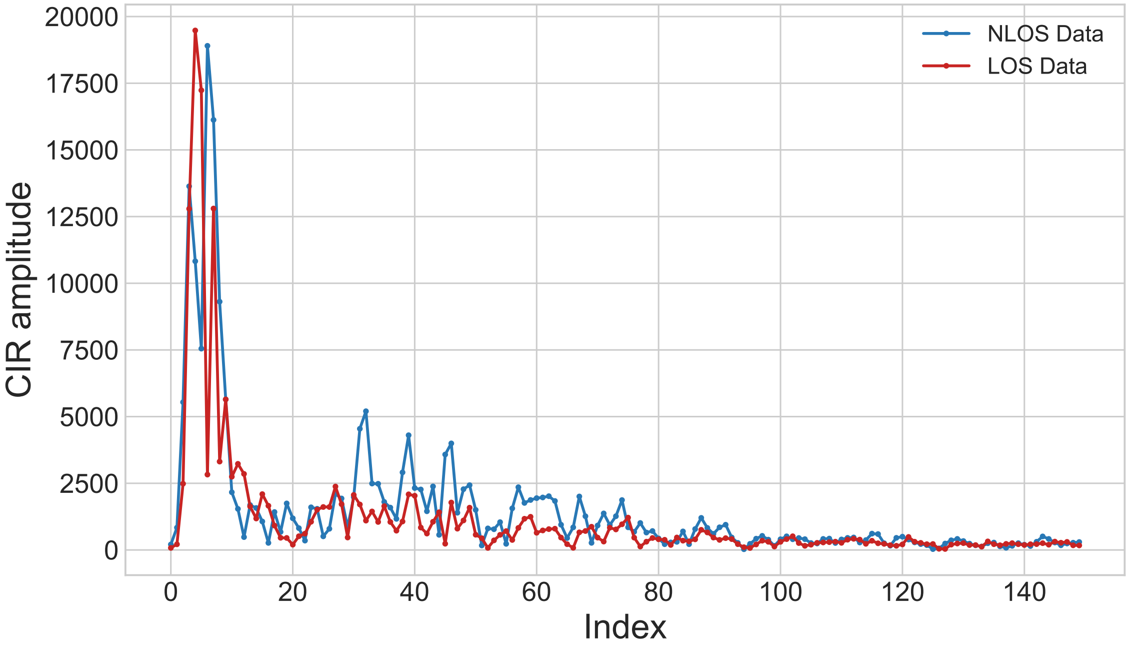


Figure 4.2: Comparison of background CIR recording, i.e., empty room, for LOS and NLOS scenarios

this work, we employed the nominal 64 MHz pulse repetition frequency corresponding to the CIR length of 1016 samples. Each sample has a 16-bit real (R) and 16-bit imaginary (I) integer. This paper uses the square root of the sum of the squares of its real and imaginary parts, i.e., $|h| = \sqrt{R^2 + I^2}$ for each sample to generate feature vectors for classification algorithms.

4.1.3.2 Signal Preprocessing

Filtering Each raw CIR snapshot of EVK1000 represents 1016 samples. The raw CIR data is noisy, and only a subset of the sequence is pertinent for activity recognition. To achieve a relatively high performance for the classification, we included the most informative subsequence among the entire CIR. Creating such a subsequence requires identifying both the rise time index and the optimal length. We employed a similar method to the rise time identification algorithm in (Marano et al. 2010; Vales et al. 2020) to explore the beginning index of the rise time n_{rise}. First, we define the threshold index n_H as follows:

$$n_H = \min\{n : |h(n)| \geq \beta |h_{max}|\} \tag{4.2}$$

where h_{max} is the maximum value of the entire processed CIR snapshot sample and $\beta \in [0,1]$ is a threshold coefficient. The objective of (4.2) is to localize the minimum index where its corresponding value exceeds the defined threshold. Then n_{rise} is expressed as:

$$n_{rise} = n_H - n_{offset} \tag{4.3}$$

where n_{offset} is an offset prior to n_{rise}. In this work, we used $n_{offset} = 3$ and $\beta = 0.4$ empirically. Consequently, the CIR snapshot $p(n)$ used for the classification algorithm is:

$$p(n) = |h(n)|, \quad n_{rise} \leq n \leq n_{rise} + L - 1 \tag{4.4}$$

where $h(n)$ represents a raw CIR snapshot, and L is the size of the cropping window for tailoring it. To summarize, $p(n)$ is generated from cropped $|h(n)|$, starting from the index n_{rise} with length L.

Concatenation During the experiment, the EVK1000 system recorded the CIR concurrently with the execution of each activity. The CIR, capturing the channel condition at the moment it is measured, effectively represents a snapshot of the activity at the corresponding timestamp t. By sequentially concatenating these CIR snapshots $p(n)$ in chronological order, we construct a continuous sequence that mirrors the process of creating a video from individual frames. This methodology allows us to transform the activity recognition task into a sequence classification problem. The concatenated sequence $\tilde{p}$ is expressed as:

$$\tilde{p} = [p_1(n), p_2(n), ..., p_t(n), ..., p_M(n)] \tag{4.5}$$

where $p_t(n)$ represents the tth processed snapshot with length L derived through (4.2), (4.3) and (4.4), M is the total number of CIR snapshots generated from recording during an activity. Consequently, the overall length of each composite feature vector, $\tilde{p}$, is given by $M \times L$. The choice of sequence length L for generating feature vectors is crucial for classification accuracy, as it directly influences the balance between the noise level and the informative content within each $p_t(n)$. Optimal selection of L depends on factors such as the complexity of the indoor environment and the system's sensitivity threshold. Similarly, the value of M is pivotal as it determines the count of CIR snapshots concatenated to form a single feature vector. Ideally, M should be set to encompass a sufficient number of the most informative CIR snapshots that implicitly contain relevant activity information. Section 4.1.5.1 delves deeper into determining the optimal values for L and M for this specific purpose.

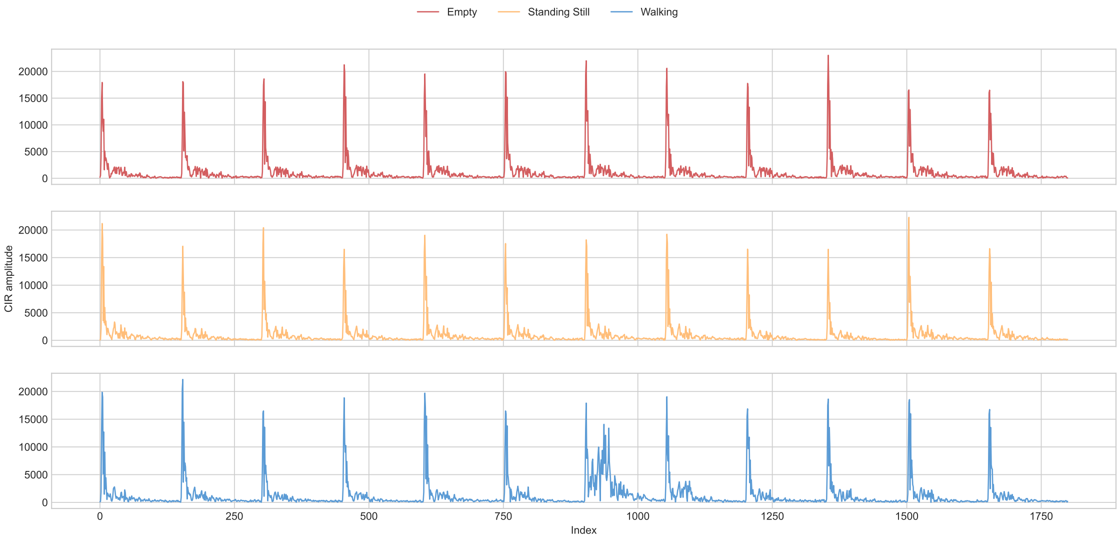


Figure 4.3: Illustration of generated feature vectors for three distinct activities.

In Figure 4.3, we present three feature vectors, corresponding to the classes of empty room, standing still, and walking, derived using equations (4.2), (4.3), (4.4), and (4.5), with parameters $\boldsymbol{L}$ and $\boldsymbol{M}$ specifically set at 150 and 12 for this illustration. It can be observed that the feature vectors for the empty room and standing still classes exhibit limited variation across consecutive CIR snapshots, aligning with the stationary nature of the two classes. Their difference can be observed in the tail region of the CIR snapshots, where the distribution of MPCs differ between the two classes. Although both the empty room and standing still scenarios correspond to temporally stationary environments, they reflect distinct physical conditions under which the signal is transmitted. In the empty room case, the tailing MPCs are dominated by static reflections from surrounding surfaces, whereas the standing still scenario introduces additional body-induced scattering and attenuation. Therefore, these consistent differences in the MPC tail profiles provide discriminative features that can be learned by the algorithm, despite the absence of dynamic motion over time. In contrast, the walking feature vector exhibits significant variations in its pattern, particularly evident from index 800 onwards. This variation is likely due to the volunteer moving closer to the line of sight during the recording, resulting in the detection of additional MPCs in the CIR snapshots at that specific moment.

4.1.3.3 Classification Models

In order to evaluate the HAR classification performance, we applied eleven supervised learning models to the recorded dataset. We performed an extensive literature review and meticulously selected these highly adopted models in HAR and time-series classification tasks. The models are further categorised into two families: four traditional machine learning (ML) and seven deep learning (DL) models. We considered Support Vector Machine (SVM), k-nearest Neighbors (kNN), Naive Bayesian Classifier (NBC) and Random Forest (RF) in ML models, and Multilayer Perceptron, One-dimensional Convolutional Neural Network, Fully Convolutional Neural Network (Wang et al. 2017b), Recurrent Neural Network, Long-short Term Memory (Hochreiter and Schmidhuber 1997), Gated Recurrent Unit (Cho et al. 2014) and Long-short Term Memory-Fully Convolutional Networks (Karim et al. 2017) in DL models. The details of DL model architectures are described as follows:

- ***Multilayer Perceptron (MLP)***
  MLP is a well-established neural network architecture that has gained popularity in classification tasks. The MLP classifier comprises one input layer, which propagates the input data into the first hidden layer for further processing. The input layer is typically followed by multiple stacked, fully connected hidden layers that apply weights to their inputs and direct the weighted inputs to non-linear activation functions. These activation functions perform non-linearity transformations on the input, enabling the network to model complex relationships between the input and output. The MLP employed in this paper consists of three fully connected hidden layers of 1024, 128, and 16 units with Rectifier (ReLU) activation function and one softmax layer as the output layer of the model.
- ***One-dimensional Convolutional Neural Network (1D-CNN)***
  CNN was initially designed for the image recognition task. By employing convolutional kernels and pooling operations, CNN learns representation from data. In this model, we used two one-dimensional convolutional layers consisting of 6 and 16 filters, respectively. The kernel size is three for all the filters. After each convolution layer, a max-pooling layer with a pooling size of 3 was applied. The final max-pooling layer's output was flattened and passed to a fully connected layer comprising 128 units. The last layer is a softmax layer for classification.
- ***Fully Convolutional Neural Network (FCN)***

Similar to the classic CNN model, FCN relies on kernels to perform convolution operations and does not use maximum pooling for downsampling. Instead, it employs a global average pooling layer before the final softmax layer to prevent overfitting, and a batch normalization layer follows each convolutional layer to improve convergence speed and generalization. The FCN model employed by this work comprises three convolutional layers with sequential connection, using kernel size of {8, 5, 3} and filter numbers of {128, 256, 128}, respectively.

- ***Recurrent Neural Network (RNN)***
  RNN is a deep learning model designed to handle sequential data by capturing temporal dependencies from memorizing previous input information. However, RNNs are limited in capturing long-term dependencies in the sequential data due to gradient vanishing, which may occur as the network gets deeper. We employed the RNN model with one layer of 8 hidden units, followed by a softmax classification layer.
- ***Long-short Term Memory (LSTM)***
  As a variant of RNN, LSTM was designed to overcome gradient vanishing by incorporating gating mechanisms within its memory cell. These mechanisms allow the LSTM to preserve or forget information from sequential input, enhancing its capability to learn long-term dependencies. We used an LSTM model with 64 memory cells as feature extractors before applying the softmax classifier in the end.
- ***Gated Recurrent Unit (GRU)***
  GRU is another modified version of the RNN, which simplifies the design of LSTM memory cells by having fewer gate functions. Despite this simplification, the GRU retains the ability to learn long-term dependencies in sequential data. The GRU model employed in this work consists of 64 GRU memory cells.
- ***Long-short Term Memory-Fully Convolutional Networks (LSTM-FCN)***
  This model leverages both LSTM and FCN architectures in one model and concatenates the outputs of LSTM and FCN blocks before the final softmax layer.

The Adam optimiser was utilised in training all the deep learning models. All of the deep learning models were trained for 100 epochs with an initial learning rate of 1e-3, which was decreased by a factor of 0.5 if no improvement in the loss function was observed after four epochs, with the final learning rate set to 1e-6.

4.1.4 System Implementation and Experiment

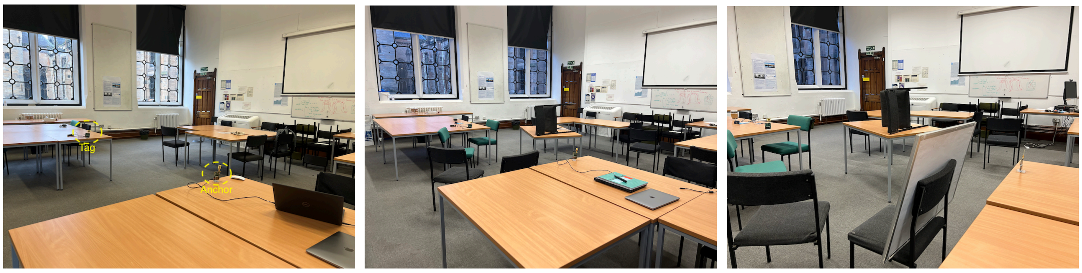
4.1.4.1 System Description

In order to implement the proposed framework, we use two boards of EVK1000 to build a UWB wireless link; one of the boards is configured as an anchor and the other as a tag. Both boards are configured with a bandwidth of 500 MHz at a centre frequency of 4.0 GHz, a data rate of 110 kbps, a pulse repetition frequency of 64 MHz, and a preamble length of 1024 bits. A Windows-running PC that is connected to the anchor via a USB cable for extracting the calculated CIRs during each activity. The two boards were installed at a distance of 2.9 m and remained fixed during the whole experiment.

4.1.4.2 Data Collection

The objective of data collection is to build a CIR dataset for assessing the classification performance of proposed models. The experiment is conducted in a large seminar room equipped with furniture. During the experiment, the activities were performed within the vicinity bounded by the EVK1000 boards. To understand how the complexity of the environment layout impacts the HAR task, we run the experiment under three different indoor settings, namely LOS, complex LOS, and NLOS. These layouts are shown in Figure 4.4.

The LOS scenario refers to the setting where an unimpeded propagation path between two boards without any objects within a two-meter range. Whereas complex LOS involves the presence of surrounding objects that induce additional MPCs in the corresponding CIR. NLOS employs the same layout as complex LOS but blocks the direct line of sight using a whiteboard.



(a) LOS (b) Complex LOS (c) NLOS

Figure 4.4: Three different layouts during the experiment: (a) LOS: No object was placed in the middle of two boards. (b) Complex LOS: The line of sight was unobstructed but with some surrounding objects placed around. (c) NLOS: Same layout as complex LOS but the line of sight was blocked by a whiteboard.

In this study, volunteers were invited to perform six activities within each layout individually: walking, waving hands, sitting down, standing up, sitting still, and standing still. These activities were selected as they encompass stationary and dynamic actions, ranging from small-scale movements such as waving hands to large-scale motions involving the entire body. Each activity's start and end of CIR recording were manually controlled using a laptop connected to one of the two boards.

More specifically, for the walking activity, volunteers were instructed to traverse the line-of-sight, delineated by two boards. The walking direction was left to the discretion of each volunteer. Waving hands was performed in standing positioning. Sitting down and standing up were antithetical activities performed on a chair placed adjacent to the line of sight. Sitting still and standing still were stationary activities that may involve only subtle body movements. Each volunteer performed each activity 40–50 times under each layout condition, with approximately three seconds of CIR data recorded for each trial. For short-duration activities such as "sitting down" and "standing up", this three-second window does not imply that the activity itself lasts for the full duration, but is designed to ensure that the complete duration and corresponding surrounding signal dynamics are captured. The receiver continuously recorded the CIR changes during these activities. Additionally, the background CIR for each layout was recorded in the absence of human interaction, serving as the seventh activity for classification purposes in the dataset. The collected dataset was subsequently divided into 80% for training and 20% for testing purposes.

4.1.5 Results

4.1.5.1 Optimal Feature Vector Dimension

To evaluate our CIR snapshot processing approach, we trained various learning models on feature vectors derived from these snapshots. These feature vectors were generated using different combinations of Length_CIR (L), and Number_CIR (M), where L denotes the cropping window length applied to each snapshot, and M represents the count of CIR snapshots utilised per feature vector. Our experimental design included a range of L values from 50 to 225 in increments of 25 and M values from 8 to 12 in increments of 2.

For each combination of L and M, and under each indoor scenario, a distinct learning model was trained, resulting in a total of 24 models per scenario. Each model's performance was evaluated using a corresponding test set, with the classification accuracies depicted via heatmaps in Figure 4.5.

A notable trend observed is an increase in model classification accuracy with the length of CIR, i.e., L, peaking at values of 100 or 125. Beyond this peak, a decline in accuracy was noted as L approached the maximum experimental value of 225. This suggests that CIRs cropped down to lengths between 1016 and 125, starting from the rise time, are optimal for capturing the most informative and least noisy samples for our HAR tasks.

Increasing the value of M from 8 to 10 generally resulted in a modest improvement in accuracy. Further, while the increments of M from 10 to 12 yielded limited enhancements, these adjustments still contributed positively to the model's performance. This trend highlights the nuanced impact of varying M, suggesting that even small changes in M can lead to discernible improvements in HAR models within indoor environments. Therefore, selecting appropriate values for both the length of individual CIR snapshots and the number of snapshots is crucial for optimizing the effectiveness of HAR models.

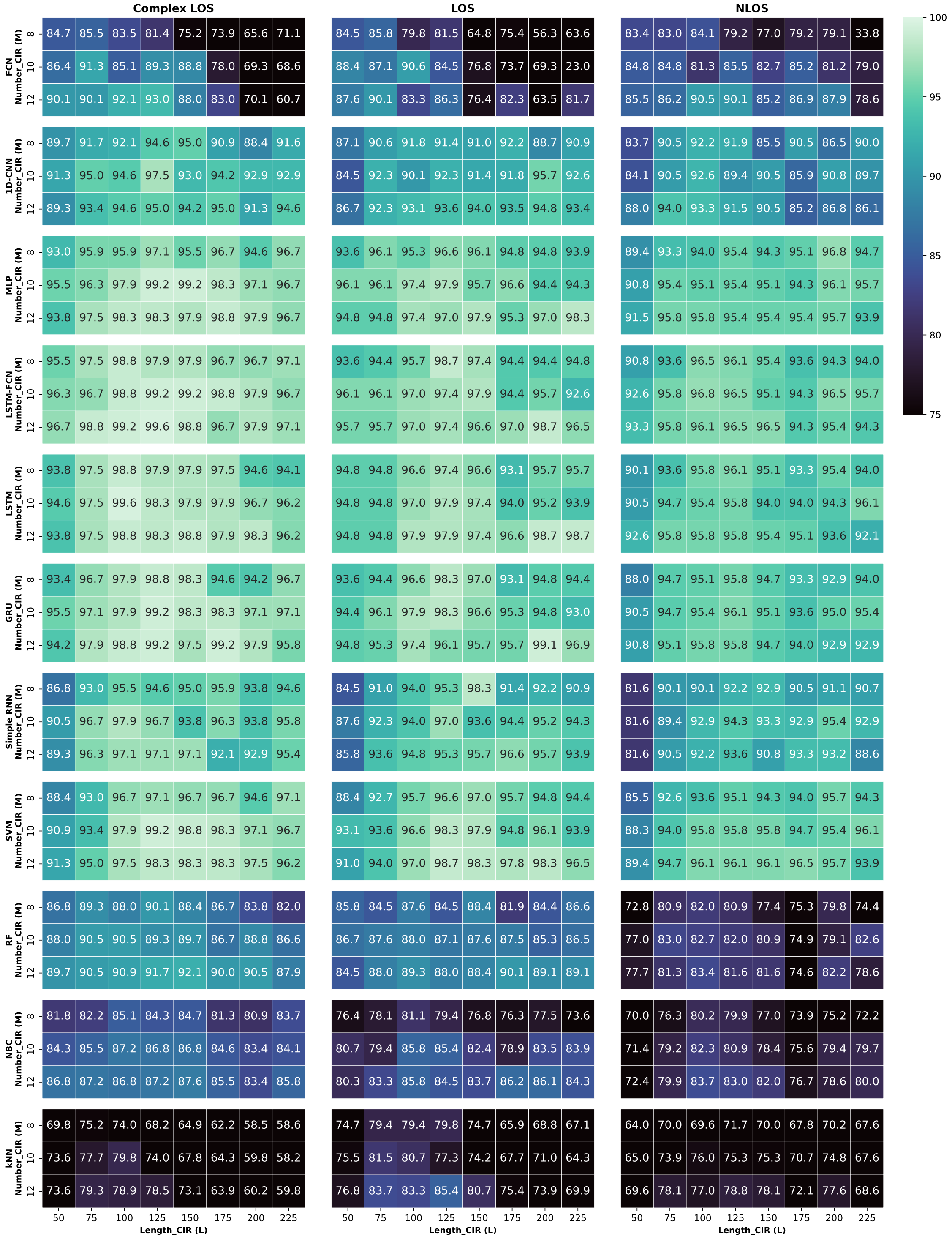


Figure 4.5: Classification accuracy of various models on test sets across three indoor scenarios: An evaluation of different Length_CIR (L) and Number_CIR (M) combinations.

In the following performance evaluation, we use the optimal choices of $L = 125$ and $M = 12$ as the cropping window size for each CIR sample and the number of CIR snapshots included in each feature vector to summarize the performance. The results are shown in figure 4.5,

4.1.5.2 Model Performance Evaluation

Our work collects the data under three indoor layouts to evaluate the model performance in diverse propagation conditions. Then, the raw CIR data recorded during activity performance are preprocessed and concatenated to generate feature sequences. Finally, we feed the dataset into eleven traditional machine learning and deep learning models to classify six human activities as well as the background scenario, i.e., layout with no human activity.

The models were individually trained on datasets collected from each layout in standalone mode. Consequently, the testing set was fed into the trained models for assessing classification accuracy. Table 4.1 summarizes the performance of all eleven models across three layouts, classified into DL and ML categories.

In general, the group of DL models outperforms ML models by their overall higher average classification accuracy on the testing set. Within the DL group, the two sequential neural networks, LSTM-FCN and LSTM, achieved the highest average classification accuracies across all three layouts, i.e. 97.83% and 97.32%, respectively. Specifically, LSTM-FCN performs the highest accuracies for complex LOS and NLOS layouts; LSTM produces the second-best classification accuracy in LOS. The sub-group of MLP, GRU, and simple RNN models produce slightly lower accuracy than the two LSTM-based models. Simple RNN produces the lowest average accuracy of 95.34% within this mini-group. FCN and 1D-CNN, two convolutional models, exhibit the lowest classification accuracy within the DL group, with FCN reporting the poorest accuracy of 86.27% in the complex LOS layout.

Table 4.1: Classification accuracy on testing set

	Models	**Complex LOS**	**LOS**	**NLOS**	**Model average**
Deep Learning	FCN	92.98%	86.27%	90.11%	89.79%
	1D-CNN	95.04%	93.56%	91.52%	93.37%
	MLP	98.35%	97.00%	95.41%	96.92%
	LSTM-FCN	**99.59%**	97.42%	**96.47%**	97.83%
	LSTM	98.35%	97.85%	95.76%	97.32%
	GRU	99.17%	96.14%	95.76%	97.02%
	Simple RNN	97.11%	95.28%	93.64%	95.34%
Deep learning average		97.23%	94.79%	94.10%	–
Machine Learning	SVM	98.35%	**98.71%**	96.11%	97.72%
	RF	91.74%	87.98%	81.63%	87.12%
	NBC	87.19%	84.55%	83.04%	84.93%
	kNN	78.51%	85.41%	78.80%	80.91%
Machine learning average		88.95%	89.16%	84.90%	–

Bold and Underline denote the highest and second-highest accuracy across all the models applied on the layout, respectively ($M = 12$, $L = 125$).

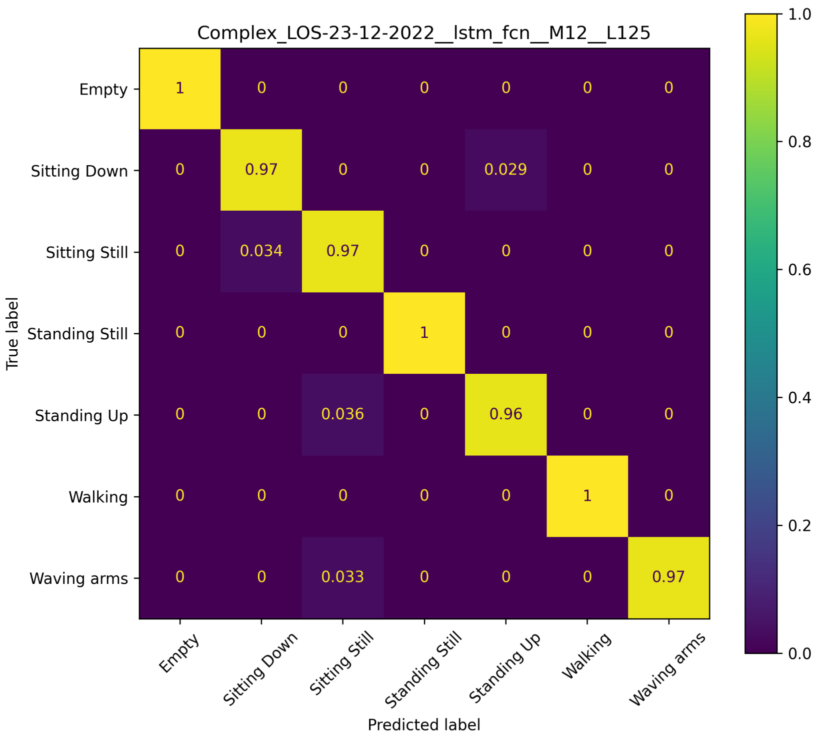


Figure 4.6: Confusion matrix for the LSTM-FCN model under the Complex-LOS scenario ($M = 12$, $L = 125$).

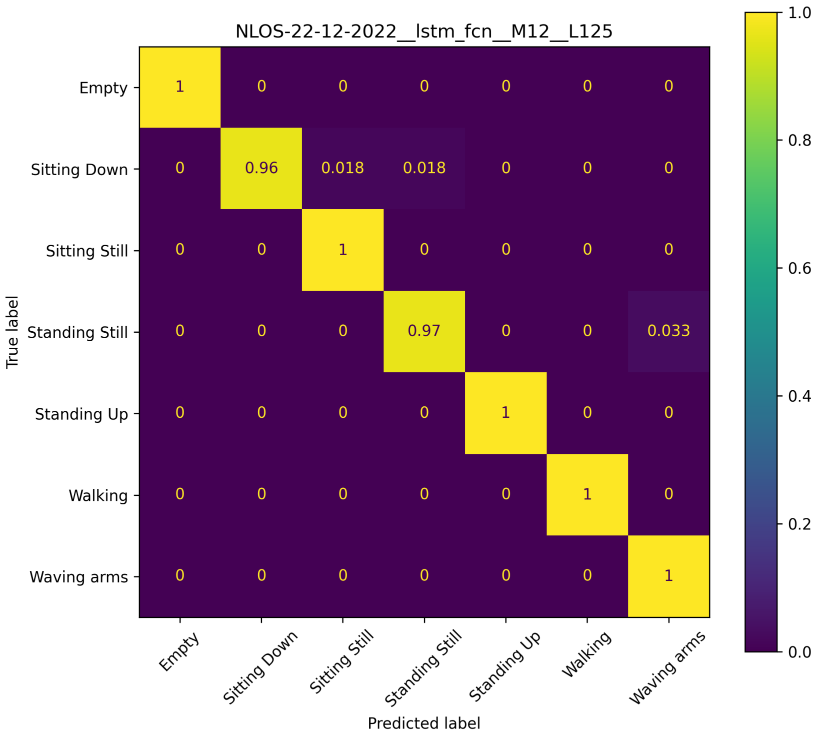


Figure 4.7: Confusion matrix for the LSTM-FCN model under the NLOS scenario ($M = 12$, $L = 125$).

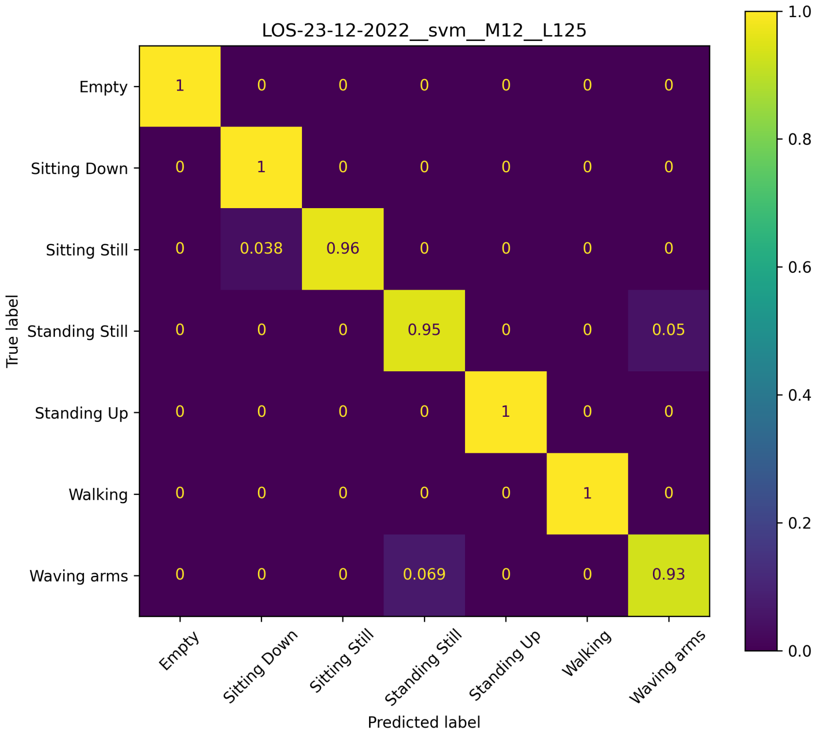


Figure 4.8: Confusion matrix for the SVM model under the LOS scenario ($M = 12$, $L = 125$).

Table 4.2: Class-wise precision, recall, and F1-score for the LSTM-FCN model under the Complex-LOS scenario ($M = 12$, $L = 125$).

Activity	Precision	Recall	F1-score
Empty	1.00	1.00	1.00
Sitting Down	1.00	0.96	0.98
Sitting Still	0.97	1.00	0.99
Standing Still	0.97	0.97	0.97
Standing Up	1.00	1.00	1.00
Walking	1.00	1.00	1.00
Waving arms	0.98	1.00	0.99

Table 4.3: Class-wise precision, recall, and F1-score for the LSTM-FCN model under the NLOS scenario ($M = 12$, $L = 125$).

Activity	Precision	Recall	F1-score
Empty	1.00	1.00	1.00
Sitting Down	0.97	1.00	0.99
Sitting Still	1.00	0.96	0.98
Standing Still	0.95	0.95	0.95
Standing Up	1.00	1.00	1.00
Walking	1.00	1.00	1.00
Waving arms	0.93	0.93	0.93

Table 4.4: Class-wise precision, recall, and F1-score for the SVM model under the LOS scenario ($M = 12$, $L = 125$).

Activity	Precision	Recall	F1-score
Empty	1.00	1.00	1.00
Sitting Down	0.97	0.97	0.97
Sitting Still	0.93	0.97	0.95
Standing Still	1.00	1.00	1.00
Standing Up	0.96	0.96	0.96
Walking	1.00	1.00	1.00
Waving arms	1.00	0.97	0.98

SVM seems to be the best-performing model across all the ML and DL models in the LOS scenarios, with achievable accuracies of 98.71%. It maintains its high performance, i.e. second best performing, in the NLOS scenario with the achievable accuracies of 96.11%. This observation may contribute to the fact that SVM could perform well on high-dimensional data (Schölkopf and Smola 2005). Random forest is the second-best model within the ML group, with an average accuracy of 87.12%. However, RF can only mark the lowest accuracy compared to DL group models. NBC and kNN demonstrate the lowest performance among all the employed models, with less than 85% of activities classified correctly. Confusion matrices (Figure 4.6, 4.7, 4.8), and corresponding class-wise evaluation tables (Table 4.2, 4.3, 4.4) are reported for three representative configurations, spanning both deep learning and machine learning models. These examples are selected to illustrate typical misclassification patterns without exhaustively enumerating all possible model–scenario combinations.

4.1.6 Conclusion & Discussion

In this paper, we have developed an innovative method for pre-processing CIR data to generate feature vectors akin to a video creation process. This methodology has been rigorously evaluated across various indoor settings. To ascertain the most effective number of CIR snapshots and the optimal length of each sample, we conducted a detailed grid search. Due to the nature of our used UWB CIR data, this stage reduced significantly the computational complexity of our proposed methodology by excluding the noise of the data and saving the most informative samples as much as possible. Then, the learning models were evaluated based on the explored optimal parameters. We have curated a comprehensive dataset of UWB CIR recordings, capturing six human activities in three distinct indoor environments, which, along with the code, is made publicly available for future work and reproducibility purposes. One advancement in our research is the transformation of HAR into a sequence classification problem using UWB CIR data. This approach simplifies the process compared to using other multi-dimensional data sources. Furthermore, we have thoroughly evaluated eleven traditional and deep learning models, with an emphasis on their ability to accurately classify activities using UWB CIR data.

Our findings revealed that sequential models, particularly those based on LSTM, consistently outperformed others in this HAR task. Their performances are attributable to their inherent ability to capture temporal dependencies within the sequential feature vectors. In contrast, models employing fully convolutional networks, such as FCN, did not yield as favourable results as those combining convolutional layers with either a fully connected layer (1D-CNN) or LSTM blocks (LSTM-FCN). The least effective models were NBC and kNN, with NBC's assumption of feature independence proving unsuitable for our dataset and kNN's struggles likely due to the high-dimensional nature of our feature vectors and the variance in CIR samples.

The consistency in model performance across different layouts suggests that despite varying propagation conditions, our data processing methodology enables the models to discern feature patterns spreading across the entire sequence rather than relying solely on information from individual snapshots. The highest achieved individual classification accuracy for complex LOS, LOS, and NLOS scenarios are 99.59%, 98.71%, and 96.47%, respectively. This finding underscores the effectiveness of our approach in adapting to diverse environmental conditions within HAR applications.

4.1.7 Future Work

Although traditional machine learning and deep learning models have shown promise in classifying human activities from UWB CIR data, they require extensive labelled datasets for training, which can be costly and time-consuming to collect. Moving forward, we aim to explore the potential of transfer learning in HAR. This involves applying a pre-trained model on a dataset from one specific layout to other datasets with different layouts, reducing the need for extensive new data labelling. Additionally, our future work includes expanding the scope of our study to encompass more complex human activities and scenarios involving multiple volunteers performing activities simultaneously. Another area of interest is enhancing data processing by exploring techniques to filter out environment-dependent background noise from the CIR samples. This will allow us to focus more on extracting and classifying information solely related to human activities, potentially increasing the accuracy and applicability of our models in varied settings. Future work will investigate principled approaches for identifying activity-relevant segments within full CIR recordings, enabling reduced input dimensionality and computational cost without requiring fixed recording windows. In addition, in the current experimental setup, the two boards were placed at a fixed distance of 2.9 m.

Varying this distance is expected to introduce changes in wireless channel conditions, which may affect the resulting CIR patterns. The impact of such distance variations on algorithm performance remains to be investigated in order to assess the robustness of the proposed approach.

4.2 Why Do Pedestrians Get Lost? A Case Study of Personal, Situational, and Environmental Factors in Greater London

Publication and Permission Statement

This chapter is based on an accepted manuscript of the following publication:

Yu Wang, Ana Basiri, Petrus Gerrits, Guy Solomon, Sascha Woelk, and Miguel Fidel Pereira (2025). *Why Do Pedestrians Get Lost? A Case Study of Personal, Situational, and Environmental Factors in Greater London. Journal of Location Based Services.*

At the time of thesis submission, this article has been accepted for publication but has not yet been formally published. The chapter reproduces the author's accepted manuscript, which is permitted for inclusion in a doctoral thesis. The content has been reformatted to conform to the thesis style.

Abstract

Despite significant advances in route optimization and navigational technologies, pedestrians still face challenges navigating complex urban spaces. This study aims to identify and quantify the multidimensional factors that may influence pedestrian navigation, with the focus on why people get lost. To understand the phenomenon of "getting lost", we collected data from an online survey in which 64 participants reported the locations and contextual information of where they had got lost. Building upon literature, expert interviews, and collected data, we identify and quantify fourteen environmental, situational, and personal factors influencing pedestrian navigation. We utilize a dual-analytical approach that combines expert-led Analytic Hierarchy Process (AHP) analysis with data-driven regression models to derive distinct weighting schemes for the factors. While the approach based on experts' opinion (i.e., AHP) demonstrates that familiarity, self-orientation skills, and access to reliable navigation tools are the most important contributing factors associated with why we get lost, data-driven models additionally highlight the significance of environmental complexity, such as angular distance between exits and number of landmarks near decision points. Importantly, expert-led and data-driven methods reveal different but complementary influences on pedestrian navigation, underscoring the value of combining specialist knowledge with insights derived directly from observed data to improve our understanding and guide the creation of more navigable urban environments.

4.2.1 Introduction

Navigating through urban environments poses unique challenges, heightened by individual differences in cognitive abilities and technological familiarity (Epstein et al. 2017; Coutrot et al. 2022). While urban spaces are designed to ease navigation, the diverse nature of human cognition and varying abilities to utilize mobile navigation applications create a spectrum of navigational efficiencies among city dwellers (Sönmez and Önder 2019). Furthermore, it remains a complicated process to create navigational applications suitable for different type of users (Hunter et al. 2016b). This disparity raises critical questions about the inclusivity of urban planning and the effectiveness of existing navigational aids, highlighting a significant problem: The need for a deeper understanding of how people interact with and navigate through complex urban spaces, and what main factors lead to people getting lost so we can design both optimal urban planning and supportive assistive technologies. The objective of this study is to identify and assess the importance of personal, situational, and environmental factors that contribute to pedestrians getting lost in Greater London. To achieve this, three research questions are formulated as follows: (1) Which factors are most associated with pedestrians getting lost in Greater London? (2) How do AHP and regression approaches compare in evaluating these factors? (3) What complementary insights do these approaches provide into the phenomenon of getting lost?

Research into human navigation has often focused on route optimization - the strategies individuals employ to travel from point A to point B (Tyagi et al. 2022). However, this approach often overlooks the multidimensional nature and factors of human navigation, which is not merely about finding the shortest path but also involves the complex interplay of cognitive processes, emotional states, and urban environment (Gath-Morad et al. 2022; Hölscher et al. 2012). Our study shifts the focus from route optimization to the phenomenon of getting lost. Extensive research from cognitive psychology, geography, and urban studies has laid the groundwork, revealing how memory, perception, and spatial reasoning affect navigation (Walkowiak et al. 2023b). The novel aspect of our research lies in its empirical foundation: a survey in which participants annotated their approximate locations of getting lost while navigating the urban environment. Rather than focusing on route scenarios, it starts with the areas where people had difficulty navigating, providing a different perspective on urban navigability challenges.

A significant body of work has investigated how the structural properties of urban environments influence navigation behaviour and error likelihood. Space syntax theory and related quantitative measures, such as integration and choice, have been widely used to describe street network connectivity and predict patterns of pedestrian movement and route choice (Hillier and Hanson 1984; Turner 2007). Turner 2007's work on axial and road-centre line representations illustrates how spatial configuration can correlate with observed movement, suggesting that configurations that are highly integrated or offer clear orientational affordances tend to support easier navigation. More recently, large-scale empirical evidence has shown that global properties of street networks, including network entropy, are systematically associated with spatial navigation ability, indicating that more topologically complex or disordered street layouts are linked to reduced navigation performance across populations (Coutrot et al. 2022). Empirical studies further indicate that navigation performance is shaped by the interaction between environmental cues and external navigational support, influencing attention, spatial knowledge acquisition, and error patterns in complex environments (Montello and Sas 2006; Golledge 1999). Cognitive research also shows that decision processes at intersections and other decision points constitute critical moments during navigation, where misinterpretation of environmental configuration can lead to increased uncertainty or mistakes (Hölscher et al. 2012). Together, these findings suggest that both the physical layout of the urban network and the interaction between spatial structure and navigational cues play important roles in navigation success and the incidence of getting lost.

In the context of pedestrian navigation, we define "getting lost" as a state of spatial disorientation, essentially a breakdown in wayfinding (the cognitive element of navigation that guides movement) where the traveller becomes uncertain of their location or the correct route to their destination (Darken and Peterson 2002b; Montello and Sas 2006). Lynch 1964 observes that even momentary disorientation in a city can provoke anxiety and that the very word "lost" implies "much more than simple geographical uncertainty", carrying overtones of "utter disaster". Also, Lynch 1964 argues that a well-structured, "legible" environment gives people a sense of security by helping them maintain orientation. Hunter et al. 2016b further defines wayfinding as the integration of cognitive and embodied processes in interaction with environmental cues, suggesting that disruptions to this dynamic (i.e., instances of "lostness") can range from minor uncertainties to severe disorientation. Wayfinding research similarly frames lostness as any lapse in navigational awareness, from minor missteps to severe confusion. Montello and Sas 2006 define geographic disorientation (i.e., getting lost) as uncertainty about one's whereabouts or direction, noting that such disorientation may be long-lasting and serious but is very often minor and temporary. Even brief episodes of being lost are common and can generate anxiety, frustration, or delays (Montello and Sas 2006). Similarly, Darken and Peterson 2002b report that disoriented travellers tend to be "anxious, uncomfortable, and generally unhappy", reinforcing that even short-lived loss of orientation is a significant event in navigation.

Accordingly, this study adopts a broad definition of "getting lost" that considers both minor wayfinding errors (e.g., a wrong turn quickly corrected) as well as prolonged disorientation. In our survey, participants were instructed to report any instance of losing track of their location or intended route, so both small detours and longer events of confusion were counted as instances of having "gotten lost". We also note that, because our study relies on self-reported getting-lost events, cultural norms may shape individuals' willingness to acknowledge or disclose disorientation, meaning that self-estimates of getting lost could vary across contexts (Walkowiak et al. 2023b).

The focal point of our research is the Greater London region, a metropolitan labyrinth known for its complexity and diversity (Maguire et al. 2006; Boeing 2019). By employing a dual-analytical approach, our study compares expert-led evaluations, grounded in Analytic Hierarchy Process (AHP) analysis (Goepel 2018), with data-driven insights derived from Ordinary Least Squares and Ridge Regression techniques. This methodological blend allows us to dissect the nature of urban navigation, examining factors such as urban complexity, self-orientation skills, personal context, and familiarity with the environment. Through this mixed-methods approach, we aim to uncover patterns and discrepancies in urban navigational efficiency and provide the basis for future targeted interventions.

This paper is structured to methodically unfold our findings and their implications. We begin by discussing the data collection process and the key factors considered, detailing the employed methodology—including the AHP and regression analyses—and presenting the results, which include analyses of multicollinearity and the evaluation of estimated weights. Finally, we reflect on the implications of our findings. Through this structured exploration, this study contributes to the interdisciplinary field of spatial cognition and computation, aiming to aid in the development of more navigable urban environments and applications.

4.2.2 Data

This research employs two established methods of analysis in parallel, in order to compare "expert-led" and "data-driven" analytical approaches to the ease of navigating urban environments. In both cases, a set of 14 factors identified from the established literature concerning navigation ability, urban complexity, and spatial cognition is used to measure environmental, situational, and personal factors which might influence the propensity of an individual to lose their way during a navigation task. These factors were identified through a focused review of relevant literature on wayfinding, disorientation, and spatial cognition in urban environments, complemented by a series of brainstorming discussions among the research team to consolidate and refine the list. This process was exploratory rather than systematic, with the purpose of deriving a comprehensive and practically relevant set of factors. Data relating to incidents where individuals "got lost" are then divided into two distinct sets: in the former case, expert-

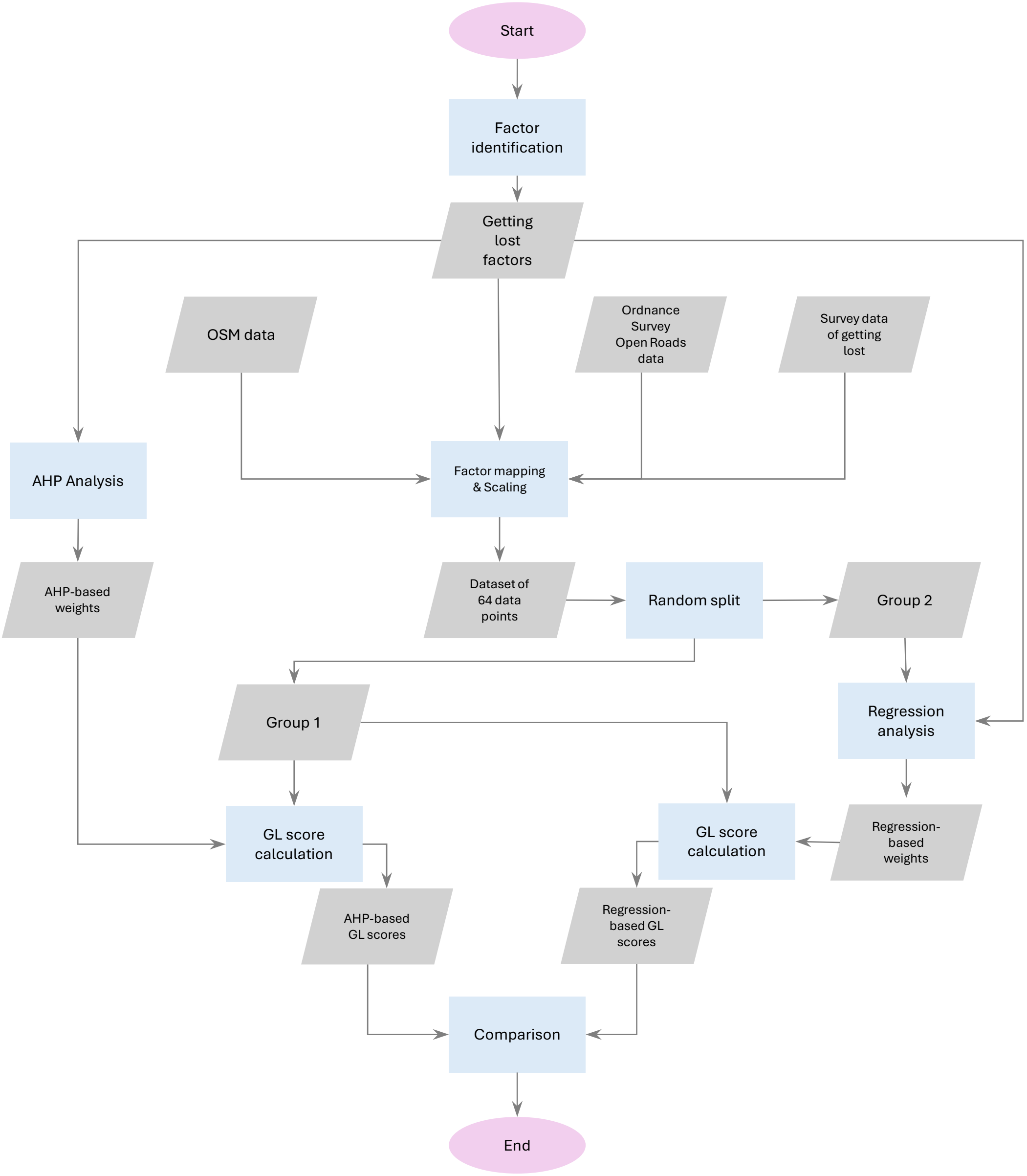


Figure 4.9: Process flowchart

led AHP (a structured approach for multi-criteria decision-making problems) is used to establish the relative importance of each factor; in the latter, OLS and Ridge Regression methods are employed to find the optimal weights for each factor. This process is outlined in Figure 4.9.

		Gender			
		Female	Male	Prefer not to say	Total
Age group	18-29	28	10	1	39
	30-39	20	11	0	31
	40-49	13	5	0	18
	50-65	9	6	0	15
	Total	70	32	1	103

Table 4.5: Survey participants by age group and gender

4.2.2.1 Data collection

To understand how and why individuals lose their way during navigation tasks, it is necessary to observe instances where this occurred. Data relating to specific incidents of getting lost were therefore collected through an online survey administered through the "Prolific" platform in 2022. To address potential recall bias and enhance the reliability of the collected data, all participants were explicitly instructed at the outset: "Please do not respond to the survey if you are not certain about the details of your getting lost event." The sample for the survey was also vetted via "Prolific": 103 individuals participated in the full survey following a screening of the 400 initial respondents, which excluded participants who offered their recollections of incidents that occurred over a year ago and were therefore likely not reliable. Responses indicating uncertainty or vague recollections were also removed from the dataset. This multi-step vetting process ensured that only recent and reliably remembered events were included in the final analysis.

The one-year recall window was selected in line with established practice in survey methodology, where 12-month reference periods are commonly used for quantitative recall (Bradburn et al. 1987; Wagenaar 1986). Prior research has demonstrated that critical details for personal events are already subject to significant memory loss after one year, with loss rates increasing steeply for longer intervals. This makes a one-year window an effective balance between sample size and data reliability. To participate in the full survey, the incident described by respondents also needed to have taken place within the "Greater London Area" (also colloquially described to potential survey participants as "within the M25") and respondents were required to be generally familiar with London (in other words, they were intended to be undertaking everyday tasks, rather than visiting purely for touristic purposes).

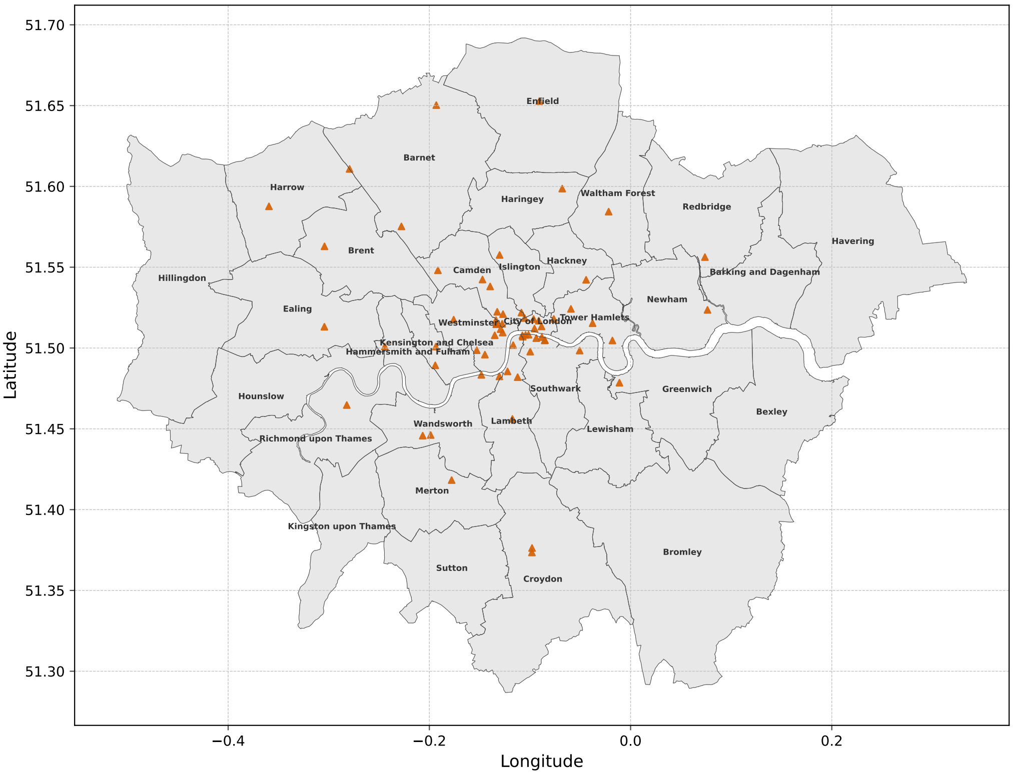


Figure 4.10: Overview of the Getting Lost Survey Points per London Borough

Participants were asked to provide their basic demographic information (age and gender, given in Table 4.5) and general travel habits (including most frequently utilized forms of transport), in addition to information regarding the location at which they got lost, how they were navigating at the time, and what they considered the relevant factors leading to them losing their way. This was primarily obtained via a combination of ordinal categorical, binary choice, and open-ended questions, but participants were also asked to mark the location at which they got lost on a map (as accurately as they were able to recall). Since this research primarily focuses on examining environmental, situational and human factors that influence navigation efficiency, with the potential to inform urban design for improved wayfinding, the subsequent analysis does not incorporate demographic variables related to instances of getting lost. The general area in which each respondent reported their incident is given in Figure 4.10.

Each survey response was assessed for the quality and consistency of the geospatial information provided. As participants submitted both a point on a map and a textual description, the latter was used to verify the former. Participants were aware of the purpose of the survey, and the descriptions they provided were often quite detailed. Therefore, where discrepancies arose, the textual description was given priority, and the geolocated point was adjusted accordingly. If a response was incomplete or could not be reliably interpreted into meaningful geospatial information (which typically occurred where the reported point and textual description were irreconcilably different), the result was excluded from further analysis.

Some participants reported instances of getting lost while driving. While this area is crucial for future research, the decision-making processes during driving may differ from those used in navigating active transportation methods, such as pedestrian navigation (Karimi et al. 2013). Accordingly, responses where the individual was driving were excluded. In addition, despite the screening of survey participants, several of the subsequently reported experiences were outside of the Greater London Area. These entries were also excluded from the analysis.

The remaining sample consists of 64 data points, each corresponding to a getting lost event. To extract related information from other data sources, each of these points was matched to the nearest junction in the Ordnance Survey (OS) Open Roads dataset by calculating the shortest distance between the observed ”getting lost” location and all available junctions. The geospatial information for each event was thus anchored to its closest junction. This process is illustrated by Figure 4.11(A).

4.2.2.2 Factors Associated with Getting Lost

In accordance with 14 categories identified from existing research, the following information was extracted from the survey data or supplementary data sources for each of the 64 points. The inclusion of any particular category should not be interpreted as confirmation of its importance in navigating urban environments but rather that there is a theoretical or empirical reason to consider the extent to which it may have an influence:

1. ***Number of exits***.

   Interpretation: More exits at a junction make it easier for an individual to get lost. Burns 1998 refers to this concept as “navigational degrees of freedom”. O’Neill 1991 expands this general principle into his “Inter Connection Density” measure, which demonstrated greater complexity was connected to lower performance in navigating between two locations.

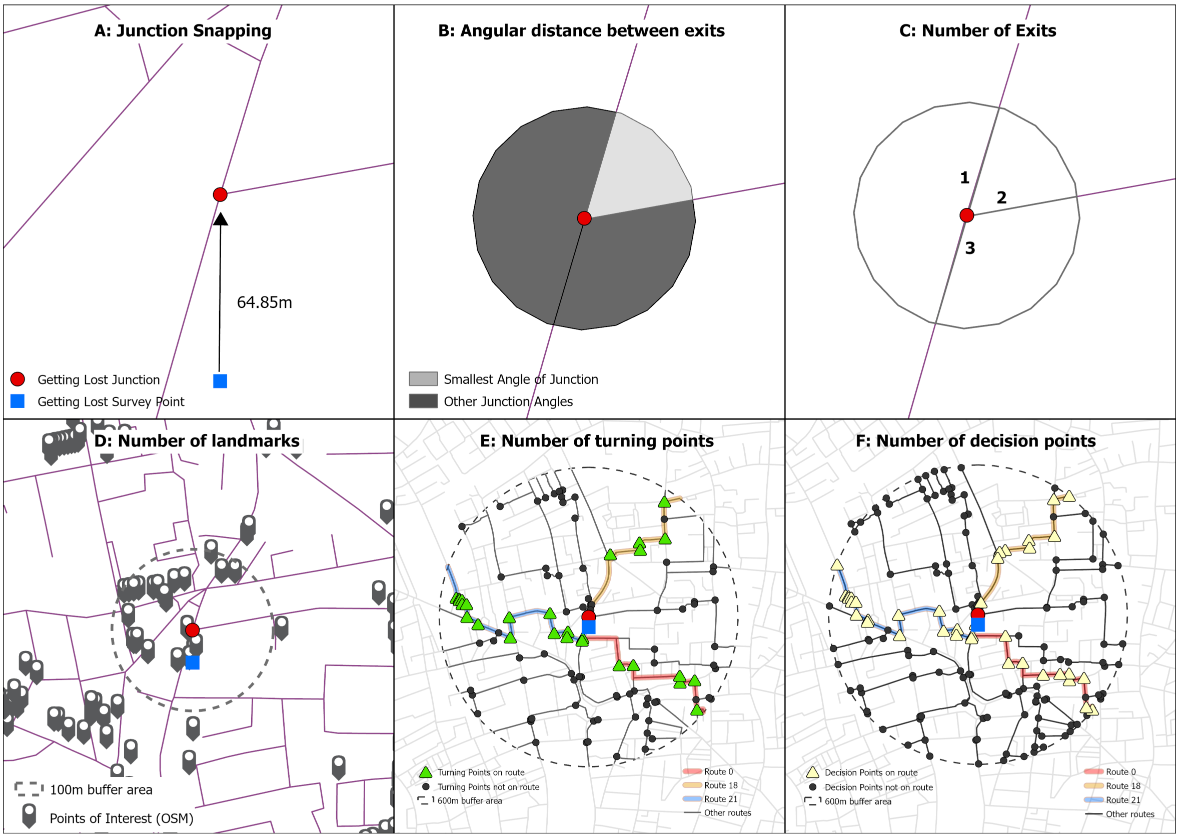


Figure 4.11: Illustrations of the complexity of spatial methodologies.
Notes. Illustration of city complexity measures for a dummy point. 4.11A: The snapping of the point at which a survey respondent reported getting lost to the nearest junction on the road network. 4.11B: The calculation of the smallest angle between two roads at the junction of the snapped point. 4.11C: The identification of the number of exits at the junction of the snapped point. 4.11D: The number of landmarks identified within a 100 m buffer of the snapped point. 4.11E: The number of turning points on the route between the edge of a 600 m buffer and the snapped point. The specifics of some routes (given in blue, orange, and red) are highlighted for illustrative purposes, but routes and turning points (given in black) from all intersections with the buffer are considered.) 4.11F: The number of decision points on the route between the edge of a 600 m buffer and the snapped point. As for 4.11E, the route and decision points between each intersection with the buffer and the snapped point are considered.

Measurement: The number of exits on the junction, as indicated by OS Open Roads. Only the junction itself is considered; turnings or branches after the junction are not included.

2. ***Angular distance between exits***.

   Interpretation: Smaller angles between exits make exits more difficult to differentiate and, therefore, make it easier for an individual to get lost. Sadalla and Montello 1989, for example, show the difficulties demonstrated by test subjects in estimating angles when moving along a pathway.

   Measurement: The minimum angle between any two adjacent exits at a junction.", as indicated by OS Open Roads.

3. ***Pedestrian flow***.

   Interpretation: Areas with heavy pedestrian traffic make it easier for an individual to get lost. The findings of Langer and Saegert 1977, for example, highlight that crowdedness may influence the performance of individuals in carrying out tasks.

   Measurement: Manually assessed using Google Street View Imagery (SVI), scored on a scale of 1 to 5 (with 1 being the lightest pedestrian flow and 5 being the heaviest pedestrian flow). The locations and corresponding scores for pedestrian flow are reported in Appendix B.

   Note: We do not use SVI as a real-time source for quantifying pedestrian and transportation flow by counting the number of pedestrians or vehicles present. Instead, contextual information provided, such as commercial density (e.g. shops, cafés), and the width and continuity of pavements are used to infer pedestrian flow; the number of traffic lanes, visible signage, intersection complexity, and road hierarchy are considered in estimating transportation flow. Additionally, SVI-based assessments were complemented with general contextual understanding of London's transport planning and urban layouts, such as road hierarchy, number of traffic lanes, and proximity to activity centres. The evaluation of pedestrian and vehicle flow therefore mainly relies on visual cues available in the SVIs, with contextual information applied consistently and based on rules to support interpretation under uncertain conditions. This assessment was further cross-validated using targeted Google Maps searches, such as place density and activity indicators, to ensure that the perceived flow estimation of each location was broadly consistent with external contextual evidence.

4. ***Transportation flow***.

   Interpretation: Areas with heavy vehicle traffic make it easier for an individual to get lost. There is evidence that navigating transport infrastructure can be a barrier to pedestrians, particularly when there is a high volume of traffic (Anciaes and Jones 2016). Measuring pedestrian and transportation flows are subject to the exact time when an individual got lost. However, given the available data, synchronising the two is impossible. This is because neither the survey participants reported their precise time of getting lost, nor the available pedestrian/transportation flow data could provide fine-grained temporal frequency data that aligns with it. Therefore, despite acknowledging the uncertainty caused by this, the best approximation could be achieved through the interpretation of the visual information contained in the available SVIs at each location where a person gets lost.

Measurement: Manually assessed using SVI, scored on a scale of 1 to 5 (with 1 being the lightest transportation flow and 5 being the heaviest transportation flow).

5. ***Visibility***.

   Interpretation: Lower visibility areas are more difficult to navigate and, therefore, make it easier for an individual to get lost (Gath-Morad et al. 2024). This draws on ideas from Kubat et al. 2012, who suggest that users in unfamiliar urban environments tend to follow visually connected routes.

   Measurement: Self-reported score (on a scale of 1 to 5, with 1 being "Not relevant" and 5 being "Highly relevant") by survey respondents, in relation to the category "Limited visibility" for the question "Which of the following made you get lost (please rate by relevance)?"

6. ***Number of decision points***.

   Interpretation: Routes involving a higher number of decision points are more complicated and, therefore, make it easier for an individual to get lost. Burns 1998 outlines that a decision point is essentially any point at which an individual encounters uncertainty along their route. Having more instances of uncertainty increases the number of choices made by the individual and may, therefore, increase the likelihood of an individual eventually making an incorrect choice.

   Measurement: The mean number of decision points for simulated routes terminating at the point at which the survey respondent got lost. Decision points were defined as the number of "maneuvers" on each route, as calculated by the "Project OSRM" routing engine (Luxen and Vetter 2011) which utilizes OSM data. Origins of the simulated routes were all intersections between a 600 m buffer around the getting lost point and the OSM road network, which were subsequently snapped to the nearest junction.

   Note: The 600 m distance is based on Transport for London's estimate that the average Londoner walks 1.2 km per day (Transport for London 2018), and reflects that survey trips were typically one-way rather than round trips. This provides a suitable estimate of the distance walked by respondents, regardless of any additional modes of transport used.

7. ***Number of landmarks***.

   Interpretation: Landmarks make it easier for individuals to navigate, so a lack of nearby landmarks makes it easier for an individual to get lost. There is notable literature regarding the interaction between landmarks and navigation. An overview is given by Chan et al. 2012 and Yesiltepe et al. 2021.

   Measurement: As outlined by Chan et al. 2012, landmarks for the purposes of navigation have been conceptualised in a number of ways. In this instance, the number of OSM points of interest within a 100 m buffer is employed as a proxy. Using this proxy, we argue that the quantity of landmarks might not make a direct significant difference, but it does increase the possibility that one of these landmarks can be of use to the lost individual. An illustration of this process is given in Figure 4.11D.

8. ***Self-orientation skills***.

   Interpretation: Individuals with poorer self-orientation skills will get lost more easily. Walkowiak et al. 2023b, for example, demonstrate different navigation abilities across individuals.

   Measurement: This factor is measured through an open-ended question included in the survey where a large proportion of participants have provided discussion on their self-orientation skills either at the moment of getting lost or in the long term. The open-ended question acts as the basis of the factor measurement, which is constructed as follows: "In your own words, please briefly describe what happened when you got lost. Why do you think you lost your way?" Four experts independently evaluated each response and assigned a score between 0 and 1, with 0 indicating the poorest and 1 the strongest orientation skills. The final score for each respondent was calculated as the average of the four expert ratings. For cases where no response was provided or where orientation ability could not reasonably be inferred, a neutral score of 0.5 was assigned to retain the observation.

9. ***Personal context***.

   Interpretation: Tired or distracted individuals will get lost more easily. Burns 1998, for example, reports "distracted attention" as the second most commonly reported cause of survey respondents losing their way.

   Measurement: The mean self-reported score across two categories for the survey question "Which of the following made you get lost (please rate by relevance)?"; "I was tired" and "Other distractions (mobile phone, listening to music, etc)". Scored on a scale of 1 to 5, with 1 being "Not relevant" and 5 being "Highly relevant".

10. ***Number of turning points***.

Interpretation: Routes involving a higher number of turning points are more complicated, and therefore make it easier for an individual to get lost. This aligns with Bailenson et al. 1998's idea of "road climbing", in that people may favor longer, straight road segments (particularly when starting a route) even if this involves travelling further, due to cognitive effort of identifying optimal routes.

Measurement: Same as the method for *Number of decision points*, but number of "turns" were counted instead of number of "maneuvers". A turn must be a maneuver, but a maneuver would not necessarily be a turn; meaning that this category reflects the number of positive actions required to navigate a route (whereas *Number of decision points* reflects the number of active and passive, such as "continue", actions combined).

11. ***City/smaller scale complexity***.

Interpretation: Areas with complex layouts make it easier for an individual to get lost. Stanitsa et al. 2023, for example, note that spatial complexity is generally regarded as related to success in reaching a destination.

Measurement: The "orientation entropy" for the appropriate London borough (as defined by OpenStreetMap 2022), utilizing OSMnx (Boeing 2017), and adapted from the method outlined in (Boeing 2019).

12. ***Access to reliable map***.

Interpretation: This factor includes all forms of external aids or tools used to support navigation in the urban environment, the accuracy of the information they provide, and the reliability of access to these aids at critical moments. We use a broad definition of "map" to include not only traditional paper or digital maps, but also wayfinding kiosks, handwritten or verbal directions, and routing applications. This reflects the diverse ways in which urban navigators access spatial information. A loss of access, inaccurate information, or interruption of service (e.g., due to technical issues or ambiguous instructions) may all increase the risk of getting lost. For instance, Groves 2011b highlights GPS positioning errors in "urban canyons".

Measurement: Binary classification.

13. ***Familiarity***.

Interpretation: Individuals are less likely to get lost on routes with which they are familiar, so a less familiar route will make it easier for an individual to get lost. O'Neill 1992 and Piccardi et al. 2011 argue for the importance of environmental familiarity in navigation tasks.

Measurement: Self-reported score (0 to 100, with 0 being not familiar and 100 being very familiar) by survey respondents, in relation to the question "How familiar were you with the route that you were on when you got lost?"

14. ***Name similarity***.

Interpretation: Streets with similar names may cause confusion, and therefore make it easier for an individual to get lost. The problem of similar-sounding street names is discussed in depth by Chan et al. 2015.

Measurement: Street name similarity was quantified using the Levenshtein distance for street names within a 100 m buffer of each getting-lost location and normalised to a similarity score between 0 and 1. A binary indicator was then derived, where a value of 1 denotes the presence of at least two street names with a similarity score above 0.7, and 0 otherwise. This representation was adopted to emphasise potentially confusing name similarity while reducing sensitivity to weak overlaps.

Each factor outlined in Section 4.2.2.2 is normalized between 0 and 1 on the basis of observed values before use in regression analysis or weighted sum calculations. Boolean variables retained their original value at zero or one, while continuous variables were min-max scaled using the observed sample range. This approach enables comparability across all factors and eliminates the influence of differing original scales on the derived weights.

4.2.3 Methodology

4.2.3.1 Analytic Hierarchy Process

In Section 4.2.2.2, we identified 14 factors that may influence pedestrian navigation and contribute to a getting lost event. However, determining the relative importance of each factor requires further analysis. In this work, we address the evaluation of the weights of getting lost factors using AHP. AHP is a structured approach for multi-criteria decision-making problems. It decomposes complex problems into a hierarchy of more easily comprehended sub-problems, each of which can be analyzed independently. Through a mathematical process, AHP subsequently synthesizes these comparisons to assign a numerical weight to each factor, representing the relative importance of the factor in contributing to the outcome.

We applied AHP to assess the impact of various factors on their contribution to a getting lost event. This involves conducting pairwise importance comparisons of the 14 factors, detailed in Section 4.2.2. To implement our AHP analysis, an online AHP survey was used which included all 14 factors and generated 91 pairwise comparison questions accordingly, covering all possible pairwise combinations of these factors (Goepel 2018). Four researchers with expertise in geospatial data science, urban studies, and navigation were invited to complete the online survey, by choosing which factor in each pair they believed was more important in getting lost events and then assigning a priority level to their choice based on their knowledge and judgment. To assess the intercoder reliability among the four experts on the 14 factors, we computed Krippendorff's Alpha (α) for interval data (Krippendorff 2018). The results of this analysis are provided in Appendix A. The total sum of the weights assigned to all factors equals one. This means each individual weight reflects the specific percentage that the factor contributes to the likelihood of a getting lost event. For our further analysis, we used the average score of each factor, calculated by taking the mean of the scores given by all four experts.

The online tool utilized for this research was AHP-OS. A full explanation of the specific implementation and underlying mathematics is outlined in (Goepel 2018). AHP-OS utilizes the method of Alonso and Lamata 2006 for calculating the consistency ratio. For this article, the weighted geometric mean aggregation of individual judgments is selected for group aggregation.

4.2.3.2 Analysis

General description Our dataset, comprising 64 data points, was randomly split into two even-sized groups for further analysis. The rationale for splitting the dataset was to ensure a fair evaluation of the weighting schemes. One-half of the data was used to derive weights via regression methods, while the other half was reserved for applying these weights and for testing purposes. This approach allows for a more robust comparison between expert-driven and data-driven weighting methods. For the first group, we computed the getting lost score for each data point. This process involved calculating the weighted sum of the individual factor scores, using the weights provided by experts through AHP analysis, as outlined in Section 4.2.3.1. Conversely, for the second group, instead of using the expert-provided AHP weights, we applied two distinct regression models to derive alternative sets of weights directly from the survey data. During this process, the target getting lost scores were uniformly set to one across all data points, and their individual factor scores were used for generating alternative weight sets. This approach was designed to evaluate different weighting schemes, potentially providing a more accurate reflection of each factor's contribution to the getting lost event. The next step involved applying the derived weights from the data points of the second group to the first group. This led to the recalculation of the getting lost scores for the first group using these new weights from the regression models. The final stage of our analysis involves a comparative review of the getting lost scores of the first group, comparing those derived from AHP weights (expert-led) with those obtained from the second group's regression models (data-driven).

In addition, we conducted both multicollinearity and correlation analyses on the getting lost matrix derived from the AHP approach, in order to further reveal the interrelationships among the factors. Multicollinearity arises when two or more factors in a regression model exhibit linear dependence, meaning that one factor can be predicted to a certain degree of accuracy using the others. Although multicollinearity does not necessarily reduce the overall predictive power of a model (O'brien 2007), it can impede the accurate estimation of individual regression coefficients. To quantify the degree of collinearity, we calculated the Variance Inflation Factor (VIF) for all 14 factors in Group 1, using it as an established indicator of multicollinearity.

We subsequently performed regression analyses using both OLS and Ridge Regression on the data points in Group 2, yielding two distinct sets of weights. These regression-derived weights, along with the initial AHP-derived weights, were each applied to the data points in Group 1, resulting in three corresponding sets of getting lost scores for Group 1. To evaluate the statistical significance of differences among these three sets of scores, we first conducted the Shapiro-Wilk test (Shapiro and Wilk 1965) to assess the normality of the getting lost score distributions under each weighting scheme. Based on the results of these normality tests, we proceeded to conduct pairwise comparisons using both the t-test (for normally distributed data) and the non-parametric Wilcoxon signed-rank test as a robustness check for cases where normality could not be assumed.

Whilst the weights calculated by the OLS and Ridge regression techniques (discussed in Section 4.2.4.2) may be negative, the weights derived from the AHP analysis must be positive. For particular factor, we therefore transform the values for each data point to account for directionality. This is outlined in Table 4.6. A value of 1 for any given factor would imply that an individual was more likely to get lost.

Factor	Direction
ID1: Number of Exits	x
ID2: Angular distance between exits	$1-x$
ID3: Pedestrian Flow	x
ID4: Transportation Flow	x
ID5: Visibility	x
ID6: Number of decision points	x
ID7: Number of Landmarks near decision points	$1-x$
ID8: Self-orientation skills	$1-x$
ID9: Personal context	x
ID10: Number of turning points	x
ID11: City/Smaller scale complexity	x
ID12: Access to reliable map	x
ID13: Familiarity	$1-x$
ID14: Name similarity	x

Table 4.6: Factor transformations and interpretation

Getting lost score of Group 1 For each data point in Group 1, we calculated the weighted sum of AHP weights and scaled factor values, resulting in the getting lost score for each data point as follows:

$$\mathbf{S_1} = \mathbf{X_1}\omega_{ahp} \tag{4.6}$$

Where $\mathbf{S_1}$ denotes the score matrix for getting lost of Group 1. $\mathbf{X_1}$ is a 32×14 matrix, where each row represents a data point and each column corresponds to one of the 14 factors. ω_{ahp} is a known column vector of dimension 14×1, containing the weights derived from our expert-led AHP online questionnaire, as described in Section 4.2.3.1.

Weights estimation of Group 2 Group 2's analysis aims to identify a distinct set of weights. These weights were optimized to maximize the overall getting lost scores for Group 2, indicating a higher likelihood of getting lost events occurring. The overall getting lost score matrix of Group 2 is calculated as follows:

$$\mathbf{S_2} = \mathbf{X_2}\omega_2 \tag{4.7}$$

Where $\mathbf{S_2}$ represents the getting lost score matrix for Group 2, with dimension 32×1, each element corresponding to the aggregated getting lost score for one of the 32 data points. The matrix $\mathbf{X_2}$ encapsulates the individual factor scores for each data point in Group 2. It is a 32×14 sized matrix, where each row corresponds to a data point, and each of the 14 columns represents one of the identified factors affecting the likelihood of getting lost. The vector $\boldsymbol{\omega_2}$ is the weight vector to be optimised. It contains the weights for the 14 factors and is determined such that, when applied to $\mathbf{X_2}$, it maximizes the getting lost scores in $\mathbf{S_2}$.

Given that each factor in our dataset is normalized to the range [0,1], and the sum of the 14 factor weights is constrained to one, the maximum possible getting lost score for any data point is capped at one. Since all data points from both Group 1 and Group 2 are derived from real-life instances of getting lost, it is reasonable to anticipate that the weighted sum of each data point's factors would approximate this maximum value at one. Therefore, to optimize the set of factor weights in a way that maximizes the likelihood of a getting event, we structured the maximized $\mathbf{S_2}$ matrix as an all-ones matrix, denoted as $\mathbf{S_{max}}$. This approach aligns with our theoretical maximum, allowing us to effectively gauge the optimal combination of weights that correspond to the highest likelihood of getting lost based on our dataset.

In our study, we applied the Ordinary Least Squares (OLS) and Ridge Regression (RR) as two regression models to determine the optimal set of weights for the factors in Group 2. This approach is framed as an optimization problem, aiming to minimize the mean squared error (MSE) between the desired maximum getting lost scores and the scores calculated through the explored weights. Mathematically, the optimal weights maximizing the getting lost score can be expressed as follows:

$$\hat{\omega}_{2_{OLS}} = \arg\min_{\omega_2} \left(\frac{1}{N} \sum_{i=1}^{N} (\mathbf{S_{max}} - \mathbf{X_2}\omega_2)^2 \right) \tag{4.8}$$

$$\hat{\omega}_{2_{RR}} = \arg\min_{\omega_2} \left(\frac{1}{N} \sum_{i=1}^{N} (\mathbf{S_{max}} - \mathbf{X_2}\omega_2)^2 + \lambda |\omega_2|_2^2 \right) \tag{4.9}$$

Where (4.8) and (4.9) represent the identified weights by using OLS and RR, respectively. Here, $\hat{\omega}_2$ represents the estimated weights vector that minimizes the MSE. $\mathbf{S_{max}}$ is the matrix of desired maximum getting lost scores for Group 2, and $\mathbf{X_2}$ is the matrix constituted by individual factor scores. The optimization process adjusts the weights

ω_2 to find the best fit between the predicted scores $\mathbf{X_2}\omega_2$ and the desired getting lost scores $\mathbf{S_{max}}$. Compared with OLS regression in (4.8), the Ridge regression model in (4.9) introduces a regularization term. This term is the sum of the L2 norm of the weights, denoted as $\|\omega_2\|_2^2$. The symbol λ denotes the regularization parameter, which serves to reduce the risk of overfitting and enables more effective handling of multicollinearity in the model.

Furthermore, we incorporated one additional constraint condition into the optimization procedure to ensure the validity and applicability of our model. The condition is expressed as $\sum_{i=1}^{14} \hat{\omega}_{2i} = 1$, mandates that the aggregate of all identified weights in the vector $\hat{\omega}_2$ must equal one. This condition, where the sum of all weights equals 1, ensures that each weight reflects the proportion of its corresponding factor in the overall model.

4.2.4 Results

4.2.4.1 Multicollinearity and correlation

As shown in Table 4.7, several factors exhibit relatively high VIFs, particularly those with values exceeding the commonly used threshold of 5. These factors include the number of exits, angular distance between exits, pedestrian flow, number of landmarks near decision points, number of decision points, and number of turning points. These factors are retained despite elevated VIF values because the regression is intended to examine relative associations rather than causal effects, with multicollinearity addressed through regularisation.

To further examine the relationships between these factors, a correlation matrix was constructed, as detailed in Figure 4.12. The results reveal a strong correlation between several pairs of factors. Notably, the number of exits and the angular distance between exits exhibit a correlation coefficient of 0.87. Pedestrian flow is correlated with the number of landmarks, with a coefficient of -0.68, and the number of decision points and the number of turning points show a correlation coefficient of 0.96.

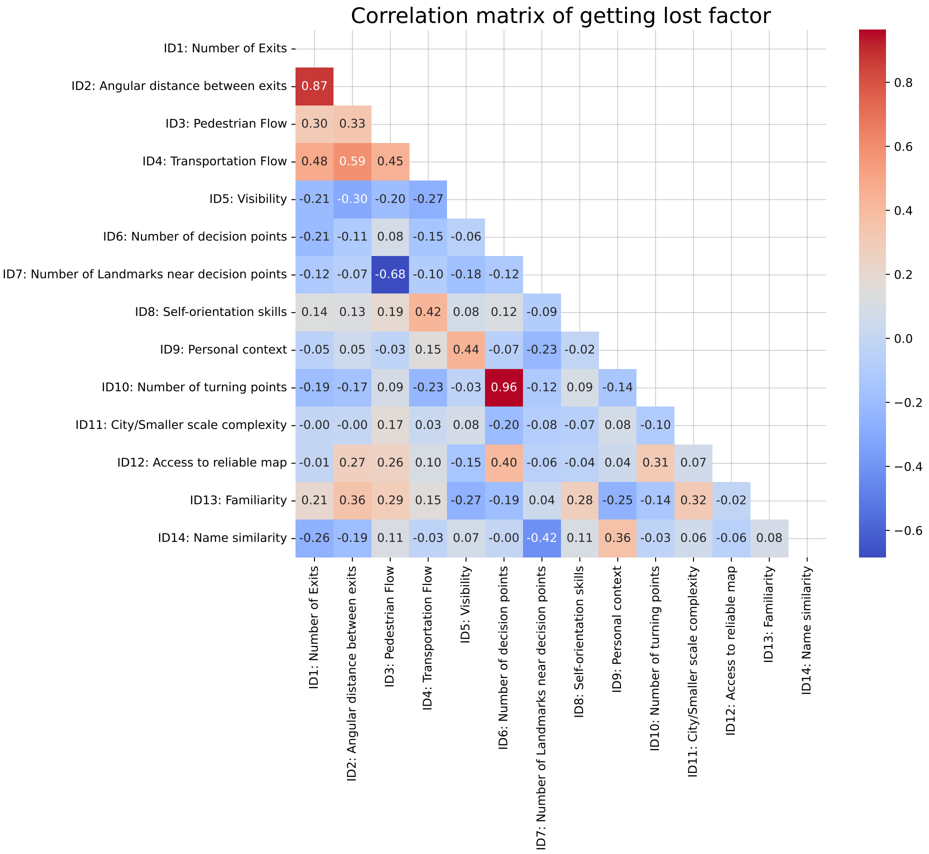


Figure 4.12: Correlation matrix among the 14 factors

Features	VIF
ID1: Number of Exits	**15.7**
ID2: Angular distance between exits	**22.9**
ID3: Pedestrian Flow	**5.8**
ID4: Transportation Flow	4.5
ID5: Visibility	2.2
ID6: Number of decision points	**42.2**
ID7: Number of Landmarks near decision points	**4.9**
ID8: Self-orientation skills	2.5
ID9: Personal context	2.2
ID10: Number of turning points	**37.2**
ID11: City/Smaller scale complexity	1.7
ID12: Access to reliable map	2.7
ID13: Familiarity	4.7
ID14: Name similarity	2.1

Table 4.7: Variance Inflation Factor (VIF) for each getting lost factor. (VIF greater than five is indicated in bold)

4.2.4.2 Expert-led and Data-driven Weights

In our analysis, we applied both OLS and Ridge regression techniques to determine the weights of factors in Group 2. The outcomes of this analysis are presented in Figure 4.13, the blue bar chart represents the AHP weights of Group 1, as specified in Section 4.2.3.1. Additionally, the weights of Group 2 derived through OLS and Ridge regression are depicted as orange and magenta bar charts, respectively.

To identify an appropriate penalty parameter λ for the Ridge regression model, we employed Ridge trace analysis (Hoerl and Kennard 1970; McDonald 2009). This technique plots the estimated regression coefficients against a sequence of λ values. When λ is small, the coefficients can fluctuate substantially, but as λ increases the curves gradually flatten, indicating that the estimates become less sensitive to further increases in λ. We selected $\lambda = 0.4$ because, as shown in Figure 4.14, the slopes of the coefficient paths begin to stabilise around this value. This suggests that the shrinkage is sufficient to control variance without introducing excessive bias. The variances of the weights estimated through OLS and Ridge regression were 0.416 and 0.009, respectively.

Method	Mean GL Score	Variance
AHP	0.59	0.01
OLS	0.94	0.07
Ridge	0.76	0.03

Table 4.8: Mean and Variance of Getting Lost Scores for Group 1 by weights from both OLS and RR

4.2.4.3 Getting Lost Scores under Different Weighting Schemes

We applied the weights estimated from OLS and Ridge regressions back to the data points of Group 1 and recalculated their getting lost scores using these new sets of weights. The mean and variance of getting lost scores for the data points in Group 1 are reported in Table 4.8. Both weights from OLS and Ridge regression produced higher getting lost scores compared to the original AHP scores in Group 1. OLS had an average score of 0.94, while Ridge regression had an average of 0.76, compared to AHP's 0.59. Among the three weighting methods, the getting lost scores derived from OLS weights exhibited the greatest variance at 0.07, in contrast to substantially lower variances for weights from AHP and Ridge, at 0.01 and 0.03, respectively.

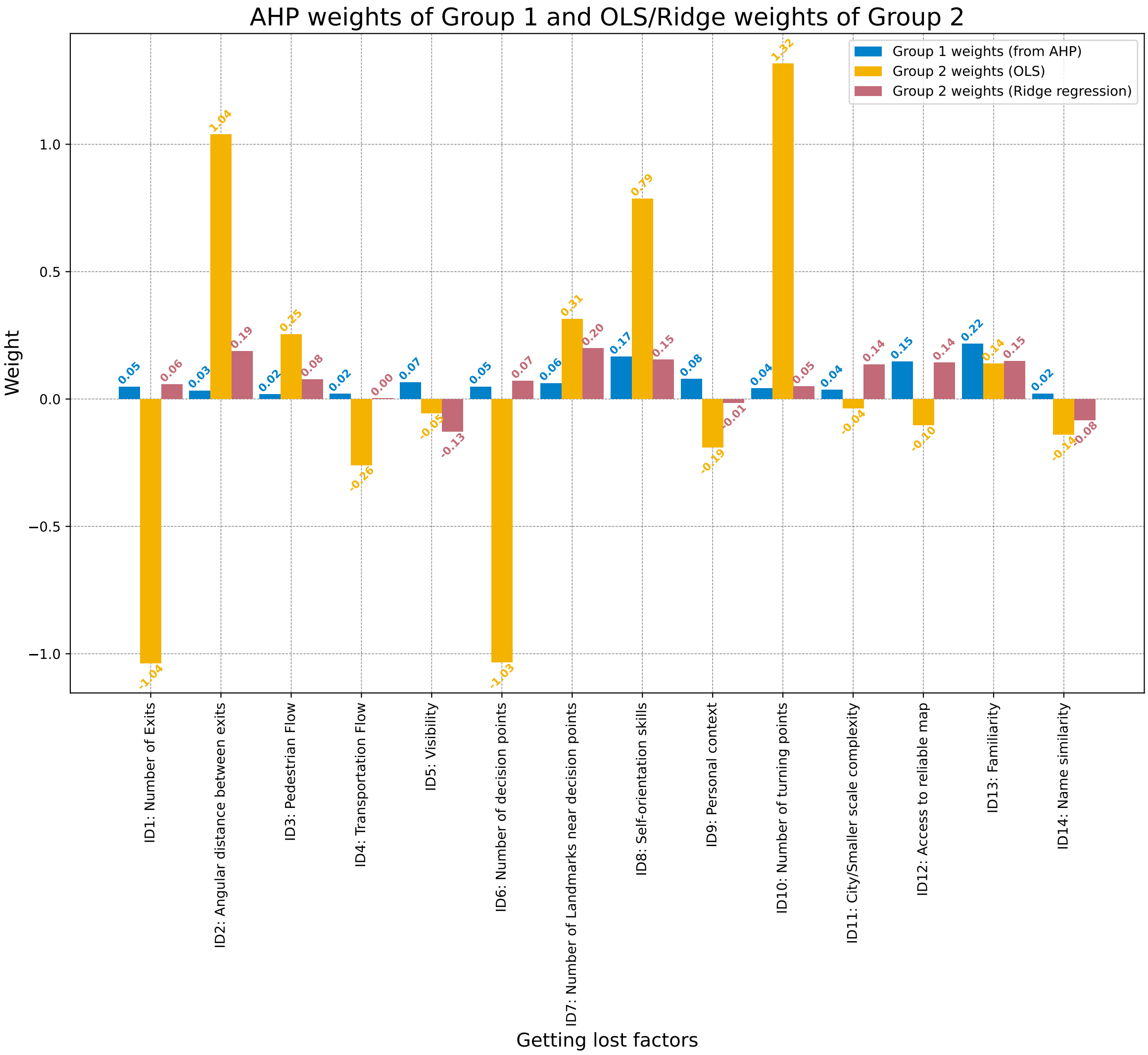


Figure 4.13: Weights from AHP analysis (on Group 1), OLS and RR (on Group 2)

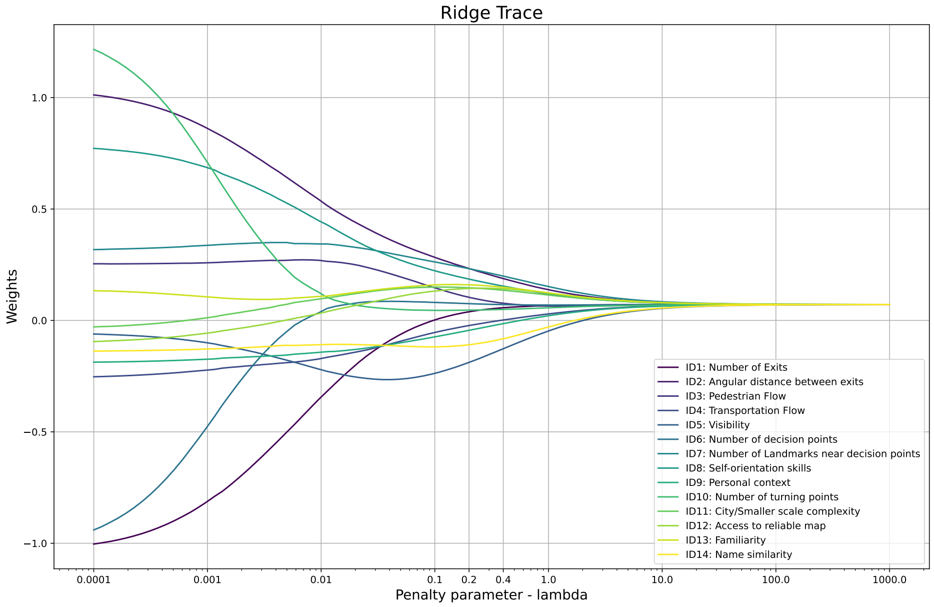


Figure 4.14: Variation of estimated weights with respect to penalty parameter

Weighting Scheme	Shapiro-Wilk Statistic	p-value
AHP	0.97	0.571
OLS	0.98	0.682
Ridge	0.94	0.083

Table 4.9: Results of the Shapiro-Wilk test for normality on Group 1 getting lost scores under different weighting schemes.

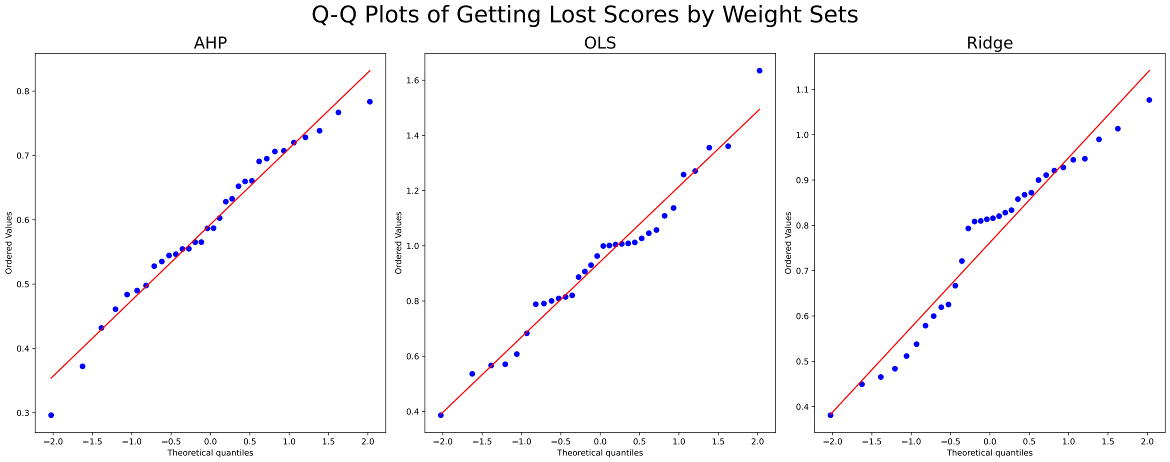


Figure 4.15: Q-Q plot of getting lost scores under different weighting schemes

The results of the Shapiro-Wilk normality test are summarized in Table 4.9. For the AHP-weighted getting lost scores, the W statistic is 0.97 with a p-value of 0.571. For the OLS-weighted scores, the W statistic is 0.98 with a p-value of 0.682. For the Ridge-weighted scores, the W statistic is 0.94 with a p-value of 0.083. To further assess and visualize the normality condition, Q-Q plots are presented in Figure 4.15. The Q-Q plot for the AHP-weighted scores displays points that largely align with the theoretical normal line, while the Q-Q plots for the OLS- and Ridge-weighted scores show greater deviations from normality.

Test	Comparison	Statistic	p-value
t-test	AHP vs. OLS	-8.8	6.0×10^{-10}
t-test	AHP vs. Ridge	-7.7	1.13×10^{-8}
t-test	OLS vs. Ridge	5.0	2.06×10^{-5}
Wilcoxon test	AHP vs. OLS	9.0	1.5×10^{-8}
Wilcoxon test	AHP vs. Ridge	22.0	2.5×10^{-7}
Wilcoxon test	OLS vs. Ridge	57.0	3.2×10^{-5}

Table 4.10: Results of t-tests and Wilcoxon tests between three weighting methods

The results of both the t-tests and Wilcoxon tests are presented in Table 4.10. For all comparisons between the AHP, OLS, and Ridge weighting methods, the p-values were found to be less than 0.001. These results indicate statistically significant differences in the getting lost scores derived from the three weighting methods.

4.2.5 Discussion and Conclusion

Navigating through complex urban areas remains challenging for pedestrians, even with the assistance of optimized routes provided by advanced navigational tools, due to the complex interplay of various factors that impact pedestrian navigation and often lead to confusion. In this work, we identify a list of factors that may cause pedestrians to get lost in urban areas, as well as their relative contributions by generating weighting systems from both expert-led and data-driven methods. Potential multicollinearity and correlation are revealed from the identified factors, leading us to employ both unregularized and regularized regression models to produce factor weights. We find that the scores for getting lost, calculated from both expert-led and data-driven methods, demonstrated distinct but complementary influences on pedestrian navigation.

Our findings from the correlation analysis are consistent with the elevated VIFs. For example, an increase in the number of exits at a junction typically corresponds to a decrease in the angular distance between these exits. Similarly, areas with a greater number of points of interest are likely to attract more pedestrians, and itineraries with a higher number of decision points may lead to an increased number of turns, explaining the strong correlation observed between these two factors. The presence of multicollinearity and correlation among factors highlights the need to utilize regularization techniques to achieve more stable and interpretable results estimates.

The OLS model produces the highest average getting lost score, but also the greatest variance, indicating differences across scores and possible overemphasis of certain factors. In contrast, Ridge regression produces more balanced and consistent estimates, with variance lower than OLS and a mean score between those of AHP and OLS. By effectively mitigating multicollinearity, Ridge regression not only achieves lower variance and higher scores than AHP, but also generates a weight distribution similar to expert-led estimates, underscoring its effectiveness in modelling getting lost events. Accordingly, because the analysis focuses on comparing relative factor weightings rather than predictive accuracy, all three weighting schemes yield average getting-lost scores above the midpoint of 0.5, indicating that they consistently capture an elevated likelihood associated with getting-lost events.

Our analysis revealed that the choice of three weighting schemes significantly influences the computed getting lost scores, as confirmed by both parametric and non-parametric significance tests. However, to answer the fundamental question—*Why do pedestrians get lost?*—we need to examine the factor weights produced by each method and their substantive interpretation. For the expert-led AHP approach, the most important factors identified were familiarity with the environment, self-orientation skills, and access to reliable navigation aids. This reflects the expert view that people-centric factors are decisive in determining the likelihood of getting lost. Familiarity represents an individual's personal knowledge of the area, while self-orientation skills capture their ability to navigate effectively in confusing environments. In addition, experts emphasise that external navigational support—such as maps or mobile navigation applications—can further strengthen a pedestrian's ability to find their way. By contrast, the OLS model places the greatest positive weights on the angular distance between exits, number of turning points, and self-orientation skills. This indicates that the model considers the structural complexity of the route—such as how many choices a pedestrian must make and how many possible paths are available at a junction—as major determinants of getting lost. While self-orientation skills still play a significant role, the prominence of environmental factors suggests that the complexity of the urban layout may strongly influence navigation outcomes. However, the presence of negative weights for certain factors points to potential issues of multicollinearity and overfitting, which can lead the model to overestimate or underestimate the true importance of some variables. Also, it is noticeable that several pairs of highly correlated factors show opposite coefficient signs, such as the number of exits versus the angular distance between exits, and the number of decision points versus the number of turning points. OLS still provides an overall unbiased fit to the data, but multicollinearity inflates coefficient variance, making individual weights unstable and sometimes counterintuitive. The regularized Ridge regression model, on the other hand, provides a more balanced interpretation. The most influential factors include number of landmarks near decision points, angular distance between exits, familiarity and self-orientation skills. Similar to the OLS-derived weighting scheme, Ridge regression assigns relatively greater importance to the underlying complexity of the urban environment. Ridge regression reduces the magnitude of negative weights, thus offering more stable and generalizable estimates. This suggests that, compared to OLS, Ridge regression, as a data-driven method, is more capable of identifying the key contributors to getting lost while mitigating the distortions introduced by highly correlated factors in the dataset. The weighting schemes from expert-led AHP and regression models are distinct yet complementary in explaining pedestrian navigation: experts emphasised human-centric factors such as familiarity and self-orientation skills which align with general cognitive explanations of why individuals become disoriented, whereas the regression models placed more weight on the environmental complexity around the getting-lost locations. In addition, this discrepancy is likely driven by the fact that the expert assessments were conducted independently of the empirical observations, leading to a generalised reasoning about how a getting lost event “should” happen, whereas the regression models were fitted to situational, location-specific data, emphasising why a particular getting lost event happens. As a result, the two weighting schemes prioritise different aspects of the navigation process. Rather than indicating inconsistency, this divergence highlights the multi-layered nature of pedestrian disorientation, where subjective perception and objective environmental context may exert influence through different mechanisms.

Taken together, the findings suggest that pedestrians become lost mainly due to two main strands of reasons. From a static view given by experts, individual reasons account for the largest share of getting lost events, while location-specific analysis from regression methods emphasises the environmental complexities. Three mechanisms of decision density, environmental complexity, and external support consistently emerge as central across methods, even though their relative weights differ between expert-led and data-driven approaches. By contrast, factors such as name similarity or transportation flows were less influential, reflecting that salience, timing, and perception may matter more than simple presence or quantity. Rather than providing definitive effect sizes, our analysis acts as an empirical case study of Greater London that illustrates how different factors contribute to pedestrians getting lost. The results highlight patterns—such as the importance of decision density, environmental complexity, and orientation support—that warrant further testing with richer data and in other contexts. While our study captures the *state* of being lost at specific locations, future work that incorporates itineraries and dynamic traces could help clarify the processes by which pedestrians gradually become disoriented.

There are several limitations in this work, which also suggest directions for future research. Methodologically, our regression models revealed issues of multicollinearity, with some factors highly correlated, which complicates the interpretation of individual factor weights and produces negative weights in the OLS model. Although Ridge regression helps mitigate these distortions, the presence of unstable or counterintuitive weights highlights the challenges of disentangling highly interrelated urban navigation factors. This reflection is important for interpreting our findings and indicates that further methodological refinement is needed, for example by exploring alternative modelling strategies that can better handle multicollinearity. A further limitation concerns the measurement of self-orientation skills, which are difficult to define and measure directly. In this study, we therefore relied on secondary proxies from open-ended questions. Future work should address this by designing a more sophisticated survey platform that enables participants to systematically evaluate their orientation skills, for example, through the Santa Barbara Sense of Direction Scale (Hegarty et al. 2002). This study adopts a generalised and route-agnostic perspective that characterises getting lost as a retrospectively reported human experience. Another limitation is that, while our survey collected the reported location at which respondents became lost, it did not include detailed itinerary information that could provide a richer context for the entire wayfinding process. As revealed by Grübel et al. 2019, the inaccuracy of pedestrians' mental maps versus the true environment can lead pedestrians to follow incorrect itineraries, ultimately causing them to get lost over time. Accordingly, future research should develop survey platforms that allow participants to capture both the locations at which they became lost and their itineraries, thereby yielding more comprehensive contextual information on getting-lost events. Additionally, the reported getting-lost locations are spatially concentrated towards the city centre, indicating a potential spatial bias in the dataset. This concentration may lead to an over-representation of centrally located urban environments while under-representing peripheral or suburban areas. Future work will therefore consider increasing sample coverage across different urban zones to achieve a more spatially balanced dataset. In addition, some of the getting-lost factors are subjectively evaluated, including visibility, pedestrian and vehicle flow. These estimates rely on visual information contained in SVIs, which capture traffic conditions at specific and often unknown times. As a result, these assessments may reflect temporal biases, for example by over- or under-representing typical traffic levels depending on the time of image acquisition. This highlights a broader issue of temporal and reporting biases in image-based urban observations, which is further examined and systematically analysed in Chapter 5. This limitation highlights the need for approaches that can quantitatively measure these factors in the future. Potential options include leveraging computer vision techniques to automatically infer the level of street activity or "busyness" from SVIs, or by using transportation statistics such as traffic flow or vehicle speed data as proxies for transportation intensity. Finally, our findings are based on a case study of the Greater London Area. However, cultural norms and urban context may shape the experience of getting lost. Thus, future studies should examine the extent to which our findings generalise to other cities or cultural settings.

To conclude, across all weighting methods, self-orientation skills, familiarity and access to reliable maps or tools consistently emerge as leading contributors to pedestrians getting lost. Notably, expert-led weights place greater emphasis on people-centric factors, such as familiarity and self-orientation skills, whereas data-driven models, particularly OLS, highlight structural features of the environment, such as the angular distance between exits. The Ridge regression results offer a middle ground between these perspectives, balancing personal and environmental influences. Additionally, the elevated weights assigned to factors like city or local-scale complexity in the regression models suggest that environmental complexity is a significant, data-driven risk for disorientation, potentially underestimated by experts. This divergence further underscores the value of integrating both expert knowledge and data-driven analysis to gain a comprehensive understanding of the factors that contribute to pedestrian getting lost events in urban environments.

Appendix A: Intercoder reliability (Krippendorff's α for interval data).

To quantify agreement among the four expert coders on the 14 factors, we computed Krippendorff's Alpha (α) for interval data (Krippendorff 2018):

$$\alpha = 1 - \frac{D_o}{D_e},$$

where D_o is the observed disagreement and D_e is the disagreement expected by chance.

Observed disagreement. Let $i = 1, \ldots, N$ index items (factors), $c = 1, \ldots, K$ index coders (experts), and $x_{ic} \in \mathbb{R}$ be coder c's weight for item i (missing values allowed), normalized at zero to one. Denote by $C_i \subseteq \{1, \ldots, K\}$ the set of coders who rated item i, with $m_i = |C_i|$. Using the squared Euclidean distance for interval data, the observed disagreement is the average pairwise squared difference within each item, aggregated over items:

$$D_o = \frac{\sum_{i=1}^{N} \sum_{\substack{c,c' \in C_i \\ c < c'}} \left(x_{ic} - x_{ic'}\right)^2}{\sum_{i=1}^{N} \binom{m_i}{2}}.$$

Expected disagreement. Let $\boldsymbol{M} = \sum_{i=1}^{N} m_i$ *be the total number of observed ratings, and* $\{z_u\}_{u=1}^{M}$ *the pooled set of all ratings across items and coders. The expected disagreement is defined analogously as the average pairwise squared difference between two ratings drawn at random (without replacement) from this pooled distribution:*

$$D_e = \frac{\displaystyle\sum_{\substack{u<v \\ u,v\in\{1,\ldots,M\}}} (z_u - z_v)^2}{\binom{M}{2}}.$$

Result and interpretation. The outcome of AHP analysis across four experts (coders) and 14 factors (items) are illustrated in Table 4.11.

Table 4.11: AHP weights derived from expert evaluation.

Factor	**Expert 1**	**Expert 2**	**Expert 3**	**Expert 4**
ID1: Number of Exits	0.03	0.04	0.04	0.09
ID2: Angular distance between exits	0.03	0.02	0.01	0.09
ID3: Pedestrian Flow	0.01	0.01	0.02	0.03
ID4: Transportation Flow	0.01	0.01	0.02	0.03
ID5: Visibility	0.09	0.03	0.07	0.09
ID6: Number of decision points	0.06	0.03	0.04	0.06
ID7: Landmarks near decision points	0.05	0.09	0.04	0.07
ID8: Self-orientation skills	0.14	0.16	0.14	0.15
ID9: Personal context	0.01	0.17	0.13	0.10
ID10: Number of turning points	0.04	0.03	0.03	0.08
ID11: City/Smaller scale complexity	0.07	0.02	0.06	0.01
ID12: Access to reliable map	0.18	0.10	0.16	0.10
ID13: Familiarity	0.26	0.25	0.24	0.09
ID14: Name similarity	0.02	0.03	0.01	0.01

Applying the above to our data (four coders, fourteen items) yields $\alpha \approx 0.69$, indicating a moderate level of agreement.

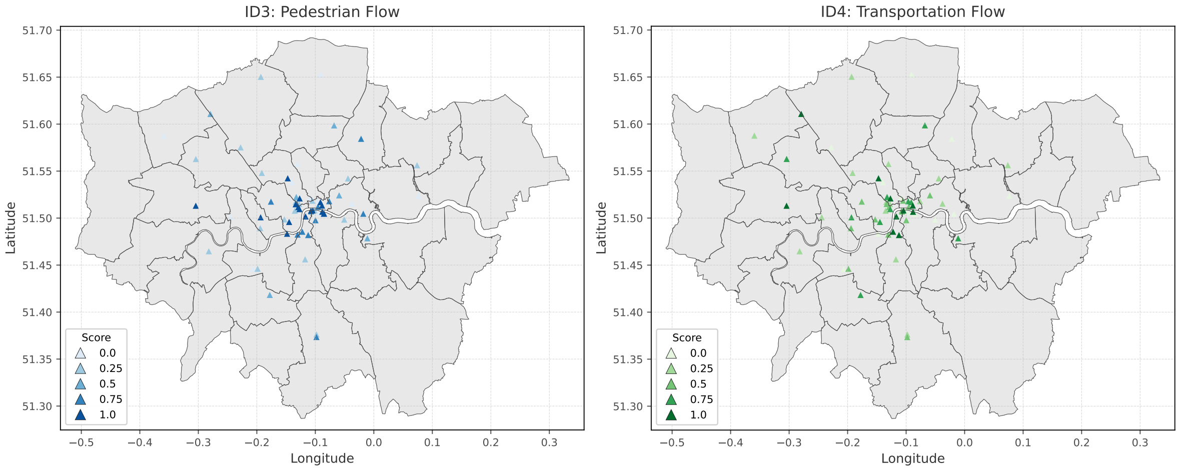


Figure A1: Geographic distribution of pedestrian flow (ID3) and vehicle flow (ID4) scores in London.

Appendix B: Pedestrian and Vehicle Flow Assessment

To enhance transparency in how pedestrian and vehicle flow were assessed, we provide here both a geographic visualisation of the assigned scores and a set of illustrative examples from Google Street View (SVI).

A. Geographic visualisation of scores Figure A1 shows the spatial distribution of the manually assigned pedestrian and vehicle flow scores across the London study area.

B. Illustrative SVI examples Figure A2 presents example SVI images illustrating how scores were assigned. Four categories are shown (0, 0.25, 0.5, 1.0), each represented with one pedestrian-flow image and one vehicle-flow image, to demonstrate the visual cues used for evaluation.

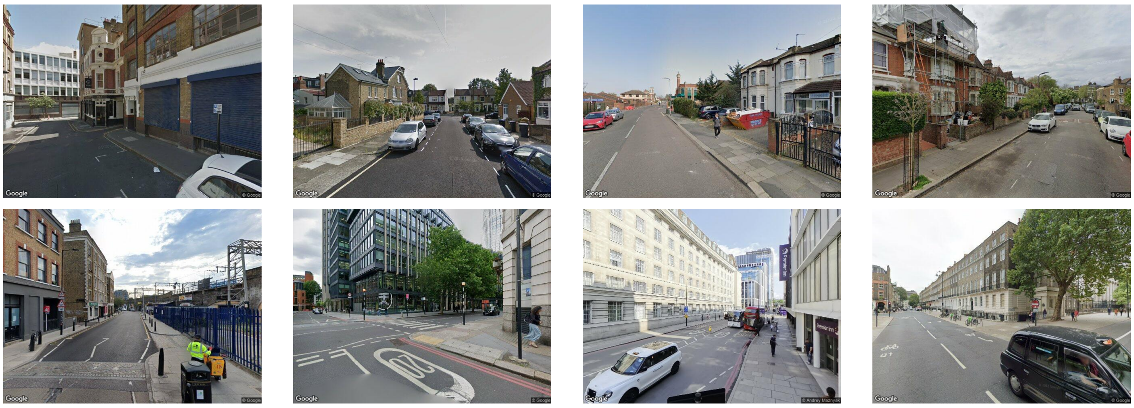


Figure A2: Illustrative SVI images used for assigning pedestrian flow (ID3) and vehicle flow (ID4) scores. Two representative images are provided for each scoring category (from left to right: 0, 0.25, 0.5, 1.0).

Chapter 5

From Static to Dynamic: Evaluating the Impact of Temporal Bias in Historical Street View Images for Cross-View Geolocalisation

Publication and Permission Statement

This chapter presents original research conducted by the author during the PhD programme. The work has not yet been published and is included here as an original contribution of the thesis.

Abstract

Street View Images (SVIs) have become a valuable source of geotagged visual data, increasingly used in tasks such as assessing socioeconomic conditions, mapping accessibility, monitoring infrastructure, and supporting navigation and perception studies. However, while the spatial coverage and utility of SVIs are well recognised, their temporal variation and its impact on the generalisability and robustness of geospatial AI models remain underexplored. To address this gap, we examine how temporal bias in SVIs affects the performance of cross-view geolocalisation models. We construct a temporally diverse dataset of historical SVIs, spatially aligned with the CVUSA validation set, enabling controlled evaluation of temporal effects. We evaluate two state-of-the-art deep learning models, TransGeo and SAFA, on this temporally diverse dataset, assessing retrieval accuracy under temporal shifts. To further explore how specific visual changes impact outcomes, we apply semantic segmentation to categorise scene types and use SHAP analysis to interpret how feature variations contribute to retrieval success. Our study establishes the importance of explicitly accounting for temporal diversity on semantic feature composition in the development and evaluation of geospatial AI methods. The results reveal a substantial decline in retrieval accuracies for both models when evaluated on the temporally diverse dataset. Semantic analysis further indicates that temporal shifts in visual features, particularly those related to urban infrastructure and natural landscape, can systematically influence model performance.

5.1 Introduction

Street view images (SVIs) offer valuable insights for a wide range of applications, including socioeconomic estimation based on built environments, geolocalisation, navigation, and perceptual analysis (Biljecki and Ito 2021; He and Li 2021; Dubey et al. 2016; Durgam et al. 2024; Zhu et al. 2021; Zhu et al. 2022; Shi et al. 2019; Wang et al. 2019a; Wang et al. 2024; Ito et al. 2024; Gu et al. 2022). Traditionally, SVIs are captured by commercial mapping firms, such as Google Maps, using dedicated camera arrays mounted on vehicles or trekker backpacks, which systematically traverse urban environments and record visual information along streets (Google 2025a). More recently, crowdsourced SVI platforms have also gained traction, allowing the public to contribute to data collection by uploading and sharing images captured with heterogeneous equipment (Huang et al. 2024). Owing to their rapidly expanding spatial coverage, especially within urban settings, SVIs have become a cost-effective and representative alternative to conventional data sources, such as on-site surveys (Dai et al. 2024; Yin et al. 2023; Clarke et al. 2010). For example, by 2019, Google had announced the collection of over 10 million miles of street view images (Google 2019). With the advancement of deep learning and computer vision (Chai et al. 2021; Marasinghe et al. 2024), SVIs have attracted significant research attention as a means to extract meaningful information using sophisticated models (Li et al. 2022e). Typically, these deep learning models are trained on benchmark SVI datasets tailored for specific tasks. For example, datasets such as Place Pulse 1.0 and 2.0 (Naik et al. 2014; Dubey et al. 2016), are widely used for studies of human-perceived indicators like safety and liveliness (Yu et al. 2024; Kang et al. 2020); semantic segmentation research relies on benchmarks such as CityScapes (Cordts et al. 2016; Gählert et al. 2020) and Mapillary Vistas (Neuhold et al. 2017); while geolocalisation models commonly employ datasets like Cross-View USA (CVUSA) (Workman et al. 2015; Zhai et al. 2017), CVACT (Liu and Li 2019b), and VIGOR (Zhu et al. 2021) to map SVIs to geo-referenced satellite images. These datasets provide SVIs spanning vast and diverse geospatial regions, thereby enabling models to learn spatial features with rich heterogeneity (Durgam et al. 2024). However, the prevailing body of research has largely focused on static images captured at single points in time, thereby neglecting the dynamic, temporal dimension of urban environments. This oversight introduces a potential temporal bias, which may significantly affect the robustness and generalisability of deep learning models trained on such data.

Alongside the rich spatial diversity provided by SVIs, recent studies have begun to explore the temporal dimension inherent in SVI data. For example, Liang et al. 2023 utilised historical SVIs to conduct spatial-temporal analysis of urban visual structure evolution; Wang et al. 2024 leveraged historical SVIs to assess changes in environmental quality over time; and Zhao et al. 2025 applied historical SVIs to evaluate urban form. These works primarily demonstrate the value of temporally diverse SVIs for specific analytic purposes. While these studies effectively leverage temporal evolution in SVIs to advance downstream tasks in urban analysis, the same temporal dynamics may pose additional challenges for certain applications. For instance, in geolocalisation tasks, SVIs collected from the same location at different times may introduce temporal bias, potentially impacting the performance of deep learning models. This gap appears to be especially pronounced in such tasks, for which, to our knowledge, little systematic evaluation of model sensitivity to temporal variation in input data has been conducted. This paper evaluates the impact of temporal bias on the outcome of some widely used pre-trained models for the purpose of cross-view geolocalisation. To do so, we construct historical SVI datasets on the locations of CVUSA, and undertake an initial empirical investigation into the evaluation and estimation of the impact of temporal bias on model performance.

This paper is organised as follows. Section 5.2 introduces the cross-view geolocalisation task and details the construction of our custom historical SVI dataset. Section 5.3 presents the methodology for evaluating the impact of temporal bias on retrieval performance, including model inference and semantic analysis. The experimental results and their spatial and temporal analyses are reported in Section 5.4. Section 5.5 provides an interpretation of these results and discusses the implications of temporal bias for the accuracy and reliability of cross-view geolocalisation methods. Finally, Section 5.6 summarises the main findings and outlines future research directions.

5.2 Cross-View Geolocalisation Task and Dataset Construction

5.2.1 Cross-View Geolocalisation Task Overview

Cross-View GeoLocalisation (CVGL) aims to identify the location where a ground-view image was taken by matching it to a geo-referenced aerial or satellite image of the same location but from a different viewpoint. CVGL has quickly become a powerful framework in geospatial AI, driven by the availability of large-scale geo-referenced images and advances in computer vision and deep learning. It infers precise locations from ground imagery by matching with aerial or satellite data, supporting applications like urban monitoring, navigation, asset management, and environmental analysis (Durgam et al. 2024). This approach is uniquely positioned to support a new generation of applications where visual context is either the primary or an essential input for decision-making. For example, in dynamic urban environments or disaster response scenarios, the ability to infer location and context from ground and aerial images underpins navigation (Kinnari et al. 2022) and search and rescue operations (Sogi et al. 2024). However, CVGL remains a challenging task due to the vast viewpoint variation between ground and aerial view images, which creates a significant domain gap and complicates the matching process. With the advent of machine learning and deep learning in recent years, geolocalisation using images has marked great strides by leveraging computer vision techniques (Durgam et al. 2024). Deep learning models can extract features from both ground and aerial view images and build an embedding space where corresponding pairs can be matched effectively (Zhu et al. 2022). Siamese networks are commonly adopted as the model architecture, using independent feature extractors for ground and aerial view images (Durgam et al. 2024). A loss layer then compares the outputs of each branch, building an embedding space where corresponding pairs are pulled closer while non-corresponding pairs are pushed farther apart. For example, Lin et al. 2015 introduced “Where-CNN”, which uses two Convolutional Neural Networks (CNNs) as feature extractors for ground and aerial inputs, and employs a Contrastive loss function (Hadsell et al. 2006) to minimise the Euclidean distance between paired inputs. Hu et al. 2018 further introduced the NetVLAD layer as a global descriptor on top of fully connected layers. Shi et al. 2019 proposed the use of a polar transform on aerial images to bring them closer to the ground domain before forwarding the inputs to a spatial-attention mechanism for embedding space learning. Liu and Li 2019b demonstrated that top-1 accuracy can be significantly enhanced with orientation information. In (Rodrigues and Tani 2023), pixel-wise keyword embeddings are derived from image semantic segmentation and used as a complementary input alongside images. Zhu et al. 2022 first introduced a pure transformer-based model for the cross-view task, which can build stronger global correlations compared to traditional CNN-based models. Shugaev et al. 2024 introduced ArcGeo, a cross-view image geolocalisation method that employs a novel batch-all angular margin loss function and large-scale pretraining strategies to significantly improve performance under limited field-of-view conditions.

5.2.2 Benchmark Datasets Overview

Alongside methodological advances enabling crossview geolocalisation, a series of benchmark datasets have been proposed for CVGL. Workman et al. 2015 introduced one of the first large-scale benchmark datasets, CVUSA (CrossView USA) dataset, containing over one million pairs of ground-level and aerial/satellite images from across the United States. The geo-referenced ground images in CVUSA originate from both Google Street View and Flickr, and the aerial images were generated using Bing Maps. Zhai et al. 2017 extended this work by leveraging the camera’s extrinsic information to warp the ground panoramas for alignment with aerial-view satellite images, resulting in 35,532 and 8,884 ground-to-aerial pairs for training and testing, respectively. Recent research on CVGL commonly relies on this subset (Zhu et al. 2022; Rodrigues and Tani 2023;

Shi et al. 2019; Sun et al. 2019), which is conventionally referred to as the "CVUSA" dataset. In the remainder of this paper, we refer to this subset as CVUSA. In (Liu and Li 2019b), the CVACT dataset was introduced, featuring a test set ten times larger than that of CVUSA, comprising 92,802 image pairs, while maintaining the same training set as CVUSA, with 35,532 samples. As an alternative to existing datasets that incorporate aligned ground-to-aerial image pairs and one-to-one retrieval for evaluation, Zhu et al. 2021 proposed the VIGOR dataset, which consists of 90,618 aerial images and 105,214 ground-view images from four major US cities of San Francisco, New York, Seattle and Chicago.

Recent and widely used CVGL datasets provide large volumes of image pairs, broad spatial coverage—from city-level (Liu and Li 2019b; Zhu et al. 2021) to national-level (Workman et al. 2015; Zhai et al. 2017)—and diverse geographic settings including both urban and rural areas. These characteristics have enabled deep learning models to achieve significant progress in CVGL tasks. However, these datasets often lack diversity in the temporal dimension and generally do not ensure temporal alignment between query and reference image pairs. As a result, important temporal variations (such as seasonal changes, differences in the time of day, or even variations within the same month) are often under-represented. This may introduce temporal bias and limit model generalisability and robustness. Despite some related research examining seasonal bias in urban form analysis tasks, such as the framework proposed by Zhao et al. 2025, there remains a need for systematic investigation into how temporal biases, especially those causing substantial scene changes like seasonal variations in street-view images, impact the accuracy and reliability of cross-view geolocalisation models. To address this gap, this study investigates whether and how street-view images captured at different times affect the performance of CVGL deep learning methods.

5.2.3 Construction of Custom Historical SVI Dataset from the CVUSA Validation Set

The geocoordinates associated with each image pair in the CVUSA validation set are provided as part of the dataset. We leveraged an open-source repository (robolyst 2025) to extract the corresponding Google Street View panorama identifiers (pano IDs). These were obtained by querying the geocoordinates associated with each image in the original CVUSA validation set. For every image in the validation set, we subsequently downloaded the associated historical SVIs by querying the extracted pano IDs using the

Google Street View Static API (Google 2025b). The corresponding metadata includes the year and month when each historical SVI was retrieved through the API. To ensure comparability during model inference, we intentionally downloaded all images for the custom dataset at the same aspect ratio and resolution as those in the original CVUSA validation set. In this paper, we utilised the same data source as CVUSA to minimise discrepancies arising from different data origins between our custom dataset and the original CVUSA validation set, thus ensuring a fair basis for comparison.

5.3 Methods

To investigate how temporal features impact cross-view geolocalisation, our framework comprises three primary stages: First, we selected the CVUSA validation set as the baseline dataset for evaluating the impact of temporal bias on widely used deep learning-based CVGL models; this set is hereafter referred to as the “val set”. The val set is conventionally employed for benchmarking the top-k accuracy across various CVGL studies, thus making it an appropriate foundation for our investigation into the effects of temporal variation in SVIs. To construct a temporally representative dataset, we performed extensive data collection in the vicinity of the original val set coordinates, expanding the dataset from its original size of 8,884 to over 43,000 panoramas. This custom dataset, hereafter referred to as the “custom set”, includes historical SVIs covering a broad range of time periods at each location, while maintaining spatial consistency with the val set. To assess spatial consistency, we calculated the absolute spatial distance between each image in the val set and its corresponding pairs within the custom set. Furthermore, similarity analysis was performed using a pre-trained CLIP model. Additionally, the statistical evaluation of the semantic components was conducted, providing a comprehensive comparison between the datasets. Second, to evaluate the impact of temporal bias on model performance, we applied two pre-trained cross-view geolocalisation models, SAFA and TransGeo, to both datasets. For each dataset, we summarised the pixel proportions of these semantic groups, facilitating a direct analysis of feature variation over time at each location. Finally, we conducted cross-view geolocalisation using the same pre-trained models on both the val and custom sets. Consistent with established conventions, top-k accuracy was adopted as the principal performance metric, enabling a systematic evaluation of model robustness and sensitivity to temporal bias as reflected in the semantic features of both datasets.

5.3.1 Distance Deviation and Similarity of Validation and Custom Dataset

SVIs captured at the same location but in different years or months may exhibit spatial deviations and are not always perfectly co-located. To quantify this spatial discrepancy, we measured the geographic distance between each original validation image and its temporally offset historical counterparts. This allows us to characterise the degree of spatial deviation introduced over time. All locations were first transformed from latitude-longitude coordinates (WGS84) into a local projected coordinate system. Specifically, for each image pair, an appropriate Universal Transverse Mercator (UTM) zone was selected based on the geographic location, and both points were projected into this local planar coordinate reference system. Distances were then computed as Euclidean distances in metres within the projected space.

Beyond distance discrepancy, SVIs collected at the same location may still reveal different patterns due to other factors, such as infrastructure renewal, demolition or reconstruction of buildings. To assess the scene-level holistic semantic similarity between historical street view images and their corresponding CVUSA validation set images, we utilised an off-the-shelf vision-language model, the ViT-B/32 CLIP model (OpenAI 2025), to extract high-level feature representations from each image pair. For every location, we extracted high-level feature representations from both the historical and its matched CVUSA image, and compared using CLIP semantic embeddings. This yielded a similarity score that captures the degree of visual and contextual resemblance, offering a quantitative measure of temporal semantic drift. After computing similarity scores for all matched pairs, we analysed their distribution by constructing an empirical CDF. Key percentiles were identified to provide an overview of similarity patterns and highlight the extent of variation between historical and reference images. The resulting CDF visualisation, together with percentile statistics, enables a nuanced assessment of how much semantic shift may occur at the same location over time.

5.3.2 Semantic Segmentation and Category-wise Pixel Proportion Calculation

SVIs captured at the same location but at different times may exhibit notable visual differences due to temporal variation. For instance, seasonal variation can lead to significant differences in the presence and density of greenery, resulting in considerable visual disparity between the images taken from the same geographic location but at different times. To evaluate the semantic composition of each street view image in both the custom and validation datasets, we employed a pre-trained semantic segmentation model, Mask2Former, which had been trained on the Cityscapes dataset. As a result, every pixel in the image was assigned to one of the 19 semantic classes defined by Cityscapes. To enable higher-level analysis and simplify downstream evaluation, the 19 original classes were consolidated into six broader semantic categories: Architectural, Greenery, Paved Surface, Sky, Urban Furniture, and Movable Objects (which include dynamic entities such as vehicles and pedestrians). The mapping from the original Cityscapes classes (Cordts et al. 2016; Gählert et al. 2020) to these regrouped categories is provided in Table 5.1. For each semantic group g, the pixel portion is defined as follows. For an image i in the val set, its proportion of pixels assigned to semantic group g is given by:

$$S_g^{(i)} = \frac{N_g^{(i)}}{N_{\text{total}}^{(i)}} \qquad (i \in \text{val set}) \tag{5.1}$$

where $N_g^{(i)}$ and $N_{\text{total}}^{(i)}$ denote the number of pixel of group g, and the total quantity of pixel of image i, respectively. For a historical image corresponding to the validation set image i, captured at time t in the custom set, the pixel portion is defined as:

$$S_g^{(i,t)} = \frac{N_g^{(i,t)}}{N_{\text{total}}^{(i,t)}} \tag{5.2}$$

where $N_g^{(i,t)}$ and $N_{\text{total}}^{(i,t)}$ denote the number of pixels belonging to group g and the total number of pixels at time t, respectively, in the image i of the val set.

This approach enabled a systematic comparison of the distribution of urban elements across images and between the custom and val sets. We aggregated the pixel proportions for each semantic category across all images and used these statistics to analyse temporal variation in scene composition. This allows us to investigate how the semantic composition of SVIs varies over time and how such variation may contribute to temporal bias in CVGL tasks.

Table 5.1: Mapping of Cityscapes semantic classes into grouped categories

Grouped Category	**Cityscapes Classes**
Architectural	building, fence, wall
Greenery	vegetation, terrain
Paved Surface	road, sidewalk
Sky	sky
Urban Furniture	pole, traffic sign, traffic light
Movable Objects	car, truck, person, bus, train, rider, bicycle, motorcycle

Cityscapes classes are grouped based on visual and functional similarity.

5.3.3 Semantic Analysis of Temporal Changes in Street View Images

To quantitatively examine the temporal variations in the semantic composition between the custom and val sets, we implemented a structured analytical framework. For the custom set, the month information was extracted from the metadata associated with each image, as described in Section 5.2.3. To further capture temporal variation, season tags were derived from the month information and assigned to each image in the custom set. As the study area covers the United States, which is located in the northern hemisphere, the months were assigned to the seasons as follows: Spring (March-May), Summer (June-August), Autumn (September-November), and Winter (December-February). This temporal annotation enabled the stratification of the data to explore both seasonal and monthly patterns in semantic changes. Comparative analyses were performed on matched image pairs from the custom and validation sets, aligned by geographic location. At each location, the absolute change in the proportion of each semantic group over time was calculated to quantify the temporal semantic variation. For a given semantic group g at location i and time t, the absolute change is defined as

$$\Delta_g^{(i,t)} = S_g^{(i,t)} - S_g^{(i)} \tag{5.3}$$

where $S_g^{(i,t)}$ denotes the proportion of pixels assigned to semantic group g in the custom set image captured at time t, and $S_g^{(i)}$ is the corresponding proportion in the val set image at location i. Seasonal and monthly trends in these semantic changes were further assessed by aggregating the absolute differences according to season and month. For each semantic group, the seasonal mean and variance were calculated. To evaluate broader temporal trends, the absolute differences were aggregated by season and month. For each semantic group, we computed the seasonal mean and variance, thereby enabling an assessment of how temporal factors influence semantic composition across different times of year.

5.3.4 Inference via SAFA and TransGeo Models

For the inference stage, we employed two state-of-the-art cross-view geolocalisation deep learning models: TransGeo (Zhu et al. 2022) and SAFA (Shi et al. 2019). The selection of these models was motivated by two primary considerations. Firstly, both TransGeo and SAFA have demonstrated high top-*k* accuracy on the CVUSA validation set, which serves as the baseline dataset in this study. This ensures that the models selected for evaluation are representative of the current best-performing approaches within the CVGL field. Secondly, TransGeo and SAFA are based on fundamentally different network architectures. TransGeo utilises a Transformer-based design, while SAFA adopts a CNN architecture. By including both a Transformer-based and a CNN-based model in our experiments, we are able to control for the potential influence of model architecture on inference outcomes. This approach provides a more balanced and comprehensive assessment of how temporal bias in SVIs affects model performance, regardless of the architectural differences.

The inference process consisted of several key stages. First, each query SVI from the custom set was systematically paired with its corresponding reference satellite image from the validation set, based on location ID. All images were pre-processed and normalised using the ImageNet mean and standard deviation values to ensure consistency with the configurations used during model training and validation, thereby mitigating input-related variance. Subsequently, feature extraction was performed using the selected model, with both the query SVIs and reference satellite images encoded into high-dimensional feature vectors. For each query image, similarity scores with all reference images were computed in the learned feature space. The retrieval results were then ranked based on these similarity scores, and standard top-*k* accuracy metrics were reported to assess retrieval performance. Specifically, after sorting the retrieval images in ascending order, if the ground-truth aerial image is ranked among the top-*k* candidates, it is considered a successful retrieval. To ensure direct comparability with the original model benchmarks, the evaluation was limited to the top-10 retrieval results during inference.

5.3.5 Semantic Composition Analysis and Model Interpretability

To systematically investigate the relationships between the temporal semantic characteristics and the retrieval performance of the model, we extracted the proportion of each semantic category as defined in Section 5.3.2. Additionally, we included the number of semantic categories per image, defined as the count of categories whose pixel proportions exceeded a minimum threshold of 10%, as an aggregated feature to represent overall scene complexity. Model retrieval success was encoded as a binary outcome: for each query SVI in the custom set, a value of one was assigned if its ground-truth reference appeared among the top-10 retrievals generated by the deep learning model, and zero otherwise. This approach allowed us to treat successful retrieval as a classification target in subsequent analyses. We also applied clustering analysis to group SVIs based on their semantic composition and to examine the variation in retrieval performance across different scene types. Following the preceding analyses, we further examined the influence of semantic scene composition on retrieval success by conducting a SHAP (SHapley Additive exPlanations) analysis to provide explanation and attribution of each component on the outcome of a model. Specifically, the goal was to quantify the contribution of individual semantic features—and an additional indicator, the number of present semantic categories—to the Top-10 retrieval accuracy for both the TransGeo and SAFA models. We then trained an XGBoost classifier (Chen and Guestrin 2016) to predict retrieval success based on the semantic proportions and category diversity of each query image. The SHAP analysis was applied to this predictive model in order to attribute and rank the relative importance of each semantic feature.

5.4 Results

5.4.1 Summary of Custom Dataset

The custom dataset comprises 43,347 panoramas, each collected from locations corresponding to those in the CVUSA validation set. For every location in the validation set, all available historical SVIs within the vicinity were retrieved, spanning a wide range of acquisition dates. As illustrated in Fig. 5.1, each geo-referenced panorama

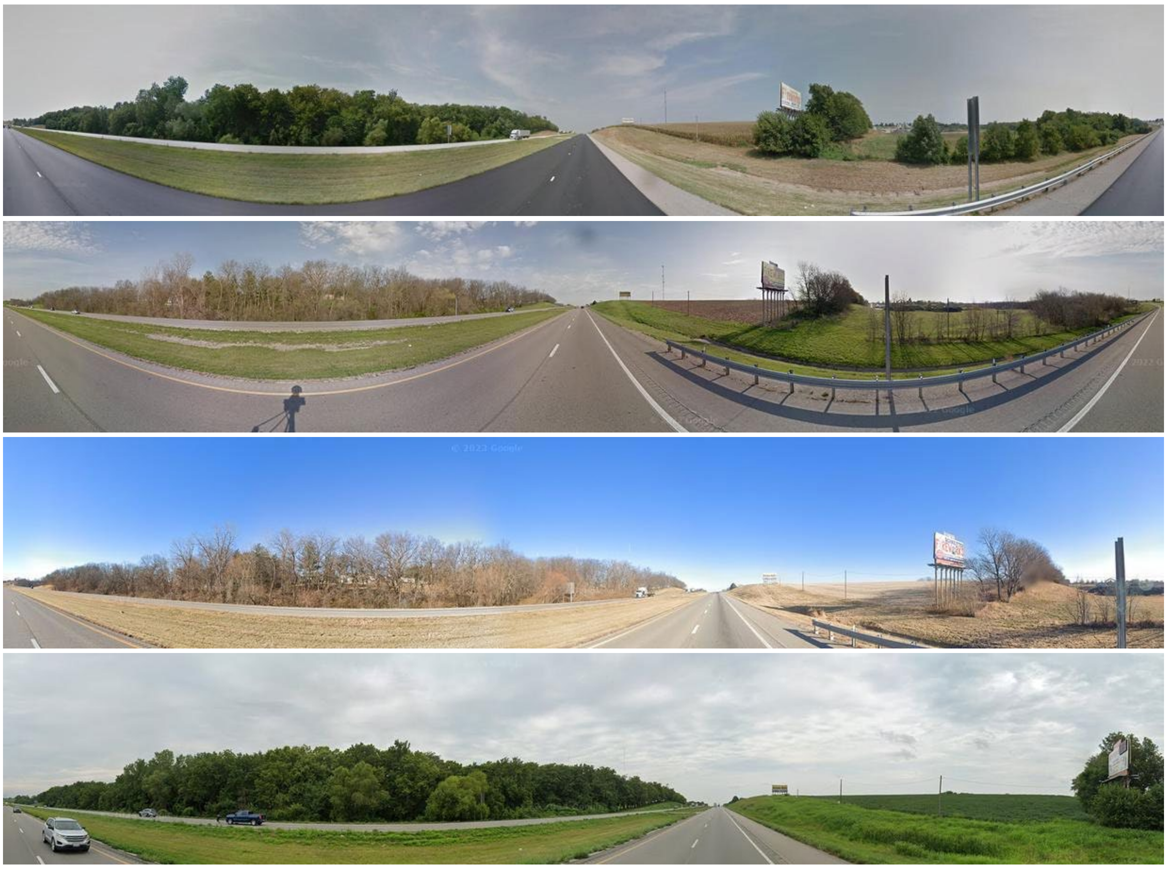


Figure 5.1: Examples of original SVIs from the CVUSA validation set (top row) and their corresponding historical SVIs from the custom dataset, collected at different times (subsequent rows).

from the CVUSA validation set (top row) is paired with its corresponding historical panoramas from the custom set, collected in the same vicinity. While the overall spatial layout remains largely stable over time, visual patterns exhibit noticeable variation due to temporal changes in scene appearance. In instances where multiple SVIs from the same location shared the same year-month timestamp, only one panorama was randomly selected to represent that temporal interval, thereby reducing redundancy. Consequently, the custom dataset contains 43,347 panoramas, in contrast to the 8,884 images in the CVUSA validation set. The seasonal and monthly distributions of the custom dataset are presented in Fig. 5.2. The number of available panoramas in the custom set peaks in September, with high counts also in August and October. In terms of seasonal distribution, summer and autumn are the most represented, each comprising over 14,000 panoramas and making up two-thirds of the total sample. In contrast, spring is less prevalent, and winter accounts for the smallest proportion of panoramas in the dataset.

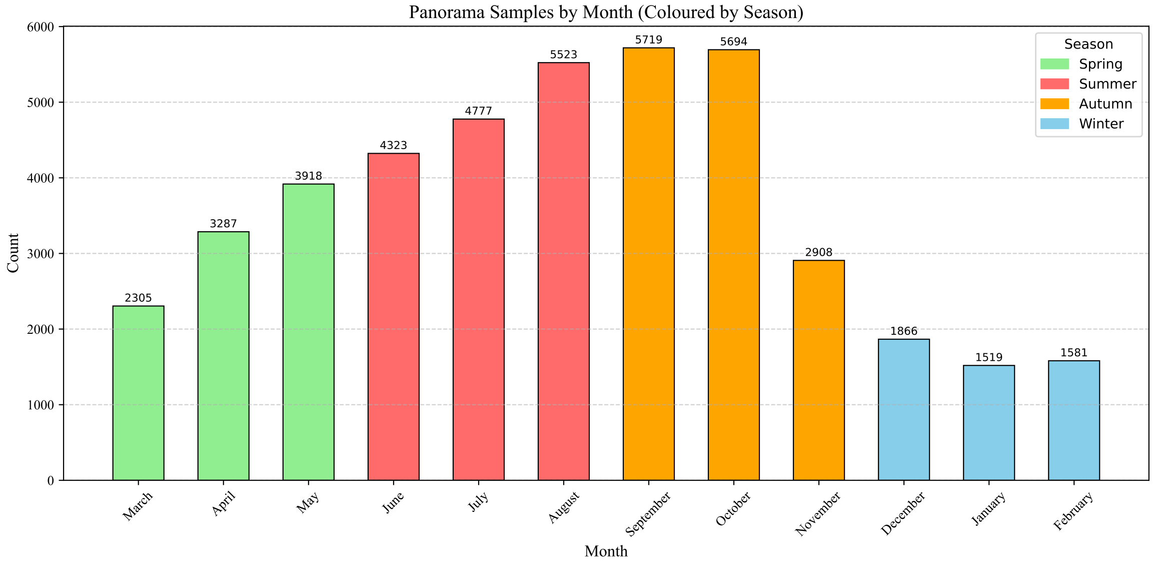


Figure 5.2: Seasonal and monthly distribution of unique street view panoramas in custom dataset.

To assess the extent of spatial deviation between historical SVIs and their corresponding reference locations, we calculated pairwise geodesic distances as described in Section 5.3.1. The maximum and minimum distance deviations between a panorama from the custom set and its corresponding pair in the val set are 184.12 metres and 0.04 metres, respectively. Across all matched locations, the mean pairwise geodesic distance is 20.74 metres, with a standard deviation of 17.30 metres. The CDF of these distances, as shown in Fig. 5.3, provides a comprehensive overview of how location shifts manifest across the dataset. Notably, half of the distance deviations are within 15.53 metres, while 95% and 99% of all deviations are within 49.69 metres and 65.43 metres, respectively. As reported in (Workman et al. 2015), the aerial images in CVUSA are collected at zoom level 19. According to Zhai et al. 2017, the size of each aerial image is 750×750 pixels, which corresponds to a ground coverage of approximately $0.298 \times 750 \approx 224$ metres per side. Therefore, the majority of SVIs in the custom dataset remain well within the same aerial image grid as their corresponding validation images, given that the 99th percentile of the CDF (65.43 metres) is substantially smaller than the aerial grid size.

The pairwise cosine similarity between images, computed using a pre-trained CLIP model as described in Section 5.3.1, is summarised in Fig. 5.4. The CDF reveals that half of the matched image pairs exhibit a cosine similarity above 0.822, while the 95th and 99th percentiles reach 0.914 and 0.945, respectively. This indicates a generally high

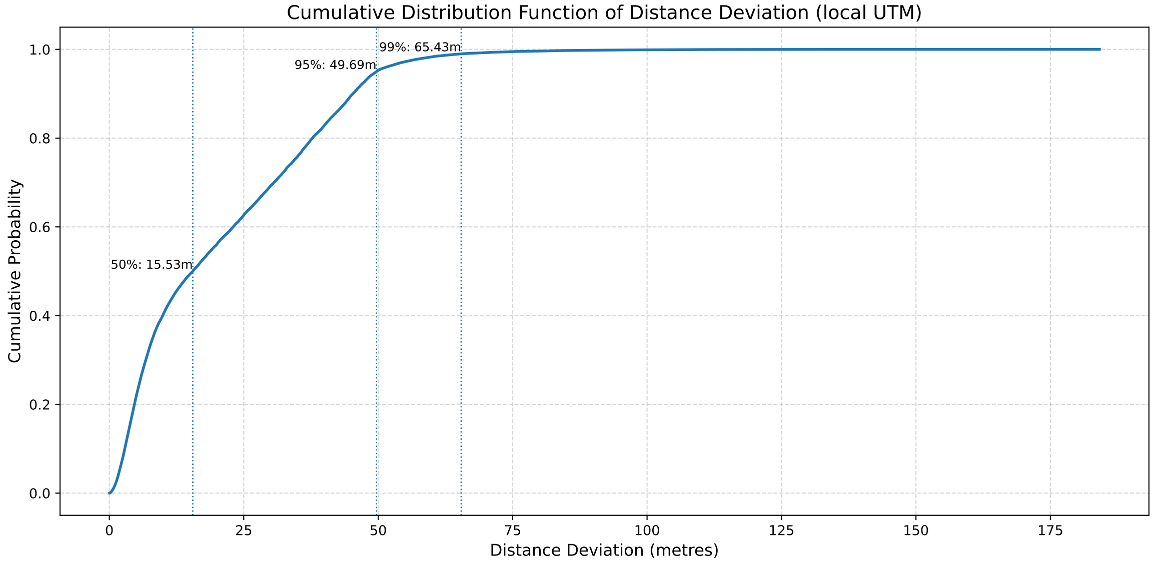


Figure 5.3: Cumulative distribution function of distance deviation, with 50%, 95% and 99% percentiles.

level of semantic similarity between historical panoramas in the custom set and their corresponding images in the CVUSA validation set. These findings further support the effectiveness of our custom set generation process, suggesting that historical panoramas largely preserve the holistic semantic content of their reference counterparts.

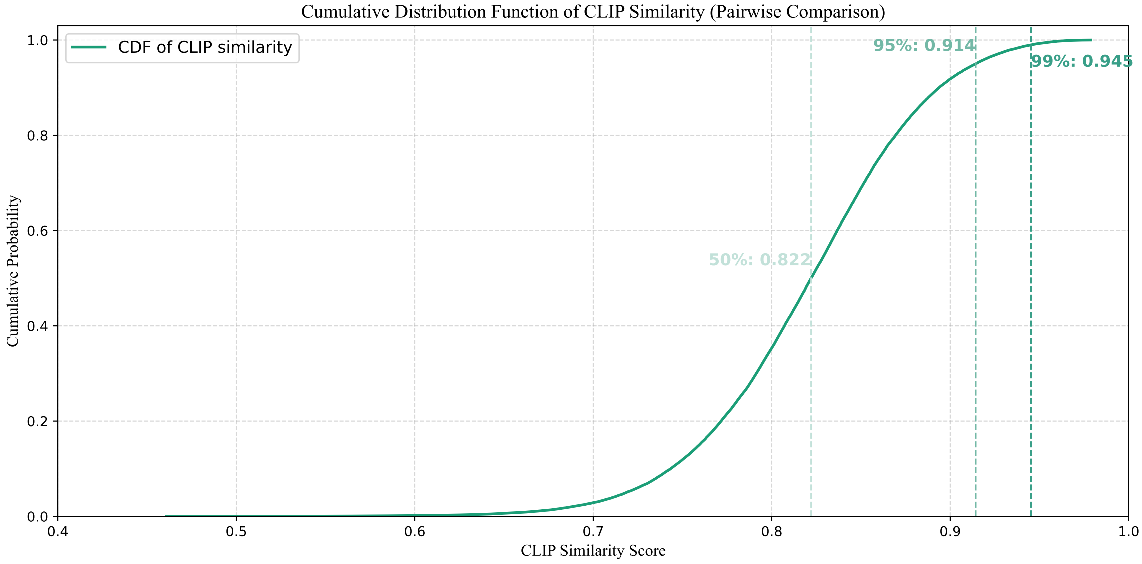


Figure 5.4: Cumulative distribution function of pairwise cosine similarity scores between historical panoramas and their matched CVUSA validation set images. The median, 95th percentile, and 99th percentile similarity scores are annotated.

To further examine the relationship between spatial distance deviation and semantic similarity, we visualised the distribution of CLIP similarity scores as a function of distance deviation percentiles, as shown in Fig. 5.5. This analysis allows us to investigate whether greater spatial offsets are associated with reduced semantic consistency between matched image pairs. The results demonstrate that CLIP similarity remains relatively stable across the entire range of distance deviations, even when binned in 20-percentile intervals. This stability further validates the comparability between the custom and validation datasets, ensuring that the historical panoramas are suitable for subsequent comparative analyses.

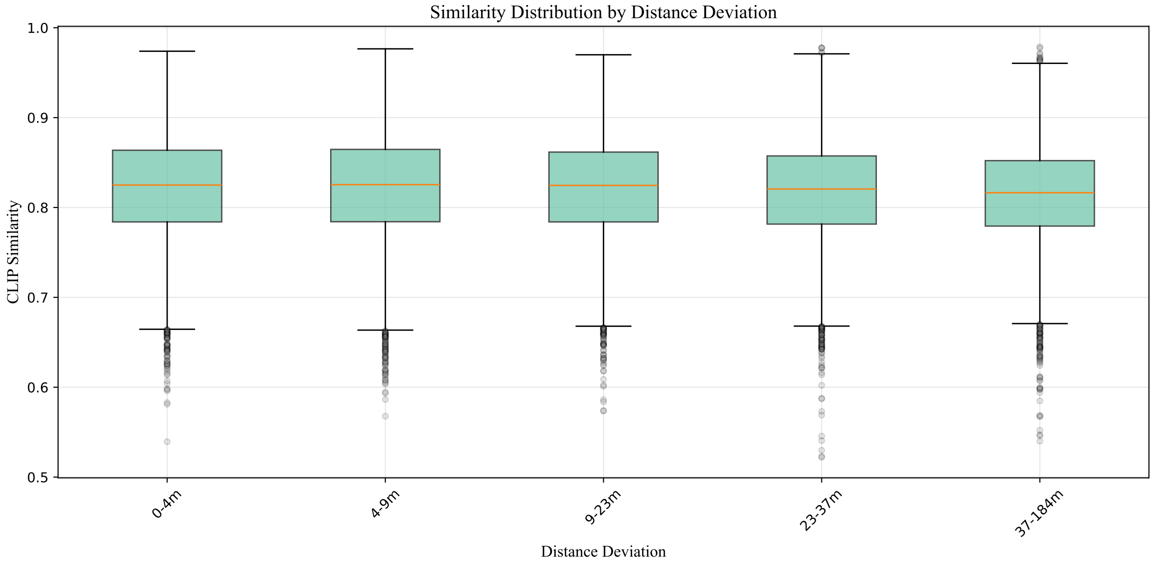


Figure 5.5: Boxplots of pair-wise CLIP cosine similarity (range [0, 1], higher values indicate greater visual similarity) across distance deviation percentiles. Each box represents similarity scores grouped into 20-percentile intervals based on the relative distance deviation between matched historical and validation panoramas. Similarity remains consistent across distance deviations.

Beyond CLIP similarity, we also illustrate the application of the Simpson Index (Simpson 1949) by computing it for the semantic segmentation images of SVIs captured at different times at the same location. The Simpson Index, defined as: $D = \sum_{i=1}^{S} p_i^2$, where S denotes the total number of semantic categories in a segmentation image, and p_i^2 is the proportion of pixels assigned to category i. Fig. 5.6 presents a representative example for a specific location. This example demonstrates how semantic diversity may vary over time, influenced by factors such as urban redevelopment (e.g., building demolition and reconstruction, or the addition or removal of signage), as well as the dynamic presence or absence of movable objects such as pedestrians and vehicles.

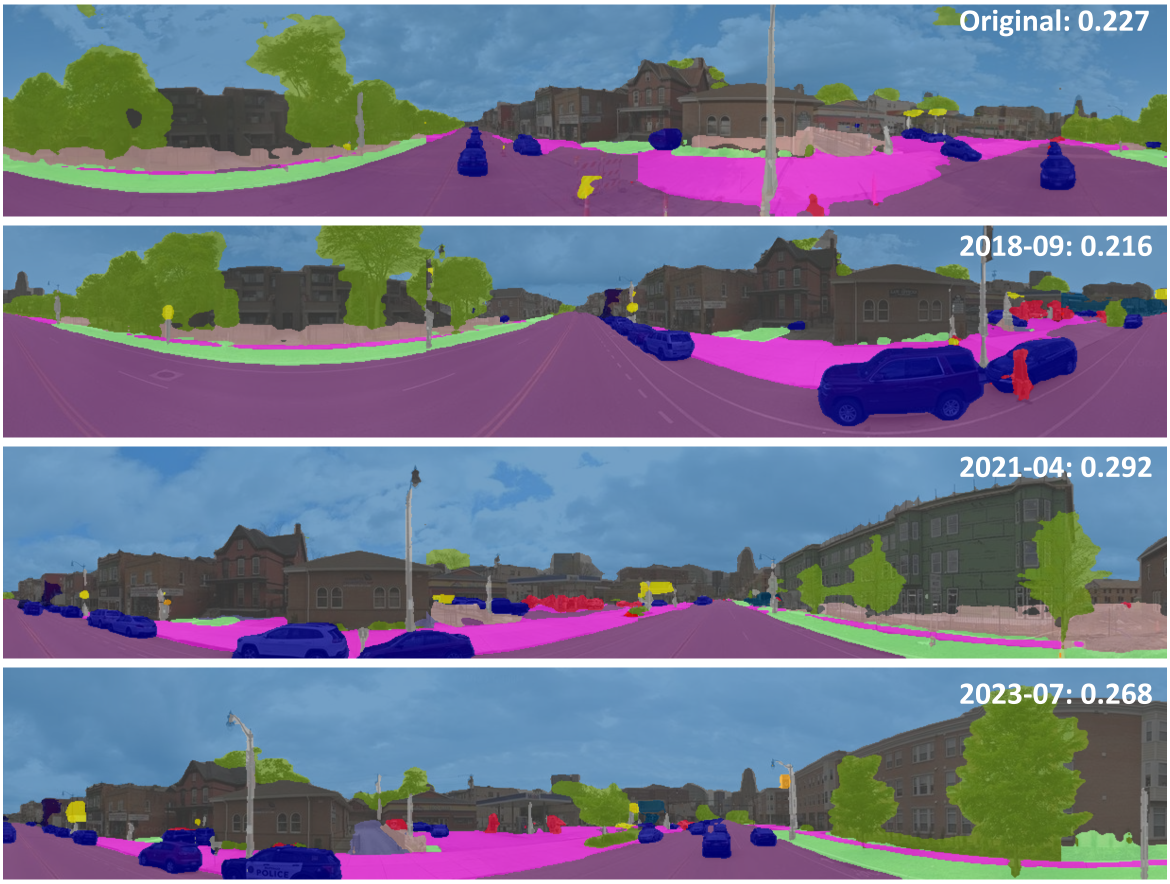
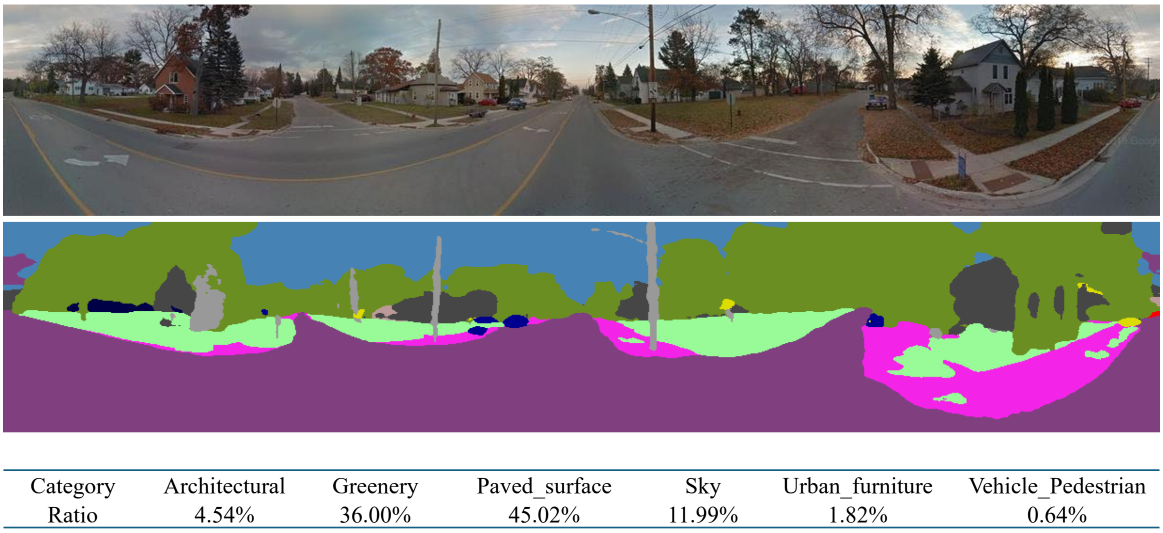


Figure 5.6: Example of the Simpson Index (range [0, 1]) computed across a sequence of panoramas captured at the same location over different time periods. Lower values indicate greater visual diversity across time, while higher values correspond to more stable visual appearance.

Using the method described in Section 5.3.2, panoramas from both the custom and validation sets were semantically segmented with the Mask2Former model pre-trained on the Cityscapes dataset. Following segmentation, the proportion of pixels belonging to each semantic category was calculated according to the regrouping scheme outlined in the methodology. As illustrated in Fig. 5.7, the example demonstrates the result of semantic segmentation for a panorama, along with the resulting proportions for each regrouped semantic category.



Category	Architectural	Greenery	Paved_surface	Sky	Urban_furniture	Vehicle_Pedestrian
Ratio	4.54%	36.00%	45.02%	11.99%	1.82%	0.64%

Figure 5.7: Example of semantic segmentation and regrouped pixel proportion calculation for a street view panorama. The top row: input panorama image; middle row: corresponding semantic segmentation map; bottom table: computed pixel proportions for each regrouped semantic category.

5.4.2 Temporal Patterns in Semantic Category Variation

For each panorama in the custom temporal dataset, the proportion of pixels belonging to each regrouped semantic category was first computed. To quantify temporal variation, these semantic proportions were then compared to those of the corresponding reference SVI image from the CVUSA validation set at the same location. The absolute difference in proportions for each category yielded a set of semantic change scores per image pair. To explore monthly and seasonal trends, semantic change scores were aggregated accordingly. For each month, the mean and standard deviation of category-wise absolute changes were computed across all matched image pairs. As illustrated in Fig. 5.8, a line chart displays the mean change per month for each regrouped category, with a shaded band denoting one standard deviation above and below the mean, reflecting inter-location or inter-year variability for each month. For seasonal trends, Fig. 5.9 presents the averaged seasonal changes as grouped bar plots, where each bar shows

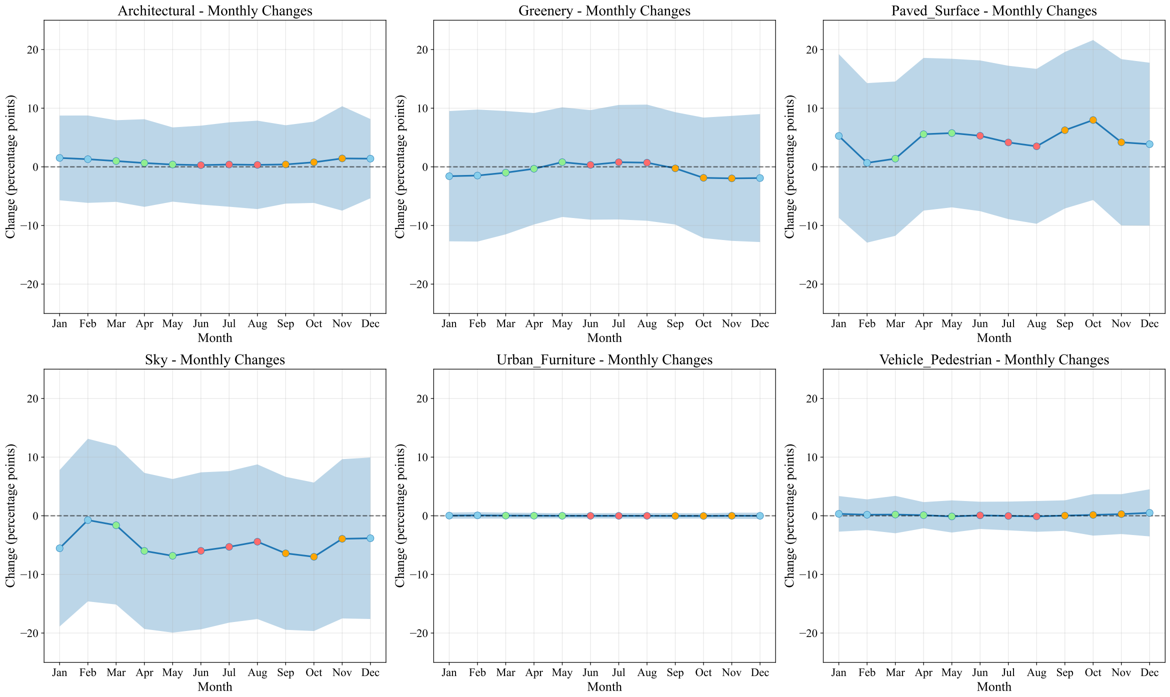


Figure 5.8: Monthly variation in the average change of each semantic category. The line represents the mean change for each month, with the shaded region indicating the mean ± standard deviation.

the mean change and its error bar denotes ± one standard deviation for that season. The distribution of semantic category changes across the custom dataset is presented in Fig. 5.10, hilighting the magnitude and direction of absolute changes in semantic composition for each category. It can be observed that both the “Architectural” and “Greenery” categories exhibit similar distributions of change, with values centred and peaked near zero, and a gradual decrease in frequency towards the extremities of the percentage axis. Furthermore, as shown in Fig. 5.8, these two categories display an inverse relationship over time, which is likely attributable to seasonal variation in foliage coverage around buildings. Specifically, greater vegetation during certain periods may obscure architectural elements, whilst sparser foliage allows for increased visibility of buildings. A similar relationship is observed between the “Paved Surface” and “Sky” categories, which respectively dominate the lower and upper portions of a panorama. Accordingly, an increase in the proportion of one is generally accompanied by a decrease in the other. Finally, both the “Urban Furniture” and “Movable Objects” categories demonstrate relatively stable pixel proportions across the entire dataset, as evidenced by their frequency distributions and patterns of temporal variation.

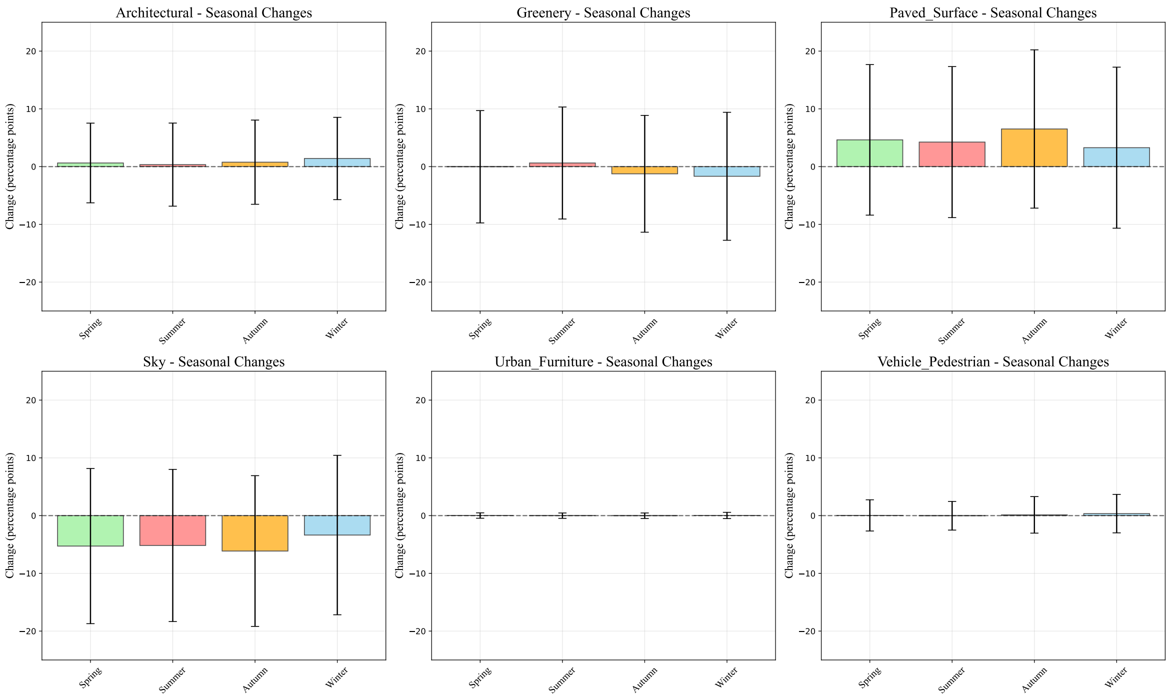


Figure 5.9: Seasonal variation in the average change of each semantic category. Bars represent the mean change for each season, with error bars indicating the mean ± standard deviation.

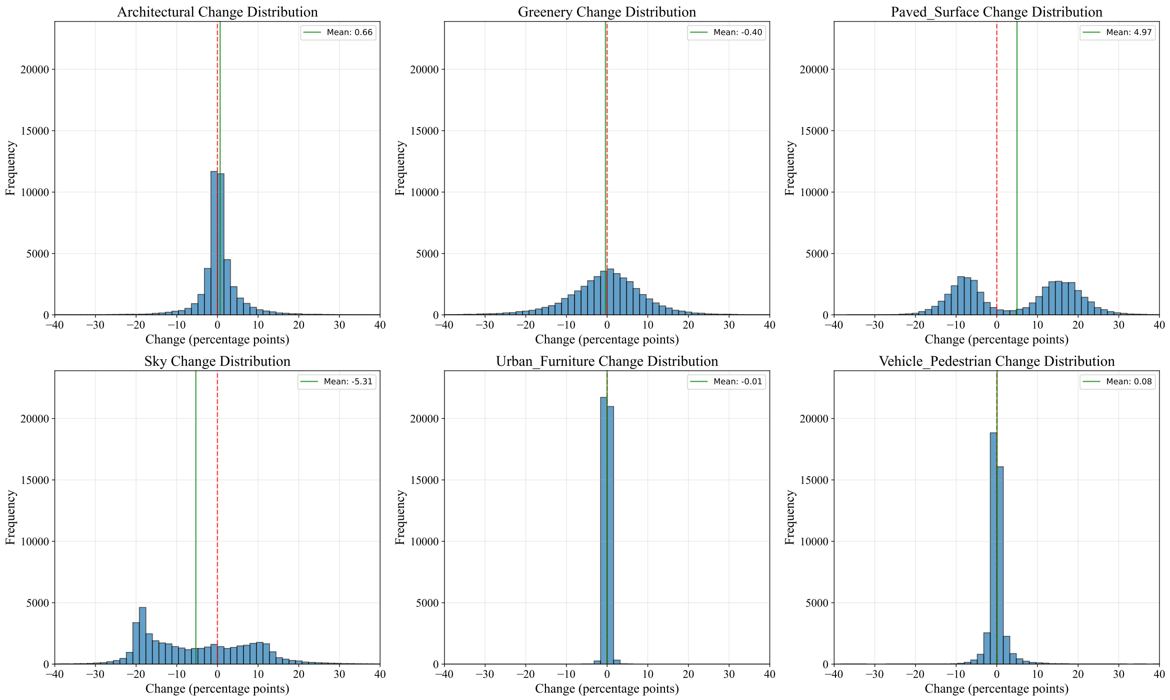


Figure 5.10: Distribution of percentage change for each semantic category in the custom dataset. The horizontal axis denotes the absolute change (in percentage points)

Table 5.2: Cross-view retrieval performance (%) of pre-trained TransGeo and SAFA models on the original CVUSA validation set and the custom dataset

Model	**CVUSA val (8.8k)**			**Custom (43k)**		
	Top-1	Top-5	Top-10	Top-1	Top-5	Top-10
TransGeo	93.39	98.26	98.98	5.77	12.82	17.51
SAFA	89.43	96.92	98.04	4.05	6.56	7.98

5.4.3 Top-k Retrieval Performance of Pre-trained Models

The top-*k* retrieval accuracy results for the pre-trained TransGeo and SAFA models are presented in Table 5.2. Both models demonstrate strong retrieval performance on the original CVUSA validation set, consistent with previously reported benchmarks. However, when evaluated on the substantially larger and more temporally diverse custom dataset, both models show a marked decline in retrieval accuracy across all top-*k* metrics. This substantial performance drop is observed not only for the top-1 retrieval, but also for top-5 and top-10, indicating that the challenge is systemic rather than restricted to a particular metric.

5.4.4 Scene Clustering by Semantic Composition

The results of KMeans clustering, performed on the semantic category proportions of panoramas in the custom dataset, are summarised in Table 5.3. The number of clusters was empirically set to four, as this configuration yielded groups with clearly distinct scene characteristics. Cluster 0 is characterised by a dominant "Sky" proportion, with an average of 54.0%, representing open sky scenes. Cluster 2, on the other hand, exhibits the highest proportion of "Greenery" at 70.5%, indicative of natural landscapes. In contrast, clusters 1 and 3 show substantially lower proportions of both "Sky" and "Greenery", while "Architectural" and "Paved Surface" categories account for a much larger share. This composition suggests that these clusters primarily correspond to built-up or urban environments, where artificial structures and paved areas prevail. Examining the cluster-wise Top-10 retrieval accuracies reported in Table 5.3, it can be seen that for both TransGeo and SAFA, the retrieval performance reaches its highest values in cluster 1 and cluster 3. For example, the Top-10 accuracy for TransGeo is 17.2% in cluster 1 and 21.1% in cluster 3, while for SAFA, these clusters yield 10.6% and 8.8%, respectively. In contrast, clusters 0 and 2 exhibit noticeably lower retrieval accuracies for both models.

Table 5.3: Mapping of Cityscapes semantic classes into grouped categories

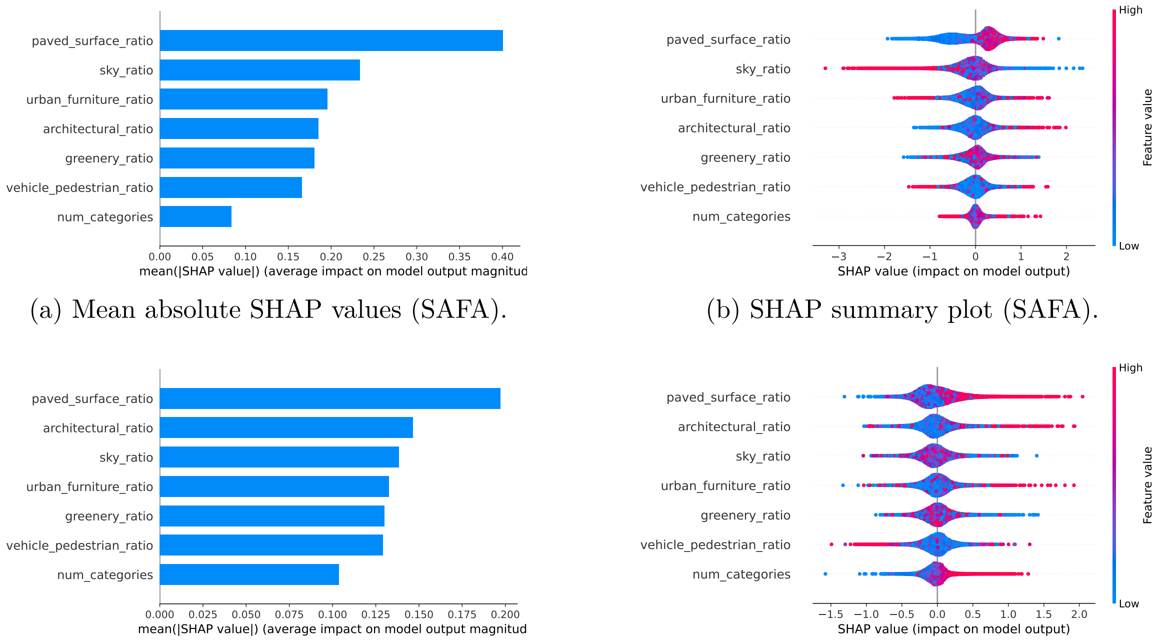
Cl.	Share	T10-TG	T10-SF	Ar.	Gr.	PS.	Sky	UF.	MO.
0	22.6%	16.6	5.2	4.5	32.3	8.2	54.0	0.4	0.6
1	36.4%	17.2	**10.6**	3.6	39.7	33.3	22.3	0.3	0.7
2	22.1%	15.8	5.7	2.8	70.5	11.7	14.2	0.3	0.6
3	19.0%	**21.1**	8.8	11.5	26.9	23.4	32.6	1.3	4.3

T10-TG and T10-SF denote top-10 retrieval accuracy (%) for TransGeo and SAFA respectively. Cl. = Cluster ID, Ar. = Architectural, Gr. = Greenery, PS. = Paved Surface, Sky = Sky, UF. = Urban Furniture, MO. = Movable Objects.

5.4.5 Interpretation of Semantic Feature Contribution via SHAP

Building on these findings, we conducted a SHAP-based interpretability analysis (Lundberg and Lee 2017) to examine how individual semantic categories—along with an additional custom indicator, the number of semantic categories present in each image—contribute to Top-10 retrieval success for both the TransGeo and SAFA models. To do this, the top-10 is converted into a binary classification label (1 if successful, 0 otherwise). An XGBoost classifier was then trained to predict the outcome based on the semantic composition of each query image. The resulting SHAP values quantify the contribution of each semantic feature to the classifier's prediction. Fig.5.11 provides an overview of SHAP results across both models. Figs.5.11a and 5.11c present the mean absolute SHAP values for each semantic attribute in SAFA and TransGeo, respectively, reflecting the overall importance of each feature in predicting retrieval success. Figs. 5.11b and 5.11d display the distribution of SHAP values for all samples, where the horizontal position indicates the direction and magnitude of each feature's effect on the prediction probability, and the colour denotes the feature value.

Among all features, "Paved Surface" emerges as the most impactful for both models, with its mean absolute SHAP value significantly surpassing those of other semantic categories. Furthermore, this feature exhibits a strong positive correlation with binary Top-10 retrieval outcome of the models. Beyond Paved Surface, both "Architectural" and "Urban Furniture" consistently rank among the top four most important features for retrieval success in both models. While their exact order of importance differs: "Urban Furniture" ranks above "Architectural" in the SAFA model, and the reverse is true for TransGeo. By contrast, both "Sky" and "Greenery" are negatively associated with retrieval success in both models, suggesting that scenes dominated by these elements are less likely to be retrieved correctly. "Movable Objects" and "Number of Categories" demonstrate the least contribution to model predictions. Notably, the directionality of their impact is less consistent: for "Movable Objects", SHAP summary



(a) Mean absolute SHAP values (SAFA).

(b) SHAP summary plot (SAFA).

(c) Mean absolute SHAP values (TransGeo).

(d) SHAP summary plot (TransGeo).

Figure 5.11: Explanability of feature importance with SHAP. (a) and (c) show feature importance rankings for SAFA and TransGeo, respectively, based on mean absolute SHAP values. (b) and (d) show SHAP value distributions across all samples for SAFA and TransGeo models.

plots reveal red dots at both extremes, indicating that both high and low values can can have positive or negative effects depending on context. For "Number of Categories", a similar effect is observed in the SAFA model, while in TransGeo, it tends to correlate positively with retrieval success.

5.5 Discussion

Existing research typically relies on street view images to draw conclusions from visual information captured at specific time points. Although SVIs constitute a rich data source for such objectives, most studies neglect the potential impact of temporal variation in visual information. In this study, we systematically investigated the impact of temporal bias in street view images on cross-view geolocalisation performance. By constructing a large-scale, temporally diverse SVI dataset that is spatially aligned with the CVUSA validation set, we evaluated two state-of-the-art deep learning models, TransGeo and SAFA, under conditions that maintain spatial correspondence but introduce substantial temporal variation, providing an out-of-distribution test scenario. Our results highlight a substantial performance drop compared to conventional benchmarks, revealing the sensitivity of current models to temporal variation. Furthermore, through a comprehensive analysis of semantic features and their influence on retrieval outcomes, we identified key scene attributes that shape model performance. This work provides new empirical evidence of the need to address temporal dynamics in the development and evaluation of cross-view geolocalisation methods that rely on SVI.

This substantial drop in retrieval accuracy primarily underscores the pronounced effect of temporal bias between the original CVUSA validation set and our temporally diverse custom dataset, even though both are spatially aligned and sampled from identical locations. The custom dataset introduces significant variation in image acquisition time, seasonal appearance, and transient urban conditions. As a result, despite their strong performance on conventional benchmarks, current cross-view models display limited generalisation capability when confronted with temporal shifts or out-of-distribution samples. The consistent performance decline observed across both model architectures suggests that the challenge of generalising across diverse spatiotemporal contexts is not model-specific. These findings highlight the need for further research into improving temporal generalisation, mitigating temporal bias, and enhancing the robustness of cross-view geolocalisation methods in real-world, time-varying environments. This work focuses on isolating temporal and semantic variation under controlled spatial matching. An important direction for future work is to examine whether and how temporal performance degradation exhibits systematic spatial structure across different urban contexts.

The SHAP analysis underscores that “Paved Surface” and “Architectural” features are highly influential for the successful geolocalisation of street view images in cross-view retrieval models. This likely stems from the dual visibility of these elements in both ground-level and aerial images, allowing the models to establish more reliable spatial correspondences. In contrast, “Greenery” and “Sky” categories contribute comparatively little to localisation performance, as their presence offers limited unique spatial cues for matching across views. It is also important to note that reductions in the proportions of “Greenery” and “Sky” often naturally correspond to increases in “Architectural” and “Paved Surface” content, as discussed in Section 5.4.2. Although “Urban Furniture” constitutes a much smaller share of SVI pixels, it contributes considerably to model predictions for both TransGeo and SAFA. These findings collectively suggest that images characterised by a higher abundance of built environment features, such as buildings, roads, and urban infrastructure, are more amenable to accurate geolocalisation. However, it is important to recognise that the relevance of different visual features may vary substantially across GeoAI tasks. Future work could extend this analysis by involving temporally indexed satellite imagery, enabling a systematic investigation of how query-side and reference-side temporal changes jointly affect cross-view retrieval performance, and whether different sources of temporal drift exhibit asymmetric or interacting impacts across urban contexts. To improve transferability to historical data, future methods could focus on learning representations that are less sensitive to transient appearance changes and more grounded in stable structural semantics. In addition, incorporating multi-temporal training signals or explicitly modelling the asymmetry between temporally variable queries and relatively stable references may help models generalise across time. This study has several limitations that suggest promising avenues for future work. First, the images in the custom dataset are not perfectly north-centred aligned, as is the case for the CVUSA validation set. While north-centring is feasible in principle, it may fail in practice due to incomplete panoramic coverage in some historical SVIs. Second, although we collected a large set of historical panoramas to enhance the temporal diversity of ground-view information, we continued to use the original aerial images from the CVUSA val set as the geo-referenced database for retrieval. Future research could further enrich the temporal alignment of cross-view datasets by constructing temporally diverse aerial images, enabling a more comprehensive assessment of temporal effects on both ground and aerial views. Furthermore, while SHAP offers valuable insights into feature importance, its interpretability is inherently tied to the chosen model and set of input features, and may not fully account for all sources of model uncertainty or context-dependent effects.

5.6 Conclusion

Most existing research relies on benchmark datasets such as CVUSA to perform cross-view geolocalisation tasks, benefiting from their rich spatial diversity. However, the potential impact of temporal bias on task performance remains underexplored. In this study, we systematically evaluated the impact of temporal feature variation in street view images on the performance of deep learning models for cross-view geolocalisation. To this end, we constructed a large-scale historical street view dataset that is spatially aligned with, but temporally distinct from, the widely used CVUSA validation set. Using this temporally diverse dataset as input, we assessed two state-of-the-art cross-view geolocalisation models and observed a substantial decline in retrieval accuracy compared to results on the original validation set. Through semantic segmentation and systematic analysis, we demonstrated that even at identical locations, visual feature shifts induced by temporal change can compromise model accuracy: both TransGeo and SAFA showed dramatic performance drops in Top-1 accuracy, from 93.39% to 5.77% and from 89.43% to 4.05%, respectively. Our explainability analysis further revealed that features with dual visibility in both ground-level and aerial views, such as roads and buildings, are particularly important for model success, whereas features such as greenery and sky contribute less to geolocalisation performance. These findings highlight the need to explicitly consider temporal diversity and feature composition when developing and evaluating geospatial AI methods, especially for applications that rely on visual data collected across varying time periods. This work establishes a foundation for the systematic evaluation of temporal bias in benchmark datasets, highlighting its impact on the performance and reliability of deep learning models in geospatial tasks. Future work should focus on building temporally diverse benchmarks and developing models that are robust to temporal and seasonal variation, thus enabling more reliable performance in real-world, dynamic urban environments.

Chapter 6

Summary

The rapid proliferation of smart devices, advances in hardware, and platforms that enable the sharing and propagation of information have given rise to new forms of urban data. These include, for example, large-scale street-level and remote sensing imagery, as well as opportunistically collected wireless measurements from everyday devices. Such data can be updated far more frequently than traditional survey-based datasets, making them valuable for capturing rapid urban change and supporting responsive decision-making. Building on this context, this thesis examines how emerging data can be leveraged to develop and enhance urban positioning and navigation-related services, with a focus on evaluating their suitability, efficiency, and reliability in complex urban environments. (Chapter 1). Specifically, the work presented in this thesis encompasses the development of innovative methodologies for 3D mapping, supporting humans in contemporary cities, and the creation of evaluation frameworks to address the frequent needs for positioning and navigation in complex urban settings. Key contributions include an innovative and economical approach for 3D mapping that utilises ubiquitous cellular signals (Chapter 3); a second strand of work that focuses on supporting humans in cities (Chapter 4), comprising a benchmark assessment of machine learning and deep learning models leveraging high-resolution UWB CIRs for human activity recognition (Section 4.1) together with a systematic identification and evaluation of factors influencing pedestrian navigation (Section 4.2); and finally, an analysis of the impact of temporal bias on vision-based geolocalisation (Chapter 5). Each chapter elaborates on the research questions, data collection strategies, and analytical approaches, while also providing insights and future directions. Taken together, this concluding chapter synthesises these findings and positions them within a broader research context, outlining a framework for advancing reliable and human-centred urban positioning and navigation.

6.1 Discussion

This thesis advances urban analytics by investigating how emerging forms of urban data can be leveraged to support human-centric urban analytics. Rather than relying on a single data modality or application, the three research strands collectively illustrate how heterogeneous, non-traditional urban data, including wireless signals, geo-tagged online questionnaire and large-scale visual datasets, can be transformed into actionable urban information. Together, these strands address key research gaps related to data availability, scalability, human-centred understanding, and bias awareness in contemporary urban analytics, as outlined in Section 1.2. This section synthesises their combined contributions, clarifies how they advance the use of new forms of urban data, and reflects on future research directions.

Outcome of Research Objective 1. *Generated accurate 3D maps using free, scalable, and globally available wireless signals as opportunistic data sources for large-scale, cost-effective 3D mapping.*

This research objective is motivated by exploring inexpensive yet effective alternative data forms for 3D mapping, without the need for dedicated equipment. 3D mapping is crucial in a wide range of urban applications, especially those related to positioning, such as GNSS improvement through shadow matching, emerging drone navigation, and autonomous driving (Biljecki et al. 2015; Groves 2011a; Wang et al. 2013a).

In this research, wireless signals are leveraged due to their prevalence in urban environments. Chapter 3 proposed a novel method for mapping 3D city models, specifically targeting the estimation of building heights, driven by the widespread cellular mobile signals as opportunistic signals. The proposed method for 3D mapping relies on ubiquitous cellular mobile infrastructures while not requiring dedicated equipment other than consumer-level mobile phones. Therefore, this work potentially reduces significantly the cost and hardware requirements for implementing a large-scale 3D mapping application. This is especially beneficial for countries in the Global South, whose resources are often more limited compared to those of countries in the Global North. RSS is the selected indicator, captured by the custom mobile app, BitToBrick, for deriving the shadowing effects brought by target buildings. The data and code for the proposed method have been made publicly accessible to enhance its applicability at scale and benefit larger research communities. In this proposed method, no prior knowledge of wireless propagation is required; in other words, the shadowing effect is purely inferred from its absolute RSS, and the building height is derived from the clustering decision boundary among distinct RSS clusters. Therefore, the computational complexity is also considerably reduced. The result demonstrates high accuracy comparable to that provided by the national mapping agency, as well as robustness in measuring building heights with limited data points.

The proposed approach for 3D mapping via cellular mobile signals, though proven to be effective in mapping tasks, is still limited by the geometric relationship among cellular antennas and target buildings. Its performance evaluation is still constrained due to the hardware limitations while conducting real-world experiments. Also, it is expected to be less effective in building height estimation under some extreme urban canyon conditions, where the detection of shadowed and unshadowed decision boundaries resulting from the target building becomes challenging. Future work will focus on the fusion of multiple sources jointly for 3D mapping generation, providing mutually beneficial and complementary application scenarios.

Therefore, the future work effort is double-folded: First, the essential effort for future 3D mapping work will be the continuous extension of my proposed mobile app into a crowdsourcing platform, which will hopefully allow users to contribute on a larger scale. This could make the approach more robust by leveraging heterogeneous data generated through a wide spectrum of hardware and mobile service providers. Additionally, future work will focus on developing a methodology for data fusion in height estimation. Specifically, the current mobile app enables the collection of information from both low-altitude cellular signals and signals from high-elevation GNSS. Joint consideration of both sources would hopefully enlarge the application scenario of the approach. Beyond the source of wireless signals, SVIs can provide more contextual information for the nearby environment, which can help refine and consolidate the results derived from wireless signals alone.

Outcome of Research Objective 2. Supported humans in cities through human-centred urban research that combines wireless sensing for activity recognition with the systematic identification of environmental, situational, and personal factors behind pedestrian disorientation.

To better support humans in urban environments, knowledge of the physical environment, gained through efforts such as 3D mapping for refined positioning, is not sufficient. Equally important is developing a human-centred understanding of how people interact with and navigate through the city. Addressing this gap, the second research objective focuses on supporting people in their daily urban lives by examining human factors directly. Here, the focus shifts from large-scale objects such as buildings towards human physical activity recognition, explored through opportunistic wireless signals. At the same time, despite the increasing availability of digital tools and navigation technologies, urban residents can still become disoriented in complex environments, pointing to additional factors that influence their wayfinding process. In response, this strand consists of two complementary streams of work: utilising wireless sensing for human activity recognition and examining the environmental, situational, and personal factors that contribute to pedestrian disorientation.

Human activities within urban environments are either researched independently or overlooked in the area of urban studies (Jiang et al. 2012). Therefore, in Section 4.1, an initial effort is made in investigating the capability of mapping human physical activities through high-resolution CIR extracted from UWB communications. The data collection was conducted with EVK1000 boards, which were designed for precise distance measurement, while they were used in the experiment to provide CIRs that describe the wireless channel variation resulting from human activities. Each of these CIRs describes a momentaneous channel condition; therefore, concatenation of these CIR snapshots recorded throughout a given activity chronologically would reveal the channel temporal pattern corresponding to it. A large bundle of deep learning and machine learning models were trained and evaluated on the dataset. As a result, recurrent models such as RNN, including its variants like GRU and LSTM-based models, produce the highest classification accuracy for human activities due to their higher capability in capturing temporal dependencies, which suits well with the proposed CIR concatenation methodologies. This benchmark research examined HAR tasks using UWB CIR under various scenarios, involving up to three environmental setups, four volunteers, eleven learning models, and a set of preprocessing strategies, with exhaustive combinations. The outcome of this work has convinced us of the capability and suitability of performing HAR tasks accurately using off-the-shelf wireless sensors, demonstrating their potential. Compared to other sensing modalities, such as vision-based HAR, using wireless signals can significantly reduce the cost and is independent of lighting conditions. Additionally, in the case where signal transmission is blocked by obstacles, i.e., a NLOS scenario, there is no noticeable drop in accuracy, as observed in the occluded scenario of the vision-based HAR experiment, further strengthening its robustness and suitability for real-world scenarios.

Meanwhile, urban residents are increasingly supported by location-based services built upon positioning and navigation technologies, such as mobile phones with integrated navigation tools, 3D city models illustrating surrounding contextual environments, and vision-based localisation. However, despite these technological advances, users often face difficulties when navigating complex urban environments, experiencing disorientation in intricate settings, and ultimately getting lost within so-called urban jungles. In Section 4.2, the factors leading to a getting lost event were explored, covering environmental, situational, and personal factors. The dataset used for this research was collected through an online survey, where volunteers reported their experiences of getting lost while navigating the Greater London Area. The volunteers detailed their getting lost locations, as well as contextual information such as their familiarity with the location, lighting conditions, and so on. After data cleansing, 64 data points remain and are included in this work. While the sample size limits statistical generalisation, it provides structured, place-specific accounts that enable an initial comparison between expert-weighted and data-driven interpretations of disorientation factors. Meanwhile, a group of geospatial experts gathered to identify a list of factors they believe may cause pedestrians to become lost, covering a wide range of personal, situational, and environmental factors. This process was conducted independently of the survey. The identified factors are weighted according to their importance, as determined by the experts' cognition, using the AHP method. Based on the provided information, responses for each data point are mapped to the identified factors by experts and scaled from 0 to 1. These data points are split into two evenly sized groups: one is used for fitting regression models to derive a distinct weighting scheme from the AHP weights (data-driven), while the other serves as the control group for testing the outcomes of the weighting schemes (for both expert-led and data-driven weights). As each reported event in the survey corresponds to a real-world scenario of getting lost, it is expected that the weighted sum of the weights and factor values will be close to one. Comparing the getting lost scores computed by both ”expert-led” and ”data-driven” methods reveals distinct yet complementary interpretations of the importance of the factors involved in getting lost. Interestingly, the expert-led weighting scheme emphasises the importance of factors related to humans, such as familiarity and self-orientation skills, as well as external support, represented by access to reliable maps. Data-driven methods attribute more weight to both human-related factors and route complexity.

In summary, this research has developed a framework for identifying the multifaceted factors that may cause pedestrians to become lost in urban environments. Relying on survey data describing disorientation cases in real-world scenarios, the work adopted both data-driven and expert-led methodologies to derive weighting systems for a list of factors contributing to disorientation, shedding light on the reasoning process behind a getting-lost event. The disparity between various weighting systems underscores the need to consider both human and environmental factors when aiming to reduce instances of pedestrians getting lost during urban navigation and in creating more inclusive and accessible urban environments.

One limitation of the “getting lost” research is the lack of itinerary information in the surveyed data, where respondents only reported their locations where they got lost. Therefore, only simulated routes around the location where respondents got lost within a buffer were leveraged for estimating itinerary-related challenges throughout their wayfinding process. Additionally, several factors that contributed to getting lost were evaluated subjectively, including visibility, pedestrian and vehicle flow. Temporal information plays a critical role in estimating such factors, while the survey data have not recorded it. To address these limitations, future work will encompass a refined design of the survey, both methodologically and technologically.

Methodologically, future data collection on disorientation aims to include itinerary information by recording both the start and end locations of a wayfinding session, as well as the route that a respondent follows. As such, each record could provide fine-grained contextual information about the route and the behavioural implications of respondents (e.g., unusual long periods of wandering around a junction, or a rapid correction of their walking direction are likely to indicate potential disorientation). Future work will also involve efforts in leveraging technological tools beyond traditional surveying. In response, an online platform is being developed that allows users to record and report their path-finding process and the challenges they encounter in a timely manner, enabling both objective and subjective evaluation of factors leading to lost events.

<u>Outcome of Research Objective 3.</u> Identified inherent biases in urban data and evaluated their impact on deep learning-based models, taking cross-view geolocalisation as a case study. The objective is achieved through investigating how temporal biases present in benchmark datasets can affect the performance of pre-trained models, ultimately enhancing their robustness and fairness.

Technological advances and the widespread use of smart devices have enabled the rapid growth of new urban data sources. Among them, images are especially prominent, generated at scale by commercial platforms, crowdsourcing, and social media, and widely used to build benchmark datasets for various urban applications. This availability has driven visual-based positioning to become a major research direction, supported by progress in AI and deep learning. This thesis provides empirical evidence that temporal bias, even in spatially balanced benchmark datasets, can significantly degrade model performance and alter feature contributions in vision-based geolocalisation.

While GNSS remains the backbone of urban positioning, the findings of this thesis demonstrate that its limitations in dense urban environments cannot be addressed by a single technological solution. Recently, vision-based geolocalisation has attracted significant research attention, and the reason is twofold in principle: first, the proliferation of visual data, from both commercial companies and social media, considerably enhances the volume of visual data, covering a wide range of geolocations. Second, advances in deep learning techniques, combined with GPU-powered computing acceleration, offer the opportunity to process a large volume of visual data, which was previously extremely challenging or even impossible. Over the past decade, visual-based geolocalisation has progressed rapidly, supported by benchmark datasets and deep architectures that learn cross-view representations for retrieval-based localisation. Researchers have proposed numerous deep learning models that rely on computer vision techniques to derive the geo-location of a query image from the visual cues it contains. One main stream of this work is cross-view geolocalisation, aiming to retrieve a corresponding geo-referenced aerial image for a given query ground-view image. A series of benchmark datasets has been proposed by the research community, using images obtained from commercial street and aerial view image providers, as well as from publicly available and geo-referenced social media photos. These datasets are intentionally generated, covering a wide range of geographical areas and providing spatial diversity. CVGL tasks, which commonly rely on computer vision techniques, generally utilise these datasets to train deep learning models of proposed architectures and have demonstrated accurate retrieval results on these benchmarks.

Even though the datasets are characterised by rich spatial diversity, a large portion of them provide only one sample image for a given location. The potential risk behind this observation suggests that the datasets used for training deep learning models commonly overemphasise spatial diversity, while overlooking temporal or seasonal variation, which may affect geolocalisation accuracy. The work detailed in Chapter 5 examines how temporal diversity at a given location affects the performance of CVGL models. To provide insights for this research objective, the CVUSA dataset, widely used for CVGL tasks, was selected as the baseline dataset. Then, a custom dataset was constructed based on the validation dataset of CVUSA. More specifically, the Google Street View Static API was queried to retrieve all available historical Street View images for each location contained in the CVUSA validation set, constructing a spatially aligned and temporally diverse custom dataset. Additionally, to minimise disturbance other than temporal variation, the exact resolution and aspect ratio were maintained as in the panoramas in the CVUSA datasets for the custom dataset. Next, two state-of-the-art pre-trained deep learning models dedicated to CVGL tasks were selected, i.e., Spatial-Aware Feature Aggregation (SAFA) and TransGeo, for the purpose of evaluating their performance on the custom dataset. As a result, both models demonstrate remarkable retrieval accuracy on the baseline dataset. To further examine the temporal variation in the set of SVIs at each location, semantic segmentation was first applied to the SVIs from the custom dataset. The pre-defined semantic categories are regrouped into new groups to simplify processing and constructing a high-level description of the elements within each panorama. Accordingly, the pixel proportion of each semantic group for each image in the custom datasets, along with its variation compared to the image's counterpart from the CVUSA validation dataset, is computed. It is observed that significant variation exists in some of the semantic groups. To understand how each semantic group contributes to retrieval accuracy, the SHAP tool for model explainability was used to estimate the influence of semantic group proportion on the outcome of the two models. The result shows that semantic groups regarding urban layouts, such as paved surface, buildings and urban furniture (e.g., traffic light, signage), contribute positively to the retrieval accuracy; while features associated with natural landscapes, such as greenery or the share of sky in a panorama, contribute negatively to the models. It is worth mentioning that these semantic groups are not independent. For example, foliage density is likely to decrease from summer to winter for a given location in the U.S. located in the Northern Hemisphere. On the other hand, for the same period of time, the proportions of other semantic groups (such as sky or buildings) at the exact location are prone to increase accordingly, due to less occlusion by the greenery.

In summary, the results from Chapter 5 indicate that variation in visual information over time greatly influences the performance of deep learning models, a factor often overlooked when building benchmark datasets for training these models. Within the context of cross-view geolocalisation research, the importance of considering both temporal diversity in visual data and feature compositions that enhance the performance of deep learning models is highlighted. More generally, to develop both fair and robust deep learning models for GeoAI applications in urban studies, attention should be paid not only to spatial diversity but also to temporal richness, ensuring that the outcomes are more representative and inclusive.

Considerations of control variates have been taken into account while evaluating the impact of temporal bias in successful CVGL tasks. These include restricting historical SVIs to be collected at the exact location as their original matching sample, maintaining an identical image aspect ratio, and ensuring a similar vertical semantic ratio. However, several subtle yet important variants remain to be explored. First, the direction alignment between the original and custom SVIs requires further investigation. Although panoramic images were used for evaluation, the impact of the central directionality on the retrieval performance remains to be examined. Future work will investigate how directional misalignment affects visual feature correspondence for improving model robustness under real-world conditions. Additionally, the retrieval accuracy was evaluated against the original satellite images, without introducing temporal diversity on the satellite side. While this reflects real-world scenarios, where a random query SVI may not be temporally aligned with its spatially matching satellite image, future work will investigate if introducing temporal diversity into satellite imagery could further improve the CVGL model's generalisation and temporal consistency.

Moving beyond the specific CVGL application, future research could focus on a broader examination of data and AI bias in geospatial applications. This work has built a foundational understanding of how temporal bias in visual data can influence model performance. Built upon this insight, subsequent research could focus on other forms of biases involved in geospatial applications, such as geographical, social and cultural biases, to enhance the reliability, robustness and fairness for a wide spectrum of spatial AI.

Viewed together, the three research strands point towards a broader research direction for urban analytics that leverages emerging urban data in an integrated, human-centred, and bias-aware manner. Future research will increasingly require the fusion of heterogeneous data modalities to jointly model urban dynamics. At the same time, greater attention must be paid to data quality, representativeness, and bias, particularly as urban analytics increasingly relies on large-scale, opportunistically collected datasets. Advancing this field therefore requires not only technical innovation, but also critical reflection on how new forms of urban data shape knowledge production and decision-making in cities.

6.2 Conclusion

Urban environments are expanding rapidly with the global trend of urbanisation, becoming increasingly diverse, dynamic, and complex ecosystems which residents live in and interact with. This complexity places increasing demands on the reliability, adaptability, and fairness of data-driven urban services that residents depend on in their daily lives. Understanding these environments poses long-standing challenges and introduces new ones as urban systems continue to evolve. Over recent decades, the proliferation of smart devices, data-sharing platforms, and advances in computational hardware have not only enabled the continuous generation of new forms of urban data but also facilitated their integration with AI and deep learning, offering substantial opportunities to develop cost-effective and scalable applications. Within this context, the work presented in this thesis advances positioning- and navigation-related urban applications through three main strands, collectively demonstrating how emerging urban data can support reliable services, but also under what conditions their limitations and biases become critical: first, the development of novel methods for 3D mapping by leveraging wireless signal sensing; second, a human-centred investigation that employs wireless signals for activity recognition and systematically evaluates pedestrian disorientation from multiple perspectives; and third, an assessment of the risks posed by biases in urban datasets, with a particular focus on temporal bias in cross-view geolocalisation.

This thesis makes contributions in three areas: advancing 3D mapping with cellular signals, supporting humans in cities through HAR and wayfinding studies, and uncovering temporal bias in urban visual data to enhance positioning and navigation. Specifically, the key contributions are detailed as follows:

- **Opportunistic wireless sensing for 3D mapping**: with the proliferation of user devices, wireless signals have become pervasive in urban environments, offering substantial potential beyond their primary communication purposes. Their potential in 3D mapping is investigated by proposing a novel method for estimating building height and facilitating 3D map creation using opportunistically collected cellular mobile signals. These experiments achieved high accuracy, demonstrating that wireless signals can serve as an alternative and complementary data source in urban studies.

- **Supported humans in cities: physical activities and disorientation**: the thesis explored the fine-grained sensing capability of wireless signals through a benchmark study on HAR. Furthermore, the thesis examined the multifaceted factors contributing to pedestrians getting lost in urban environments, moving beyond single-aspect analyses to include environmental, individual, and contextual dimensions. By quantifying the contribution of each factor, the work provides a more holistic understanding of urban disorientation. These insights inform the design of more inclusive and navigable urban spaces, contributing to practical strategies for mitigating the risk of disorientation.
- **Revealed temporal bias in cross-view geo-localisation**: visual-data-based urban geo-localisation often depends on benchmark datasets covering large spatial extents, yet little is known about their inherent biases. A framework was developed to assess temporal bias in CVGL tasks by collecting historical street view imagery and evaluating state-of-the-art deep learning models under temporal variation. The results show that visual changes over time can significantly degrade geo-localisation performance. Further analysis revealed that shifts in semantic content play a critical role in these performance drops, highlighting the importance of accounting for both spatial and temporal diversity when designing and evaluating visual localisation systems.

In conclusion, this thesis demonstrates the potential of emerging forms of urban data to advance positioning- and navigation-based applications, aiming to improve the quality of location-based services and foster timely and responsive urban environments. It contributes an innovative and economical approach to 3D mapping by leveraging opportunistic wireless signals, a capability essential for modern urban positioning. It further advances a human-centred perspective by recognising physical activities through wireless sensing and systematically evaluating the environmental, situational, and personal factors that influence pedestrian wayfinding and disorientation, thereby underscoring the need to integrate human and environmental considerations in inclusive urban design. Finally, it reveals the risks posed by temporal bias in vision-based geo-localisation and calls for future research that incorporates both spatial and temporal diversity into system design. Collectively, these contributions expand the methodological toolkit for leveraging new data forms for urban positioning and navigation-related applications, offering both practical solutions and theoretical insights into how cities can be more effectively understood, navigated, and designed.

Bibliography

Adachi, K., P. Lago, T. Okita and S. Inoue (2021). ‘Improvement of Human Action Recognition Using 3D Pose Estimation’. In: *Activity and Behavior Computing*, pp. 21–37. DOI: 10.1007/978-981-15-8944-7_2.

Adjrad, M. and P. D. Groves (2016). ‘Intelligent Urban Positioning Using Shadow Matching and GNSS Ranging Aided by 3D Mapping’. In: *Proceedings of the 29th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS+ 2016)*, pp. 534–553. DOI: 10.33012/2016.14845.

Agarwal, M., M. Sun, C. Kamath, A. Muslim, P. Sarker, J. Paul, H. Yee, M. Sieniek, K. Jablonski, Y. Mayer et al. (2024). ‘General Geospatial Inference with a Population Dynamics Foundation Model’. Preprint. arXiv: 2411.07207.

Ahuja, K., Y. Jiang, M. Goel and C. Harrison (2021). ‘Vid2doppler: Synthesizing Doppler Radar Data from Videos for Training Privacy-Preserving Activity Recognition’. In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–10. DOI: 10.1145/3411764.3445138.

Aichner, T., M. Grünfelder, O. Maurer and D. Jegeni (2021). ‘Twenty-Five Years of Social Media: A Review of Social Media Applications and Definitions from 1994 to 2019’. In: *Cyberpsychology, Behavior, and Social Networking* 24.4, pp. 215–222. DOI: 10.1089/cyber.2020.0134.

Ajayi, O. G. and A. Ojima (2022). ‘Performance Evaluation of Selected Cloud Occlusion Removal Algorithms on Remote Sensing Imagery’. In: *Remote Sensing Applications: Society and Environment* 25, p. 100700. DOI: 10.1016/j.rsase.2022.100700.

Al Tareq, A., M. J. Rana, M. R. Mostofa and M. S. Rahman (2024). ‘Impact of IoT and Embedded System on Semiconductor Industry a Case Study’. In: *Control Systems and Optimization Letters* 2.2, pp. 211–216. DOI: 10.59247/csol.v2i2.111.

Alahi, M. E. E., A. Sukkuea, F. W. Tina, A. Nag, W. Kurdthongmee, K. Suwannarat and S. C. Mukhopadhyay (2023). ‘Integration of IoT-enabled Technologies and Artificial Intelligence (AI) for Smart City Scenario: Recent Advancements and Future Trends’. In: *Sensors* 23.11, p. 5206. DOI: 10.3390/s23115206.

Alonso, J. A. and M. T. Lamata (2006). 'Consistency in the Analytic Hierarchy Process: A New Approach'. In: *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 14.04, pp. 445–459. DOI: 10.1142/S0218488506004114.

Alsafery, W., O. Rana and C. Perera (2023). 'Sensing within Smart Buildings: A Survey'. In: *ACM Computing Surveys* 55 (13s), pp. 1–35. DOI: 10.1145/3596600.

Alzubaidi, L., J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie and L. Farhan (2021). 'Review of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, Future Directions'. In: *Journal of Big Data* 8.1, p. 53. DOI: 10.1186/s40537-021-00444-8.

Amiruzzaman, M., A. Curtis, Y. Zhao, S. Jamonnak and X. Ye (2021). 'Classifying Crime Places by Neighborhood Visual Appearance and Police Geonarratives: A Machine Learning Approach'. In: *Journal of Computational Social Science* 4.2, pp. 813–837. DOI: 10.1007/s42001-021-00107-x.

Anciaes, P. R. and P. Jones (2016). 'Effectiveness of Changes in Street Layout and Design for Reducing Barriers to Walking'. In: *Transportation Research Record* 2586.1, pp. 39–47. DOI: 10.3141/2586-05.

Ang, L.-M. and K. P. Seng (2016). 'Big Sensor Data Applications in Urban Environments'. In: *Big Data Research* 4, pp. 1–12. DOI: 10.1016/j.bdr.2015.12.003.

Angel, S. (2023). 'Urban Expansion: Theory, Evidence and Practice'. In: *Buildings & Cities* 4.1, pp. 124–138. DOI: 10.5334/bc.348.

Apple Maps (2025). *Apple Maps*. URL: https://maps.apple.com/ (visited on 31/12/2025).

Arandjelovic, R., P. Gronat, A. Torii, T. Pajdla and J. Sivic (2016). 'NetVLAD: CNN Architecture for Weakly Supervised Place Recognition'. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5297–5307. DOI: 10.1109/CVPR.2016.572.

Arribas-Bel, D. (2014). 'Accidental, Open and Everywhere: Emerging Data Sources for the Understanding of Cities'. In: *Applied Geography* 49, pp. 45–53. DOI: 10.1016/j.apgeog.2013.09.012.

Bagheri, H., M. Schmitt, P. d'Angelo and X. X. Zhu (2018). 'A Framework for SAR-optical Stereogrammetry over Urban Areas'. In: *ISPRS journal of photogrammetry and remote sensing* 146, pp. 389–408. DOI: 10.1016/j.isprsjprs.2018.10.003.

Bai, X., W. Wen and L.-T. Hsu (2022). 'Time-Correlated Window-Carrier-Phase-Aided GNSS Positioning Using Factor Graph Optimization for Urban Positioning'. In: *IEEE Transactions on Aerospace and Electronic Systems* 58.4, pp. 3370–3384. DOI: 10.1109/TAES.2022.3149730.

Bailenson, J. N., M. S. Shum and D. H. Uttal (1998). 'Road Climbing: Principles Governing Asymmetric Route Choices on Maps'. In: *Journal of Environmental Psychology* 18.3, pp. 251–264. ISSN: 0272-4944. DOI: 10.1006/jevp.1998.0095.

Balado, J., R. Garozzo, L. Winiwarter and S. Tilon (2025). 'A Systematic Literature Review of Low-Cost 3D Mapping Solutions'. In: *Information Fusion* 114, p. 102656. DOI: 10.1016/j.inffus.2024.102656.

Baldwin, D. (2003). 'Wayfinding Technology: A Road Map to the Future'. In: *Journal of Visual Impairment & Blindness* 97.10, pp. 612–620. DOI: 10.1177/0145482X0309701006.

Baltsavias, E. P. (1999). 'Airborne Laser Scanning: Basic Relations and Formulas'. In: *ISPRS Journal of photogrammetry and remote sensing* 54.2–3, pp. 199–214. DOI: 10.1016/S0924-2716(99)00015-5.

Barranquero, M., A. Olmedo, J. Gómez, A. Tayebi, C. J. Hellín and F. Saez de Adana (2023). 'Automatic 3D Building Reconstruction from OpenStreetMap and LiDAR Using Convolutional Neural Networks'. In: *Sensors* 23.5, p. 2444. DOI: 10.3390/s23052444.

Basiri, A., T. Lines and M. Fidel Pereira (2023). 'Scalable 3D Mapping of Cities Using Computer Vision and Signals of Opportunity'. In: *International Journal of Geographical Information Science* 37.7, pp. 1470–1495. DOI: 10.1080/13658816.2023.2191674.

Batty, M. (2013). *The New Science of Cities.* Cambridge, Massachusetts, United States: MIT press.

Batty, M. (2016). 'Urban Resilience: How Cities Need to Adapt to Unanticipated and Sudden Change'. In: *Perspectives on Complex Global Challenges: Education, Energy, Healthcare, Security and Resilience: Education, Energy, Healthcare, Security and Resilience*, pp. 169–171. DOI: 10.1002/9781118984123.ch25.

Batty, M. (2018). *Inventing Future Cities.* Cambridge, Massachusetts, United States: MIT press.

Batty, M. (2020). 'Defining Smart Cities: High and Low Frequency Cities, Big Data and Urban Theory'. In: *The Routledge Companion to Smart Cities.* Routledge, pp. 51–60.

Batty, S., S. Davoudi and A. Layard (2012). *Planning for a Sustainable Future.* London, UK: Routledge.

Bayat, A. and P. Kawalek (2023). 'Digitization and Urban Governance: The City as a Reflection of Its Data Infrastructure'. In: *International Review of Administrative Sciences* 89.1, pp. 21–38. DOI: 10.1177/00208523211033205.

Beddiar, D. R., B. Nini, M. Sabokrou and A. Hadid (2020). 'Vision-Based Human Activity Recognition: A Survey'. In: *Multimedia Tools and Applications* 79, pp. 30509–30555. DOI: 10.1007/s11042-020-09004-3.

Belhadi, A., Y. Djenouri, G. Srivastava, D. Djenouri, J. C.-W. Lin and G. Fortino (2021). 'Deep Learning for Pedestrian Collective Behavior Analysis in Smart Cities: A Model of Group Trajectory Outlier Detection'. In: *Information Fusion* 65, pp. 13–20. DOI: 10.1016/j.inffus.2020.08.003.

Ben-Afia, A., L. Deambrogio, D. Salós, A.-C. Escher, C. Macabiau, L. Soulier and V. Gay-Bellile (2014). 'Review and Classification of Vision-Based Localisation Techniques in Unknown Environments'. In: *IET Radar, Sonar & Navigation* 8.9, pp. 1059–1072. DOI: 10.1049/iet-rsn.2013.0389.

Benedetti, A. C., C. Costantino, R. Gulli and G. Predari (2022). 'The Process of Digitalization of the Urban Environment for the Development of Sustainable and Circular Cities: A Case Study of Bologna, Italy'. In: *Sustainability* 14.21, p. 13740. DOI: 10.3390/su142113740.

Bernard, J., E. Bocher, E. Le Saux Wiederhold, F. Leconte and V. Masson (2022). 'Estimation of Missing Building Height in OpenStreetMap Data: A French Case Study Using GeoClimate 0.0. 1'. In: *Geoscientific Model Development* 15.19, pp. 7505–7532. DOI: 10.5194/gmd-15-7505-2022.

Berton, G., C. Masone and B. Caputo (2022). 'Rethinking Visual Geo-Localization for Large-Scale Applications'. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4878–4888. DOI: 10.1109/cvpr52688.2022.00483.

Bibbò, L., R. Carotenuto and F. Della Corte (2022). 'An Overview of Indoor Localization System for Human Activity Recognition (HAR) in Healthcare'. In: *Sensors* 22.21, p. 8119. DOI: 10.3390/s22218119.

Bibri, S. E. (2019). 'On the Sustainability of Smart and Smarter Cities in the Era of Big Data: An Interdisciplinary and Transdisciplinary Literature Review'. In: *Journal of Big Data* 6.1, p. 25. DOI: 10.1186/s40537-019-0182-7.

Biljecki, F. (2020). 'Exploration of Open Data in Southeast Asia to Generate 3D Building Models'. In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 6, pp. 37–44. DOI: 10.5194/isprs-annals-VI-4-W1-2020-37-2020.

Biljecki, F. and K. Ito (2021). 'Street View Imagery in Urban Analytics and GIS: A Review'. In: *Landscape and Urban Planning* 215, p. 104217. DOI: 10.1016/j.landurbplan.2021.104217.

Biljecki, F., J. Stoter, H. Ledoux, S. Zlatanova and A. Çöltekin (2015). 'Applications of 3D City Models: State of the Art Review'. In: *ISPRS International Journal of Geo-Information* 4.4, pp. 2842–2889. DOI: 10.3390/ijgi4042842.

Bing Maps (2025). *Bing Maps*. URL: https://www.bing.com/maps (visited on 31/12/2025).

Bitelli, G., V. A. Girelli, A. Lambertini et al. (2018). 'Integrated Use of Remote Sensed Data and Numerical Cartography for the Generation of 3D City Models'. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 42.2, pp. 97–102. DOI: 10.5194/isprs-archives-xlii-2-97-2018.

Blaschke, T., G. J. Hay, Q. Weng and B. Resch (2011). ‘Collective Sensing: Integrating Geospatial Technologies to Understand Urban Systems—an Overview’. In: *Remote Sensing* 3.8, pp. 1743–1776. DOI: 10.3390/rs3081743.

Bocus, M. J., K. Chetty and R. J. Piechocki (2021). ‘UWB and WiFi Systems as Passive Opportunistic Activity Sensing Radars’. In: *2021 IEEE Radar Conference (RadarConf21)*. IEEE, pp. 1–6. DOI: 10.1109/RadarConf2147009.2021.9455175.

Boeing, G. (2017). ‘OSMnx: New Methods for Acquiring, Constructing, Analyzing, and Visualizing Complex Street Networks’. In: *Computers, Environment and Urban Systems* 65, pp. 126–139. ISSN: 0198-9715. DOI: 10.1016/j.compenvurbsys.2017.05.004.

Boeing, G. (2019). ‘Urban Spatial Order: Street Network Orientation, Configuration, and Entropy’. In: *Applied Network Science* 4.1, pp. 1–19. ISSN: 2364-8228. DOI: 10.1007/s41109-019-0189-1.

Bok, Y., D.-G. Choi and I. S. Kweon (2014). ‘Sensor Fusion of Cameras and a Laser for City-Scale 3D Reconstruction’. In: *Sensors* 14.11, pp. 20882–20909. DOI: 10.3390/s141120882.

Bouchabou, D., S. M. Nguyen, C. Lohr, B. LeDuc and I. Kanellos (2021). ‘A Survey of Human Activity Recognition in Smart Homes Based on IoT Sensors Algorithms: Taxonomies, Challenges, and Opportunities with Deep Learning’. In: *Sensors* 21.18, p. 6037. DOI: 10.3390/s21186037.

Boyle, D. E., D. C. Yates and E. M. Yeatman (2013). ‘Urban Sensor Data Streams: London 2013’. In: *IEEE Internet Computing* 17.6, pp. 12–20. DOI: 10.1109/MIC.2013.85.

Bradburn, N. M., L. J. Rips and S. K. Shevell (1987). ‘Answering Autobiographical Questions: The Impact of Memory and Inference on Surveys’. In: *Science* 236.4798, pp. 157–161. DOI: 10.1126/science.3563494.

Bradley, V. C., S. Kuriwaki, M. Isakov, D. Sejdinovic, X.-L. Meng and S. Flaxman (2021). ‘Unrepresentative Big Surveys Significantly Overestimated US Vaccine Uptake’. In: *Nature* 600.7890, pp. 695–700. DOI: 10.1038/s41586-021-04198-4.

Brenner, N. and C. Schmid (2015). ‘Towards a New Epistemology of the Urban?’ In: *City* 19.2–3, pp. 151–182. DOI: 10.1080/13604813.2015.1014712.

Bromley, J., I. Guyon, Y. LeCun, E. Säckinger and R. Shah (1993). ‘Signature Verification Using a “Siamese” Time Delay Neural Network’. In: *Advances in Neural Information Processing Systems* 6.

Brunner, D., G. Lemoine, L. Bruzzone and H. Greidanus (2009). ‘Building Height Retrieval from VHR SAR Imagery Based on an Iterative Simulation and Matching Technique’. In: *IEEE Transactions on Geoscience and Remote Sensing* 48.3, pp. 1487–1504. DOI: 10.1109/TGRS.2009.2031910.

Budiyono, A. (2012). 'Principles of GNSS, Inertial, and Multi-Sensor Integrated Navigation Systems'. In: *Industrial Robot: An International Journal* 39.3. DOI: 10.1108/ir.2012.04939caa.011.

Burns, P. C. (1998). 'Wayfinding Errors While Driving'. In: *Journal of Environmental Psychology* 18.2, pp. 209–217. ISSN: 0272-4944. DOI: 10.1006/jevp.1998.0077.

Cai, S., Y. Guo, S. Khan, J. Hu and G. Wen (2019). 'Ground-to-Aerial Image Geo-Localization with a Hard Exemplar Reweighting Triplet Loss'. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 8391–8400. DOI: 10.1109/ICCV.2019.00848.

Calabrese, F., L. Ferrari and V. D. Blondel (2014). 'Urban Sensing Using Mobile Phone Network Data: A Survey of Research'. In: *ACM computing Surveys (CSUR)* 47.2, pp. 1–20. DOI: 10.1145/2655691.

Cândido, R. L., M. Steinmetz-Wood, P. Morency and Y. Kestens (2018). 'Reassessing Urban Health Interventions: Back to the Future with Google Street View Time Machine'. In: *American Journal of Preventive Medicine* 55.5, pp. 662–669. DOI: 10.1016/j.amepre.2018.04.047.

Cao, Z., T. Simon, S.-E. Wei and Y. Sheikh (2017). 'Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields'. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7291–7299. DOI: 10.1109/cvpr.2017.143.

Carminati, M., G. R. Sinha, S. Mohdiwale and S. L. Ullo (2021). 'Miniaturized Pervasive Sensors for Indoor Health Monitoring in Smart Cities'. In: *Smart Cities* 4.1, pp. 146–155. DOI: 10.3390/smartcities4010008.

Castaldo, F., A. Zamir, R. Angst, F. Palmieri and S. Savarese (2015). 'Semantic Cross-View Matching'. In: *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 9–17. DOI: 10.1109/iccvw.2015.137.

Cellmapper (2025). *Cellmapper Homepage.* URL: https://www.cellmapper.net (visited on 31/12/2025).

Chai, J., H. Zeng, A. Li and E. W. Ngai (2021). 'Deep Learning in Computer Vision: A Critical Review of Emerging Techniques and Application Scenarios'. In: *Machine Learning with Applications* 6, p. 100134. DOI: 10.1016/j.mlwa.2021.100134.

Chambers, J. and J. Evans (2020). 'Informal Urbanism and the Internet of Things: Reliability, Trust and the Reconfiguration of Infrastructure'. In: *Urban Studies* 57.14, pp. 2918–2935. DOI: 10.1177/0042098019890798.

Chan, E., O. Baumann, M. A. Bellgrove and J. B. Mattingley (2012). 'From Objects to Landmarks: The Function of Visual Location Information in Spatial Navigation'. In: *Frontiers in Psychology* 3, pp. 304–304. ISSN: 1664-1078. DOI: 10.3389/fpsyg.2012.00304.

Chan, K., M. Vasardani and S. Winter (2015). ‘Getting Lost in Cities: Spatial Patterns of Phonetically Confusing Street Names’. In: *Transactions in GIS* 19.4, pp. 535–562. DOI: 10.1111/tgis.12093.

Chen, D. and G. X. Gao (2019). ‘Probabilistic Graphical Fusion of LiDAR, GPS, and 3D Building Maps for Urban UAV Navigation’. In: *Navigation* 66.1, pp. 151–168. DOI: 10.1002/navi.298.

Chen, K., D. Zhang, L. Yao, B. Guo, Z. Yu and Y. Liu (2021). ‘Deep Learning for Sensor-Based Human Activity Recognition: Overview, Challenges, and Opportunities’. In: *ACM Computing Surveys (CSUR)* 54.4, pp. 1–40. DOI: 10.1145/3447744.

Chen, S., P. He, B. Yu, D. Wei and Y. Chen (2024a). ‘The Challenge of Noise Pollution in High-Density Urban Areas: Relationship between 2D/3D Urban Morphology and Noise Perception’. In: *Building and Environment* 253, p. 111313. DOI: 10.1016/j.buildenv.2024.111313.

Chen, T. and C. Guestrin (2016). ‘Xgboost: A Scalable Tree Boosting System’. In: *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, pp. 785–794. DOI: 10.1145/2939672.2939785.

Chen, Z., B. Yang, R. Zhu and Z. Dong (2024b). ‘City-Scale Solar PV Potential Estimation on 3D Buildings Using Multi-Source RS Data: A Case Study in Wuhan, China’. In: *Applied Energy* 359, p. 122720. DOI: 10.1016/j.apenergy.2024.122720.

Cheng, L., A. Zhao, K. Wang, H. Li, Y. Wang and R. Chang (2020). ‘Activity Recognition and Localization Based on UWB Indoor Positioning System and Machine Learning’. In: *2020 11th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*. IEEE, pp. 0528–0533. DOI: 10.1109/IEMCON51383.2020.9284937.

Cho, K., B. Van Merriënboer, D. Bahdanau and Y. Bengio (2014). ‘On the Properties of Neural Machine Translation: Encoder-decoder Approaches’. Preprint. arXiv: 1409.1259.

Choudhury, S., A. R. Kreidieh, I. Kuznetsov and N. Arora (2024). ‘Towards a Trajectory-Powered Foundation Model of Mobility’. In: *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Spatial Big Data and AI for Industrial Applications*, pp. 1–4. DOI: 10.1145/3681766.3699610.

Christodoulides, A., G. K. Tam, J. Clarke, R. Smith, J. Horgan, N. Micallef, J. Morley, N. Villamizar and S. Walton (2025). ‘Survey on 3D Reconstruction Techniques: Large-Scale Urban City Reconstruction and Requirements’. In: *IEEE Transactions on Visualization and Computer Graphics* 31.10, pp. 9343–9367. DOI: 10.1109/TVCG.2025.3540669.

Chung, S., J. Lim, K. J. Noh, G. Kim and H. Jeong (2019). ‘Sensor Data Acquisition and Multimodal Sensor Fusion for Human Activity Recognition Using Deep Learning’. In: *Sensors* 19.7, p. 1716. DOI: 10.3390/s19071716.

Cinnamon, J. and L. Jahiu (2021). 'Panoramic Street-Level Imagery in Data-Driven Urban Research: A Comprehensive Global Review of Applications, Techniques, and Practical Considerations'. In: *ISPRS International Journal of Geo-Information* 10.7, p. 471. DOI: 10.3390/ijgi10070471.

Clarke, P., J. Ailshire, R. Melendez, M. Bader and J. Morenoff (2010). 'Using Google Earth to Conduct a Neighborhood Audit: Reliability of a Virtual Audit Instrument'. In: *Health & place* 16.6, pp. 1224–1229. DOI: 10.1016/j.healthplace.2010.08.007.

Comber, A., M. Umezaki, R. Zhou, Y. Ding, Y. Li, H. Fu, H. Jiang and A. Tewkesbury (2012). 'Using Shadows in High-Resolution Imagery to Determine Building Height'. In: *Remote Sensing Letters* 3.7, pp. 551–556. DOI: 10.1080/01431161.2011.635161.

Commission, F. C. et al. (2002). 'Revision of Part 15 of the Commission's Rules Regarding Ultra WideBand Transmission Systems'. In: *First report and order*, FCC–02.

Cordts, M., M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth and B. Schiele (2016). 'The Cityscapes Dataset for Semantic Urban Scene Understanding'. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3213–3223. DOI: 10.1109/cvpr.2016.350.

Coutrot, A., R. A. Kievit, S. J. Ritchie, E. Manley, J. M. Wiener, C. Hölscher, R. C. Dalton, M. Hornberger and H. J. Spiers (July 2025). 'Education Is Positively and Causally Linked With Spatial Navigation Ability Across the Lifespan'. In: *Open Mind: Discoveries in Cognitive Science* 9, pp. 926–939. DOI: 10.1162/opmi.a.13.

Coutrot, A., E. Manley, S. Goodroe, C. Gahnstrom, G. Filomena, D. Yesiltepe, R. C. Dalton, J. M. Wiener, C. Hölscher, M. Hornberger et al. (2022). 'Entropy of City Street Networks Linked to Future Spatial Navigation Ability'. In: *Nature* 604.7904, pp. 104–110. DOI: 10.1038/s41586-022-04486-7.

Crooks, A., D. Pfoser, A. Jenkins, A. Croitoru, A. Stefanidis, D. Smith, S. Karagiorgou, A. Efentakis and G. Lamprianidis (2015). 'Crowdsourcing Urban Form and Function'. In: *International Journal of Geographical Information Science* 29.5, pp. 720–741. DOI: 10.1080/13658816.2014.977905.

Crooks, A. and L. See (2022). 'Leveraging Street Level Imagery for Urban Planning'. In: *Environment and Planning B: Urban Analytics and City Science* 49.3 (3), pp. 773–776. DOI: 10.1177/23998083221083364.

Cui, Z., L. Mei, S. Pei, B. Li and X. Zhou (2024). 'Privacy-Preserving Human Activity Recognition via Video-Based Range-Doppler Synthesis'. In: *2024 27th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*. IEEE, pp. 649–654. DOI: 10.1109/CSCWD61410.2024.10580232.

Dai, S., Y. Li, A. Stein, S. Yang and P. Jia (2024). 'Street View Imagery-Based Built Environment Auditing Tools: A Systematic Review'. In: *International Journal of Geographical Information Science* 38.6, pp. 1136–1157. DOI: 10.1080/13658816.2024.2336034.

Dalal, N. and B. Triggs (2005). 'Histograms of Oriented Gradients for Human Detection'. In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. Vol. 1. Ieee, pp. 886–893. DOI: 10.1109/CVPR.2005.177.

Dang, L. M., K. Min, H. Wang, M. J. Piran, C. H. Lee and H. Moon (2020). 'Sensor-Based and Vision-Based Human Activity Recognition: A Comprehensive Survey'. In: *Pattern Recognition* 108, p. 107561. DOI: 10.1016/j.patcog.2020.107561.

Dao, M.-S. (2022). 'Multimodal and Crossmodal AI for Smart Data Analysis'. Preprint. arXiv: 2209.01308.

Darken, R. P. and B. Peterson (2002a). 'Spatial Orientation, Wayfinding, and Representation'. In: *Handbook of Virtual Environments*. CRC Press, pp. 533–558.

Darken, R. P. and B. Peterson (2002b). 'Spatial Orientation, Wayfinding, and Representation'. In: *Handbook of Virtual Environments*. CRC Press, pp. 533–558.

Darwish, A. and K. I. Lakhtaria (2011). 'The Impact of the New Web 2.0 Technologies in Communication, Development, and Revolutions of Societies'. In: *Journal of Advances in Information Technology* 2.4, pp. 204–216. DOI: 10.4304/jait.2.4.204-216.

Davies, C. and E. Pederson (2001). 'Grid Patterns and Cultural Expectations in Urban Wayfinding'. In: *International Conference on Spatial Information Theory*. Springer, pp. 400–414. DOI: 10.1007/3-540-45424-1_27.

De Bellefon, M.-P., P.-P. Combes, G. Duranton, L. Gobillon and C. Gorin (2021). 'Delineating Urban Areas Using Building Density'. In: *Journal of Urban Economics* 125, p. 103226. DOI: 10.1016/j.jue.2019.103226.

De Leonardis, G., S. Rosati, G. Balestra, V. Agostini, E. Panero, L. Gastaldi and M. Knaflitz (2018). 'Human Activity Recognition by Wearable Sensors: Comparison of Different Classifiers for Real-Time Applications'. In: *2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*. IEEE, pp. 1–6. DOI: 10.1109/MeMeA.2018.8438750.

Delikostidis, I., C. P. van Elzakker and M.-J. Kraak (2016). 'Overcoming Challenges in Developing More Usable Pedestrian Navigation Systems'. In: *Cartography and Geographic Information Science* 43.3, pp. 189–207. DOI: 10.1080/15230406.2015.1031180.

Demrozi, F., G. Pravadelli, A. Bihorac and P. Rashidi (2020). 'Human Activity Recognition Using Inertial, Physiological and Environmental Sensors: A Comprehensive Survey'. In: *IEEE Access* 8, pp. 210816–210836. DOI: 10.1109/ACCESS.2020.3037715.

Deng, J. S., K. Wang, Y. Hong and J. G. Qi (2009). 'Spatio-Temporal Dynamics and Evolution of Land Use Change and Landscape Pattern in Response to Rapid Urbanization'. In: *Landscape and Urban Planning* 92.3–4, pp. 187–198. DOI: 10.1016/j.landurbplan.2009.05.001.

Deuser, F., W. Mansour, H. Li, K. Habel, M. Werner and N. Oswald (2025). 'Temporal Resilience in Geo-Localization: Adapting to the Continuous Evolution of Urban and Rural Environments'. In: *Proceedings of the Winter Conference on Applications of Computer Vision*, pp. 479–488. DOI: 10.1109/WACVW65960.2025.00055.

Dirgová Luptáková, I., M. Kubovčík and J. Pospíchal (2022). 'Wearable Sensor-Based Human Activity Recognition with Transformer Model'. In: *Sensors* 22.5, p. 1911. DOI: 10.3390/s22051911.

Do, P. N. B. and Q. C. Nguyen (2019). 'A Review of Stereo-Photogrammetry Method for 3-D Reconstruction in Computer Vision'. In: *2019 19th International Symposium on Communications and Information Technologies (ISCIT)*. IEEE, pp. 138–143. DOI: 10.1109/ISCIT.2019.8905144.

Dollár, P., V. Rabaud, G. Cottrell and S. Belongie (2005). 'Behavior Recognition via Sparse Spatio-Temporal Features'. In: *2005 IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*. IEEE, pp. 65–72. DOI: 10.1109/VSPETS.2005.1570899.

Dong, X., J. Cao and W. Zhao (2024). 'A Review of Research on Remote Sensing Images Shadow Detection and Application to Building Extraction'. In: *European Journal of Remote Sensing* 57.1, p. 2293163. DOI: 10.1080/22797254.2023.2293163.

Dorninger, P. and N. Pfeifer (2008). 'A Comprehensive Automated 3D Approach for Building Extraction, Reconstruction, and Regularization from Airborne Laser Scanning Point Clouds'. In: *Sensors* 8.11, pp. 7323–7343. DOI: 10.3390/s8117323.

Dowman, I. J. (2004). 'Integration of LiDAR and IFSAR for Mapping'. In: *International Archives of Photogrammetry and Remote Sensing* 35.B2, pp. 90–100.

Du, H., T. Jin, Y. He, Y. Song and Y. Dai (2020). 'Segmented Convolutional Gated Recurrent Neural Networks for Human Activity Recognition in Ultra-Wideband Radar'. In: *Neurocomputing* 396, pp. 451–464. DOI: 10.1016/j.neucom.2018.11.109.

Dubey, A., N. Naik, D. Parikh, R. Raskar and C. A. Hidalgo (2016). 'Deep Learning the City: Quantifying Urban Perception at a Global Scale'. In: *European Conference on Computer Vision*. Springer, pp. 196–212. DOI: 10.1007/978-3-319-46448-0_12.

Durgam, A., S. Paheding, V. Dhiman and V. Devabhaktuni (2024). 'Cross-View Geo-Localization: A Survey'. In: *IEEE Access*, pp. 192028–192050. DOI: 10.1109/ACCESS.2024.3507280.

Dyer, G. M., S. Khomenko, D. Adlakha, S. Anenberg, J. Angelova, M. Behnisch, G. Boeing, X. Chen, M. Cirach, K. de Hoogh et al. (2024). 'Commentary: A Road Map for Future Data-Driven Urban Planning and Environmental Health Research'. In: *Cities* 155, p. 105340. DOI: 10.1016/j.cities.2024.105340.

Elaksher, A., T. Ali and A. Alharthy (2023). 'A Quantitative Assessment of LIDAR Data Accuracy'. In: *Remote Sensing* 15.2, p. 442. DOI: 10.3390/rs15020442.

Elhanashi, A., P. Dini, S. Saponara and Q. Zheng (2024). 'Advancements in TinyML: Applications, Limitations, and Impact on IoT Devices'. In: *Electronics* 13.17, p. 3562. DOI: 10.3390/electronics13173562.

Enge, P. K. (1994). 'The Global Positioning System: Signals, Measurements, and Performance'. In: *International Journal of Wireless Information Networks* 1.2, pp. 83–105. DOI: 10.1007/BF02106512.

Epstein, R. A., E. Z. Patai, J. B. Julian and H. J. Spiers (2017). 'The Cognitive Map in Humans: Spatial Navigation and Beyond'. In: *Nature Neuroscience* 20.11, pp. 1504–1513. DOI: 10.1038/nn.4656.

Esrafilian, O. and D. Gesbert (2017). '3D City Map Reconstruction from UAV-based Radio Measurements'. In: *GLOBECOM 2017-2017 IEEE Global Communications Conference*. IEEE, pp. 1–6. DOI: 10.1109/GLOCOM.2017.8254657.

Ester, M., H.-P. Kriegel, J. Sander, X. Xu et al. (1996). 'A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise.' In: *Kdd-The Second International Conference on Knowledge Discovery and Data Mining*. Vol. 96. AAAI, pp. 226–231. DOI: 10.5120/739-1038.

Fan, C., X. Zhong and J. Wei (2019). 'BS-to-ground Channel Reconstruction with 3D Obstacle Map Based on RSS Measurements'. In: *IEEE Access* 7, pp. 99633–99641. DOI: 10.1109/ACCESS.2019.2930556.

Fan, Z., C.-C. Feng and F. Biljecki (2025). 'Coverage and Bias of Street View Imagery in Mapping the Urban Environment'. In: *Computers, Environment and Urban Systems* 117, p. 102253. DOI: 10.1016/j.compenvurbsys.2025.102253.

Farr, A. C., T. Kleinschmidt, P. Yarlagadda and K. Mengersen (2012). 'Wayfinding: A Simple Concept, a Complex Process'. In: *Transport Reviews* 32.6, pp. 715–743. DOI: 10.1080/01441647.2012.712555.

Feichtenhofer, C., A. Pinz and A. Zisserman (2016). 'Convolutional Two-Stream Network Fusion for Video Action Recognition'. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1933–1941. DOI: 10.1109/cvpr.2016.213.

Feng, J., S. Wang, T. Liu, Y. Xi and Y. Li (2025). 'UrbanLLaVA: A Multi-Modal Large Language Model for Urban Intelligence with Spatial Reasoning and Understanding'. Preprint. arXiv: 2506.23219.

Flanagin, A. J. and M. J. Metzger (2008). 'The Credibility of Volunteered Geographic Information'. In: *GeoJournal* 72.3, pp. 137–148. DOI: 10.1007/s10708-008-9188-y.

Frantz, D., F. Schug, A. Okujeni, C. Navacchi, W. Wagner, S. van der Linden and P. Hostert (2021). 'National-Scale Mapping of Building Height Using Sentinel-1 and Sentinel-2 Time Series'. In: *Remote Sensing of Environment* 252, p. 112128. DOI: 10.1016/j.rse.2020.112128.

Frolking, S., R. Mahtta, T. Milliman, T. Esch and K. C. Seto (2024). 'Global Urban Structural Growth Shows a Profound Shift from Spreading out to Building Up'. In: *Nature Cities* 1.9, pp. 555–566. DOI: 10.1038/s44284-024-00100-1.

Fu, B., N. Damer, F. Kirchbuchner and A. Kuijper (2020). 'Sensing Technology for Human Activity Recognition: A Comprehensive Survey'. In: *IEEE Access* 8, pp. 83791–83820. DOI: 10.1109/ACCESS.2020.2991891.

Fukushima, K. (1980). 'Neocognitron: A Self-Organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position'. In: *Biological Cybernetics* 36.4, pp. 193–202. DOI: 10.1007/BF00344251.

Gählert, N., N. Jourdan, M. Cordts, U. Franke and J. Denzler (2020). 'Cityscapes 3d: Dataset and Benchmark for 9 Dof Vehicle Detection'. Preprint. arXiv: 2006.07864.

Galbrun, E., K. Pelechrinis and E. Terzi (2016). 'Urban Navigation beyond Shortest Route: The Case of Safe Paths'. In: *Information Systems* 57, pp. 160–171. DOI: 10.1016/j.is.2015.10.005.

Gao, S., Y. Liu, Y. Kang and F. Zhang (2021). 'User-Generated Content: A Promising Data Source for Urban Informatics'. In: *Urban Informatics*. Springer, pp. 503–522.

Gath-Morad, M., J. Grübel, K. Steemers, K. Sailer, L. Ben-Alon, C. Hölscher and L. Aguilar (2024). 'The Role of Strategic Visibility in Shaping Wayfinding Behavior in Multilevel Buildings'. In: *Scientific Reports* 14.1, p. 3735. DOI: 10.1038/s41598-024-53420-6.

Gath-Morad, M., L. E. A. Melgar, R. Conroy-Dalton and C. Hölscher (2022). 'Beyond the Shortest-Path: Towards Cognitive Occupancy Modeling in BIM'. In: *Automation in Construction* 135, p. 104131. DOI: 10.1016/j.autcon.2022.104131.

Geravesh, S. and V. Rupapara (2023). 'Artificial Neural Networks for Human Activity Recognition Using Sensor Based Dataset'. In: *Multimedia Tools and Applications* 82.10, pp. 14815–14835. DOI: 10.1007/s11042-022-13716-z.

Ghahramani, M., M. Zhou and G. Wang (2020). 'Urban Sensing Based on Mobile Phone Data: Approaches, Applications, and Challenges'. In: *IEEE/CAA Journal of Automatica Sinica* 7.3, pp. 627–637. DOI: 10.1109/JAS.2020.1003120.

Glaeser, E. L., S. D. Kominers, M. Luca and N. Naik (2018). 'Big Data and Big Cities: The Promises and Limitations of Improved Measures of Urban Life'. In: *Economic Inquiry* 56.1, pp. 114–137. DOI: 10.1111/ecin.12364.

Goepel, K. D. (2018). 'Implementation of an Online Software Tool for the Analytic Hierarchy Process (AHP-OS)'. In: *International Journal of the Analytic Hierarchy Process* 10.3. DOI: 10.13033/ijahp.v10i3.590.

Golledge, R. G. (1999). *Wayfinding Behavior: Cognitive Mapping and Other Spatial Processes*. Baltimore, United States: JHU press.

Goodchild, M. F. (2007). 'Citizens as Sensors: The World of Volunteered Geography'. In: *GeoJournal* 69.4, pp. 211–221. DOI: 10.1007/s10708-007-9111-y.

Goodchild, M. F. and L. Li (2012). 'Assuring the Quality of Volunteered Geographic Information'. In: *Spatial statistics* 1, pp. 110–120. DOI: 10.1016/j.spasta.2012.03.002.

Google (2001). *Google Earth Homepage*. URL: https://www.google.com/earth/ (visited on 31/12/2025).

Google (2019). *Google Maps 101: How Imagery Powers Our Map*. URL: https://blog.google/products/maps/google-maps-101-how-imagery-powers-our-map/ (visited on 31/12/2025).

Google (2025a). *Discover When, Where and How We Collect 360 Imagery*. URL: https://www.google.co.uk/intl/en_uk/streetview/how-it-works/ (visited on 31/12/2025).

Google (2025b). *Street View Static API*. URL: https://developers.google.com/maps/documentation/streetview (visited on 31/12/2025).

Google Maps (2025). *Google Maps*. URL: https://maps.google.com (visited on 31/12/2025).

Gorji, A., A. Bourdoux, H. Sahli et al. (2021). 'On the Generalization and Reliability of Single Radar-Based Human Activity Recognition'. In: *IEEE Access* 9, pp. 85334–85349. DOI: 10.1109/ACCESS.2021.3088452.

Gorjian, M. (2025). 'From Deductive Models to Data-Driven Urban Analytics: A Critical Review of Statistical Methodologies, Big Data, and Network Science in Urban Studies'. In: *Preprint]. Preprints. https://doi. org/10.20944/preprints202508* 523, p. v1. DOI: https://doi.org/10.20944/preprints202508.

Grekousis, G. (2019). 'Artificial Neural Networks and Deep Learning in Urban Geography: A Systematic Review and Meta-Analysis'. In: *Computers, Environment and Urban Systems* 74, pp. 244–256. DOI: 10.1016/j.compenvurbsys.2018.10.008.

Groves, P. (2013). *Principles of GNSS, Inertial, and Multisensor Integrated Navigation Systems, Second Edition*. Artech House. ISBN: 978-1-60807-005-3.

Groves, P. D. (2011a). 'Shadow Matching: A New GNSS Positioning Technique for Urban Canyons'. In: *The Journal of Navigation* 64.3, pp. 417–430. DOI: 10.1017/S0373463311000087.

Groves, P. D. and M. Adjrad (2019). 'Performance Assessment of 3D-mapping–Aided GNSS Part 1: Algorithms, User Equipment, and Review'. In: *Navigation* 66.2, pp. 341–362. DOI: 10.1002/navi.288.

Groves, P. D., Z. Jiang, L. Wang and M. K. Ziebart (2012). 'Intelligent Urban Positioning Using Multi-Constellation GNSS with 3D Mapping and NLOS Signal Detection'. In: *Proceedings of the 25th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS 2012)*, pp. 458–472.

Groves, P. D. (2011b). 'Shadow Matching: A New GNSS Positioning Technique for Urban Canyons'. In: *Journal of Navigation* 64.3, pp. 417–430. DOI: 10.1017/S0373463311000087.

Grübel, J., S. Wise, T. Thrash and C. Hölscher (2019). 'A Cognitive Model for Routing in Agent-Based Modelling'. In: *AIP Conference Proceedings*. Vol. 2116. 1. AIP Publishing LLC, p. 250005. DOI: 10.1063/1.5114245.

Gu, J., E. Stefani, Q. Wu, J. Thomason and X. E. Wang (2022). 'Vision-and-Language Navigation: A Survey of Tasks, Methods, and Future Directions'. Preprint. arXiv: 2203.12667.

Gupta, M., A. Abdolrahmani, E. Edwards, M. Cortez, A. Tumang, Y. Majali, M. Lazaga, S. Tarra, P. Patil, R. Kuber et al. (2020). 'Towards More Universal Wayfinding Technologies: Navigation Preferences across Disabilities'. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–13. DOI: 10.1145/3313831.3376581.

Hadsell, R., S. Chopra and Y. LeCun (2006). 'Dimensionality Reduction by Learning an Invariant Mapping'. In: *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*. Vol. 2. IEEE, pp. 1735–1742. DOI: 10.1109/CVPR.2006.100.

Han, S., Z. Gong, W. Meng, C. Li and X. Gu (2016). 'Future Alternative Positioning, Navigation, and Timing Techniques: A Survey'. In: *IEEE Wireless Communications* 23.6, pp. 154–160. DOI: 10.1109/MWC.2016.1500181RP.

Hasan, S., X. Zhan and S. V. Ukkusuri (2013). 'Understanding Urban Human Activity and Mobility Patterns Using Large-Scale Location-Based Data from Online Social Media'. In: *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing*, pp. 1–8. DOI: 10.1145/2505821.2505823.

He, N. and G. Li (2021). 'Urban Neighbourhood Environment Assessment Based on Street View Image Processing: A Review of Research Trends'. In: *Environmental Challenges* 4, p. 100090. DOI: 10.1016/j.envc.2021.100090.

Hecht, B. and M. Stephens (2014). 'A Tale of Cities: Urban Biases in Volunteered Geographic Information'. In: *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 8. 1, pp. 197–205. DOI: 10.1609/icwsm.v8i1.14554.

Hegarty, M., A. E. Richardson, D. R. Montello, K. Lovelace and I. Subbiah (2002). 'Development of a Self-Report Measure of Environmental Spatial Ability'. In: *Intelligence* 30.5, pp. 425–447. DOI: 10.1016/S0160-2896(02)00116-2.

Heipke, C. (2010). 'Crowdsourcing Geospatial Data'. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 65.6, pp. 550–557. DOI: 10.1016/j.isprsjprs.2010.06.005.

Herman, L. and T. Řezník (2015). '3D Web Visualization of Environmental Information–Integration of Heterogeneous Data Sources When Providing Navigation and Interaction'. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 40, pp. 479–485. DOI: 10.5194/isprsarchives-XL-3-W3-479-2015.

Hillier, B. and J. Hanson (1984). *The Social Logic of Space*. Cambridge University Press. DOI: 10.1017/CBO9780511597237.

Hochreiter, S. and J. Schmidhuber (1997). 'Long Short-Term Memory'. In: *Neural computation* 9.8, pp. 1735–1780. DOI: 10.1162/neco.1997.9.8.1735.

Hoerl, A. E. and R. W. Kennard (1970). 'Ridge Regression: Biased Estimation for Nonorthogonal Problems'. In: *Technometrics* 12.1, pp. 55–67. DOI: 10.2307/1271436.

Hofmann-Wellenhof, B., H. Lichtenegger and E. Wasle (2008). *GNSS—Global Navigation Satellite Systems: GPS, GLONASS, Galileo, and More*. Springer Vienna: Springer Vienna.

Hölscher, C., M. Brösamle and G. Vrachliotis (2012). 'Challenges in Multilevel Wayfinding: A Case Study with the Space Syntax Technique'. In: *Environment and Planning B: Planning and Design* 39.1, pp. 63–82. DOI: 10.1068/b34050t.

Holte, M. B., C. Tran, M. M. Trivedi and T. B. Moeslund (2012). 'Human Pose Estimation and Activity Recognition from Multi-View Videos: Comparative Explorations of Recent Developments'. In: *IEEE Journal of Selected Topics in Signal Processing* 6.5, pp. 538–552. DOI: 10.1109/JSTSP.2012.2196975.

Hrabar, S. and G. Sukhatme (2009). 'Vision-Based Navigation through Urban Canyons'. In: *Journal of Field Robotics* 26.5, pp. 431–452. DOI: 10.1002/rob.20284.

Hsieh, C.-F., Y.-C. Chen, C.-Y. Hsieh and M.-L. Ku (2020). 'Device-Free Indoor Human Activity Recognition Using Wi-Fi RSSI: Machine Learning Approaches'. In: *2020 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-taiwan)*. IEEE, pp. 1–2. DOI: 10.1109/icce-taiwan49838.2020.9258097.

Hsu, L.-T., Y. Gu, Y. Huang and S. Kamijo (2015). 'Urban Pedestrian Navigation Using Smartphone-Based Dead Reckoning and 3-D Map-Aided GNSS'. In: *IEEE Sensors Journal* 16.5, pp. 1281–1293. DOI: 10.1109/JSEN.2015.2496621.

Hu, C., Q. Yang, Y. Liu, T. Röddiger, K.-J. Butkow, M. Ciliberto, A. L. Pullin, J. Stuchbury-Wass, M. Hassan, C. Mascolo et al. (2025). 'A Survey of Earable Technology: Trends, Tools, and the Road Ahead'. Preprint. arXiv: 2506.05720.

Hu, D. and J. Minner (2023). 'UAVs and 3D City Modeling to Aid Urban Planning and Historic Preservation: A Systematic Review'. In: *Remote Sensing* 15.23, p. 5507. DOI: 10.3390/rs15235507.

Hu, J., Y. Gao, X. Wang and Y. Liu (2023). 'Recognizing Mixed Urban Functions from Human Activities Using Representation Learning Methods'. In: *International Journal of Digital Earth* 16.1, pp. 289–307. DOI: 10.1080/17538947.2023.2170482.

Hu, S., M. Feng, R. M. Nguyen and G. H. Lee (2018). 'CVM-Net: Cross-View Matching Network for Image-Based Ground-to-Aerial Geo-Localization'. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7258–7267. DOI: 10.1109/cvpr.2018.00758.

Hu, S. and G. H. Lee (2020). ‘Image-Based Geo-Localization Using Satellite Imagery’. In: *International Journal of Computer Vision* 128.5, pp. 1205–1219. DOI: 10.1007/s11263-019-01186-0.

Hu, T., S. Wang, B. She, M. Zhang, X. Huang, Y. Cui, J. Khuri, Y. Hu, X. Fu, X. Wang et al. (2021). ‘Human Mobility Data in the COVID-19 Pandemic: Characteristics, Applications, and Challenges’. In: *International Journal of Digital Earth* 14.9, pp. 1126–1147. DOI: 10.1080/17538947.2021.1952324.

Huang, H., N.-C. Lin, L. Barrett, D. Springer, H.-C. Wang, M. Pomplun and L.-F. Yu (2017). ‘Automatic Optimization of Wayfinding Design’. In: *IEEE Transactions on Visualization and Computer Graphics* 24.9, pp. 2516–2530. DOI: 10.1109/TVCG.2017.2761820.

Huang, H., Y. Cheng and R. Weibel (2019). ‘Transport Mode Detection Based on Mobile Phone Network Data: A Systematic Review’. In: *Transportation Research Part C: Emerging Technologies* 101, pp. 297–312. DOI: 10.1016/j.trc.2019.02.008.

Huang, H., X. A. Yao, J. M. Krisp and B. Jiang (2021). ‘Analytics of Location-Based Big Data for Smart Cities: Opportunities, Challenges, and Future Directions’. In: *Computers, Environment and Urban Systems* 90, p. 101712. DOI: 10.1016/j.compenvurbsys.2021.101712.

Huang, X., S. Wang, D. Yang, T. Hu, M. Chen, M. Zhang, G. Zhang, F. Biljecki, T. Lu, L. Zou et al. (2024). ‘Crowdsourcing Geospatial Data for Earth and Human Observations: A Review’. In: *Journal of Remote Sensing* 4, p. 0105. DOI: 10.34133/remotesensing.0105.

Hunter, R. H., L. A. Anderson and B. L. Belza (2016a). ‘Introduction to Community Wayfinding’. In: *Community Wayfinding: Pathways to Understanding.* Springer, pp. 3–16.

Hunter, R. H., L. A. Anderson and B. L. Belza (2016b). ‘Introduction to Community Wayfinding’. In: *Community Wayfinding: Pathways to Understanding.* Cham: Springer International Publishing, pp. 3–16. ISBN: 978-3-319-31072-5. DOI: 10.1007/978-3-319-31072-5_1.

Hussain, Z., Q. Z. Sheng and W. E. Zhang (2020). ‘A Review and Categorization of Techniques on Device-Free Human Activity Recognition’. In: *Journal of Network and Computer Applications* 167, p. 102738. DOI: 10.1016/j.jnca.2020.102738.

Iftikhar, H., P. Shah and Y. Luximon (2021). ‘Human Wayfinding Behaviour and Metrics in Complex Environments: A Systematic Literature Review’. In: *Architectural Science Review* 64.5, pp. 452–463. DOI: 10.1080/00038628.2020.1777386.

Ilci, V. and C. Toth (2020). ‘High Definition 3D Map Creation Using GNSS/IMU/LiDAR Sensor Integration to Support Autonomous Vehicle Navigation’. In: *Sensors* 20.3, p. 899. DOI: 10.3390/s20030899.

Inoue, M., S. Inoue and T. Nishida (2018). 'Deep Recurrent Neural Network for Mobile Human Activity Recognition with High Throughput'. In: *Artificial Life and Robotics* 23.2, pp. 173–185. DOI: 10.1007/s10015-017-0422-x.

Iossifova, D., C. Doll and A. Gasparatos (2017). 'Defining the Urban: Why Do We Need Definitions?' In: *Defining the Urban: Interdisciplinary and Professional Perspectives*. Routledge.

Ishikawa, T. (2019). 'Satellite Navigation and Geospatial Awareness: Long-Term Effects of Using Navigation Tools on Wayfinding and Spatial Orientation'. In: *The Professional Geographer* 71.2, pp. 197–209. DOI: 10.1080/00330124.2018.1479970.

Ito, K., Y. Kang, Y. Zhang, F. Zhang and F. Biljecki (2024). 'Understanding Urban Perception with Visual Data: A Systematic Review'. In: *Cities* 152, p. 105169. DOI: 10.1016/j.cities.2024.105169.

Javed, A. R., R. Faheem, M. Asim, T. Baker and M. O. Beg (2021). 'A Smartphone Sensors-Based Personalized Human Activity Recognition System for Sustainable Smart Cities'. In: *Sustainable Cities and Society* 71, p. 102970. DOI: 10.1016/j.scs.2021.102970.

Ji, S., Y. Zheng and T. Li (2016). 'Urban Sensing Based on Human Mobility'. In: *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 1040–1051. DOI: 10.1145/2971648.2971735.

Jiang, S., J. Ferreira Jr and M. C. Gonzalez (2012). 'Discovering Urban Spatial-Temporal Structure from Human Activity Patterns'. In: *Proceedings of the ACM SIGKDD International Workshop on Urban Computing*, pp. 95–102. DOI: 10.1145/2346496.2346512.

Juhász, L. and H. H. Hochmair (2016). 'User Contribution Patterns and Completeness Evaluation of Mapillary, a Crowdsourced Street Level Photo Service'. In: *Transactions in GIS* 20.6, pp. 925–947. DOI: 10.1111/tgis.12190.

Kadhim, N. and M. Mourshed (2017). 'A Shadow-Overlapping Algorithm for Estimating Building Heights from VHR Satellite Images'. In: *IEEE Geoscience and Remote Sensing Letters* 15.1, pp. 8–12. DOI: 10.1109/LGRS.2017.2762424.

Kafi, M. A., Y. Challal, D. Djenouri, M. Doudou, A. Bouabdallah and N. Badache (2013). 'A Study of Wireless Sensor Networks for Urban Traffic Monitoring: Applications and Architectures'. In: *Procedia Computer Science* 19, pp. 617–626. DOI: 10.1016/j.procs.2013.06.082.

Kamar, E., A. Kapoor and E. Horvitz (2015). 'Identifying and Accounting for Task-Dependent Bias in Crowdsourcing'. In: *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*. Vol. 3, pp. 92–101. DOI: 10.1609/hcomp.v3i1.13238.

Kandt, J. and M. Batty (2021). 'Smart Cities, Big Data and Urban Policy: Towards Urban Analytics for the Long Run'. In: *Cities* 109, p. 102992. DOI: 10.1016/j.cities.2020.102992.

Kang, Y., F. Zhang, S. Gao, H. Lin and Y. Liu (2020). ‘A Review of Urban Physical Environment Sensing Using Street View Imagery in Public Health Studies’. In: *Annals of GIS* 26.3, pp. 261–275. DOI: 10.1080/19475683.2020.1791954.

Kaplan, E. D. and C. Hegarty (2017). *Understanding GPS/GNSS: Principles and Applications*. Norwood, Massachusetts, United States: Artech house.

Kapoor, R., S. Ramasamy, A. Gardi and R. Sabatini (2017). ‘UAV Navigation Using Signals of Opportunity in Urban Environments: A Review’. In: *Energy Procedia* 110, pp. 377–383. DOI: 10.1016/j.egypro.2017.03.156.

Karim, F., S. Majumdar, H. Darabi and S. Chen (2017). ‘LSTM Fully Convolutional Networks for Time Series Classification’. In: *IEEE Access* 6, pp. 1662–1669. DOI: 10.1109/ACCESS.2017.2779939.

Karimi, H. A., M. Jiang and R. Zhu (2013). ‘Pedestrian Navigation Services: Challenges and Current Trends’. In: *Geomatica* 67.4, pp. 259–271. DOI: 10.5623/cig2013-052.

Karpathy, A., G. Toderici, S. Shetty, T. Leung, R. Sukthankar and L. Fei-Fei (2014). ‘Large-Scale Video Classification with Convolutional Neural Networks’. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1725–1732. DOI: 10.1109/cvpr.2014.223.

KartaView Contributors (2025). *KartaView: A Street-Level Imagery Platform*. URL: https://kartaview.org/landing (visited on 31/12/2025).

Kaseris, M., I. Kostavelis and S. Malassiotis (2024). ‘A Comprehensive Survey on Deep Learning Methods in Human Activity Recognition’. In: *Machine Learning and Knowledge Extraction* 6.2, pp. 842–876. DOI: 10.3390/make6020040.

Kay, W., J. Carreira, K. Simonyan, B. Zhang, C. Hillier, S. Vijayanarasimhan, F. Viola, T. Green, T. Back, P. Natsev et al. (2017). ‘The Kinetics Human Action Video Dataset’. Preprint. arXiv: 1705.06950.

Kim, D. and J. Choi (2025). ‘An Analysis of 119 Emergency Rescue Service Accessibility Based on 3D Buildings’. In: *Journal of the Korean Geographical Society* 60.3, pp. 371–380. DOI: 10.22776/kgs.2025.60.3.371.

Kim, J. and K. M. Jang (2023). ‘An Examination of the Spatial Coverage and Temporal Variability of Google Street View (GSV) Images in Small-and Medium-Sized Cities: A People-Based Approach’. In: *Computers, Environment and Urban Systems* 102, p. 101956. DOI: 10.1016/j.compenvurbsys.2023.101956.

Kinnari, J., F. Verdoja and V. Kyrki (2022). ‘Season-Invariant GNSS-denied Visual Localization for Uavs’. In: *IEEE Robotics and Automation Letters* 7.4, pp. 10232–10239. DOI: 10.1109/LRA.2022.3191038.

Kitchin, R. (2014). *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. London, UK: Sage.

Koletsis, E., C. P. van Elzakker, M.-J. Kraak, W. Cartwright, C. Arrowsmith and K. Field (2017). 'An Investigation into Challenges Experienced When Route Planning, Navigating and Wayfinding'. In: *International Journal of Cartography* 3.1, pp. 4–18. DOI: 10.1080/23729333.2017.1300996.

Korada, L. (2021). 'Unlocking Urban Futures: The Role of Big Data Analytics and AI in Urban Planning–a Systematic Literature Review and Bibliometric Insight'. In: *Migration Letters* 18.6, pp. 775–795.

Koukiou, G. (2024). 'SAR Features and Techniques for Urban Planning—a Review'. In: *Remote Sensing* 16.11, p. 1923. DOI: 10.3390/rs16111923.

Koutra, S. and C. S. Ioakimidis (2022). 'Unveiling the Potential of Machine Learning Applications in Urban Planning Challenges'. In: *Land* 12.1, p. 83. DOI: 10.3390/land12010083.

Krippendorff, K. (2018). *Content Analysis: An Introduction to Its Methodology*. Thousand Oaks, CA, United States: Sage publications.

Kubat, A. S., A. Ozbil Torun, O. Ozer and H. Ekinoglu (2012). 'The Effect of Built Space on Wayfinding in Urban Environments: A Study of the Historical Peninsula in Istanbul'. In: *Proceedings: Eighth International Space Syntax Symposium*. Chile: Pontificia Universidad Catolica de Chile, 8029:1–20. ISBN: 978-956-345-862-6.

Kumar, P., S. Chauhan and L. K. Awasthi (2024). 'Human Activity Recognition (HAR) Using Deep Learning: Review, Methodologies, Progress and Future Research Directions'. In: *Archives of Computational Methods in Engineering* 31.1, pp. 179–219. DOI: 10.1007/s11831-023-09986-x.

Kustu, T. and A. Taskin (2023). 'Deep Learning and Stereo Vision Based Detection of Post-Earthquake Fire Geolocation for Smart Cities within the Scope of Disaster Management: İstanbul Case'. In: *International Journal of Disaster Risk Reduction* 96, p. 103906. DOI: 10.1016/j.ijdrr.2023.103906.

Kwan, M.-P. and J. Lee (2005). 'Emergency Response after 9/11: The Potential of Real-Time 3D GIS for Quick Emergency Response in Micro-Spatial Environments'. In: *Computers, Environment and Urban Systems* 29.2, pp. 93–113. DOI: 10.1016/j.compenvurbsys.2003.08.002.

Lab, E. (2020). *Canopy Height Models, Digital Surface Models & Digital Elevation Models - Work with LiDAR Data in Python*. Ofcom London.

Lafontaine, V., K. Bouchard, J. Maítre and S. Gaboury (2023). 'Denoising UWB Radar Data for Human Activity Recognition Using Convolutional Autoencoders'. In: *IEEE Access* 11, pp. 81298–81309. DOI: 10.1109/ACCESS.2023.3300224.

Langer, E. J. and S. Saegert (1977). 'Crowding and Cognitive Control'. In: *Journal of Personality and Social Psychology* 35.3, pp. 175–182. ISSN: 0022-3514. DOI: 10.1037/0022-3514.35.3.175.

Langley, R. B., P. J. Teunissen and O. Montenbruck (2017). 'Introduction to GNSS'. In: *Springer Handbook of Global Navigation Satellite Systems*. Springer, pp. 3–23.

Laptev, I., M. Marszalek, C. Schmid and B. Rozenfeld (2008). 'Learning Realistic Human Actions from Movies'. In: *2008 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, pp. 1–8. DOI: 10.1109/CVPR.2008.4587756.

Lau, B. P. L., S. H. Marakkalage, Y. Zhou, N. U. Hassan, C. Yuen, M. Zhang and U.-X. Tan (2019). 'A Survey of Data Fusion in Smart City Applications'. In: *Information Fusion* 52, pp. 357–374. DOI: 10.1016/j.inffus.2019.05.004.

Lau, B. P. L., N. Wijerathne, B. K. K. Ng and C. Yuen (2017). 'Sensor Fusion for Public Space Utilization Monitoring in a Smart City'. In: *IEEE Internet of Things Journal* 5.2, pp. 473–481. DOI: 10.1109/JIOT.2017.2748987.

Lawrence, V. (2004). 'The Role of National Mapping Organizations'. In: *The Cartographic Journal* 41.2, pp. 117–122. DOI: 10.1179/000870404X12581.

LeCun, Y., Y. Bengio and G. Hinton (2015). 'Deep Learning'. In: *Nature* 521.7553, pp. 436–444. DOI: 10.1038/nature14539.

LeCun, Y., B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard and L. D. Jackel (1989). 'Backpropagation Applied to Handwritten Zip Code Recognition'. In: *Neural Computation* 1.4, pp. 541–551. DOI: 10.1162/neco.1989.1.4.541.

Lentzas, A. and D. Vrakas (2020). 'Non-Intrusive Human Activity Recognition and Abnormal Behavior Detection on Elderly People: A Review'. In: *Artificial Intelligence Review* 53.3, pp. 1975–2021. DOI: 10.1007/s10462-019-09724-5.

Li, B., W. Cui, W. Wang, L. Zhang, Z. Chen and M. Wu (2021). 'Two-Stream Convolution Augmented Transformer for Human Activity Recognition'. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 35. 1, pp. 286–293. DOI: 10.1609/aaai.v35i1.16103.

Li, J., X. Huang, L. Tu, T. Zhang and L. Wang (2022a). 'A Review of Building Detection from Very High Resolution Optical Remote Sensing Images'. In: *GIScience & Remote Sensing* 59.1, pp. 1199–1225. DOI: 10.1080/15481603.2022.2101727.

Li, S., A. Balatsoukas-Stimming and A. Burg (2022b). 'Device-Free Movement Tracking Using the UWB Channel Impulse Response with Machine Learning'. In: *2022 IEEE 23rd International Workshop on Signal Processing Advances in Wireless Communication (SPAWC)*. IEEE, pp. 1–5. DOI: 10.1109/spawc51304.2022.9833950.

Li, T., T. Xia, H. Wang, Z. Tu, S. Tarkoma, Z. Han and P. Hui (2022c). 'Smartphone App Usage Analysis: Datasets, Methods, and Applications'. In: *IEEE Communications Surveys & Tutorials* 24.2, pp. 937–966. DOI: 10.1109/COMST.2022.3163176.

Li, X., H. Ning, X. Huang, B. Dadashova, Y. Kang and A. Ma (2022d). 'Urban Infrastructure Audit: An Effective Protocol to Digitize Signalized Intersections by Mining Street View Images'. In: *Cartography and Geographic Information Science* 49.1, pp. 32–49. DOI: 10.1080/15230406.2021.1992299.

Li, X., C. Ratti and I. Seiferling (2018). ‘Quantifying the Shade Provision of Street Trees in Urban Landscape: A Case Study in Boston, USA, Using Google Street View’. In: *Landscape and Urban Planning* 169, pp. 81–91. DOI: 10.1016/j.landurbplan.2017.08.011.

Li, X., Y. Zhou, P. Gong, K. C. Seto and N. Clinton (2020). ‘Developing a Method to Estimate Building Height from Sentinel-1 Data’. In: *Remote Sensing of Environment* 240, p. 111705. DOI: 10.1016/j.rse.2020.111705.

Li, Y., L. Peng, C. Wu and J. Zhang (2022e). ‘Street View Imagery (SVI) in the Built Environment: A Theoretical and Systematic Review’. In: *Buildings* 12.8, p. 1167. DOI: 10.3390/buildings12081167.

Li, Y. and J. Ibanez-Guzman (2020). ‘LiDAR for Autonomous Driving: The Principles, Challenges, and Trends for Automotive Lidar and Perception Systems’. In: *IEEE Signal Processing Magazine* 37.4, pp. 50–61. DOI: 10.1109/msp.2020.2973615.

Li, Y., L. Zhao, Y. Chen, N. Zhang, H. Fan and Z. Zhang (2023). ‘3D LiDAR and Multi-Technology Collaboration for Preservation of Built Heritage in China: A Review’. In: *International Journal of Applied Earth Observation and Geoinformation* 116, p. 103156. DOI: 10.1016/j.jag.2022.103156.

Liang, X., T. Zhao and F. Biljecki (2023). ‘Revealing Spatio-Temporal Evolution of Urban Visual Environments with Street View Imagery’. In: *Landscape and Urban Planning* 237, p. 104802. DOI: 10.1016/j.landurbplan.2023.104802.

Lin, C., Y. Wang, B. Gong, H. Liu and H. Liu (2025). ‘Roadside LiDAR Deployment Optimization for Vehicle-Infrastructure Cooperative Perception in Urban Occlusion Environments’. In: *IEEE Transactions on Instrumentation and Measurement* 74, pp. 1–14. DOI: 10.1109/tim.2025.3558231.

Lin, T., Y. Wang, X. Liu and X. Qiu (2022). ‘A Survey of Transformers’. In: *AI Open* 3, pp. 111–132. DOI: 10.1016/j.aiopen.2022.10.001.

Lin, T.-Y., Y. Cui, S. Belongie and J. Hays (June 2015). ‘Learning Deep Representations for Ground-to-Aerial Geolocalization’. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5007–5015.

Lines, T. and A. Basiri (2021). ‘3D Map Creation Using Crowdsourced GNSS Data’. In: *Computers, Environment and Urban Systems* 89, p. 101671. DOI: 10.1016/j.compenvurbsys.2021.101671.

Ling, H., R.-C. Chou and S.-W. Lee (1989). ‘Shooting and Bouncing Rays: Calculating the RCS of an Arbitrarily Shaped Cavity’. In: *IEEE Transactions on Antennas and propagation* 37.2, pp. 194–205. DOI: 10.1109/8.18706.

Liu, C., Y. Huang, Z. Liu and Z. Wu (2025). ‘Progress of Urban Informatics in Urban Planning Research, Education, and Practice’. In: *Urban Informatics* 4.1, p. 6. DOI: 10.1007/s44212-025-00070-2.

Liu, C., B. Shi, X. Yang, N. Li and H. Wu (2013). ‘Automatic Buildings Extraction from LiDAR Data in Urban Area by Neural Oscillator Network of Visual Cortex’. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 6.4, pp. 2008–2019. DOI: 10.1109/JSTARS.2012.2234726.

Liu, H., Y. Tong, J. Han, P. Zhang, X. Lu and H. Xiong (2020). ‘Incorporating Multi-Source Urban Data for Personalized and Context-Aware Multi-Modal Transportation Recommendation’. In: *IEEE Transactions on Knowledge and Data Engineering* 34.2, pp. 723–735. DOI: 10.1109/TKDE.2020.2985954.

Liu, I., R. M. Levy, J. J. Barton and G. Iaria (2011). ‘Age and Gender Differences in Various Topographical Orientation Strategies’. In: *Brain Research* 1410, pp. 112–119. DOI: 10.1016/j.brainres.2011.07.005.

Liu, J., H. Liu, Y. Chen, Y. Wang and C. Wang (2019). ‘Wireless Sensing for Human Activity: A Survey’. In: *IEEE Communications Surveys & Tutorials* 22.3, pp. 1629–1645. DOI: 10.1109/COMST.2019.2934489.

Liu, L. and H. Li (2019a). ‘Lending Orientation to Neural Networks for Cross-View Geo-Localization’. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5624–5633. DOI: 10.1109/cvpr.2019.00577.

Liu, L. and H. Li (June 2019b). ‘Lending Orientation to Neural Networks for Cross-View Geo-Localization’. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5617–5626.

Liu, Y., J. Ding, G. Deng, Y. Li, T. Zhang, W. Sun, Y. Zheng, J. Ge and Y. Liu (2024a). ‘Image-Based Geolocation Using Large Vision-Language Models’. Preprint. arXiv: 2408.09474.

Liu, Y., A. Obukhov, J. D. Wegner and K. Schindler (2024b). ‘Point2Building: Reconstructing Buildings from Airborne LiDAR Point Clouds’. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 215, pp. 351–368. DOI: 10.1016/j.isprsjprs.2024.07.012.

Liu, Z. and K. He (2024). ‘A Decade’s Battle on Dataset Bias: Are We There Yet?’ Preprint. arXiv: 2403.08632.

Liu, Z., K. Janowicz, L. Cai, R. Zhu, G. Mai and M. Shi (2022). ‘Geoparsing: Solved or Biased? An Evaluation of Geographic Biases in Geoparsing’. In: *AGILE: GIScience Series* 3, p. 9. DOI: 10.5194/agile-giss-3-9-2022.

Lloyd, S. (1982). ‘Least Squares Quantization in PCM’. In: *IEEE transactions on information theory* 28.2, pp. 129–137. DOI: 10.1109/TIT.1982.1056489.

Long, Y. and L. Liu (2016). ‘Transformations of Urban Studies and Planning in the Big/Open Data Era: A Review’. In: *International Journal of Image and Data Fusion* 7.4, pp. 295–308. DOI: 10.1080/19479832.2016.1215355.

Lopez-Fuentes, L., J. van de Weijer, M. González-Hidalgo, H. Skinnemoen and A. D. Bagdanov (2018). 'Review on Computer Vision Techniques in Emergency Situations'. In: *Multimedia Tools and Applications* 77.13, pp. 17069–17107. DOI: 10.1007/s11042-017-5276-7.

Lu, Y., H. Ma, E. Smart and H. Yu (2021). 'Real-Time Performance-Focused Localization Techniques for Autonomous Vehicle: A Review'. In: *IEEE Transactions on Intelligent Transportation Systems* 23.7, pp. 6082–6100. DOI: 10.1109/TITS.2021.3077800.

Lundberg, S. M. and S.-I. Lee (2017). 'A Unified Approach to Interpreting Model Predictions'. In: *Advances in neural information processing systems* 30, pp. 4768–4777. DOI: 10.5555/3295222.3295230.

Luo, H., J. Zhang, X. Liu, L. Zhang and J. Liu (2024). 'Large-Scale 3D Reconstruction from Multi-View Imagery: A Comprehensive Review'. In: *Remote Sensing* 16.5, p. 773. DOI: 10.3390/rs16050773.

Luusua, A., J. Ylipulli, M. Foth and A. Aurigi (2023). 'Urban AI: Understanding the Emerging Role of Artificial Intelligence in Smart Cities'. In: *AI & society* 38.3, pp. 1039–1044. DOI: 10.1007/s00146-022-01537-5.

Luxen, D. and C. Vetter (2011). 'Real-Time Routing with OpenStreetMap Data'. In: *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. Gis '11. Chicago, Illinois: ACM, pp. 513–516. ISBN: 978-1-4503-1031-4. DOI: 10.1145/2093973.2094062.

Lynch, K. (1964). *The Image of the City*. Harvard-MIT Joint Center for Urban Studies Series. Cambridge, Massachusetts, United States: MIT Press. ISBN: 978-0-262-62001-7.

Lynch, K. (2023). 'The Image of the City (1960)'. In: *Anthologie Zum Städtebau. Band III: Vom Wiederaufbau Nach Dem Zweiten Weltkrieg Bis Zur Zeitgenössischen Stadt*. Gebr. Mann Verlag, pp. 481–488. DOI: 10.5771/9783786175247-481.

Maguire, E. A., R. Nannery and H. J. Spiers (2006). 'Navigation around London by a Taxi Driver with Bilateral Hippocampal Lesions'. In: *Brain* 129.11, pp. 2894–2907. DOI: 10.1093/brain/awl286.

Mahbub, U. and M. A. R. Ahad (2022). 'Advances in Human Action, Activity and Gesture Recognition'. In: *Pattern Recognition Letters* 155, pp. 186–190. DOI: 10.1016/j.patrec.2021.11.003.

Mai, G., W. Huang, J. Sun, S. Song, D. Mishra, N. Liu, S. Gao, T. Liu, G. Cong, Y. Hu et al. (2023). 'On the Opportunities and Challenges of Foundation Models for Geospatial Artificial Intelligence'. Preprint. arXiv: 2304.06798.

Mai, G., Y. Xie, X. Jia, N. Lao, J. Rao, Q. Zhu, Z. Liu, Y.-Y. Chiang and J. Jiao (2025). 'Towards the next Generation of Geospatial Artificial Intelligence'. In: *International Journal of Applied Earth Observation and Geoinformation* 136, p. 104368. DOI: 10.1016/j.jag.2025.104368.

Malajner, M., P. Planinšič and D. Gleich (2015). 'UWB Ranging Accuracy'. In: *2015 International Conference on Systems, Signals and Image Processing (IWSSIP)*. IEEE, pp. 61–64. DOI: 10.1109/IWSSIP.2015.7314177.

Manandhar, P., H. Rafiq and E. Rodriguez-Ubinas (2023). 'Current Status, Challenges, and Prospects of Data-Driven Urban Energy Modeling: A Review of Machine Learning Methods'. In: *Energy Reports* 9, pp. 2757–2776. DOI: 10.1016/j.egyr.2023.01.094.

Mandal, A., S. Leavy and S. Little (2021). 'Dataset Diversity: Measuring and Mitigating Geographical Bias in Image Search and Retrieval'. In: *Proceedings of the 1st International Workshop on Trustworthy AI for Multimedia Computing*, pp. 19–25. DOI: 10.1145/3475731.3484956.

Manley, E. and A. Dennett (2019). 'New Forms of Data for Understanding Urban Activity in Developing Countries'. In: *Applied Spatial Analysis and Policy* 12.1, pp. 45–70. DOI: 10.1007/s12061-018-9264-8.

Mapillary (2025). *Mapillary*. URL: https://www.mapillary.com/ (visited on 31/12/2025).

Marano, S., W. M. Gifford, H. Wymeersch and M. Z. Win (2010). 'NLOS Identification and Mitigation for Localization Based on UWB Experimental Data'. In: *IEEE Journal on selected areas in communications* 28.7, pp. 1026–1035. DOI: 10.1109/JSAC.2010.100907.

Marasinghe, R., T. Yigitcanlar, S. Mayere, T. Washington and M. Limb (2024). 'Computer Vision Applications for Urban Planning: A Systematic Review of Opportunities and Constraints'. In: *Sustainable Cities and Society* 100, p. 105047. DOI: 10.1016/j.scs.2023.105047.

Marques, L. and J. Roca (2021). 'Three-Dimensional Modelling for Cultural Heritage'. In: *Methods and Applications of Geospatial Technology in Sustainable Urbanism*. IGI Global, pp. 418–443.

Martí, P., L. Serrano-Estrada and A. Nolasco-Cirugeda (2019). 'Social Media Data: Challenges, Opportunities and Limitations in Urban Studies'. In: *Computers, Environment and Urban Systems* 74, pp. 161–174. DOI: 10.1016/j.compenvurbsys.2018.11.001.

Masone, C. and B. Caputo (2021). 'A Survey on Deep Visual Place Recognition'. In: *IEEE Access* 9, pp. 19516–19547. DOI: 10.1109/ACCESS.2021.3054937.

Mayor of London (2026). *Tall Buildings Statement*. URL: https://www.london.gov.uk/sites/default/files/52._tall_buidings_statement_2018.pdf (visited on 13/01/2026).

Mazhar, F., M. G. Khan and B. Sällberg (2017). 'Precise Indoor Positioning Using UWB: A Review of Methods, Algorithms and Implementations'. In: *Wireless Personal Communications* 97.3, pp. 4467–4491. DOI: 10.1007/s11277-017-4734-x.

McDonald, G. C. (2009). 'Ridge Regression'. In: *Wiley Interdisciplinary Reviews: Computational Statistics* 1.1, pp. 93–100. DOI: 10.1002/wics.14.

Mekruksavanich, S., N. Hnoohom and A. Jitpattanakul (2018). 'Smartwatch-based sitting detection with human activity recognition for office workers syndrome'. In: *2018 International ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI-NCON)*, pp. 160–164. DOI: 10.1109/ECTI-NCON.2018.8378302.

Meng, X.-L. (2018). 'Statistical Paradises and Paradoxes in Big Data (I): Law of Large Populations, Big Data Paradox, and the 2016 US Presidential Election'. In: *The Annals of Applied Statistics* 12.2, pp. 685–726. DOI: 10.1214/18-AOAS1161SF.

Merenda, M., C. Porcaro and D. Iero (2020). 'Edge Machine Learning for AI-enabled IoT Devices: A Review'. In: *Sensors* 20.9, p. 2533. DOI: 10.3390/s20092533.

Michalina, D., P. Mederly, H. Diefenbacher and B. Held (2021). 'Sustainable Urban Development: A Review of Urban Sustainability Indicator Frameworks'. In: *Sustainability* 13.16, p. 9348. DOI: 10.3390/su13169348.

Milioto, A., I. Vizzo, J. Behley and C. Stachniss (2019). 'RangeNet++: Fast and Accurate LiDAR Semantic Segmentation'. In: *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 4213–4220. DOI: 10.1109/IROS40897.2019.8967762.

Mill, T., A. Alt and R. Liias (2013). 'Combined 3D Building Surveying Techniques–Terrestrial Laser Scanning (TLS) and Total Station Surveying for BIM Data Management Purposes'. In: *Journal of Civil Engineering and Management* 19 (sup1), S23–S32. DOI: 10.3846/13923730.2013.795187.

Milojevic-Dupont, N., F. Wagner, F. Nachtigall, J. Hu, G. B. Brüser, M. Zumwald, F. Biljecki, N. Heeren, L. H. Kaack, P.-P. Pichler et al. (2023). 'EUBUCCO v0.1: European Building Stock Characteristics in a Common and Open Database for 200+ Million Individual Buildings'. In: *Scientific Data* 10.1, p. 147. DOI: 10.1038/s41597-023-02040-2.

Minoli, D. and B. Occhiogrosso (2018). 'Ultrawideband (UWB) Technology for Smart Cities IoT Applications'. In: *2018 IEEE International Smart Cities Conference (ISC2)*. IEEE, pp. 1–8. DOI: 10.1109/ISC2.2018.8656958.

Miriam Daniel, Google Maps (2025). *20 Things You Didn't Know You Could Do with Google Maps*. URL: https://blog.google/products/maps/20-years-google-maps-20-features/? (visited on 01/09/2025).

Moghaddam, M. G., A. A. N. Shirehjini and S. Shirmohammadi (2025). 'Device-Free Human Activity Recognition: A Systematic Literature Review'. In: *IEEE Open Journal of Instrumentation and Measurement* 4, pp. 1–34. DOI: 10.1109/ojim.2024.3502885.

Mohtadifar, M., M. Cheffena and A. Pourafzal (2022). 'Acoustic-and Radio-Frequency-Based Human Activity Recognition'. In: *Sensors* 22.9, p. 3125. DOI: 10.3390/s22093125.

Molisch, A. F. (2009). ‘Ultra-Wide-Band Propagation Channels’. In: *Proceedings of the IEEE* 97.2, pp. 353–371. DOI: 10.1109/JPROC.2008.2008836.

Montello, D. and C. Sas (Mar. 2006). ‘Human Factors of Wayfinding in Navigation’. In: *International Encyclopedia of Ergonomics and Human Factors, Second Edition - 3 Volume Set.* Ed. by W. Karwowski. CRC Press, pp. 2051–2056. ISBN: 978-0-415-30430-6. DOI: 10.1201/9780849375477.ch394.

Montello, D. R. (2015). ‘Spatial Cognition’. In: *International Encyclopedia of the Social & Behavioral Sciences (Second Edition).* Elsevier, pp. 111–115.

Morgan, M. and A. Habib (2002). ‘Interpolation of LiDAR Data and Automatic Building Extraction’. In: *ACSM-ASPRS Annual Conference Proceedings.* Citeseer, pp. 432–441.

Moshiri, P. F., R. Shahbazian, M. Nabati and S. A. Ghorashi (2021). ‘A CSI-based Human Activity Recognition Using Deep Learning’. In: *Sensors* 21.21, p. 7225. DOI: 10.3390/s21217225.

Muhmad Kamarulzaman, A. M., W. S. Wan Mohd Jaafar, M. N. Mohd Said, S. N. M. Saad and M. Mohan (2023). ‘UAV Implementations in Urban Planning and Related Sectors of Rapidly Developing Nations: A Review and Future Perspectives for Malaysia’. In: *Remote Sensing* 15.11, p. 2845. DOI: 10.3390/rs15112845.

Munir, F., S. Azam, A. M. Sheri, Y. Ko and M. Jeon (2019). ‘Where Am I: Localization and 3D Maps for Autonomous Vehicles.’ In: *VEHITS* 2019, pp. 452–457. DOI: 10.5220/0007718404520457.

Münzner, S., P. Schmidt, A. Reiss, M. Hanselmann, R. Stiefelhagen and R. Dürichen (2017). ‘CNN-based Sensor Fusion Techniques for Multimodal Human Activity Recognition’. In: *Proceedings of the 2017 ACM International Symposium on Wearable Computers*, pp. 158–165. DOI: 10.1145/3123021.3123046.

Mutegeki, R. and D. S. Han (2020). ‘A CNN-LSTM Approach to Human Activity Recognition’. In: *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC).* IEEE, pp. 362–366. DOI: 10.1109/ICAIIC48513.2020.9065078.

Myers, G. (2021). ‘Urbanisation in the Global South’. In: *Urban Ecology in the Global South.* Springer, pp. 27–49.

Nadia, A., S. Lyazid, K. Okba and C. Abdelghani (2023). ‘A CNN-MLP Deep Model for Sensor-Based Human Activity Recognition’. In: *2023 15th International Conference on Innovations in Information Technology (IIT).* IEEE, pp. 121–126. DOI: 10.1109/IIT59782.2023.10366481.

Naik, N., J. Philipoom, R. Raskar and C. Hidalgo (2014). ‘Streetscore-Predicting the Perceived Safety of One Million Streetscapes’. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 779–785. DOI: 10.1109/CVPRW.2014.121.

Neuhold, G., T. Ollmann, S. Rota Bulo and P. Kontschieder (2017). 'The Mapillary Vistas Dataset for Semantic Understanding of Street Scenes'. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 4990–4999. DOI: 10.1109/iccv.2017.534.

Neupane, B., T. Horanont and J. Aryal (2021). 'Deep Learning-Based Semantic Segmentation of Urban Features in Satellite Images: A Review and Meta-Analysis'. In: *Remote Sensing* 13.4, p. 808. DOI: 10.3390/rs13040808.

Ng, H.-F. and L.-T. Hsu (2021). '3D Mapping Database-Aided GNSS RTK and Its Assessments in Urban Canyons'. In: *IEEE Transactions on Aerospace and Electronic Systems* 57.5, pp. 3150–3166. DOI: 10.1109/TAES.2021.3069271.

Niu, H. and E. A. Silva (2020). 'Crowdsourced Data Mining for Urban Activity: Review of Data Sources, Applications, and Methods'. In: *Journal of Urban Planning and Development* 146.2, p. 04020007. DOI: 10.1061/(ASCE)UP.1943-5444.0000566.

O'Neill, M. J. (1991). 'Evaluation of a Conceptual Model of Architectural Legibility'. In: *Environment and Behavior* 23.3, pp. 259–284. DOI: 10.1177/0013916591233001.

O'Neill, M. J. (1992). 'Effects of Familiarity and Plan Complexity on Wayfinding in Simulated Buildings'. In: *Journal of Environmental Psychology* 12.4, pp. 319–327. ISSN: 0272-4944. DOI: 10.1016/s0272-4944(05)80080-5.

O'brien, R. M. (2007). 'A Caution Regarding Rules of Thumb for Variance Inflation Factors'. In: *Quality & Quantity* 41, pp. 673–690. DOI: 10.1007/s11135-006-9018-6.

Ofcom (2020). *Connected Nations 2023*. Ofcom London.

Ofcom (2024). *Connected Nations Update: Spring 2024*. URL: https://www.ofcom.org.uk/phones-and-broadband/coverage-and-speeds/connected-nations-update-spring-2024/ (visited on 31/12/2025).

OpenAI (2025). *CLIP*. URL: https://github.com/openai/CLIP (visited on 31/12/2025).

OpenCellID (2008). *OpenCellID Homepage*. URL: https://opencellid.org/ (visited on 31/12/2025).

OpenStreetMap (2009). *Simple 3D Buildings*. URL: https://wiki.openstreetmap.org/wiki/Simple_3D_Buildings (visited on 31/12/2025).

OpenStreetMap (2021). *SPARQL_examples*. <a href="{https://wiki.openstreetmap.org/wiki/

OpenStreetMap (2022). *London Borough Boundaries*. URL: https://wiki.openstreetmap.org/wiki/London_borough_boundaries (visited on 31/12/2025).

OpenStreetMap contributors (2017). *Planet Dump Retrieved from https://planet.osm.org*. URL: https://www.openstreetmap.org (visited on 31/12/2025).

Oughton, E. J. and Z. Frias (2018). 'The Cost, Coverage and Rollout Implications of 5G Infrastructure in Britain'. In: *Telecommunications Policy* 42.8, pp. 636–652. DOI: 10.1016/j.telpol.2017.07.009.

Over, M., A. Schilling, S. Neubauer and A. Zipf (2010). 'Generating Web-Based 3D City Models from OpenStreetMap: The Current Situation in Germany'. In: *Computers, Environment and urban systems* 34.6, pp. 496–507. DOI: 10.1016/j.compenvurbsys.2010.05.001.

Paiva, S., M. A. Ahad, G. Tripathi, N. Feroz and G. Casalino (2021). 'Enabling Technologies for Urban Smart Mobility: Recent Trends, Opportunities and Challenges'. In: *Sensors* 21.6, p. 2143. DOI: 10.3390/s21062143.

Pan, H. (2020). 'Cloud Removal for Remote Sensing Imagery via Spatial Attention Generative Adversarial Network'. Preprint. arXiv: 2009.13015.

Pan, Y., Y. Tian, X. Liu, D. Gu and G. Hua (2016). 'Urban Big Data and the Development of City Intelligence'. In: *Engineering* 2.2, pp. 171–178. DOI: 10.1016/J.ENG.2016.02.003.

Papadimitratos, P. and A. Jovanovic (2008). 'Protection and Fundamental Vulnerability of GNSS'. In: *2008 IEEE International Workshop on Satellite and Space Communications*. IEEE, pp. 167–171. DOI: 10.1109/IWSSC.2008.4656777.

Pareek, P. and A. Thakkar (2021). 'A Survey on Video-Based Human Action Recognition: Recent Updates, Datasets, Challenges, and Applications'. In: *Artificial Intelligence Review* 54.3, pp. 2259–2322. DOI: 10.1007/s10462-020-09904-8.

Park, Y. and J.-M. Guldmann (2019). 'Creating 3D City Models with Building Footprints and LiDAR Point Cloud Classification: A Machine Learning Approach'. In: *Computers, Environment and Urban Systems* 75, pp. 76–89. DOI: 10.1016/j.compenvurbsys.2019.01.004.

Peng, B., W. Jing, Z. Zheng, X. Wen, Z. Lu and Z. Wang (2022). '3D City Map Reconstruction from LEO Communication Satellite SNR Measurements'. In: *2022 IEEE/CIC International Conference on Communications in China (ICCC)*. IEEE, pp. 268–273. DOI: 10.1109/ICCC55456.2022.9880764.

Peng, Y. and S. Guo (2020). 'Detailed Feature Representation and Analysis of Low Frequency UWB Radar Range Profile for Improving Through-wall Human Activity Recognition'. In: *2020 IEEE Radar Conference (RadarConf20)*. IEEE, pp. 1–6. DOI: 10.1109/radarconf2043947.2020.9266629.

Piccardi, L., M. Risetti and R. Nori (2011). 'Familiarity and Environmental Representations of a City: A Self-Report Study'. In: *Psychological Reports* 109.1, pp. 309–326. DOI: 10.2466/01.13.17.PR0.109.4.309-326.

Pienaar, S. W. and R. Malekian (2019). 'Human Activity Recognition Using LSTM-RNN Deep Neural Network Architecture'. In: *2019 IEEE 2nd Wireless Africa Conference (WAC)*. IEEE, pp. 1–5. DOI: 10.1109/africa.2019.8843403.

Prandi, C., B. R. Barricelli, S. Mirri and D. Fogli (2023). 'Accessible Wayfinding and Navigation: A Systematic Mapping Study'. In: *Universal Access in the Information Society* 22.1, pp. 185–212. DOI: 10.1007/s10209-021-00843-x.

Psiaki, M. L. and T. E. Humphreys (2016). 'GNSS Spoofing and Detection'. In: *Proceedings of the IEEE* 104.6, pp. 1258–1270. DOI: 10.1109/JPROC.2016.2526658.

Psyllidis, A. (2020). 'Sensing the City through New Forms of Urban Data'. In: *Seeing the City: Interdisciplinary Perspectives on the Study of the Urban*, pp. 56–69. DOI: 10.2307/j.ctv1b741xh.7.

Qin, R. and T. Liu (2022). 'A Review of Landcover Classification with Very-High Resolution Remotely Sensed Optical Images—Analysis Unit, Model Scalability and Transferability'. In: *Remote Sensing* 14.3, p. 646. DOI: 10.3390/rs14030646.

Qorvo (2025). *DW1000 User Manuel.* URL: https://www.qorvo.com/products/d/da007967 (visited on 31/12/2025).

Qorvo (n.d.). *EVK1000 User Manuel.* URL: https://www.qorvo.com/products/d/da007993.

Radwin, D. (2009). 'High Response Rates Don't Ensure Survey Accuracy'. In: *Chronicle of Higher Education* 56.7, B8–B9.

Ramanujam, E., T. Perumal and S. J. I. S. J. Padmavathi (2021). 'Human Activity Recognition with Smartphone and Wearable Sensors Using Deep Learning Techniques: A Review'. In: *IEEE Sensors Journal* 21.12, pp. 13029–13040. DOI: 10.1109/JSEN.2021.3069927.

Ramírez-Moreno, M. A., S. Keshtkar, D. A. Padilla-Reyes, E. Ramos-López, M. García-Martínez, M. C. Hernández-Luna, A. E. Mogro, J. Mahlknecht, J. I. Huertas, R. E. Peimbert-García et al. (2021). 'Sensors for Sustainable Smart Cities: A Review'. In: *Applied Sciences* 11.17, p. 8198. DOI: 10.3390/app11178198.

Rao, J. N. and I. Molina (2015). *Small Area Estimation.* Hoboken, NJ, USA: John Wiley & Sons.

Rapoport, A. (2013). *Human Aspects of Urban Form: Towards a Man—Environment Approach to Urban Form and Design.* Reprint of the 1977 edition. Amsterdam: Elsevier.

Reades, J., F. Calabrese, A. Sevtsuk and C. Ratti (2007). 'Cellular Census: Explorations in Urban Data Collection'. In: *IEEE Pervasive Computing* 6.3, pp. 30–38. DOI: 10.1109/MPRV.2007.53.

Regmi, K. and A. Borji (2018). 'Cross-View Image Synthesis Using Conditional GANs'. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3501–3510. DOI: 10.1109/cvpr.2018.00369.

Rehan, H. (2023). 'Internet of Things (IoT) in Smart Cities: Enhancing Urban Living through Technology'. In: *Journal of Engineering and Technology* 5.1, pp. 1–16.

Resch, B. and M. Szell (11th Dec. 2019). 'Human-Centric Data Science for Urban Studies'. In: *ISPRS International Journal of Geo-Information* 8.12. DOI: 10.3390/ijgi8120584.

Reuther, A., P. Michaleas, M. Jones, V. Gadepally, S. Samsi and J. Kepner (2021). 'AI Accelerator Survey and Trends'. In: *2021 IEEE High Performance Extreme Computing Conference (HPEC)*. IEEE, pp. 1–9. DOI: 10.1109/HPEC49654.2021.9622867.

robolyst (2025). *Streetview*. URL: https://github.com/robolyst/streetview (visited on 31/12/2025).

Rodrigues, R. and M. Tani (2023). 'Semgeo: Semantic Keywords for Cross-View Image Geo-Localization'. In: *2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, pp. 1–5. DOI: 10.1109/icassp49357.2023.10094763.

Rohil, M. K. and Y. Ashok (2022). 'Visualization of Urban Development 3D Layout Plans with Augmented Reality'. In: *Results in Engineering* 14, p. 100447. DOI: 10.1016/j.rineng.2022.100447.

Roser, M. (2023). 'Technology over the Long Run: Zoom out to See How Dramatically the World Can Change within a Lifetime'. In: *Our World in Data*.

Rulff, J., G. Pereira, M. Hosseini, M. Lage and C. Silva (2024). 'Towards Data-Informed Interventions: Opportunities and Challenges of Street-Level Multimodal Sensing'. Preprint. arXiv: 2410.22092.

Rustam, F., A. A. Reshi, I. Ashraf, A. Mehmood, S. Ullah, D. M. Khan and G. S. Choi (2020). 'Sensor-Based Human Activity Recognition Using Deep Stacked Multilayered Perceptron Model'. In: *IEEE Access* 8, pp. 218898–218910. DOI: 10.1109/ACCESS.2020.3041822.

Sadalla, E. K. and D. R. Montello (1989). 'Remembering Changes in Direction'. In: *Environment and Behavior* 21.3, pp. 346–363. DOI: 10.1177/0013916589213006.

Saleh, S., A. Elmezayen, Q. Bader, M. Elhabiby and A. Noureldin (2022). 'Would Future mmWave Wireless Networks Be an Alternative Positioning Technique to GNSS-based High Precision Positioning?' In: *2022 IEEE 95th Vehicular Technology Conference:(VTC2022-spring)*. IEEE, pp. 1–5.

Salih, S., A. Abdelmaboud, O. Husain, A. Motwakel, H. Elshafie, M. Sharif and M. Hamdan (2025). 'IoT in Urban Development: Insight into Smart City Applications, Case Studies, Challenges, and Future Prospects'. In: *PeerJ Computer Science* 11, e2816. DOI: 10.7717/peerj-cs.2816.

Sarsodia, T., U. R. Bhatt, R. Upadhyay and V. Bhat (2022). 'A Survey on Different Application Areas Based on RSS (Received Signal Strength) and Possible Hardware and Software Tools for the Collection of RSS'. In: *2022 IEEE International Conference on Current Development in Engineering and Technology (CCET)*. IEEE, pp. 1–6. DOI: 10.1109/CCET56606.2022.10080843.

Schläpfer, M., J. Lee and L. Bettencourt (2015). 'Urban Skylines: Building Heights and Shapes as Measures of City Size'. Preprint. arXiv: 1512.00946.

Schölkopf, B. and A. Smola (2005). 'Support Vector Machines and Kernel Algorithms'. In: *Encyclopedia of Biostatistics*. Wiley, pp. 5328–5335.

Schön, S., K.-N. Baasch, L. Icking, A. KarimiDoona, Q. Lin, F. Ruwisch, A. Schaper and J. Su (2022). 'Towards Integrity for GNSS-based Urban Navigation–Challenges and Lessons Learned'. In: *2022 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, pp. 1774–1781. DOI: 10.1109/IV51971.2022.9827402.

Senousi, A. M., J. Zhang, W. Shi and X. Liu (2021). 'A Proposed Framework for Identification of Indicators to Model High-Frequency Cities'. In: *ISPRS International Journal of Geo-Information* 10.5, p. 317. DOI: 10.3390/ijgi10050317.

Shah, M. and N. Sureja (2025). 'A Comprehensive Review of Bias in Deep Learning Models: Methods, Impacts, and Future Directions'. In: *Archives of Computational Methods in Engineering* 32.1, pp. 255–267. DOI: 10.1007/s11831-024-10134-2.

Shahroudy, A., J. Liu, T.-T. Ng and G. Wang (2016). 'NTU RGB+D: A Large Scale Dataset for 3D Human Activity Analysis'. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1010–1019. DOI: 10.1109/CVPR.2016.115.

Shahtahmassebi, A. R., C. Li, Y. Fan, Y. Wu, M. Gan, K. Wang, A. Malik, G. A. Blackburn et al. (2021). 'Remote Sensing of Urban Green Spaces: A Review'. In: *Urban Forestry & Urban Greening* 57, p. 126946. DOI: 10.1016/j.ufug.2020.126946.

Shapiro, S. S. and M. B. Wilk (1965). 'An Analysis of Variance Test for Normality (Complete Samples)'. In: *Biometrika* 52.3–4, pp. 591–611. DOI: 10.2307/2333709.

Sharma, S., H. Mohammadmoradi, M. Heydariaan and O. Gnawali (2019). 'Device-Free Activity Recognition Using Ultra-Wideband Radios'. In: *2019 International Conference on Computing, Networking and Communications (ICNC)*. IEEE, pp. 1029–1033. DOI: 10.1109/ICCNC.2019.8685504.

Shavit, Y. and I. Klein (2021). 'Boosting Inertial-Based Human Activity Recognition with Transformers'. In: *IEEE Access* 9, pp. 53540–53547. DOI: 10.1109/ACCESS.2021.3070646.

Shi, W. (2021). 'Introduction to Urban Sensing'. In: *Urban Informatics*. Springer, pp. 311–314.

Shi, Y. and H. Li (2022). 'Beyond Cross-View Image Retrieval: Highly Accurate Vehicle Localization Using Satellite Image'. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17010–17020. DOI: 10.1109/cvpr52688.2022.01650.

Shi, Y., L. Liu, X. Yu and H. Li (2019). 'Spatial-Aware Feature Aggregation for Image Based Cross-View Geo-Localization'. In: *Advances in Neural Information Processing Systems* 32, pp. 10090–10100.

Shi, Y., X. Yu, L. Liu, T. Zhang and H. Li (2020). 'Optimal Feature Transport for Cross-View Image Geo-Localization'. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 34. 07, pp. 11990–11997. DOI: 10.1609/aaai.v34i07.6875.

Shibasaki, R., S. Hori, S. Kawamura and S. Tani (2020). 'Integrating Urban Data with Urban Services'. In: *Society 5.0: A People-Centric Super-Smart Society*. Springer Singapore, pp. 67–83. DOI: 10.1007/978-981-15-2989-4_4. URL: https://doi.org/10.1007/978-981-15-2989-4_4.

Shin, D.-H. (2009). 'Ubiquitous City: Urban Technologies, Urban Infrastructure and Urban Informatics'. In: *Journal of Information Science* 35.5, pp. 515–526. DOI: 10.1177/0165551509100832.

Shirky, C. (2010). *Cognitive Surplus: How Technology Makes Consumers into Collaborators*. London, UK: Penguin.

Shugaev, M., I. Semenov, K. Ashley, M. Klaczynski, N. Cuntoor, M. W. Lee and N. Jacobs (Jan. 2024). 'ArcGeo: Localizing Limited Field-of-View Images Using Cross-View Matching'. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 209–218.

Silva, B. N., M. Khan and K. Han (2018). 'Towards Sustainable Smart Cities: A Review of Trends, Architectures, Components, and Open Challenges in Smart Cities'. In: *Sustainable Cities and Society* 38, pp. 697–713. DOI: 10.1016/j.scs.2018.01.053.

Simonyan, K. and A. Zisserman (2014). 'Two-Stream Convolutional Networks for Action Recognition in Videos'. In: *Advances in Neural Information Processing Systems* 27, pp. 568–576.

Simpson, E. H. (1949). 'Measurement of Diversity'. In: *nature* 163.4148, pp. 688–688. DOI: 10.1038/163688a0.

Skondras, A., E. Karachaliou, I. Tavantzis, N. Tokas, E. Valari, I. Skalidi, G. A. Bouvet and E. Stylianidis (2022). 'UAV Mapping and 3D Modeling as a Tool for Promotion and Management of the Urban Space'. In: *Drones* 6.5, p. 115. DOI: 10.3390/drones6050115.

Sogi, N., T. Shibata, M. Terao, K. Senzaki, M. Tani and R. Rodrigues (2024). 'Disaster Damage Visualization by VLM-based Interactive Image Retrieval and Cross-View Image Geo-Localization'. In: *IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, pp. 1746–1749. DOI: 10.1109/IGARSS53475.2024.10640928.

Sönmez, B. E. and D. E. Önder (2019). 'The Influence of GPS-based Navigation Systems on Perception and Image Formation: A Case Study in Urban Environments'. In: *Cities* 86, pp. 102–112. DOI: 10.1016/j.cities.2018.12.018.

Spiers, H. J., A. Coutrot and M. Hornberger (Jan. 2023). 'Explaining World-Wide Variation in Navigation Ability from Millions of People: Citizen Science Project Sea Hero Quest'. In: *Topics in Cognitive Science* 15.1, pp. 120–138. DOI: 10.1111/tops.12590.

Stanitsa, A., S. H. Hallett and S. Jude (2023). 'Investigating Pedestrian Behaviour in Urban Environments: A Wi-Fi Tracking and Machine Learning Approach'. In: *Multimodal Transportation* 2.1, p. 100049. ISSN: 2772-5863. DOI: 10.1016/j.multra.2022.100049.

Steiniger, S., M. Neun and A. Edwardes (2006). 'Foundations of Location Based Services'. In: *Lecture Notes on LBS* 1.272, p. 2.

Stoter, J., H. De Kluijver and V. Kurakula (2008). '3D Noise Mapping in Urban Areas'. In: *International Journal of Geographical Information Science* 22.8, pp. 907–924. DOI: 10.1080/13658810701739039.

Straczkiewicz, M., P. James and J.-P. Onnela (2021). 'A Systematic Review of Smartphone-Based Human Activity Recognition Methods for Health Research'. In: *NPJ Digital Medicine* 4.1, p. 148. DOI: 10.1038/s41746-021-00514-4.

Stuchbury-Wass, J., A. Ferlini and C. Mascolo (2022). 'Multimodal Attention Networks for Human Activity Recognition from Earable Devices'. In: *Adjunct Proceedings of the 2022 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2022 ACM International Symposium on Wearable Computers*, pp. 258–260. DOI: 10.1145/3544793.3563422.

Sun, B., C. Chen, Y. Zhu and J. Jiang (2019). 'Geocapsnet: Ground to Aerial View Image Geo-Localization Using Capsule Network'. In: *2019 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, pp. 742–747. DOI: 10.1109/ICME.2019.00133.

Sun, S., A. A. Folarin, Y. Ranjan, Z. Rashid, P. Conde, C. Stewart, N. Cummins, F. Matcham, G. Dalla Costa, S. Simblett et al. (2020). 'Using Smartphones and Wearable Devices to Monitor Behavioral Changes during COVID-19'. In: *Journal of medical Internet research* 22.9, e19992. DOI: 10.2196/19992.

Sun, W. and R. Wang (2018). 'Fully Convolutional Networks for Semantic Segmentation of Very High Resolution Remotely Sensed Images Combined with DSM'. In: *IEEE Geoscience and Remote Sensing Letters* 15.3, pp. 474–478. DOI: 10.1109/LGRS.2018.2795531.

Sung, S., H. Kim and J.-I. Jung (2023). 'Accurate Indoor Positioning for UWB-based Personal Devices Using Deep Learning'. In: *IEEE Access* 11, pp. 20095–20113. DOI: 10.1109/ACCESS.2023.3250180.

Taylor, W., S. A. Shah, K. Dashtipour, A. Zahid, Q. H. Abbasi and M. A. Imran (2020). 'An Intelligent Non-Invasive Real-Time Human Activity Recognition System for next-Generation Healthcare'. In: *Sensors* 20.9, p. 2653.

Thi, T. H., J. Zhang, L. Cheng, L. Wang and S. Satoh (2010). 'Human Action Recognition and Localization in Video Using Structured Learning of Local Space-Time Features'. In: *2010 7th IEEE International Conference on Advanced Video and Signal Based Surveillance*. IEEE, pp. 204–211. DOI: 10.1109/AVSS.2010.76.

Tommasi, T., N. Patricia, B. Caputo and T. Tuytelaars (2017). 'A Deeper Look at Dataset Bias'. In: *Domain Adaptation in Computer Vision Applications*. Springer, pp. 37–55.

Torralba, A. and A. A. Efros (2011). 'Unbiased Look at Dataset Bias'. In: *Cvpr 2011*. IEEE, pp. 1521–1528. DOI: `10.1109/CVPR.2011.5995347`.

Toschi, I., E. Nocerino, F. Remondino, A. Revolti, G. Soria and S. Piffer (2017). 'Geospatial Data Processing for 3D City Model Generation, Management and Visualization'. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 42, pp. 527–534. DOI: `10.5194/isprs-archives-XLII-1-W1-527-2017`.

Toth, C. and G. Jóźków (2016). 'Remote Sensing Platforms and Sensors: A Survey'. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 115, pp. 22–36. DOI: `10.1016/j.isprsjprs.2015.10.004`.

Townsend, A. (2015). 'Cities of Data: Examining the New Urban Science'. In: *Public Culture* 27.2, pp. 201–212. DOI: `10.1215/08992363-2841808`.

Tran, D., L. Bourdev, R. Fergus, L. Torresani and M. Paluri (2015). 'Learning Spatiotemporal Features with 3D Convolutional Networks'. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 4489–4497. DOI: `10.1109/iccv.2015.510`.

Transport for London (2018). *Strategic Walking Analysis: TfL City Planning Strategic Analysis*. URL: `https://content.tfl.gov.uk/strategic-walking-analysis.pdf` (visited on 31/12/2025).

Turner, A. (2007). 'From Axial to Road-Centre Lines: A New Representation for Space Syntax and a New Model of Route Choice for Transport Network Analysis'. In: *Environment and Planning B: Planning and Design* 34.3, pp. 539–555. DOI: `10.1068/b32067`.

Tyagi, N., J. Singh and S. Singh (2022). 'A Review of Routing Algorithms for Intelligent Route Planning and Path Optimization in Road Navigation'. In: *Recent Trends in Product Design and Intelligent Manufacturing Systems: Select Proceedings of IPDIMS 2021*, pp. 851–860. DOI: `10.1007/978-981-19-4606-6_78`.

Tyagi, N., J. Singh, S. Singh and S. S. Sehra (2024). 'A 3D Model-Based Framework for Real-Time Emergency Evacuation Using GIS and IoT Devices.' In: *ISPRS International Journal of Geo-Information* 13.12, p. 445. DOI: `10.3390/ijgi13120445`.

U.S. Department of Defense (2020). *Global Positioning System Standard Positioning Service Performance Standard, 5th Edition*. Performance Standard. U.S. Government. URL: `https://www.gps.gov/sites/default/files/2025-07/2020-SPS-performance-standard.pdf`.

United Nations (2022). *World Cities Report 2022: Envisaging the Future of Cities*. URL: `https://unhabitat.org/world-cities-report-2022-envisaging-the-future-of-cities` (visited on 31/12/2025).

United Nations (2024). *World Cities Report 2024 - Cities and Climate Action*. URL: https://unhabitat.org/wcr/ (visited on 31/12/2025).

United Nations Economic Commission for Europe, Conference of European Statisticians Bureau (2024). *In-Depth Review of Timeliness, Frequency and Granularity of Official Statistics*. URL: https://unece.org/sites/default/files/2024-05/ECE_CES_2024_06_E.PDF (visited on 01/09/2025).

Usman, M., M. R. Asghar, I. S. Ansari, F. Granelli and K. A. Qaraqe (2018). 'Technologies and Solutions for Location-Based Services in Smart Cities: Past, Present, and Future'. In: *IEEE Access* 6, pp. 22240–22248. DOI: 10.1109/ACCESS.2018.2826041.

Vaez, S., M. Burke and T. Alizadeh (2016). 'Urban Form and Wayfinding: Review of Cognitive and Spatial Kknowledge for iIndividuals' Navigation'. In: *ATRF 2016*.

Vales, V. B., T. Domínguez-Bolaño, C. J. Escudero and J. A. Garcia-Naya (2020). 'Using the Power Delay Profile to Accelerate the Training of Neural Network-Based Classifiers for the Identification of LOS and NLOS UWB Propagation Conditions'. In: *IEEE Access* 8, pp. 220205–220214. DOI: 10.1109/ACCESS.2020.3043503.

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser and I. Polosukhin (2017). 'Attention Is All You Need'. In: *Advances in Neural Information Processing Systems* 30, pp. 6000–6010.

Vidya, B. and P. Sasikumar (2022). 'Wearable Multi-Sensor Data Fusion Approach for Human Activity Recognition Using Machine Learning Algorithms'. In: *Sensors and Actuators A: Physical* 341, p. 113557. DOI: 10.1016/j.sna.2022.113557.

Vinutha, H. P., B. Poornima and B. M. Sagar (2018). 'Detection of Outliers Using Interquartile Range Technique from Intrusion Dataset'. In: *Information and Decision Sciences: Proceedings of the 6th International Conference on FICTA*. Springer, pp. 511–518. DOI: 10.1007/978-981-10-7563-6_53.

Vo, N. N. and J. Hays (2016). 'Localizing and Orienting Street Views Using Overhead Imagery'. In: *European Conference on Computer Vision*. Springer, pp. 494–509. DOI: 10.1007/978-3-319-46448-0_30.

Vrigkas, M., C. Nikou and I. A. Kakadiaris (2015). 'A Review of Human Activity Recognition Methods'. In: *Frontiers in Robotics and AI* 2, p. 28. DOI: 10.3389/frobt.2015.00028.

Wagenaar, W. A. (1986). 'My Memory: A Study of Autobiographical Memory over Six Years'. In: *Cognitive Psychology* 18.2, pp. 225–252. DOI: 10.1016/0010-0285(86)90013-7.

Walkowiak, S., A. Coutrot, M. Hegarty, P. F. Velasco, J. M. Wiener, R. C. Dalton, C. Hölscher, M. Hornberger, H. J. Spiers and E. Manley (5th July 2023a). 'Cultural Determinants of the Gap between Self-Estimated Navigation Ability and Wayfinding Performance: Evidence from 46 Countries'. In: *Scientific Reports* 13.1, p. 10844. DOI: 10.1038/s41598-023-30937-w.

Walkowiak, S., A. Coutrot, M. Hegarty, P. F. Velasco, J. M. Wiener, R. C. Dalton, C. Hölscher, M. Hornberger, H. J. Spiers and E. Manley (2023b). 'Cultural Determinants of the Gap between Self-Estimated Navigation Ability and Wayfinding Performance: Evidence from 46 Countries'. In: *Scientific Reports* 13.1, pp. 10844–10844. ISSN: 2045-2322. DOI: `10.1038/s41598-023-30937-w`.

Wan, S., L. Qi, X. Xu, C. Tong and Z. Gu (2020). 'Deep Learning Models for Real-Time Human Activity Recognition with Smartphones'. In: *mobile networks and applications* 25.2, pp. 743–755. DOI: `10.1007/s11036-019-01445-x`.

Wang, D., J. Yang, W. Cui, L. Xie and S. Sun (2022a). 'CAUTION: A Robust WiFi-Based Human Authentication System via Few-Shot Open-Set Recognition'. In: *IEEE Internet of Things Journal*, pp. 17323–17333. DOI: `10.1109/JIOT.2022.3156099`.

Wang, J., X. Zhang, Q. Gao, H. Yue and H. Wang (2016). 'Device-Free Wireless Localization and Activity Recognition: A Deep Learning Approach'. In: *IEEE Transactions on Vehicular Technology* 66.7, pp. 6258–6267. DOI: `10.1109/TVT.2016.2635161`.

Wang, L., P. D. Groves and M. K. Ziebart (2013a). 'GNSS Shadow Matching: Improving Urban Positioning Accuracy Using a 3D City Model with Optimized Visibility Scoring Scheme'. In: *NAVIGATION: Journal of the Institute of Navigation* 60.3, pp. 195–207. DOI: `10.1002/navi.38`.

Wang, L., P. D. Groves and M. K. Ziebart (2013b). 'Urban Positioning on a Smartphone: Real-Time Shadow Matching Using GNSS and 3D City Models'. In: *Proceedings of the 26th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS+ 2013)*. The Institute of Navigation, pp. 1606–1619.

Wang, L., X. Geng, X. Ma, F. Liu and Q. Yang (2018a). 'Cross-City Transfer Learning for Deep Spatio-Temporal Prediction'. Preprint. arXiv: `1802.00386`.

Wang, Q., H. Luo, J. Wang, L. Sun, Z. Ma, C. Zhang, M. Fu and F. Zhao (2022b). 'Recent Advances in Pedestrian Navigation Activity Recognition: A Review'. In: *IEEE Sensors Journal* 22.8, pp. 7499–7518. DOI: `10.1109/JSEN.2022.3153610`.

Wang, R., J. Peethambaran and D. Chen (2018b). 'LiDAR Point Clouds to 3-D Urban Models: A Review'. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 11.2, pp. 606–627. DOI: `10.1109/JSTARS.2017.2781132`.

Wang, R., Y. Liu, Y. Lu, J. Zhang, P. Liu, Y. Yao and G. Grekousis (2019a). 'Perceptions of Built Environment and Health Outcomes for Older Chinese in Beijing: A Big Data Approach with Street View Images and Deep Learning Technique'. In: *Computers, Environment and Urban Systems* 78, p. 101386. DOI: `10.1016/j.compenvurbsys.2019.101386`.

Wang, T., J. Zhao, M. Yatskar, K.-W. Chang and V. Ordonez (2019b). 'Balanced Datasets Are Not Enough: Estimating and Mitigating Gender Bias in Deep Image Representations'. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 5310–5319. DOI: `10.1109/iccv.2019.00541`.

Wang, W., A. X. Liu, M. Shahzad, K. Ling and S. Lu (2017a). 'Device-Free Human Activity Recognition Using Commercial WiFi Devices'. In: *IEEE Journal on Selected Areas in Communications* 35.5, pp. 1118–1131. DOI: 10.1109/JSAC.2017.2679658.

Wang, X. (2023). *Alternative Navigation Methods in GNSS Challenged Environments*. Columbus, Ohio, United States: The Ohio State University.

Wang, Y., Q. Chen, Q. Zhu, L. Liu, C. Li and D. Zheng (2019c). 'A Survey of Mobile Laser Scanning Applications and Key Techniques over Urban Areas'. In: *Remote Sensing* 11.13, p. 1540. DOI: 10.3390/rs11131540.

Wang, Y. and A. Basiri (2024). 'Advancing Human Activity Recognition Using Ultra-Wideband Channel Impulse Response Snapshots'. In: *2024 International Conference on Activity and Behavior Computing (ABC)*. IEEE, pp. 1–10. DOI: 10.1109/ABC61795.2024.10651886.

Wang, Y. and A. Basiri (2025). 'Bit to Brick: From Cellular Mobile Signals to 3D City Map Creation'. In: *Big Earth Data*, pp. 1–25. DOI: 10.1080/20964471.2025.2561319.

Wang, Y. and M. Li (2019). 'Urban Impervious Surface Detection from Remote Sensing Images: A Review of the Methods and Challenges'. In: *IEEE Geoscience and Remote Sensing Magazine* 7.3, pp. 64–93. DOI: 10.1109/MGRS.2019.2927260.

Wang, Z., K. Ito and F. Biljecki (2024). 'Assessing the Equity and Evolution of Urban Visual Perceptual Quality with Time Series Street View Imagery'. In: *Cities* 145, p. 104704. DOI: 10.1016/j.cities.2023.104704.

Wang, Z., H. Li and R. Rajagopal (2020). 'Urban2vec: Incorporating Street View Imagery and Pois for Multi-Modal Urban Neighborhood Embedding'. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 34. 01, pp. 1013–1020. DOI: 10.1609/aaai.v34i01.5450.

Wang, Z. and M. Menenti (2021). 'Challenges and Opportunities in LiDAR Remote Sensing'. In: *Frontiers in Remote Sensing* 2, p. 641723. DOI: 10.3389/frsen.2021.641723.

Wang, Z., W. Yan and T. Oates (2017b). 'Time Series Classification from Scratch with Deep Neural Networks: A Strong Baseline'. In: *2017 International Joint Conference on Neural Networks (IJCNN)*. IEEE, pp. 1578–1585. DOI: 10.1109/IJCNN.2017.7966039.

Warburg, F., S. Hauberg, M. Lopez-Antequera, P. Gargallo, Y. Kuang and J. Civera (2020). 'Mapillary Street-Level Sequences: A Dataset for Lifelong Place Recognition'. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2626–2635. DOI: 10.1109/cvpr42600.2020.00270.

Wazny, K. (2017). '"Crowdsourcing" Ten Years in: A Review'. In: *Journal of Global Health* 7.2, p. 020602. DOI: 10.7189/jogh.07.020601.

Webber, M. and R. F. Rojas (2021). 'Human Activity Recognition with Accelerometer and Gyroscope: A Data Fusion Approach'. In: *IEEE Sensors Journal* 21.15, pp. 16979–16989. DOI: 10.1109/JSEN.2021.3079883.

Wegener, M., F. Gnad and M. Vannahme (1986). 'The Time Scale of Urban Change'. In: *Advances in Urban Systems Modelling*. Ed. by B. Hutchinson and M. Batty. Amsterdam: North-Holland, pp. 175–197.

Wei, Y., Y. Zheng and Q. Yang (2016). 'Transfer Knowledge between Cities'. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1905–1914. DOI: 10.1145/2939672.2939830.

Wen, W. W. and L.-T. Hsu (2022). '3D LiDAR Aided GNSS NLOS Mitigation in Urban Canyons'. In: *IEEE Transactions on Intelligent Transportation Systems* 23.10, pp. 18224–18236. DOI: 10.1109/TITS.2022.3167710.

Weyand, T., A. Araujo, B. Cao and J. Sim (2020). 'Google Landmarks Dataset v2–a Large-Scale Benchmark for Instance-Level Recognition and Retrieval'. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2575–2584. DOI: 10.1109/cvpr42600.2020.00265.

Willenborg, B., M. Sindram and T. H. Kolbe (2017). 'Applications of 3D City Models for a Better Understanding of the Built Environment'. In: *Trends in spatial analysis and modelling: decision-support and planning strategies*, pp. 167–191. DOI: 10.1007/978-3-319-52522-8_9.

Wilson, D., X. Zhang, W. Sultani and S. Wshah (2021). 'Visual and Object Geo-Localization: A Comprehensive Survey'. Preprint. arXiv: 2112.15202.

Wong, K. K. Y. (2018). 'Towards a National 3D Mapping Product for Great Britain'. PhD thesis. UCL (University College London).

Workman, S., R. Souvenir and N. Jacobs (2015). 'Wide-Area Image Geolocalization with Aerial Reference Imagery'. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3961–3969. DOI: 10.1109/iccv.2015.451.

Wu, J., Y. Fu, D. Chen, H. Zhang, E. Xue, J. Shao, L. Tang, B. Zhao, C. Lai and Z. Ye (2023). 'Sedentary behavior patterns and the risk of non-communicable diseases and all-cause mortality: A systematic review and meta-analysis'. In: *International journal of nursing studies* 146, p. 104563. DOI: 10.1016/j.ijnurstu.2023.104563.

Xia, K., J. Huang and H. Wang (2020). 'LSTM-CNN Architecture for Human Activity Recognition'. In: *IEEE Access* 8, pp. 56855–56866. DOI: 10.1109/ACCESS.2020.2982225.

Xu, F., Q. Wang, E. Moro, L. Chen, A. Salazar Miranda, M. C. González, M. Tizzoni, C. Song, C. Ratti, L. Bettencourt et al. (2025). 'Using Human Mobility Data to Quantify Experienced Urban Inequalities'. In: *Nature Human Behaviour*, pp. 1–11. DOI: 10.1038/s41562-024-02079-0.

Xue, H., W. Jiang, C. Miao, F. Ma, S. Wang, Y. Yuan, S. Yao, A. Zhang and L. Su (2020). 'DeepMV: Multi-View Deep Learning for Device-Free Human Activity Recognition'. In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4.1, pp. 1–26. DOI: 10.1145/3380980.

Yan, L., R. Zhu, M.-P. Kwan, W. Luo, D. Wang, S. Zhang, M. S. Wong, L. You, B. Yang, B. Chen et al. (2023). 'Estimation of Urban-Scale Photovoltaic Potential: A Deep Learning-Based Approach for Constructing Three-Dimensional Building Models from Optical Remote Sensing Imagery'. In: *Sustainable Cities and Society* 93, p. 104515. DOI: 10.1016/j.scs.2023.104515.

Yan, S., Y. Xiong and D. Lin (2018). 'Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition'. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 32. 1, pp. 7444–7452. DOI: 10.1609/aaai.v32i1.12328.

Yan, X., Z. Jiang, P. Luo, H. Wu, A. Dong, F. Mao, Z. Wang, H. Liu and Y. Yao (2024). 'A Multimodal Data Fusion Model for Accurate and Interpretable Urban Land Use Mapping with Uncertainty Analysis'. In: *International Journal of Applied Earth Observation and Geoinformation* 129, p. 103805. DOI: 10.1016/j.jag.2024.103805.

Yan, Y. and B. Huang (2022). 'Estimation of Building Height Using a Single Street View Image via Deep Neural Networks'. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 192, pp. 83–98. DOI: 10.1016/j.isprsjprs.2022.08.006.

Yang, C. and A. Soloviev (2020). 'Mobile Positioning with Signals of Opportunity in Urban and Urban Canyon Environments'. In: *2020 IEEE/ION Position, Location and Navigation Symposium (PLANS)*. IEEE, pp. 1043–1059. DOI: 10.1109/PLANS46316.2020.9109876.

Yang, J., M. N. Nguyen, P. P. San, X. Li, S. Krishnaswamy et al. (2015). 'Deep Convolutional Neural Networks on Multichannel Time Series for Human Activity Recognition.' In: *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI 2015)*. Vol. 15. Buenos Aires, Argentina, pp. 3995–4001. DOI: 10.5555/2832747.2832806.

Yang, J., X. Chen, H. Zou, C. X. Lu, D. Wang, S. Sun and L. Xie (2023a). 'SenseFi: A Library and Benchmark on Deep-Learning-Empowered WiFi Human Sensing'. In: *Patterns* 4.3, p. 100703. DOI: 10.1016/j.patter.2023.100703.

Yang, R., W. Zhang, N. Tiwari, H. Yan, T. Li and H. Cheng (2022). 'Multimodal Sensors with Decoupled Sensing Mechanisms'. In: *Advanced Science* 9.26, p. 2202470. DOI: 10.1002/advs.202202470.

Yang, Y., A. Pentland and E. Moro (2023b). 'Identifying Latent Activity Behaviors and Lifestyles Using Mobility Data to Describe Urban Dynamics'. In: *EPJ Data Science* 12.1, p. 15. DOI: 10.1140/epjds/s13688-023-00390-w.

Yesiltepe, D., A. O. Torun, A. Coutrot, M. Hornberger, H. Spiers and R. C. Dalton (2021). 'Computer Models of Saliency Alone Fail to Predict Subjective Visual Attention to Landmarks during Observed Navigation'. In: *Spatial Cognition & Computation* 21.1, pp. 39–66. DOI: 10.1080/13875868.2020.1830993.

Yi, S., X. Li, W. Tu and T. Zhao (2025). 'Planning for Cooler Cities: A Multimodal AI Framework for Predicting and Mitigating Urban Heat Stress through Urban Landscape Transformation'. Preprint. arXiv: 2507.23000.

Yin, C., N. Peng, Y. Li, Y. Shi, S. Yang and P. Jia (2023). 'A Review on Street View Observations in Support of the Sustainable Development Goals'. In: *International Journal of Applied Earth Observation and Geoinformation* 117, p. 103205. DOI: 10.1016/j.jag.2023.103205.

Yin, J., J. Dong, N. A. Hamm, Z. Li, J. Wang, H. Xing and P. Fu (2021). 'Integrating Remote Sensing and Geospatial Big Data for Urban Land Use Mapping: A Review'. In: *International Journal of Applied Earth Observation and Geoinformation* 103, p. 102514. DOI: 10.1016/j.jag.2021.102514.

Yin, Y., L. Xie, Z. Jiang, F. Xiao, J. Cao and S. Lu (2024). 'A Systematic Review of Human Activity Recognition Based on Mobile Devices: Overview, Progress and Trends'. In: *IEEE Communications Surveys & Tutorials* 26.2, pp. 890–929. DOI: 10.1109/COMST.2024.3357591.

Ying, S., P. Van Oosterom and H. Fan (2023). 'New Techniques and Methods for Modelling, Visualization, and Analysis of a 3D City'. In: *Journal of Geovisualization and Spatial Analysis* 7.2, p. 26. DOI: 10.1007/s41651-023-00157-x.

Yoshida, T., K. Kano, K. Higashiura, K. Yamaguchi, K. Takigami, K. Urano, S. Aoki, T. Yonezawa and N. Kawaguchi (2022). 'A Data-Driven Approach for Online Pre-Impact Fall Detection with Wearable Devices'. In: *Sensor-and Video-Based Activity and Behavior Computing: Proceedings of 3rd International Conference on Activity and Behavior Computing (ABC 2021).* Springer, pp. 133–147. DOI: 10.1007/978-981-19-0361-8_8.

Yu, D. and C. Fang (2023). 'Urban Remote Sensing with Spatial Big Data: A Review and Renewed Perspective of Urban Studies in Recent Decades'. In: *Remote Sensing* 15.5, p. 1307. DOI: 10.3390/rs15051307.

Yu, H., S. Cang and Y. Wang (2016). 'A Review of Sensor Selection, Sensor Devices and Sensor Deployment for Wearable Sensor-Based Human Activity Recognition Systems'. In: *2016 10th International Conference on Software, Knowledge, Information Management & Applications (Skima).* IEEE, pp. 250–257. DOI: 10.1109/SKIMA.2016.7916228.

Yu, K., Y. Chen, D. Wang, Z. Chen, A. Gong and J. Li (2019). 'Study of the Seasonal Effect of Building Shadows on Urban Land Surface Temperatures Based on Remote Sensing Data'. In: *remote sensing* 11.5, p. 497. DOI: 10.3390/rs11050497.

Yu, X., J. Ma, Y. Tang, T. Yang and F. Jiang (2024). ‘Can We Trust Our Eyes? Interpreting the Misperception of Road Safety from Street View Images and Deep Learning’. In: *Accident Analysis & Prevention* 197, p. 107455. DOI: 10.1016/j.aap.2023.107455.

Yue, H. (2025). ‘Investigating Streetscape Environmental Characteristics Associated with Road Traffic Crashes Using Street View Imagery and Computer Vision’. In: *Accident Analysis & Prevention* 210, p. 107851. DOI: 10.1016/j.aap.2024.107851.

Yurtsever, E., J. Lambert, A. Carballo and K. Takeda (2020). ‘A Survey of Autonomous Driving: Common Practices and Emerging Technologies’. In: *IEEE Access* 8, pp. 58443–58469. DOI: 10.1109/ACCESS.2020.2983149.

Zafari, F., A. Gkelias and K. K. Leung (2019). ‘A Survey of Indoor Localization Systems and Technologies’. In: *IEEE Communications Surveys & Tutorials* 21.3, pp. 2568–2599. DOI: 10.1109/COMST.2019.2911558.

Zaman, M., N. Puryear, S. Abdelwahed and N. Zohrabi (2024). ‘A Review of IoT-based Smart City Development and Management’. In: *Smart Cities* 7.3, pp. 1462–1501. DOI: 10.3390/smartcities7030061.

Zanella, A., N. Bui, A. Castellani, L. Vangelista and M. Zorzi (2014). ‘Internet of Things for Smart Cities’. In: *IEEE Internet of Things journal* 1.1, pp. 22–32. DOI: 10.1109/JIOT.2014.2306328.

Zangenehnejad, F. and Y. Gao (2021). ‘GNSS Smartphones Positioning: Advances, Challenges, Opportunities, and Future Perspectives’. In: *Satellite navigation* 2.1, p. 24. DOI: 10.1186/s43020-021-00054-y.

Zeng, X., Y. Yu, S. Yang, Y. Lv and M. N. I. Sarker (2022). ‘Urban Resilience for Urban Sustainability: Concepts, Dimensions, and Perspectives’. In: *Sustainability* 14.5, p. 2481. DOI: 10.3390/su14052481.

Zhai, M., Z. Bessinger, S. Workman and N. Jacobs (2017). ‘Predicting Ground-Level Scene Layout from Aerial Imagery’. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 867–875. DOI: 10.1109/cvpr.2017.440.

Zhang, A., C. Xia and W. Li (2022a). ‘Exploring the Effects of 3D Urban Form on Urban Air Quality: Evidence from Fifteen Megacities in China’. In: *Sustainable Cities and Society* 78, p. 103649. DOI: 10.1016/j.scs.2021.103649.

Zhang, F., A. Salazar-Miranda, F. Duarte, L. Vale, G. Hack, M. Chen, Y. Liu, M. Batty and C. Ratti (2024). ‘Urban Visual Intelligence: Studying Cities with Artificial Intelligence and Street-Level Imagery’. In: *Annals of the American Association of Geographers* 114.5, pp. 876–897. DOI: 10.1080/24694452.2024.2313515.

Zhang, G., P. Xu, H. Xu and L.-T. Hsu (2021). ‘Prediction on the Urban GNSS Measurement Uncertainty Based on Deep Learning Networks with Long Short-Term Memory’. In: *IEEE Sensors Journal* 21.18, pp. 20563–20577. DOI: 10.1109/JSEN.2021.3098006.

Zhang, J. and X. Lin (2017). 'Advances in Fusion of Optical Imagery and LiDAR Point Cloud Applied to Photogrammetry and Remote Sensing'. In: *International Journal of Image and Data Fusion* 8.1, pp. 1–31. DOI: 10.1080/19479832.2016.1160960.

Zhang, J. and S. He (2020). 'Smart Technologies and Urban Life: A Behavioral and Social Perspective'. In: *Sustainable Cities and Society* 63, p. 102460. DOI: 10.1016/j.scs.2020.102460.

Zhang, Q., W. Wang and S.-C. Zhu (2018a). 'Examining CNN Representations with Respect to Dataset Bias'. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 32. 1, pp. 4464–4473. DOI: 10.1609/aaai.v32i1.11833.

Zhang, S., Y. Li, S. Zhang, F. Shahabi, S. Xia, Y. Deng and N. Alshurafa (2022b). 'Deep Learning in Human Activity Recognition with Wearable Sensors: A Review on Advances'. In: *Sensors* 22.4, p. 1476. DOI: 10.3390/s22041476.

Zhang, S., S. Lo, Y.-H. Chen, T. Walter and P. Enge (2018b). 'GNSS Multipath Detection in Urban Environment Using 3D Building Model'. In: *2018 IEEE/ION Position, Location and Navigation Symposium (PLANS)*. IEEE, pp. 1053–1058. DOI: 10.1109/PLANS.2018.8373486.

Zhang, X., X. Li, W. Sultani, Y. Zhou and S. Wshah (2023). 'Cross-View Geo-Localization via Learning Disentangled Geometric Layout Correspondence'. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 37. 3, pp. 3480–3488. DOI: 10.1609/aaai.v37i3.25457.

Zhang, X. Q. (2016). 'The Trends, Promises and Challenges of Urbanisation in the World'. In: *Habitat International* 54, pp. 241–252. DOI: 10.1016/j.habitatint.2015.11.018.

Zhang, Y., X. Xiong, S. Yang, Q. Zhang, M. Chi, X. Wen, X. Zhang and J. Wang (2025). 'Enhancing the Visual Environment of Urban Coastal Roads through Deep Learning Analysis of Street-View Images: A Perspective of Aesthetic and Distinctiveness'. In: *PLOS one* 20.1, e0317585. DOI: 10.1371/journal.pone.0317585.

Zhang, Z. and L. Zhu (2023). 'A Review on Unmanned Aerial Vehicle Remote Sensing: Platforms, Sensors, Data Processing Methods, and Applications'. In: *Drones* 7.6, p. 398. DOI: 10.3390/drones7060398.

Zhao, C., Q. Weng and A. M. Hersperger (2020). 'Characterizing the 3-D Urban Morphology Transformation to Understand Urban-Form Dynamics: A Case Study of Austin, Texas, USA'. In: *Landscape and Urban Planning* 203, p. 103881. DOI: 10.1016/j.landurbplan.2020.103881.

Zhao, F., C. Zhang and B. Geng (2024). 'Deep Multimodal Data Fusion'. In: *ACM Computing Surveys* 56.9, pp. 1–36. DOI: 10.1145/364944.

Zhao, T., X. Liang, F. Biljecki, W. Tu, J. Cao, X. Li and S. Yi (2025). 'Quantifying Seasonal Bias in Street View Imagery for Urban Form Assessment: A Global Analysis of 40 Cities'. In: *Computers, Environment and Urban Systems* 120, p. 102302. DOI: 10.1016/j.compenvurbsys.2025.102302.

Zheng, Y., L. Capra, O. Wolfson and H. Yang (2014). 'Urban Computing: Concepts, Methodologies, and Applications'. In: *ACM Transactions on Intelligent Systems and Technology (TIST)* 5.3, pp. 1–55. DOI: 10.1145/262959.

Zhou, H., Y. Zhao, Y. Liu, S. Lu, X. An and Q. Liu (2023). 'Multi-Sensor Data Fusion and CNN-LSTM Model for Human Activity Recognition System'. In: *Sensors* 23.10, p. 4750. DOI: 10.3390/s23104750.

Zhou, Z., Y. Xi, S. Xing and Y. Chen (2024). 'Cultural Bias Mitigation in Vision-Language Models for Digital Heritage Documentation: A Comparative Analysis of Debiasing Techniques'. In: *Artificial Intelligence and Machine Learning Review* 5.3, pp. 28–40. DOI: 10.69987/AIMLR.2024.50303.

Zhu, N., J. Marais, D. Bétaille and M. Berbineau (2018). 'GNSS Position Integrity in Urban Environments: A Review of Literature'. In: *IEEE Transactions on Intelligent Transportation Systems* 19.9, pp. 2762–2778. DOI: 10.1109/TITS.2017.2766768.

Zhu, S., M. Shah and C. Chen (2022). 'Transgeo: Transformer Is All You Need for Cross-View Image Geo-Localization'. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1162–1171. DOI: 10.1109/CVPR52688.2022.00123.

Zhu, S., T. Yang and C. Chen (2021). 'VIGOR: Cross-View Image Geo-Localization beyond One-to-One Retrieval'. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3640–3649. DOI: 10.1109/cvpr46437.2021.00364.

Zhu, X. X., D. Tuia, L. Mou, G.-S. Xia, L. Zhang, F. Xu and F. Fraundorfer (2017). 'Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources'. In: *IEEE Geoscience and Remote Sensing Magazine* 5.4, pp. 8–36. DOI: 10.1109/MGRS.2017.2762307.

Zidan, J., E. I. Adegoke, E. Kampert, S. A. Birrell, C. R. Ford and M. D. Higgins (2020). 'GNSS Vulnerabilities and Existing Solutions: A Review of the Literature'. In: *IEEE Access* 9, pp. 153960–153976. DOI: 10.1109/ACCESS.2020.2973759.