A New Picture of the City: Volunteered Geographic Image Information and the Cities

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy by

Meixu Chen

Supervised by Dr Daniel Arribas-Bel and Prof Alex Singleton

Department of Geography and Planning, School of Environmental Sciences

University of Liverpool

March 2021

Statement of Originality and Contributions to Publications

I, Meixu Chen, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

I confirm that published papers contained in this thesis were mostly contributed by myself. My contributions included data collection, pre-processing, analysis and modelling designation and operation, results validation, evaluation and discussion and paper writing.

My supervisors contributed to supervision, research discussion and modification of the written papers.

Abstract

The urbanisation process continuously influences human life, causing long-term challenges for the planning and management of urban areas. In recent years, with the emergence of new forms of data and advances in techniques, the ways of managing and governing this process have evolved and formed a new research field: urban analytics. A growing number of human behaviours can be traced through quantities of data, which enables attributes of the urban environment to be managed more efficiently, potentially beneficial to complex decision-making processes by stakeholders. As such, how to extract useful information from new data and provide more suitable methods requires careful consideration.

The question of how human activity relates to the built environment has been an important topic in the sensing of cities. Existing ways to perceive the city either focus on environmental aspects that cover historical, social, or cultural dimensions of urban space through surveys, interviews, or mobility data (e.g., social media data), or extract visible features from georeferenced images to gain perceptions of the city. However, both approaches are often disconnected and lack dynamic consideration.

The main aim of this thesis is to address these challenges and gaps within urban analytics. It develops a methodological framework to leverage user-generated geotagged images and modern analytical techniques to obtain insights. Such framework is designed to mine spatial, temporal and image attributes of the Flickr images, which combines multiple dimensions including spatiotemporal dynamic analysis, computer vision models, summary statistics, and varying machine learning algorithms that allow understanding of human interactions with the built environment.

The overall analysis and results enrich our current understanding of how user-generated urban pictures represent but also shape the city. This is especially important given the growing popularity of volunteered geographic information and urban

analytics over the last decade. Their rapid growth has facilitated debates worldwide, but there is still a large potential of volunteered geographic information such as geotagged image information which has been underestimated in most circumstances. The findings presented in this thesis offer richer evidence that aims to help the improvement of strategic planning systems, and empowering policymakers to make smarter decisions in terms of urban governance.

Keywords: volunteered geographic information, geotagged Flickr images, urban perception, urban areas of interest, machine learning, housing

Acknowledgement

To begin, I would like to express my greatest appreciation to my first supervisor and daily friend, Dr Daniel Arribas-Bel, for his invaluable advice, positive stimulation, patient guidance during my PhD research. Particularly appreciated his valuable time and insightful feedback relating to my thesis writing. Additionally, Daniel has provided several teaching and research opportunities for me, helping me enrich my academic experience and broaden my knowledge. I would also like to thank my second supervisor, Professor Alex Singleton, for the expertise he has offered over the years and especially thank him for the powerful support to my post-PhD research. I am very grateful to have such supervisors to support my PhD study.

A big thanks to my viva examiners Dr Jamie Goodwin-White and Dr James Howarth, who were willing to spend their valuable time to review my thesis during summer and offered me highly useful suggestions and comments. I would also like to thank my colleagues and alumni at our Geographic Data Science Lab and other institution of the University of Liverpool. A special thank you to Dr Francisco Rowe and peer Dominik Fahrner for collaborating in contribution to promote reproducibility and open science. Also, to Dr Les Dolega and professor Guanpeng Dong who continuously offered me teaching opportunities over the years. I am thankful to my peer PhD Sam Comber and Dr Jacob Macdonald, who provided useful advice to improve my work, and to Lidan Shang, who shared my happiness and pressure to balance my study and life.

Thank you everything to my parents, for sharing every story of my PhD study, and offering funding to support me selflessly. A giant appreciation and apology to my mother for her sacrifice on my behalf, isolating myself from the truth of her illness for a while to avoid distracting my study, and support me to continue to carry my research on abroad even she needed me most at that period. I am also thankful to my families for helping me look after my mother. I cannot finish my PhD study without support from them.

Moreover, I cannot thank my boyfriend Dr Yunzhe Liu enough, who built his dream on me to start our PhD journeys together. He is the one to chat with about my study, stress, happiness, confusion, anxiety and every feeling in my study and life. He is my largest mental support system.

Lastly, I would like to thank my little friends, Cambridge and Oxford, two lovely guinea pigs who brought me much joy and company especially during COVID-19 pandemic. And for the countless people who have supported me along my 4-year PhD journey. I will never forget all these moments.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ALE: Accumulated Local Effects

API: Application Programming Interface

CNN: Convolutional Neural Networks

COCO: Common Objects in Context

CV: Cross-Validation

DBSCAN: Density-Based Spatial Clustering of Applications with Noise

HDBSCAN: Hierarchical Density-Based Spatial Clustering of Applications with Noise

GAN: generative adversarial networks

GIS: Geographic Information System

GPS: Global Positioning System

GPU: Graphic Processing Unit

GBM: Gradient Boosting Machine

GSM: Georeferenced-Social Media

GSV: Google Street View

GWR: Geographically Weighted Regression

HPM: Hedonic Price model

ILSVRC-2012: ImageNet Large Scale Visual Recognition Challenge 2012

KDE: Kernel Density Estimation

LBSN: Location-Based Social Network

MSE: Mean Squared Estimation

MSOA: Middle Layer Super Output Area

OA: Output Area

OPTICS: Ordering Points to Identify the Clustering Structure

OSM: Open Street Map

PDP: Partial Dependence Plots

POI: Point of Interest

PSS: Participatory Sensing System

R-CNN: Region Convolutional Neural Network

RF: Random Forest

R^2: Coefficient of Determination

UAOI: Urban Areas of Interest

UGC: User-Generated Content

VGI: Volunteered Geographic Information

VIF: Variance Inflation Factor

YOLO: You Only Look Once

1. Introduction

1.1 Sensing the City Through Urban Analytics

A city possesses various functions for human living covering at least four fundamental categories: dwelling (e.g., houses), work and service (e.g., offices, schools, stores), entertainment (e.g., galleries, parks, plazas), and transportation (e.g., railway stations; Ittelson, 1978). Given their composition, variability, and complexity, cities can hardly be understood at a single scale (Singleton et al., 2018). Within various perspectives, a city can be viewed as a single entity, or abstracted as an object that is spatially located at bounded extends. Over the past 100 years, humans have witnessed worldwide urbanisation under economic, cultural, social, and technical forces. Although the forces that drive urbanisation are quite different among different cities of the world, a general trend is that people are attracted by the proximity to infrastructure, employment opportunities, and cultural districts (Singleton et al., 2018). The transition to urbanised living continuously influences human interactions which are complex and dynamic, jointly determining urban morphology, structure, and functions (Batty, 2013) and posing challenges for the planning and management of urban areas.

Data are the carriers of digital footprints that shape the world into numbers, characters, symbols, images, sounds, bits, and more, constituting the structural elements for the creation of information and knowledge (Kitchin, 2014). The unfolding and transformed digital world have changed every aspect of humans' daily life in a city (Goodchild, 2007). A growing range of human behaviours can be traced back through numerous digital footprints that can expose emerging patterns if aggregated and modelled (Arribas-Bel, 2014). This societal change allows urban researchers to investigate new forms of data at a more detailed level rather than having to rely entirely on traditional data sources (Singleton et al., 2018). Although new forms of data may be less representative and comprehensive than traditional data sources, their distinct

characteristics make them worth exploring. First, these new data are available at high frequencies enabled by sensors such as mobile phones, digital cameras, or computers (Manyika et al., 2011). Second, most of these sources are freely available to researchers through application programming interfaces (APIs) or open data at governmental institutions. Finally, many new datasets such as social media data are generated by individuals and can not only quantify human activity but capture human perceptions of cities. As such, these data can help reduce the error of location measurement observations, avoid discretisation of continuous urban research problems, and fill gaps where traditional data are lacking (Arribas-Bel, 2014).

Unlike traditional sources, new forms of data are often generated as a by-product which is unstructured and large and can pose challenges to extracting useful patterns. As a response, new forms of data require the development of techniques to process this data (Kitchin, 2016). In fact, Data are undergoing an innovative transition to exploratory (i.e., data-driven) science, from the mode of complex phenomena simulation to new, data-intensive analytical methods (Lynch, 2009). The premise of data-driven science is to employ theory-guided knowledge discovery methods to recognise hypotheses worth being investigated and tested further (Kitchin, 2014). This data-driven science enables the storage, analysis and presentation of new forms of data through evolutionary computer hardware (e.g., central processing unit and graphics processing unit), more user-friendly software and open-access programming language (e.g., Python or R; Singleton et al., 2018).

Many core approaches in data-driven science are machine learning tasks that happen in an automated way without any human intervention (Singleton & Arribas-Bel, 2021).

Recently, machine learning has permeated social science more, including quantitative analytics in urban geography (Kandt & Batty, 2021). Machine learning is usually constituted of unsupervised and supervised learning (Hastie et al., 2009). Primary precedents for unsupervised applications within quantitative geography are geodemographic analysis or regionalisation and zone design (Openshaw & Openshaw,

1997; Singleton & Longley, 2009), which often generate the sociodemographic groups based on statistical similarity as well as varying geographic dimensions. Applications in supervised learning include spatial econometrics or geographically weighted regression (Anselin, 1989; Brunsdon et al., 1998), which integrate space in a regression model to automatically build the representations of phenomena. These approaches are implemented through different data analytics software, enabling researchers and stakeholders to process, analyse, and visualise these new forms of data for problem-solving at fine spatiotemporal resolutions and frequently for urban governance (Kitchin, 2016). This is the so-called new term in the field of urban studies: urban analytics that uses the new data sources from social media, crowdsourcing and sensor networks, and rely on the power of quantitative modelling from the description, prediction and explanation (Singleton et al., 2018). This new field enables attributes of the urban environment to be managed more efficiently, which is potentially beneficial to complex decision-making processes and different stakeholders. The detailed information on urban analytics is stated in Section 2.2.1.

1.2 Current Opportunities and Challenges

As the statement of the research context has shown, urban analytics make it possible to ask new or complex questions about cities to gain insights using new forms of data. Illustrations of urban analytics include integrating these new data in policymaking to offer richer evidence for the improvement of strategic planning systems or empowering policymakers to make smarter decisions in terms of governance of urban areas (Singleton et al., 2018). However, a certain degree of error, uncertainty, or bias embedded in all data cannot be neglected. As such, careful considerations are required in asking the right questions and using appropriate tools when working with these data sources.

The question about how human activity relates to the built environment has been a popular topic over the years, playing an important role in understanding and governing cities. Many proposals have been attempted to explore ties between human activity

and the built environment from surveys to systematic observations or Geographic Information System (GIS)-based analytics (Brownson et al., 2009) at an early stage to more data-driven approaches in recent years (Lloyd & Cheshire, 2017; Sulis et al., 2018; Zhang et al., 2019). Applications vary from unpacking the environmental impacts on human activity, such as the effects of infrastructure accessibility on leisure activity (Duncan et al., 2005) or proximity to destinations on transportation activity (Humpel et al., 2002), to mining human-environment interactions such as travel behaviour exploration (O'Brien et al., 2014), event detection (Kisilevich et al., 2010), and urban representations recognition (Zhou et al., 2014). Although the rapid expansion of urban analytics has provided ample opportunities for stakeholders to gain insight into how cities operate, the question of how to better use data and what methods are suitable remains open.

Based on the literature, there are usually two main perspectives to perceive the city: one analyses environmental preferences using surveys or interviews to extract historical, social, and cultural attributes of cities (Frank & Engelke, 2001; Hristova et al., 2018; Sulis et al., 2018); the other one focuses on identifying perceived attributes such as safety, wealth, and the uniqueness of the urban environment from images (Dubey et al., 2016; Khosla et al., 2014; Naik et al., 2014). The former perspective could explore multiple aspects of the environment but very limited involvement in the perceived attributes. The latter tends to extract urban perceived attributes mostly from Street View imagery, which is often collected by street view fleets (Google Maps Street View, 2020) and thus underestimating the importance of human cognition in sensing the city. Integrating both approaches to gain a more comprehensive understanding of human–(built) environment interactions is challenging. Moreover, high mobility of human activity between cities or within the city on a daily or monthly basis leads to people's perception of the city changing dynamically. This is also a challenge for stakeholders governing and managing the city.

1.3 Research Objectives

This research seizes the opportunities presented by the rise of urban analytics and contributes to closing some of the gaps outlined above, aiming to answer the following primary research question: “How can the perception of the city be better understood by volunteered geographic image information?” Using image data about Inner London from the Flickr platform, the following research objectives were formulated to address different dimensions of the relationship between built and experienced environment:

1. To identify urban areas of interest (UAOIs) in the city from Flickr images and profile their dynamic spatiotemporal attributes. *Chapter 3*
2. To recognise human-perceived scene features from Flickr images to investigate the driving factors for the varying popularity levels of urban areas. *Chapter 4*
3. To mine dynamic perceived scene features to uncover the temporal variations and the formation of UAOIs. *Chapter 4*
4. To analyse impacts of perceived scene features on the urban environment using housing prices as a case study *Chapter 5*
5. To explore the comparative merits in terms of both performance and interpretability of modern machine learning models, as compared to traditional linear regression approaches, in the application of hedonic models of housing prices. *Chapter 5*

1.4 Thesis Structure

To achieve the above objectives, this thesis provides an in-depth analysis of particular volunteered geographic information (VGI): geotagged Flickr images. To leverage these images’ potential, the thesis uses a wide range of modern urban analytics methods ranging from exploratory spatial data techniques to state-of-the-art computer vision algorithms. The overall analysis and results enrich the current understanding of how user-generated urban pictures shape the city. This is of special relevance given

the growing popularity of VGI and urban analytics over the last decade. Their rapid growth has facilitated debates worldwide, but there is still substantial untapped potential in VGI, which has been underestimated in most circumstances. Thus, a novel research framework is proposed and developed to make full use of geotagged images and modern analytical techniques to enrich the research of urban perception field and provide richer evidence for stakeholders (see flowchart in Figure 1.1). Such a framework is designed to mine spatial, temporal, and image attributes of the Flickr data, combining spatiotemporal dynamic analysis, image recognition models, and a variety of machine learning algorithms that enhance the understanding of human interactions with the built environment through multiple dimensions.

This thesis is divided into six chapters that together elaborate on how volunteered geographic image information can be used to understand cities. This section briefly describes what each chapter focuses on, thus providing a holistic overview of the thesis.

Chapter 1 provides an overview of the thesis. It begins with the research motivation to understand the cities, followed by an introduction of new forms of data and data-driven science, and then explains urban analytics field which combines new forms of data with modern analytical methods together to sense the city. Based on that, a few opportunities and challenges are discussed to raise our main research question and research objectives. The structure of the thesis is then provided to help readers navigate the entire document. Finally, several key contributions of the thesis to theory and practice are outlined.

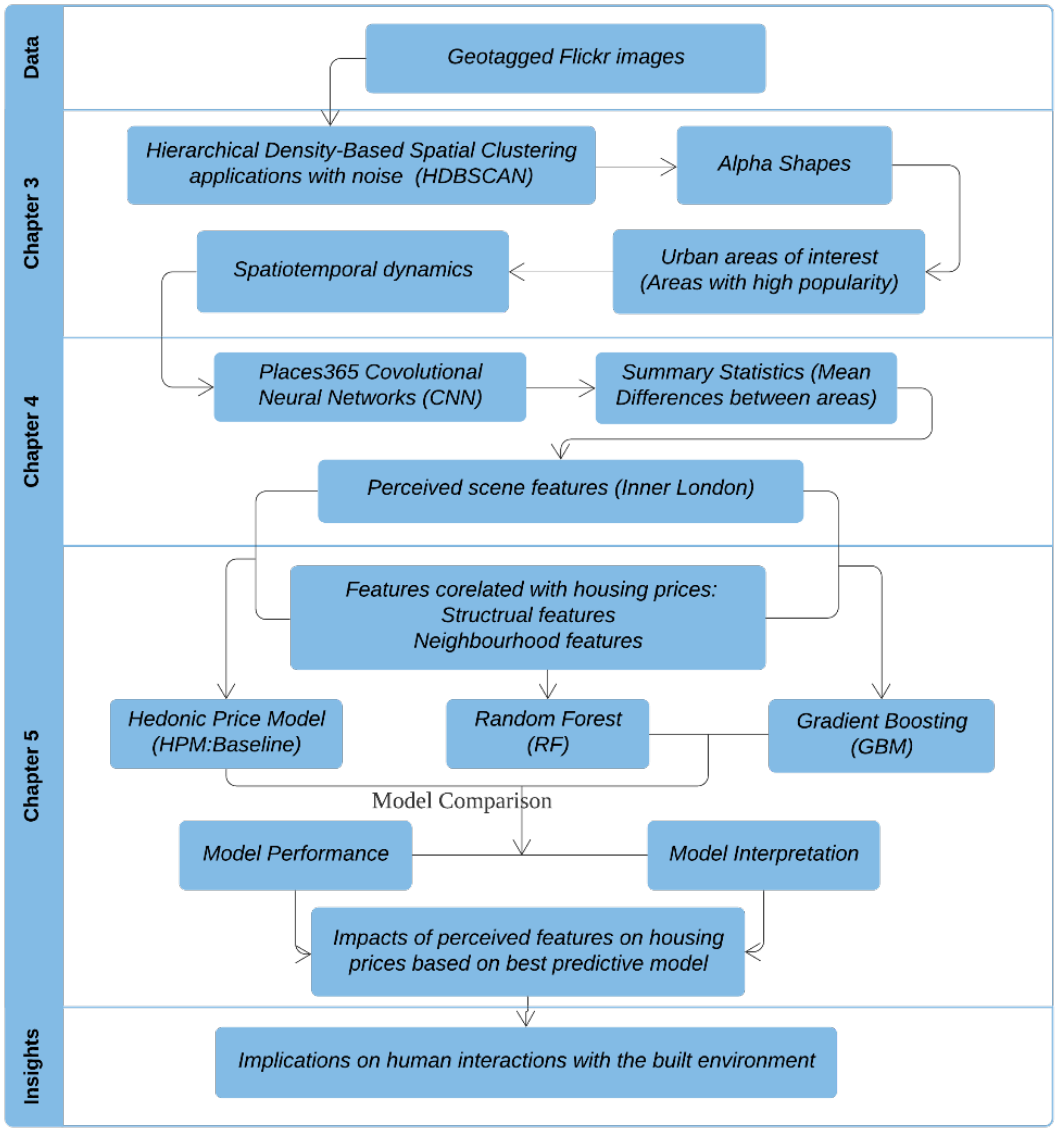


Figure 1.1 Method Framework Designed to Sense the City Through Geotagged Flickr Images

Chapter 2 reviews the relevant literature and theoretical context of the thesis. This includes the following domains: human interaction and the built environment, new data sources that use humans as sensors, and the potential of images in cities. Specifically, it first explores the theoretical context of urban perception with a detailed definition and an account of its development over time. The chapter then summarises work on the relationship between human activity and the built environment and different ways to understand human interaction in the city. VGI is introduced, discussed, and explained concerning the thesis, first covering social sensing, then

social media data, and finally geotagged Flickr data. The main advantages and limitations of these sources are thus evaluated in the context of the study. This chapter concludes with an illustration of the potential of imagery in urban research and the main algorithms currently used to leverage the value of images.

Chapters 3 through 5 consist of three case studies that explore different directions and approaches to understanding the city using volunteered geographic image information (i.e., Flickr images). All of these maintain consistency and progression since the same study area and time range of data are used. Both Chapters 3 and 4 focus on UAOIs, a concept that provides functional definitions of a city's spatial structure. Chapter 5 uses insights from the previous two chapters to exploring the potential of images for helping to establish a better understanding of housing prices.

Chapter 3 is intended to contribute to the understanding of the dynamics of UAOIs geographically and temporally, developing a methodological framework that combines a clustering algorithm with a technique to delineate tight boundaries around clusters of points and an innovative visualisation to profile the dynamics of various UAOIs. The results illustrate the interactions between human activity and the built environment and visualise how the popularity of certain regions is influenced by time and how its levels differ across different areas.

Chapter 4 extends the work of Chapter 3 and focuses on the image information in the Flickr data. It characterises different attractiveness levels of urban areas based on an aggregation of information extracted from images collected within the boundaries of those areas. An advanced image recognition model was utilised to extract features from millions of images from Inner London in the period from 2013 to 2015. The identified characteristics were then integrated into different urban areas and time ranges followed by a series of visualisation techniques such as feature importance plots and accumulated local effects plots. The findings demonstrate that urban areas with higher population densities cover more iconic landmarks and leisure zones, while

others are more related to daily life scenes. The dynamic nature of these results suggests that season determines human preferences for travel and activity modes.

Chapter 5 further identifies the relationships between urban perceptual features and the surrounding housing market. This analysis is based on the image features recognised in the previous chapter and property transaction records on a monthly basis. Combined with ancillary datasets and built around a traditional housing price model (i.e., hedonic price model [HPM]), structural, neighbourhood, and perceived scene characteristics were identified to uncover their impacts on housing prices. Two machine learning algorithms – random forest and gradient boosting machines – were harnessed to compare their performance and interpretability with the baseline model. The results corroborate that volunteered geographic image information could be added as an additional data source when analysing the housing market. Furthermore, machine learning algorithms are shown to be comparable to traditional HPM in terms of performance and interpretability.

Chapter 6 provides a set of concluding remarks, including a summary that highlights the main findings and contributions from both theoretical and technical perspectives as well as limitations, and suggests further extensions to the work presented in the document. The implications of these findings are discussed with a focus on how they can be used to inform policies associated with urban planning and design. Implications include insights related to human interactions with the built environment, business marketing in terms of store site selection, and real estate appraisal concerning the adjustment of housing prices and construction of attractive neighbourhoods. The chapter concludes that volunteered geographic image information (i.e., Flickr images in this study) is a valuable data source for sensing human interactions with the built environment and that it can be used to expand and improve upon the previous research on urban perception.

1.5 Key Contributions

The core contribution of this thesis is to develop a methodology framework to better understand urban perception through volunteered geographic image information, the explicit contributions are presented separately in Chapters 3 through 5.

The methods and results in Chapter 3 are of interest in several fields and domains. They could help urban planners develop better strategies in tourism planning such as more efficient resource allocation. Local authorities may also benefit from the results in terms of police patrol and traffic monitoring. Researchers and practitioners could consider UAOIs as an additional geographic layer to understand the use of the urban built environment. Furthermore, part of the relevance of the thesis' contribution lies in the fact that it can be deployed using data that are available almost in real-time. Unlike more traditional data sources, geotagged Flickr images are constantly added to services, thus providing an opportunity to study the evolution of UAOIs not only retrospectively but as they evolve. This holds distinct value for practitioners such as urban planners and policymakers.

Chapter 4 contributes to the research field of urban perception. It recognises the value of perceived scene features from Flickr images, instead of relying on traditional methods that use tags or other image sources like street-level images. More importantly, the work bridges the research gap between image recognition techniques and urban perception analytics, implying that local scales and dynamic characteristics are important to the study of urban perception. In terms of practical significance, the regular and dynamic characteristics of the urban environment provide new decision-making insights for policymakers. The regular characteristics are informative for urban planners to have a macroscopic understanding of urban areas and aid them in formulating relevant policies such as target investments in certain areas to stimulate consumption for economic growth. The dynamic characteristics of perceived scene features can help transport planners regulate trip frequency in various seasons, for instance with greater trip frequency in the winter than in the summer. Moreover,

retailers may also be inspired by the dynamic perceived scene features to better design personalised advertisements at specific places and times or expand their open hours in the summer.

Chapter 5 integrates urban perception into a housing study, which expands and improves upon the previous literature since little research has explicitly considered volunteered geographic image information to explain and understand the housing market. Additionally, the chapter shows that models to investigate the impacts of features on housing prices built on machine learning techniques are superior and more flexible in performance than the traditional HPM but remain interpretable, avoiding the common black-box problems attributed to several machine learning algorithms. The findings provide a reference for stakeholders to consider user-generated images as an additional dataset for real estate appraisal. This data source can capture human interactions with the urban environment, reflecting their interests and perceptions of urban scenes. The patterns would also be informative to real estate developers for early-stage site selection for the construction of residential buildings. Furthermore, the government should pay more attention to the adjustment and design of housing development based on various facilities and surrounding urban features, which could assist in improving the vitality of the area surrounding a property. This subsequently influences people's willingness to buy that property.

In summary, Chapter 1 introduced the research background of this thesis and outlined current opportunities and challenges in urban analytics and urban perception. A range of objectives were proposed to address our research question on "How can the perception of the city be better understood by volunteered geographic image information?", followed by a detailed structure and key contributions of the thesis.

2. Literature Review

Abstract: This chapter explains the conceptual and empirical knowledge that motivated the underlying research question proposed within this thesis. To begin, the theoretical concept of urban perception is introduced, and its development is discussed to identify the gaps that human cognitions are overlooked in most cases. Different dimensions of cities are then outlined to explain why they are relevant to this study and corresponding challenges are identified. The next two sections illustrate the data from two perspectives: Section 2.3 discusses the data from human participation perspective and Section 2.4 unfolds the data from an image perspective. A range of opportunities and challenges concerning data quality, data applications, data sources, data-driven methods are then proposed. Thus, this thesis sought to build research on these gaps to bring theoretical and practical meanings to the knowledge reviewed in this chapter.

2.1 Urban Perception

2.1.1 Background and Origins

Research on urban perception covers many aspects compared with the traditional definition of perception, which includes more general cognition with perception, thinking, imagery, and emotion, intention, and assessment (Ittelson, 1978). A direct and common perspective to understand urban perception was based on the idea of "image" initially proposed by Boulding (Boulding, 1957) and later developed in two directions. One direction claimed that individuals form generalised mental images when they experience environmental physical elements, which perspective was represented by the seminal work of Lynch (1960), while the other emphasised the nature of the cognitive image or map and information extraction in terms of people's perception of their living conditions (Milgram, 1976; Tuan et al., 1975).

Another direction in which to explore urban perception included studies of environmental values, preferences, and aesthetics. An early study involved the concept of landscape and considered historical and cultural variations of a city as a reflection and determinant of social and cultural values in the city (Lowenthal, 1968). More empirically oriented studies examined landscape preferences and scenic beauty to assess urban environment (Daniel & Boster, 1976; Herzog et al., 1976). This direction was then developed into a formal study of environmental or urban aesthetics as an empirical topic in Wohlwill's (1976) work. These works shared a common claim that urban value and aesthetics constitute all parts of urban perception.

It is difficult to neglect the social, cultural, and personal learning contexts, which not only influenced environmental elements or aesthetics but also the cognition involved (Rapoport & Hawkes, 1970). Urban activist Jane Jacobs claimed that urban streets work as principal visual scenes in a city and emphasises the connections between the neighbourhoods of inner cities and the social interaction of urban dwellers (Jacobs, 1961). Environmental psychologist Ittelson (1978) highlighted that the definitions of the environment vary in the direction of emphasis among different individuals and

groups, but social, cultural, and physical environments cannot be apart. In other words, no person acquired any source of information unless through participation in the environment. Therefore, the concept of urban perception was complex and rich and included studies involving perceptual, cognitive, imaginal, emotional, and meaning aspects of a city analysed using a series of methodologies and techniques (Ittelson, 1978). Except physical, interpersonal, and cultural aspects of urban environment, urban perception also highlighted the significance of human conditions, including demands, actions, motivation, cognition, and so on, as typically treated in a variety of literature sources (Craik & Zube, 1976; Holcomb & Saarinen, 1977; Ittelson, 1978; Lawson & Ittelson, 1977; Stokols & Moos, 1979).

2.1.2 The Development and Evolution of urban perception: From Qualitative to Quantitative

Studies of urban perception have existed since at least the 1960s, though these studies have mostly focused on qualitative analysis instead of a quantitative perspective (Rapoport & Hawkes, 1970). These qualitative analyses were concerned with understanding theories, perspectives, or phenomena through unstructured and nonnumerical data collected from participant observation and interviews (Strauss, 1988), while quantitative analysis relies on numerical data to understand underlying patterns and empirical relationships and generalised results to a wider population via statistical, mathematical, or computational techniques (Blaikie, 2003). Urban planners and architects managed to measure urban perception through a range of evaluative dimensions on the basis of visual surveys or interviews (Herzog et al., 1976; Nasar, 1990; Schroeder & Anderson, 1984; Scott, 1998). For instance, Herzog et al. (1976) conducted a nonmetric factor analysis on urban scenes rated by college students, demonstrating that the five urban dimensions cultural, contemporary, commercial, entertainment, and campus reflect people’s preferences regarding urban environments. Nasar (1990) found by interviewing 440 participants that naturalness, upkeep, openness, order, and historical significance are significant features according to which

the public evaluate a city. In terms of the connection between visual appearance and human perception, empty buildings and graffiti were linked with low safety (Schroeder & Anderson, 1984) while disorders such as trash and abandoned properties and cars were associated with a perception of the breakdown of social order and fear of crime (Skogan, 1990). However, the use of surveys or interviews were criticised because they were costly, time-consuming, and largely based on traditional expert assessments, which resulted in limitations for contextual behaviour evaluation and interaction (Zube et al., 1982).

Qualitative studies remained dominant in the study of urban perception until the popularity of new sources of data and the advancement of computer vision techniques several years ago (Kandt & Batty, 2021). Since then, there has been an abundance of literature exploring urban perception using a quantitative approach. Some concerned the identification of visual representations of the city, including evaluation of landscape preference by extracting numeric variables from land cover and elevation data at the very beginning of the century (Wherrett, 2000), and concentration of the discovery of geographically or temporally representative image elements through machine learning methods (Doersch et al., 2012; Kennedy & Naaman, 2008; Lee et al., 2015; Lee et al., 2013; Zhou et al., 2014, Zhang et al., 2018). Similar works focused more on quantifying perceptual characteristics of the city. An important study in this context that enabled further studies was that of Salesses and colleagues, who proposed a reproducible quantitative measure for the urban perception of safety, class, and uniqueness which relied on a survey with pairwise comparisons of the geotagged images of a few cities (Salesses et al., 2013). Various researchers later exploited this dataset to identify urban perceptual characteristics such as safety, uniqueness, and wealth using varying statistical techniques (Naik et al., 2014; Ordonez & Berg, 2014), which both measured six perceptual attributes provided by Dubey and his colleagues: safe, lively, beautiful, wealthy, boring, and depressing (Dubey et al., 2016). Instead of identifying the visible characteristics, some studies further explored the relationships between urban perceptual features and other attributes. Arietta et al. (2014) trained a

predictor to discover relationships between the visible appearance of a city and its nonvisual attributes (e.g., crime statistics and population density). Khosla et al.(2014) demonstrated that it is possible to infer the distance to surrounding scenes based only on visible attributes far from them. Li et al.(2015) analysed the relationship between urban greenery and perceived safety based on the Place Pulse 1.0 dataset. Naik et al. (2017) connected changes in urban perceptual characteristics with socioeconomic attributes such as population density and proximity to the city centre to predict neighbourhood improvement. Zhang et al. (2018) identified 150 scene categories segmented from the street view images as being correlated with the six perceptual measures identified by Dubey and his colleagues (2016).

Although many studies attempted to quantify urban perception, they mostly relied on street-level imagery (Arietta et al., 2014; Doersch et al., 2012; Lee et al., 2015; Li et al., 2015; Naik et al., 2017; Zhang et al., 2018) where a mass of data was collected from professional street-view fleets. This led to the perception of the city not being captured by different individuals, which weakened the significance of human cognitions in the definition of urban perception. Additionally, previous research on connecting perceptual features and surrounding nonvisual socioeconomic attributes were still limited and mainly focused on crime rate (Arietta et al., 2014; Khosla et al., 2014; Naik et al., 2017). Therefore, this work intends to reduce these gaps to gain some new insights into urban perception.

2.2 Understanding the Multiple Dimensions of Cities

2.2.1 The Emergence of Urban Analytics

Urban analytics is a new term that emerged in the 2010s and is beginning to gain traction. There are currently two mainstream definitions: Batty (2019) claimed that the term is derived from urban analysis but a wider scope for the term ‘analytics’, suggesting a series of tools to cope with issues of big data, urban simulation, and geodemographics and grasp and estimate the features of cities. Goodchild defined

urban analytics as new types of urban research that take advantage of new data sources such as social media, crowdsourcing, and sensor networks as well as the power of computer technology (Singleton et al., 2018). As such, urban analytics creates many opportunities and challenges. Opportunities come from recently increasing new forms of data and associated computational techniques (Arribas-Bel, 2014) that enable more accurate measurement and the extraction of actionable insights; challenges primarily remain in the nature of urban complexity and how to take full advantage of the opportunities new methods provide to help urban planners improve citizens' quality of life (Batty, 2019).

Understanding a city is a complex process that generally requires multiple sources of information such as social, economic, cultural, political, and technical scales. For decades, researchers have considered cities as a whole or as coarse aggregations, examples of which include but were not limited to modelling urban systems (Batty, 2005), profiling geodemographic statistics (Singleton & Longley, 2009), unpacking human mobility and migration (González et al., 2008), and exploring population health and diseases (Ng et al., 2014). These studies mostly analysed cities using longitudinal analysis (e.g., UK census data) which was incapable of capturing small details at a finer level of resolution. However, real-time and disaggregate data have increasingly become available over the last decade, allowing a number of works to analyse high spatiotemporal resolution. As a result, urban analytics has become extremely prevalent and powerful.

2.2.2 Human Activity and the Built Environment

In this context, human activity refers to any bodily movement that generates energy consumption in one's daily life, which could be categorised into leisure-oriented (e.g., exercise), occupational, household, or other activities (Caspersen et al., 1985; Frank & Engelke, 2001). In the social sciences, the term 'built environment' refers to the places and spaces humans created and in which they live, work, and access daily, varying in scale from buildings to cities and larger areas (Roof & Oleru, 2008). The built

environment includes land use patterns, multiscale built and natural characteristics such as architectural details, and infrastructure that connects one place to another, such as transportation systems (Brownson et al., 2009). Connections between different elements of the built environment and human activity have been identified in previous works through surveys (e.g., Duncan et al., 2005; Humpel et al., 2002); for example, leisure activity was mostly influenced by accessibility to facilities, and transportation activity was associated with the proximity to destinations. Hence, to understand the relationships between the built environment and human activity, it is necessary to better understand each of these separately as well as the meeting points where they intersect.

Over the last few decades, studies have attempted to explore the connections between the built environment and human activity using various approaches. At a relatively early stage, these measurements primarily consisted of three categories: surveys, systematic observations, and archival datasets (Brownson et al., 2009). The first category focused on examining which elements of the built environment are most likely to affect human activity, such as availability of transport, infrastructure, and natural features for activity through telephone interviews or self-administered questionnaires; the second was mainly used to quantify attributes that are best evaluated through observational measures, such as street layout patterns, quality of public space, and sidewalk quality; and the third used GIS-derived metrics and analytics to assess relationships between the built environment features and human activity derived from existing datasets that have a spatial reference. Assessed variables primarily included population density, land use mix, accessibility to facilities, and street pattern (Brownson et al., 2009). Unlike the first two measurements, which had problems with low response rates and time-consuming collection processes, GIS-based analytics were considered the only feasible way to obtain objective measures when studies concerned dispersed individuals or neighbourhoods within a large geographic area (Boarnet, 2003). However, a very small centralised national repository

of GIS data existed, and the time dimension was mostly absent from previous studies at the early stage of the measurements (Boarnet, 2003; Forsyth et al., 2016).

More recently, an increasing number of studies have used varying geographic granularities (e.g., census tracts, postcode areas, or points from GIS data) since the popularity and power of urban analytics to measure relationships between human activity and the built environment. Examples include mining human mobility patterns using census data in administrative areas (Cao et al., 2018; McCollum et al., 2020); exploring human travel patterns using different transport data such as taxis, subways, and bicycles at local or global geographic granularities (El-Assi et al., 2017; Liu & Cheng, 2020; O'Brien et al., 2014); detecting human events and behaviours through social media and mobile data in bounded or vague areas (Kisilevich et al., 2010; Li et al., 2013; Papadopoulos et al., 2011); and linking human activity to the housing market by accessibility to facilities and natural landscape at the household or district level (Baker et al., 2016; Hamilton & Morgan, 2010). Although these measures facilitated advancements in understanding correlations of human activity and the built environment, challenges remain, such as more potential attributes of the existing data, more novel and reliable methods, and more active interaction between humans and the built environment.

2.2.3 Urban Areas of Interest

Points of interest (POIs) is a widely used term describing the significance and popularity of the population concedes specific places of cities (Hu et al., 2015). An UAOI, which Hu et al. (2015) first introduced, was a part of the urban built environment that could be identified and delineated through aggregations of human activity. Such areas included varying POIs such as tourist attractions, iconic landmarks, business buildings, or recreational areas that were of interest to large numbers of people (McKenzie et al., 2014). UAOIs also referred to the areas that simply offered places for people to view the landscape rather than contain famous POIs, such as the areas that provide a good view of an urban landmark (Hu et al., 2015). The typical

geometric representations of UAOIs are polygons instead of points. In terms of computation, operations performed on polygons were more efficient than a set of points (Akdag et al., 2014). Additionally, such polygons created geographic representations of areas of interest that are simpler and more accessible to understand. In addition to the ability to capture the physical spaces of the city, the concept of UAOI also offered opportunities to capture the functional aspects of the social morphology (Crooks et al., 2016). As such, UAOIs were also part of the perceptual spaces that emphasised the significance of human cognition and could thus be utilised to explore interactions between people and places (McCullough, 2005). For UAOIs to be useful, they should emerge from the aggregation of activities of different people and vary among people of different contexts, ages, and cultures.

Prior to the definition of UAOI, similar studies mostly focused on the investigation of POI and fuzzy areas (i.e., areas without clear boundaries). For instance, some studies concentrated on personalised POI recommendations (Crandall et al., 2009; Ye et al., 2011), while others mined spatial patterns from fuzzy areas built from the aggregation of large quantities of people (Hollenstein & Purves, 2010; Li et al., 2013). These works highlighted the interactions between human activity and the built environment. However, in these frameworks, it was difficult to capture the overall characteristics of areas and their relatively long-term spatiotemporal changes due to a lack of clear geographic boundaries. The concept of UAOI was necessary to fill in this gap.

2.2.4 Housing in the Built Environment

The pursuit of a good life is easier in a built environment that was equipped with facilities enabling links to other communities, encouraging human activity for self-fulfilment and a healthy lifestyle (Molinsky & Forsyth, 2018). Within this context, housing and neighbourhood environments are particularly important in that they are the locations where people spend most of their time and which have a significant impact on physical, psychological, and social health. From the urban planning perspective, the characteristics of housing and neighbourhood can be viewed as

measurements of human progress and quality of life (Molinsky & Forsyth, 2018). The embodiment of these characteristics is housing prices, which are outcomes of the interaction of multiple parties (Law et al., 2019). Housing prices not only reflect housing characteristics such as varying property types and locations but also factor in neighbourhood features such as accessibility to transportation and facilities (Kong et al., 2007; Lu, 2018; Powe et al., 1995; Wilhelmsson, 2009). Accordingly, these features jointly influence the housing market and people's willingness to purchase, creating challenges for urban planners, urban designers, and practitioners in terms of regulation, construction, or evaluation.

Given the importance of housing and neighbourhood in the built environment, many studies were implemented to analyse their characteristics that impact housing prices (e.g., Hamilton & Morgan, 2010; Jim & Chen, 2006; Wen & Tao, 2015; Zhang & Dong, 2018). Traditional data sources included an official or commercial statistical database, proprietary listings, and questionnaire surveys (Granziera & Kozicki, 2015), and traditional methods relied on the Hedonic Price Modelling (HPM) to uncover the intrinsic value of a single attribute based on the prediction of the marginal changes in observed prices (Palmquist, 1984; Rosen, 1974). For example, Wen and Tao (2015) utilised HPM to examine the influence of urban structure on housing prices where the housing data was provided by a real estate agent company and neighbourhood data was obtained from a field survey. However, these data sources were labour intensive in collection and management, and they may not be freely available to the public. Furthermore, HPM had limitations on strong assumptions in terms of linear relation and spatial heterogeneity issues (Anglin & Gençay, 1996; Dubé & Legros, 2014). Although alternative methods such as spatial econometrics, geographically weighted regression, and machine learning (Choumert et al., 2014; Huang et al., 2017; Park & Bae, 2015) as well as more recent data such as social media (Liu & Long, 2016; Rae & Sener, 2016) were employed to reduce the limitations, challenges still exist regarding the interpretation of methods and mining of valuable information from data. As such, how to apply urban analytics (i.e., new data and new methods) to characterise

the housing and neighbourhood of the built environment to improve the research is an important remaining challenge.

2.3 Citizens as Sensors: Human Participation

2.3.1 Volunteered Geographic Information

In the middle of the first decade of the new century, an evolving phenomenon gained popularity: increasing numbers of individuals began to voluntarily engage in the creation of geographic information that has significant impacts on the relationship between geography and the public. This phenomenon was termed VGI (Goodchild, 2007). Specifically, VGI referred to large numbers of citizen volunteers acting as sensors to create, collect, and disseminate geographic data for mapping. The process of VGI was enabled and facilitated by modern web technologies (Web 2.0). On the one hand, Web 2.0 allowed users to supply information to websites instead of only downloading the content and even enabled them to edit content generated by others, which was so-called user-generated content (UGC). On the other hand, the turn of the century saw the appearance of large numbers of digital devices, particularly mobile phones, embedded with technologies to accelerate the process of VGI, such as location identification with coordinates enabled by the global positioning system (GPS), high-quality graphics, and accessible internet broadband connection. The emergence of VGI contributed to some developments and applications that relied on UGC (Flanagin & Metzger, 2008), such as the online map platform Wikimapia (Wikimapia.org) that provided possibilities for users to add location-based places and annotations and OpenStreetMap (https://www.openstreetmap.org/), which allowed individuals to create detailed and up-to-date maps based upon their GPS data and view, edit, and use the maps (Goodchild, 2007). More recently, various mobile apps such as Twitter and Flickr made it more convenient for users to play the role of active sensors to contribute geographical information (Li et al., 2013).

VGI has become increasingly important in recent years. A variety of location-based data and georeferenced images of VGI allowed monitoring of geographic phenomena at the high spatiotemporal resolution, which significantly improved environmental knowledge and geographers' understanding of Earth (Goodchild, 2007; Flanagin & Metzger, 2008). Additionally, VGI was less expensive than any other alternative because individuals voluntarily provided it, so the data was available to all users. More importantly, compared to the information produced by traditional professional institutions and government agencies, VGI had an advantage in terms of the information that required native experience and up-to-date local circumstances since individuals were in the best position to offer users emotional value and social interaction (Flanagin & Metzger, 2008; Parker et al., 2013). As a result, the most valuable aspect of VGI was its ability to represent local life and activities in various areas that were overlooked by the global media (Goodchild, 2007).

The motivations of individuals' contribution to the contents of VGI were self-promotion and social networking, which pursued personal satisfaction and convenient connections with communities (Goodchild, 2007). Hence, a few limitations associated with VGI should also be given careful notice. First, data provided by volunteers were quite heterogeneous, and the data quality varied among different people, which was a critical issue for the reliability of VGI (Flanagin & Metzger, 2008). The content of VGI was provided by its creator without any citation, reference, or other authority, and most volunteers were not trained in or familiar with the intricacies of geographic information generation (Goodchild, 2007). As a result, before using VGI, the data content must be carefully considered in terms of whether there is a straightforward statement of the data source; whether the metadata is open to all; and whether the data is structured, finished, or correct (Brown et al., 2013; Flanagin & Metzger, 2008). According to Goodchild and Li (2012), intrinsic approaches can be used to determine the data quality of VGI in three domains: crowdsourcing revision (assure data quality by various contributors), social measures (estimate data quality by evaluating the contributors), and spatial consistency (analyse the consistency of contributed entities).

It is worth mentioning that although many studies investigated VGI quality, there was still no solid framework to assess crowdsourced spatial data. On the one hand, the nature of VGI can vary if handled by various geospatial experts; on the other hand, geographic information retrieval techniques can extract various VGI content from the internet that was distinct from conventional spatial data (i.e., it was difficult to make a comparison between VGI data and traditional data to assess VGI data quality; Antoniou & Skopeliti, 2015).

2.3.2 Social Sensing

The term 'social sensing' represents a set of individual-level spatiotemporal big data (i.e., GPS trajectory data and social media check-in data) and associated methods and applications to capture human behaviours including activity and movement, social ties, and perceptions in order to detect socioeconomic characteristics of the environment to complement remote sensing data (Liu et al., 2015). Compared with VGI, social sensing emphasised the importance of mining socioeconomic characteristics instead of a new data collection approach (Liu et al., 2015). Furthermore, social sensing can capture people's perception, which was helpful to model both activities and social ties, while some conventional VGI such as POIs and street lines contained little such information. Hence, social sensing was a subset of VGI that was generally related to particular objectives such as disaster response and social networking (Goodchild & Glennon, 2010; Steiger et al., 2015).

Based on the concept of social sensing, several attributes can be listed. First, on the collective level, social sensing data can be used to analyse the geographical influence on the observed patterns, such as research on land use detection (Vyron Antoniou et al., 2016), human-environment interaction (Luo et al., 2016), and place semantics (Hu et al., 2015). Second, the rich temporal information of social sensing data enables the identification of specific events and monitoring of dynamic variations. Examples of previous research include the measurement of urban vitality from dynamic mobility data (Sulis et al., 2018), deriving retail centre locations and catchments based on tweets

posted at a different time (Lloyd & Cheshire, 2017), and linking typical behaviours to observable characteristics in the context of geodemographics from the 24-hour weekday cycle of Twitter (Lansley & Longley, 2016). Finally, social sensing created interactions between individuals and places that helped build spatial networks, such as the works about human mobility (Kou & Cai, 2019; Luo et al., 2016; Steiger et al., 2015).

However, social sensing data could suffer from several issues which required much attention. First, the study area should be discretised into regular or irregular units when analysing the temporal variations of activities (Liu et al., 2015). The local spatial heterogeneity of various activities resulted in sharp distribution gradients if a high spatial resolution was adopted. Hence, a coarser resolution should be chosen to smooth the activity distributions. However, most activities extracted from social sensing data were correlated with population density (Kang et al., 2012), which was quite different between urban areas and rural areas. Therefore, a varying resolution scheme would be more reasonable instead of the regular rasterisation that most existing studies have adopted (Reades et al., 2009; Sun et al., 2011; Toole et al., 2012). Furthermore, the activity rhythms would be different as seasons change, while most research using social sensing data only covered short periods such as one month, which made it less possible to focus on long-term dynamics (Liu et al., 2015).

2.3.3 Social Media Data

Social media data is a special kind of VGI and core data source of social sensing (Goodchild, 2007; Liu et al., 2015). Platforms for social media are virtual communities and digital networks where people build, post, exchange, and comment on content (Ahlqvist, 2008); they include Twitter, Foursquare, Flickr, and Instagram. According to Ahlqvist et al. (2008), social media is built on three pillars: content, communities, and Web 2.0. Users generate different formats of content such as check-in, short text, or images to post on the internet. As such, communities are created so that individuals can communicate with their peers or people who have common interests. Both content

and communities are enabled by Web 2.0 (Goodchild, 2007). People can share information about their activities and interactions in relation to the built environment at any time, and such social media is also known as a participatory sensing system (PSS; Burke et al., 2006).

Social media data was expanded further by introducing a spatial dimension through location-based services, which built connections between cyberspace and the actual physical environment. Geotagged social media (GSM) referred to social media that allowed people to connect geospatial information with shared content; GSM differed slightly from a location-based social network (LBSN; Zheng et al., 2014). An LBSN was a social networking platform that heavily relied on the interactions between individuals and locations (Roick & Heuser, 2013). On the contrary, GSM focused more on human activity and experiences with their surroundings, where the geographic locations of the content can be attached or not. For instance, Flickr was a GSM platform rather than an LBSN platform, while Foursquare was a LBSN which could also be considered a GSM application. Millions of individuals interacted with places in their everyday lives through different forms of GSM, such as uploading geotagged images to Flickr, checking in at a location on Foursquare, or tweeting about a local event.

Researchers examined the general principles and applicability of GSM data in spatial and social sciences (Batty, 2016; Elwood et al., 2012; Silva et al., 2013; Sui & Goodchild, 2011). GSM data was individual-level data that represented one's behaviour at a finer spatial and temporal precision, creating new possibilities for urban study. As a response, it could also be viewed as an extremely comprehensive lens from which to perceive cities (Arribas-Bel, 2014).

2.3.4 Geotagged Flickr Data

Flickr was a website that allowed users to import, browse, archive, organise, and post images and videos from anywhere in the world (https://www.flickr.com/). According

to Goodchild (2007), Flickr was a good demonstration of VGI because it was a platform that helps users locate images on the Earth's surface using latitude and longitude coordinates. Since Flickr was launched in 2004, there have been over 100 million registered users, 10 billion images were shared on the website, and nearly one in every 30 photos was geotagged (Catt, 2009; Smith, 2021). The large numbers of users and available geotagged images showed the consistency and popularity of Flickr.

One of the advantages of Flickr data is that most Flickr images are taken in an urban built environment (Hollenstein & Purves, 2010). Furthermore, Flickr metadata is accessible for retrieval through its public API, and users have been able to access this data from anywhere in the world since 2004 (Hu et al., 2015). This feature makes it possible to unpack people's behaviours and how they evolve, which set it apart from other LSBNs like Twitter and Foursquare, which have limited long-term data. More importantly, Flickr data is more targeted compared to Twitter data. For instance, people at a specific location may tweet random content such as common activities and their emotional expressions, which means that unrelated information was contained at the particular geolocations from which the individuals tweeted. By contrast, Flickr is a platform designed for users to upload and share photos, suggesting that the information contained in photos is more targeted and associated with the actual geographical environment.

Given geotagged Flickr images contained plenty of spatial, temporal, text, and image information, a range of applications were intensively researched aiming to address urban issues. Some explored features of important events such as where and when events took place that using spatial and temporal attributes of the data (Kisilevich et al., 2010; Rattenbury et al., 2007). Similar research focused on landmark detection or POIs that grouped spatial information through aggregations of geotagged Flickr images and found representative photographs based on the text attribute (Crandall et al., 2009; Lee et al., 2014; Papadopoulos et al., 2011; Sun et al., 2015). Further works paid attention to the attractive regions instead of places, which extended the use of similar analysis to bounded areas enclosed by convex hull technique or vague areas

drawn by kernel density estimation approach (KDE; Hollenstein & Purves, 2010; Li et al., 2013). Alternatively, a few other works such as land cover classification (Vyron Antoniou et al., 2016), important places definition (Li & Goodchild, 2012), cultural ecosystem investigation (Hristova et al., 2018), and human-environment correlation (Ahlfeldt, 2013) were also conducted to mine potential of geotagged Flickr images. The methods commonly used in these works included spatiotemporal analysis, semantic analysis, statistical analysis, clustering analysis, and prediction analysis (Ahlfeldt, 2013; Kennedy & Naaman, 2008).

Most of the research conducted relied on spatial, temporal, and semantic attributes of geotagged Flickr images, whereas the content of images was limited mined and utilised which was the key pillar of the data (Lee et al., 2014; Li & Goodchild, 2012; Miah et al., 2017; Sun et al., 2015). Given data text attribute is highly heterogeneous (Goodchild, 2007), the titles and tags embedded in each image were not necessarily associated with the image itself. As a result, the reliability of the data would be lower without looking into the content of images. This work seeks to fill in this gap in terms of the rarely used image attribute of geotagged Flickr images.

2.4 The Potential of Images in Urban-Related Issues

2.4.1 Primary Image Data Sources

An image – an artificial entity that depicts people's visual perception – is generally a photograph or other two-dimensional (2D) picture. Much information could be identified from images when sensing cities: (1) the appearance such as colour, shape, and coverage of the objects or people involved can reflect the beauty of the natural landscape (Seresinhe et al., 2017); (2) the high temporal frequency and spatial resolution can capture certain patterns that are currently difficult to measure in other ways, including crime surveillance and greenery coverage and detection (Collins et al., 2000; Stubbings et al., 2019); and (3) the global coverage enabled forecasting of weather conditions and air pollution (Ferrare et al., 1990; Jedlovec, 2013). These

embedded signals had great potential for mining the socioeconomic characteristics of the cities. However, images were undervalued until the early 2010s when advances were made in image recognition via deep learning and computer vision (Lecun et al., 2015). The diversity of image data sources and various computer vision algorithms allowed an increasing number of researchers to assess a wide range of issues since the 2010s (Dubey et al., 2016; Naik et al., 2014; Richards et al., 2018; Salesses et al., 2013).

There were primarily three types of image data sources to understand cities at different geographical granularities. The first one consisted of remote sensing data from satellite, plane, and drone images, which allowed the understanding of the physical appearance of the built environment from above (Liu & Long, 2016; Zhang et al., 2017). The second category covered street-level imagery and user-generated imagery that can not only capture physical appearance but also provided a direction on how people perceived and experienced the built environment (Doersch et al., 2012; Dubey et al., 2016; Law et al., 2020; Zhang et al., 2018; Zhou et al., 2014). The third type mainly focused on the real estate market, which involved indoor imagery that offered a perspective to uncover household preferences from inside spaces for residence, entertainment, and working (Ahmed & Moustafa, 2016; Zhang et al., 2017). These types of image data sources have been confirmed as applicable in addressing complex research questions in cities through different computer vision algorithms over the last decade.

2.4.2 Dominant Image Recognition Algorithms

Computer vision involved the task of quantifying the representation of visual elements in raw form, where the computer understood a scene from a few presented image samples (Lecun et al., 2015). The dominant technique for computer vision has been convolutional neural networks (CNN) since a significant success during an ImageNet competition in 2012. CNN was designed to process data in the form of multiple arrays, such as colour image data which consisted of three 2D arrays presented as pixel values

in the three colour channels. A typical CNN architecture comprised convolutional, pooling, and fully connected layers, which can be trained differently with different algorithms based on various image recognition tasks (Guo et al., 2016). The dominant CNN algorithms that tackled urban-related issues can primarily be divided into three categories: (1) image classification, (2) image segmentation and localisation, and (3) generative models.

Image classification categorised the content of an image into a single or multiple features from a fixed set of categories trained from an imagery database (Karpathy, 2016). Large numbers of accurately pre-trained CNN models have been developed since 2010; notable examples included AlexNet (Krizhevsky et al., 2012), VGGNet (Simonyan & Zisserman, 2015), GoogLeNet (Szegedy et al., 2015), ResNet (K. He et al., 2016), and DenseNet (Huang et al., 2017). These models were mostly trained on 1,000 object categories from an ImageNet image dataset. The applications of image classification algorithms were mostly twofold: one was to directly extract categories from a pre-trained model, such as recognising scenic categories to quantify the beauty of outdoor places (Seresinhe et al., 2017), and the other was to finetune a few parameters of the layers from a labelled image dataset based on pre-trained models, such as evaluation of the activeness of street frontage (Law et al., 2020) and urban landcover or land use classification (Kang et al., 2018).

Object detection and image segmentation (i.e., semantic segmentation) enabled identifying and localising multiple objects in an image. Object detection identifies different sub-images and generates a bounding box around a recognised object, while image segmentation partitions an image into objects or parts with accurate boundaries (Gandhi, 2018). The notable models for this category included region-based CNN (R-CNN; Girshick et al., 2014), fast R-CNN (Ren et al., 2017), you only look once (Redmon & Farhadi, 2017), and mask R-CNN (K. He et al., 2017). These models were generally trained on the common objects in context (COCO) dataset, a large-scale image dataset with 80 categories released for segmentation and localisation tasks (COCO, 2018). Examples of this type of image recognition task included localising

building polygons in the given satellite images (Zhao et al., 2018); detecting the amount of vegetation from street-level imagery (Stubbings et al., 2019); and quantifying the perception related to safe, lively, boring, wealthy, depressing, and beautiful from a new crowdsourced imagery dataset (Dubey et al., 2016).

Generative models outputted the representation of images in an unsupervised way without being shown labelled inputs, which provided a less data-intensive alternative to CNNs that did not require the assembly of numerous labelled images to train the networks (Comber et al., 2020; Ibrahim et al., 2020). Typical models for this image recognition task included autoencoders, deep belief networks, and generative adversarial networks (GANs; Goodfellow et al., 2016). For instance, GANs model generates synthetic and new graphical data in an unsupervised way, enabling the creation of unique objects or scenes relied on the underlying features of the trained images. Applications for this task included extracting visual features from leisure and retail environments (Comber et al., 2020) and exploring urban forms based on a large set of street networks collected from satellite imagery across millions of cities (Moosavi, 2017).

Although different image data sources and computer vision algorithms have been used to address urban issues related to the built environment, such as urban representations (Zhou et al., 2014), land use classification (Zhang et al., 2017), urban safety (Naik et al., 2014), urban perception (Khosla et al., 2014), and so on, challenges remained in a few aspects. First, most studies were implemented on a global and city scale, while limited research applied and scaled up such algorithms to certain areas of a city (e.g., UAOI introduced in the previous subsection 2.2.3). As a response, representations of small areas in a city were rarely analysed. Second, the task of capturing rapid urban changes (e.g., weekly, or monthly) was complex and remained under explored. This was probably because intensive research relied heavily on street-level images, which was not real-time and was often released at frequencies longer than one year (Google Maps Street View, 2019) Although several satellite images could take images from above every few days, applications in the built environment were generally restricted

in land use classification (Hu et al., 2016; Zhang et al., 2017). The last challenge was algorithm implementation. The CNN algorithms had a few common restrictions, including high computational expense that usually required a good graphic processing unit (GPU) to process the image recognition tasks, and trained on intensive training data to acquire well-fitted models in most cases. As such, this thesis intends to use volunteered geographic image information to capture rapid urban change characteristics at a neighbourhood level through a pre-trained CNN model which was trained on our dataset and has relatively high accuracy (Zhou et al., 2018).

2.5 Summary

This chapter started with reviewing urban perception in Section 2.1, where its concept and development in the last few years was introduced and discussed. The research gaps, that human cognitions into urban perception were underestimated in most cases, were identified and hence the key research question of this thesis was motivated and proposed. Within this theoretical framework, different dimensions of cities were then outlined in Section 2.2 to explain why they were used in this study. To begin with, urban analytics as a new and emerging research field has enabled a number of works to analyse finer spatiotemporal resolution. The research on the relationships between human activity and the built environment, as a result, benefited from urban analytics through identifying more attributes from new forms of data and more novel methods. A specific research orientation termed UAOI was then introduced and stated as an object to explore the correlation between human activity and the built environment, which has the capacity to capture the areal characteristics and long-term geo-temporal dynamics but was overlooked before. The final subsection discussed housing which was a specific embodiment of the study of human activity and the environment. By reviewing previous works and methods issues associated with housing prices, the research opportunities were identified to bring human-generated images into urban perception studies.

The next two sections illustrated new forms of data used in this research from two aspects. Section 2.3 introduced and discussed commonly used terms and notions related to the data, illustrating that this kind of data was using citizens as sensors through how they experienced the environment. Furthermore, data attributes, data quality, data-driven approaches and data applications were listed and reviewed to obtain an implication that user-generated geotagged images have been underestimated but had great potential. The last section of this chapter focused only on image data. A range of widely used image data sources and image recognition models were reviewed and discussed to determine which data source and model would be utilised in this study and why they were selected.

3. Understanding the Dynamics of Urban Areas of Interest through Volunteered Geographic Information

N.B. The research presented in this chapter is an adapted version of the publication:

Chen, M., Arribas-Bel, D., & Singleton, A. (2019). Understanding the dynamics of urban areas of interest through volunteered geographic information. Journal of Geographical Systems. https://doi.org/10.1007/s10109-018-0284-3

Abstract: Obtaining insights into the dynamics of urban structure is crucial to the planning and management of urban areas. This chapter focuses on urban areas of interest (UAOIs), a concept that provides functional definitions of a city's spatial structure. Traditional sources of social data can rarely capture these aspects at scale while spatial information on the city alone does not capture how the population values different parts of the city and in different ways. Hence, we leverage Volunteered Geographic Information (VGI) to overcome some of the limits of traditional sources in providing urban structural and functional insights. We use a special type of VGI - metadata from geotagged Flickr images - to identify UAOIs and exploit their temporal and spatial attributes. To do this, we propose a methodological strategy that combines Hierarchical Density-Based Spatial Clustering for Applications with Noise (HDBSCAN) and the 'α-shape' algorithm to quantify the dynamics of UAOIs in Inner London for a period 2013 to 2015, and develop an innovative visualisation of UAOI profiles from which UAOI dynamics can be explored. The results expand and improve upon the previous literature on this topic, and provide a useful reference for urban practitioners who might wish to include more timely information when making decisions.

3.1 Introduction

As stated in Chapter 1, the rapid growth of urban populations across the globe is resulting in new kinds of technical, physical, material, and social challenges and constraints (Chourabi et al., 2012). To tackle such issues, how to better plan, govern and manage a city to improve its sustainability, optimise processes and maximise the provision of collective public and private services have become a significant urban strategy in many developed and developing regions of the world (Harrison et al., 2010; Washburn et al., 2010; Batty, 2017). In terms of operationalisation, it is often relevant to obtain timely insights into the dynamics of urban population at a temporal granularity finer than that of traditional surveys, which can be enhanced by or provided through digital technologies.

It is within this context that the present chapter engages with the concept of UAOIs, which refer to parts of the urban built environment that can be delineated in their extent through the clustering of human activity. Such areas may contain business zones, tourist attractions, iconic landmarks, recreational zones, or other attractors (Hu et al., 2015). The notion of a UAOI is, therefore, a combination of morphological features including buildings and streets, and ‘Points of Interest’, as defined by the relevance the population concedes specific parts of cities. As such, a UAOI can be viewed as a perceptual space, which is captured by the social morphology of the city, albeit rooted in physical space (Crooks et al., 2016). Accordingly, a UAOI should emerge from the activities of a large collection of different people to avoid very individual conceptions. Furthermore, such definitions are complex, as unlike well-defined geographic divisions or administrative districts, the delineation of a UAOI may vary between people in different contexts, ages, and cultures.

Identifying and understanding UAOIs has applications in multiple fields. For spatial planning, they may assist in identifying areas with greater public priority in the context of limited resource availability (Gandy, 2006). For retailing, they can help identify areas where people cluster, and how these have evolved, which might aid in-store

location or for targeting advertisements more effectively. For transport planning, they may help prioritise traffic flows or the provision of public transport; for statistical agencies, they may provide useful reference distributions in comparison with official geographical divisions.

The challenge of defining UAOIs over time resides in the need for granular spatiotemporal data recorded within cities. Although traditional data sources used in urban studies, such as remotely sensed data, have a lengthy history of application and can be used to characterise urban morphology, they do not capture human dynamics beyond expansion or contraction of the built form. Alternatively, survey or census data might be utilized to inform the discovery of UAOIs, but these are usually costly to administer and may be of limited temporal granularity (Shi et al., 2014; Tasse and Hong, 2014). A third alternative has emerged in the last few years. Several new forms of digital data derived from urban activity through passive or active forms of data collection capture urban form and/or social functional geography (Arribas-Bel, 2014; Crooks et al., 2016). Such data are referred to as Volunteered Geographic Information (VGI; Goodchild, 2007), which includes the use of digital devices by communities or individuals to create, accumulate, upload, and communicate geographic information, typically through contemporary web technology. Commonly designated as VGI is a variety of content from social media networks, which often support geolocation of assets and include networks such as Twitter, Facebook, Flickr, and Instagram. Data derived from these networks have been used in a variety of contexts to explore spatial, temporal, and even semantic information about human activities (Jiang et al., 2015; Lansley and Longley, 2016; Lloyd and Cheshire, 2017; Gao et al., 2017).

In this chapter, we examine the potential of data derived from the online photo management and sharing website Flickr to extract and understand Urban Areas of Interest. Although there are inherent biases associated with geotagged Flickr data, a number of studies have utilised these data effectively to explore various issues within urban contexts (Hollenstein and Purves 2010; Lee et al., 2013; Hu et al., 2015; Gao et al., 2017). Flickr offers an attractive proposition as a data source for a number of

reasons. The scale of the Flickr network is extensive and, as of 2016, Flickr had 122 million users with more than 10 billion images shared, demonstrating a large degree of penetration (Smith, 2016). Secondly, the metadata of each Flickr photo is available through its public Application Programming Interface, which can be retrieved back to 2004, making it possible to consider the temporal dimension of imagery. These features are in contrast to other sources of VGI from social networks, which have rather limited data retrieval limits (e.g., Foursquare only allows one month; Foursquare for Developer, 2017). Finally, studies have suggested that Flickr photos, in most cases, are taken in the urban built environment, and as such, enhance their suitability as a source to identify UAOIs (Crandall et al., 2009; Hollenstein and Purves 2010).

Our goal is, therefore, to present a new method of extracting UAOIs and to provide new insights into their fine-grained spatiotemporal evolution and characteristics. We used geotagged Flickr data from three recent years (2013-2015), and have focused particularly on the seasonal variability of the UAOIs. A recent hierarchical algorithm was used to extract clusters, reducing many of the drawbacks of traditional, previously used, density-based methods. An 'α-shape' algorithm was then utilised to construct boundaries identifying the UAOI extents. Once built, we conduct further analysis on the spatial and temporal patterns associated with the identified UAOIs, and propose an approach to build a spatiotemporal profile for each UAOI.

The structure of this chapter is organised as follows. The next section discusses related work about points of interest and areas of interest, as well as techniques for the analysis of geotagged photo data. Section 3.3 describes the data collection, data bias and pre-processing stages. Section 3.4, the core of the chapter, proposes a methodological framework to extract and understand UAOIs, including an approach to validate the number of Flickr users in the extracted UAOIs. This is followed by a discussion of the spatiotemporal dynamics of the identified UAOIs. Finally, section 3.6 concludes the chapter and suggests future extensions to this research.

3.2 Literature Review

There is a growing body of research that uses geotagged photos, which has examined both the attributes contained within the metadata and the image itself. Most studies have focused on exploring landmark detection involving travel route recommendations, which generally integrate some aspect of movement/trajectory analysis (Zheng et al., 2012; Sun et al., 2015). Alternative approaches examine geotagged photos to address the question of where and when events take place (Rattenbury et al., 2007; Kisilevich et al., 2010b; Papadopoulos et al., 2011). There is also further work that combines both image analysis with exploration of the metadata, and applies it across a range of topics including detecting cultural differences (Yanai et al., 2009), land cover classification and validation (Antoniou et al., 2016), and definition of significant places on the basis of people's interaction with their surroundings (Li and Goodchild, 2012).

The most prevalent use is how geotagged social media are used to extract points of interest are based upon exploiting the locational aspect of semantics (e.g., crowd-sourced tagging). For example, Crandall et al. (2009) presented techniques that can automatically identify popular places through representative images and textual labels from Flickr. Lee et al. (2013) proposed a framework to extract points of interest and their agglomerations from geotagged photos. Andrienko and Andrienko (2013) extended such work through the additional consideration of the time of geotagged social media through a number of space-time visual analytic approaches. Other related work has extended the use of similar analysis techniques to include the exploration of attractive regions. For example, Kisilevich et al. (2010a) proposed a systematic framework for the exploration of points of interest obtained from Flickr and Panoramio, utilizing a convex hull to create boundaries of concentrated areas for visualization. Hollenstein and Purves (2010) also linked the derivation of data-driven density surfaces to the extraction of urban boundaries; this was extended by Li, Goodchild, and Xu (2013) who constructed spatial boundaries using KDE, which was utilized to approximate the number of place occurrences per unit area. In some sense, a generality

between all of these was an aim of creating clusters of geotagged data. However, with limited exception, this line of inquiry has rarely focused on spatiotemporal changes, and acutely so over a multi-year period. Furthermore, although the popular non-parametric density estimation technique –KDE has examples of use to construct and visualise attractive aggregations of points of interest, this approach is not designed to delineate specific boundary lines of clusters, a valuable and necessary feature when identifying areas of interest.

One alternative to KDE is density-based clustering algorithms. This family of techniques has more recently been applied to identify points of interest or attractive areas (Kisilevich et al. 2010a; Kisilevich et al., 2010b; Lee 2013; Andrienko and Andrienko, 2013; Gao et al., 2017). The most widely used approach in this category is DBSCAN (Density-based spatial clustering for applications with noise, Ester et al., 1996), which involves two parameters: the search radius (epsilon) and the minimum number of points (MinPts). Once both are specified, the algorithm identifies clusters of at least MinPts observations using epsilon as the maximum distance for the neighbour search. However, both parameters need to be finely tuned, typically requiring manual experimentation in both cases before an appropriate value can be selected. In addition, DBSCAN only uses a global (single) density threshold to extract a flat partition, which fails to distinguish clusters of different densities. The OPTICS (Ordering Points to Identify the Clustering Structure, Ankerst et al., 1999) algorithm presents an improvement to DBSCAN as it only requires the MinPts parameter to be specified while also producing a hierarchical result. However, this approach still relies on a global density threshold, which is unable to find the most significant clusters based on different density levels (Campello et al., 2013).

A recent application of DBSCAN with particular relevance to this chapter was proposed by Hu et al. (2015), and presented a methodological approach that extracts UAOIs for six cities based on ten years' worth of geotagged Flickr photos. Building on this work, our research provides new insights into spatiotemporal changes in UAOIs by proposing a number of extensions. First, we focus on finer temporal scales,

which allows us to consider seasonal variability. We demonstrate that this degree of resolution matters because it can capture seasonal UAOIs that emerge and disappear rapidly. Secondly, in terms of UAOI discovery, we introduce a more advanced method (Campello et al., 2013) than DBSCAN called HDBSCAN for extracting UAOIs. As discussed later in the methodological framework section, this approach overcomes some of the main drawbacks of other density-based clustering methods. Third, we propose the creation of spatiotemporal profiles based on small-scale geographic areas and use these to quantify the characteristics of spatiotemporal change in the UAOIs.

3.1 Data

3.3.1 Data Description

Greater London was used as the study area because the regional boundaries contain a very large volume of geotagged photographs. Flickr data can be retrieved and downloaded using a public Application Programming Interface (API, https://www.flickr.com/services/api/) through the Python interface (Stüvel, 2016). Among the 10 billion images shared on Flickr, 3.33% contain geographic information (Smith, 2016; Catt, 2009). We used a bounding box to collect all geotagged data uploaded for Greater London. Dates between 1 January 2013 and 31 December 2015 were selected, as these three years have the highest number of Flickr photos since Flickr was launched (Michel, 2017), and the data volume is decreasing due to storage limitation of images for each Flickr user. As this study focuses on the geo-temporal exploration of UAOIs, only locational and temporal metadata were retrieved and used in this study[1]. Our data set contained a total of 1,575,200 entries contributed by 34,615 unique users, with the following attributes: user ID, geographic coordinates, and

[1] The use of only location and timestamp does not mean we do not recognise the value of other forms of metadata such as tags and text, or even content from each photo itself. These are fundamentally different sources of information that do not directly relate to the identification of UAOIs, and is hence beyond the scope of this paper. Their use is illustrated, for example, in Crandall et al. (2009), Li, Goodchild and Xu (2013) and Gao et al. (2017).

timestamp (i.e., the time when the photo was taken). Table 3.1 displays an extract from the data.

Table 3.1 A sample of georeferenced Flickr metadata in London

User ID	Latitude	Longitude	Date
1	51.507577	-0.099349	2015-01-19 19:40:42
2	51.500504	-0.127419	2015-03-07 15:36:50
3	51.499434	-0.163905	2013-11-29 16:00:27
4	51.51353	-0.113193	2013-03-13 12:30:09
5	51.500461	-0.138487	2013-03-09 12:23:55
6	51.516541	-0.097525	2013-01-21 08:06:11
7	51.530199	-0.125688	2014-03-30 15:27:43
8	51.55811	-0.282823	2013-01-26 21:37:44

As with other geotagged social media data, the Flickr sample had a few issues related to data quality. Data quality has been defined as data that are fit for use by data consumers (Wang and Strong, 1996). In our case, “fit for use” involves allowing us to identify the spatial dynamics of the urban population. Data quality cannot be assured since it varies among different contributors, leading to data sources that are quite heterogeneous (Goodchild, 2007; Imran et al., 2015). For example, as Flickr provides users with manual geotagging; the photo might be geotagged in a place by one user that differs from where it was taken in practice. In this case, the results would not be able to accurately identify areas of interest. In addition, social media use is self-selecting; users may not necessarily be representative of everyone who lives in or visits a city. For example, the primary user age group of Flickr is between 35 and 44 (Kahootz Media, 2018). In addition, usage of the Flickr service is rather uneven, with more active users contributing to a larger number of photos (Davies, 2016; Li, Hollenstein and Purves, 2010). Such issues imply that research results may be focused on a particular segment of the population, and thus warrants caution when drawing conclusions. However, the degree of penetration and popularity of Flickr is such that we argue our results are still meaningful and can help us better understand urban dynamics from the perspective of people who experience the city through these lenses.

3.3.2 Data Pre-processing

Data obtained directly from the API was preprocessed before analysis in two main stages: (1) subdividing the data set, and (2) eliminating noise.

A visualisation of the spatial distribution of the downloaded geotagged photo locations in London can be seen in Figure 3.1, where KDE (O'Sullivan and Unwin, 2014) is applied. The darker the red the higher the density, thus implying that more photographs were taken in central London relative to peripheral areas. Indeed, 73.5% of our Flickr data are located within the Inner London[2] definition. Thus, the study extent was narrowed to that of Inner London. In terms of the time dimension, our interest is set on the seasonal variability at the monthly level. Thus, we further divided the data into 36 monthly slices covering the periods between the first and last day of each month.

[2] The Inner London definition comprises a series of centrally located London Boroughs within the Greater London Authority Extent (Mayor of London, 2017)

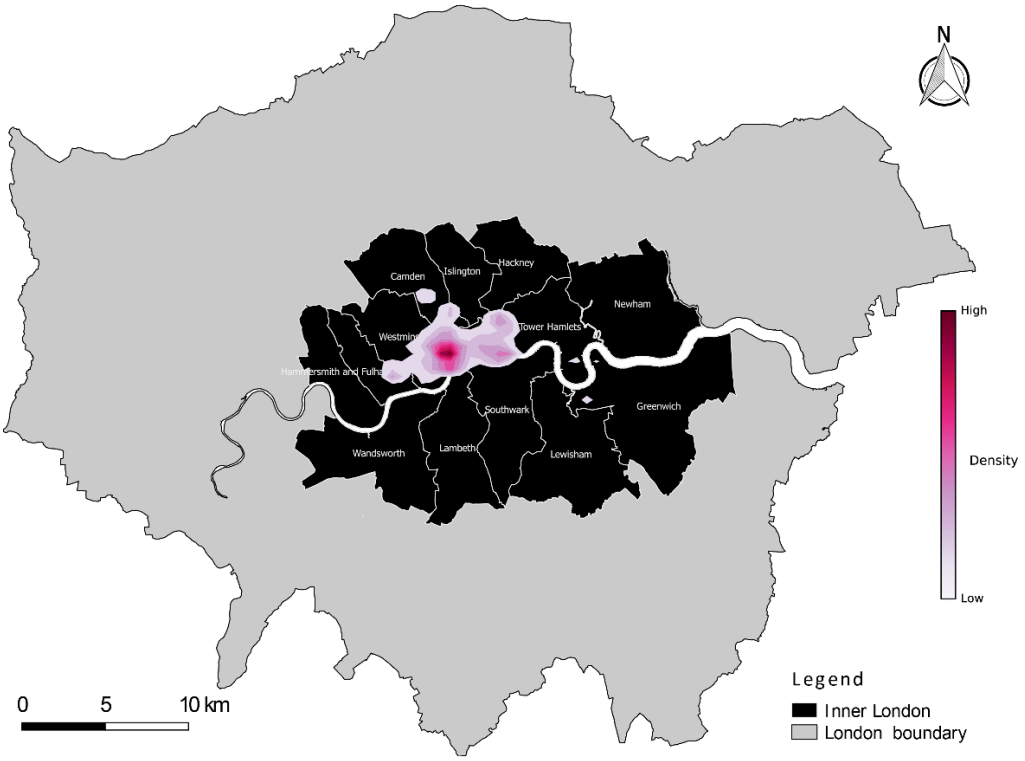


Figure 3.1 Distribution of the spatial density of Flickr photos in London from 2013 to 2015 using a kernel density visualisation

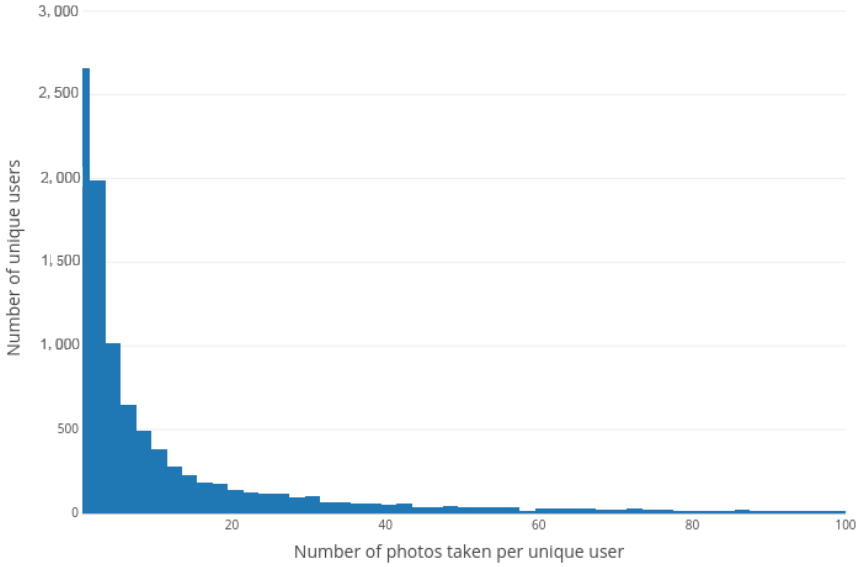


Figure 3.2 Relationship between the number of Flickr photos taken by each unique user and the number of unique users in London (2013-2015).

Next, we needed to identify erroneous or noisy records. First, we considered those cases where a user uploaded a few photos at identical geographic locations (i.e., at least two photos geocoded with the same longitude and latitude by the same user) in a month. Many photographs at the same longitude and latitude (given the recorded degree of precision of the coordinates shown in Table 3.1) are quite unlikely, and as such are classified as erroneous in terms of the location attribute, perhaps as a result of faulty hardware. To remove this effect, only one record for each of these users was maintained. A similar case arises when a user takes multiple photos in the same second of one day at different places; we also removed these cases. More importantly, Figure 3.2 shows that a small group of users contributes a large proportion of the photos. These are known as 'active users' by Hollenstein and Purves (2010) and Hu et al. (2015). The figure shows that a single user may upload hundreds of photos in a year, while most users only upload dozens. In our definition of a UAOI, we argue that these entities should be agreed upon by many people, and the dominance of an active user may lead to bias in extracted UAOIs. An overemphasis on contributed content from any one user or subset of users (and their associated interests) will also influence the generality of the definitions of the UAOIs. To reduce the impact that such active users may have on shaping the outcomes of the analysis, we implemented a further set of cleaning routines that reduced the proportion of photos from active users, by keeping only one photo for each user based on tags used and the time when the photo was taken. Specifically, if a user took several photos in a minute but with the same tags, only one photo was retained. The rationale for this approach was to remove photos within a limited spatial extent on the hypothesis that people's average walking speed is 5km/h (Onaverage, 2017). On this basis, the maximum walking distance within a minute is approximately 83m. Within this distance, only a single user's photo that has the same text is retained. The specific data pre-processing steps are summarised in Table 3.2, showing how many photos and users are removed following each step. After this process, an average number of 12,228 photos and 2,275 unique users remained in each month.

Table 3.2 The number of photos and users at different stages of data pre-processing

(1) Subdividing dataset	The total number of photos			The total number of unique users		
Raw data collected in bounding box	1,579,694			39,531		
Data within Inner London	1,162,891			34,700		
(2) Eliminating noise	The number of photos in each month			The number of unique users in each month		
	Mean	Median	Standard Deviation	Mean	Median	Standard Deviation
Subdivide data into 36 time slices	32,221	32,991	5,070.98	2,287	2,248	402.16
Remove photos geocoded with identical coordinates of one user	16,116	15,899	2,618.94	2,287	2,248	402.16
Remove systematic outliers	15,493	15,595	2,342.96	2,275	2,244	393.02
Remove dominance of active users	12,228	11,913	1,794.09	2,265	2,233	392.69

3.4 Methodological Framework

In the following section, we present a systematic framework designed to extract and map the evolution of UAOIs from the subset of geotagged Flickr photos outlined in the previous section. Our methodology consists of two main parts: cluster detection and boundary delineation.

3.4.1 Extracting Urban Areas of Interest by the Hierarchical Density-Based Spatial Clustering for Applications with Noise algorithm

We define UAOIs as those areas where multiple Flickr users have gathered and taken large numbers of spatiotemporally clustered photos, reflecting a consensual view that some aspect of the urban environment is of interest. The extraction of such areas can be understood as a clustering problem, in particular, as one that has the aim of identifying robust, non-overlapping, and dense concentrations of points. Following recent advances in the literature, we selected a density-based method. The advantages of such an approach are that they can produce results without pre-specification of cluster frequency, and are robust to arbitrary shapes and the presence of outliers/noise deviating away from the main spatial distribution (Hans-Peter et al., 2011).

We applied the HDBSCAN (Hierarchical Density-Based Spatial Clustering for Applications with Noise; Campello et al., 2013) as our clustering method as this overcomes several of the major drawbacks of other density-based algorithms, which fails to distinguish clusters of different densities that only use a single density threshold. Contrary to more traditional algorithms, there is only one parameter to tune in HDBSCAN, with the other key parameter in the original DBSCAN implementation, i.e., the minimum cluster size (MinPts), which is endogenously determined by the method. This approach represents a step forward in the direction of more robust, automated, and data-driven techniques for the delineation of UAOIs. McInnes et al. (2017) describe the HDBSCAN process as comprising five steps:

1) Transform the space based on the estimates of density by defining a 'mutual reachability distance', which is a new distance metric to spread apart points with low density;

2) Build a minimum spanning tree to implement single-linkage clustering, which is a core feature of this algorithm;

3) Construct a cluster hierarchy of connected components by iteratively sorting the edges of the tree by distance in increasing order. The result can be viewed as a dendrogram that shows where robust single linkage stops;

4) Condense the cluster hierarchy shown in the dendrogram into a smaller tree by attaching more data to each node;

5) Extract clusters that persist and are robust from the condensed tree.

Operationally, various epsilon values are generated automatically by the different density levels resulting from the single-linkage hierarchy, which allows HDBSCAN to find clusters of various densities. Also, it ensures improvements over OPTICS and DBSCAN by providing a clustering hierarchy, where a simplified tree of the most significant clusters (i.e., maximised stability) can be easily extracted.

When using HDBSCAN, the only parameter to specify is the minimum cluster size (mclSize), representing the minimum number of points (i.e., Flickr photos) required for a UAOI to exist. In order to select an appropriate mclSize, we extensively explored the sensitivity of the final solution to changes in the parameter. A few thresholds representing the minimal number of Flickr photos, from 10 to 1,000, were set as the minimum cluster size (mclSize) parameter, which was applied in all time slots. Figure 3.3 presents example outputs from this sensitivity analysis. We can see that if the mclSize is small (e.g.,10 or 50), more UAOIs are identified but there are also greater numbers of points labelled as noise (i.e., not part of any clusters); if the mclSize is larger (e.g., 500 or 1,000), more robust results emerge, although clusters are significantly larger, causing potentially interesting smaller areas to be missed. Furthermore, due to the number of Flickr photos and users varying between months, it could be argued as being inappropriate to assign an absolute value for all time sequences. To handle these issues, values of 1% to 4% of the Flickr photos in each month were assigned to mclSize across different iterations as discussed previously in order to produce appropriate frequencies of groups that fit the definition of a UAOI. After multiple experimental results, 1% of Flickr photos in each month were used as the value for the minimum cluster size parameter, ensuring a higher number of UAOIs but also being cognizant of smaller clusters that may be of relevance.

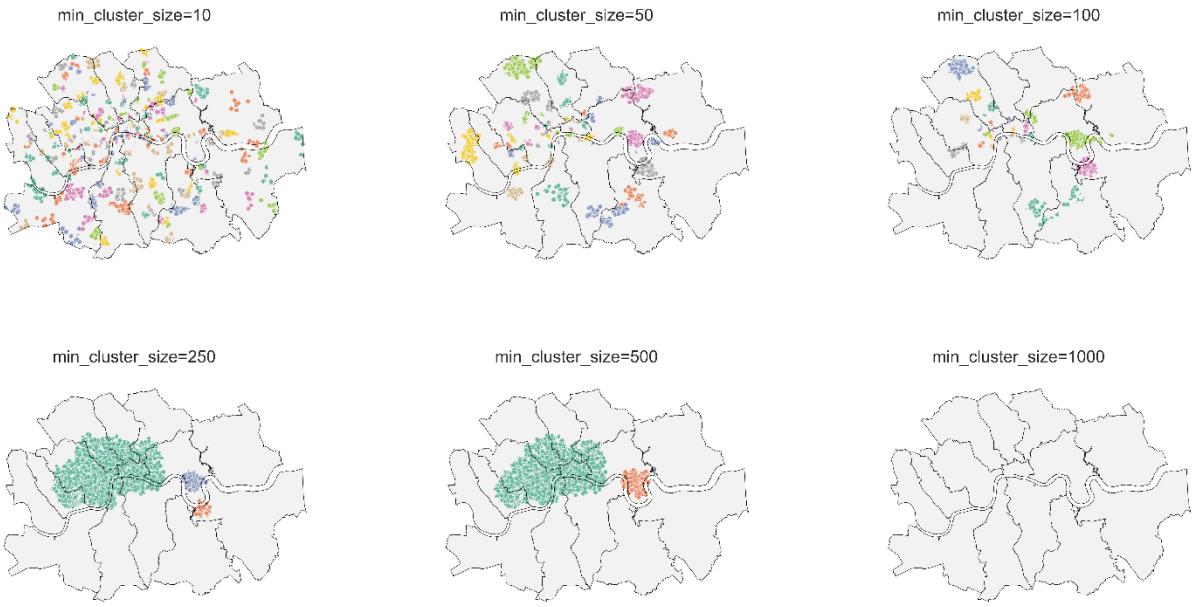
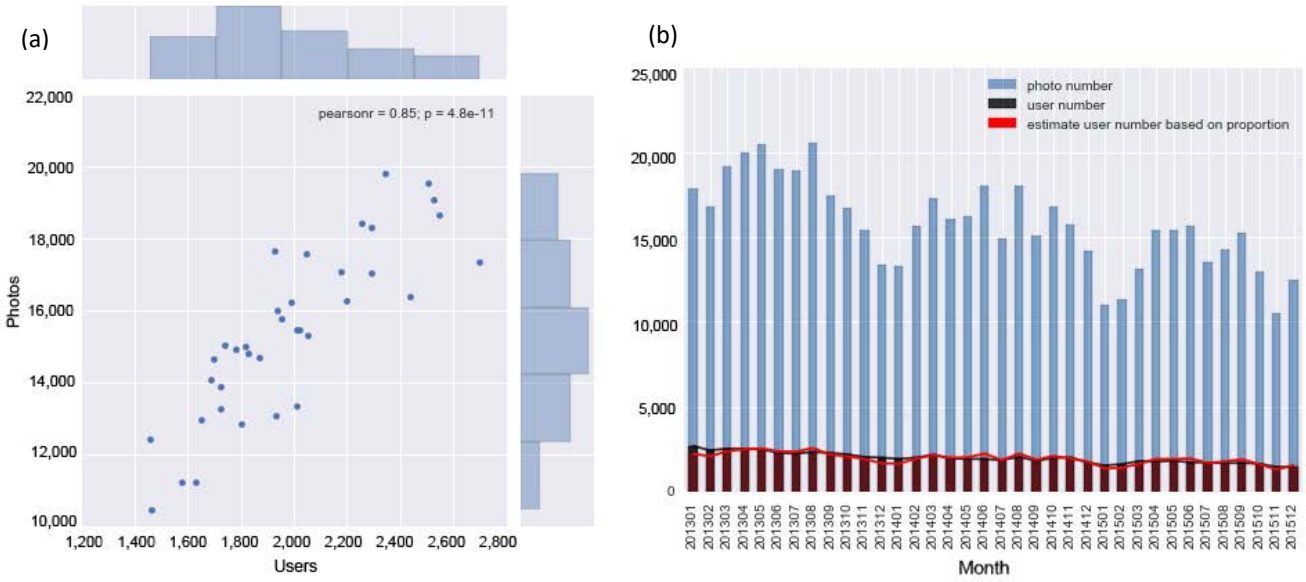


Figure 3.3 Different urban areas of interest extracted by different minimum cluster size (min_cluster_size) values in one month. Colours indicate the location of clusters.

As UAOIs should be formed through the collective actions of multiple users within each specific time slice, the 1% parameter selection does not ensure that a set number of Flickr users are captured in each UAOI. As such, it was then necessary to verify the practical significance of the extracted UAOIs. An intuitive approach is to examine the relationship between the number of Flickr photos and the number of users in each month. If they are correlated, then we can estimate the number of Flickr users by the number of photos per month. Specifically, the scatter plot in Figure 3.4 (a) shows that there is a high positive correlation between the two variables, with a Pearson coefficient of 0.85, implying that as the number of photos increases, so too does the number of users in a given UAOI. A linear regression model was then fitted using these two variables so that the user number could be estimated based on the number of photos in each month. The resulting R-Squared was 0.725 with a p-value for the coefficient value below 0.05, implying that the model is statistically significant and 72.5% of the variation in photo numbers could be explained by the model. Figure 3.4 (b) is a graph presenting the number of photos, users, and the calculated user number in various time sequences. The red line fluctuates slightly around the black line,

meaning that the 1% photo number as the HDBSCAN parameter value can be interpreted as having at least 1% of users in each UAOI, which satisfies our definition of a UAOI. Therefore, we adopt these clustering results for the next stage of the analysis, which turns clusters of points into polygon boundaries.

Figure 3.4 Exploring the relationship between Flickr photographs and users to ensure each urban area of interest contains multiple users. (a) Correlation analysis and (b) estimated proportion



3.4.2 Constructing a Perceptual Boundary to Enclose the Extracted Urban Areas of Interest

The clusters from the method described above are represented as a group of points. However, within this study, we are interested in extracting largely non-overlapping UAOIs that refer to an area within a specific border. In other words, we are interested in identifying polygons rather than sets of points. The reason behind this procedure is twofold. First, as mentioned when introducing the concept, a UAOI was defined as a section of the city with an extraordinarily large density of images. Under this definition, two overlapping UAOIs would simply be merged into one. Secondly, our focus is to quantify spatiotemporal changes in the shape and extent of these polygons. In this

context, even though a UAOI is identified with fixed borders at each point in time, its definition over time is much vaguer and is allowed to change, evolve, and morph in line with changes to its underlying structure.

As such, the next step involves the construction of boundaries that enclose all geotagged images identified as part of a UAOI cluster. To delineate these shapes, we adopted a variant of the concave hull algorithm: the alpha shapes (Edelsbrunner, Kirkpatrick & Seidel, 1983). Alpha shapes are a widely used, robustly tested algorithm that creates a tighter boundary as compared to the traditional convex hull method, which may produce large empty areas that do not belong to the original point data set (Akdag et al., 2014).

An alpha shape, which is a geometric concept, is a linear approximation of an original shape. It is a generalisation of the convex hull, and a subgraph of the Delaunay Triangulation (Edelsbrunner, Kirkpatrick & Seidel, 1983). It establishes a connection between each point and nearby points and removes the furthest triangles that are away from their neighbours. In this context, α is a parameter that controls the desired level of detail, ranging from the standard "crude" convex hull ($\alpha = \infty$) to the set of points itself ($\alpha = 0$, Da, 2018). The algorithm first computes a Delaunay triangulation of the set of points (S) and for each Delaunay edge, it computes the values α-min (e) and α-max (e). Next, for each edge e, if α-min (e) $\leq \alpha \leq \alpha$-max(e), the edge is kept in the α-shape of S. We have tailored this general method to our application by developing a technique to find the most appropriate alpha value for each cluster. Like the parameter selection in HDBSCAN clustering, an absolute alpha value for all point clusters would not be suitable in that some areas would contain more empty areas in the range from 0.001 to 0.005. We then identified the first case where a single point was excluded from the main polygon, and selected the previous value of alpha. This strategy resulted in the tightest polygon that still contained every point in the cluster. As an illustration, Figure 3.5 represents three examples of different UAOIs produced with varying alpha values. In this case, 0.003 excludes a point (which in the original algorithm is still linked through an edge, but not an area), and 0.001 implies too sparse a solution

compared to 0.002, which allows a tighter shape that still includes all points in the cluster within the same polygon. Hence, the value selected for this case is 0.002.

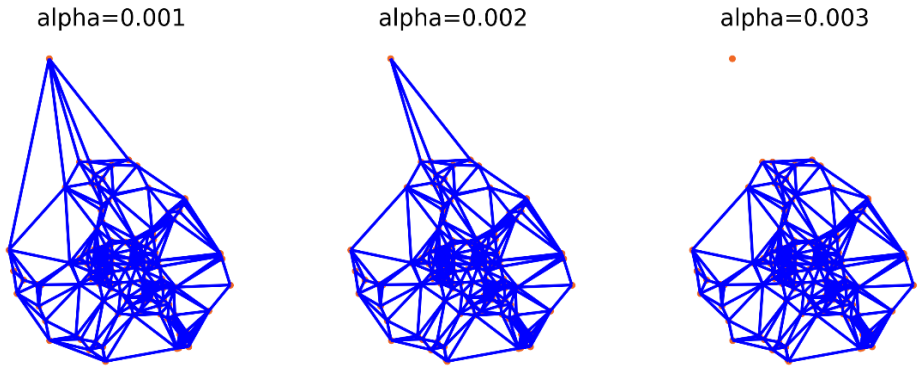


Figure 3.5 An example of one Urban Area of Interest that changes with different alpha values for one month of data

3.5 Results

After applying the method described above to the geotagged Flickr photo data set of inner London from 2013 to 2015, the UAOIs were extracted for the 36 monthly slices to capture spatiotemporal characteristics presented in this section.

We begin from a purely spatial perspective, “compressing” the temporal dimension. This approach allowed us to gain an idea of the stability of different parts of the city in being identified as UAOIs. Figure 3.6 presents each UAOI together in a single map. Figure 3.6 is produced by overlaying all UAOIs from different time sequences with a large degree of transparency to show the spatial distribution of the more stable UAOIs. Areas in darker pink are thus consistently identified as being of interest during the three-year period, including Trafalgar Square, St. Pancras International and tube station, King’s Cross, Jubilee Gardens, Westminster Pier, Borough Market, Millennium Bridge, Tower Bridge, the Canary Wharf financial centre, and the museums located on Museum Lane. These represent popular tourist attractions, cultural venues, business centres, and locations with intense traffic

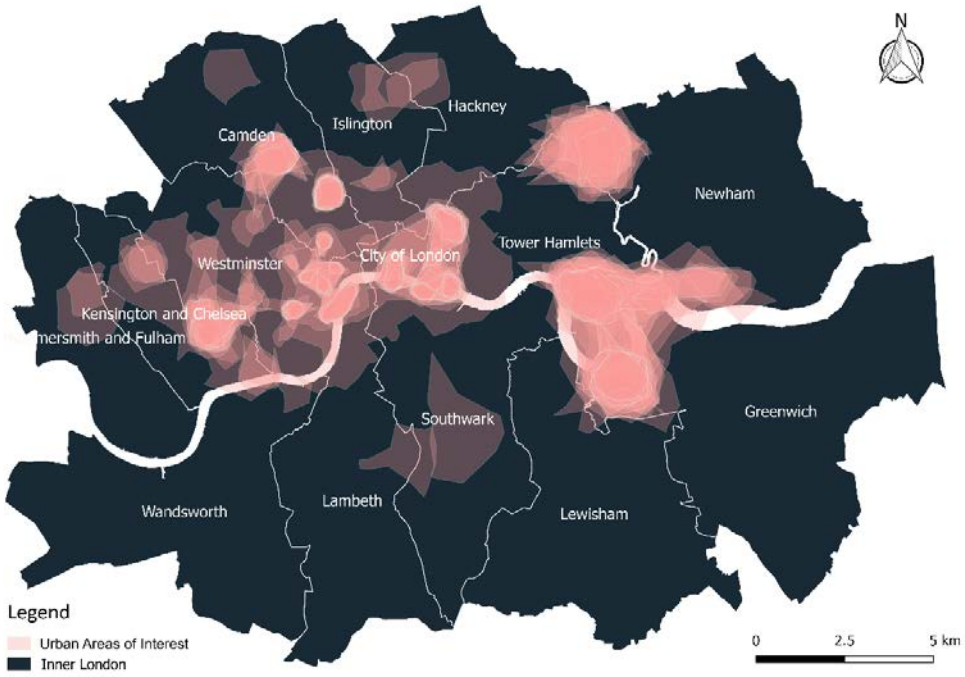


Figure 3.6 All Urban Areas of Interest extracted in inner London from 2013 to 2015 showing the most stable and popular spatial zones

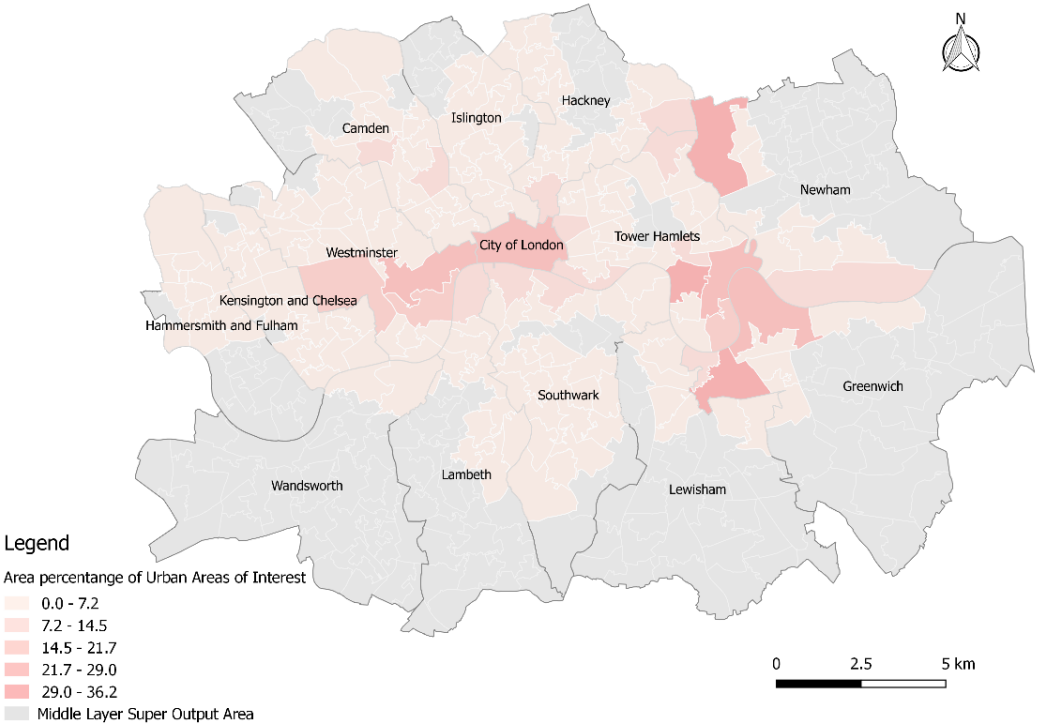


Figure 3.7 The overall spatial distribution of the total area of the Urban Areas of Interest in each Middle Layer Super Output Area

Figure 3.7 is generated by aggregating the results of our analysis at the administrative boundary level, i.e., the Middle Layer Super Output Area (MSOA). MSOAs are designed to improve the reporting of small area (neighbourhood) statistics and are built from a hierarchy of Output Areas (OAs; Office for National Statistics, 2018). These areas are intermediate in size between Output Areas and local authorities. Our intention with Figure 3.7 was to transfer the extent to which a given part of the city belongs to a UAOI into fixed geography that can be analysed over time. The map displays the total area identified as a UAOI in each MSOA over the entire period considered. The map effectively represents those small-scale areas that are more popular, shifting the attention from the organically evolving shapes of UAOIs to the more stable boundaries of MSOAs. The overall pattern displayed is similar to that in Figure 3.6, showing higher values in the northwest of Newham, the border of Tower Hamlets and Greenwich, the City of London, and the middle of Westminster borough, implying a higher degree of attention in these districts.

Although by the nature of the analysis and the source of data employed, it is very hard to carry out a formal validation of the results, the patterns displayed in Figures 3.6 and Figure 3.7 are well aligned with established knowledge from the literature. Both maps result from the interaction between the urban built environment and human behaviour, and highlight popular areas generally covering business centres, public entertainment (theatre, Art Centre, and Sports Centre) and food markets, as well as open spaces. They also illustrate that people are more likely to take photos in those regions where most of the significant landmarks and unique buildings are located. A good example is the City of London, which contains a historical centre with historical buildings as well as modern skyscrapers, and serves as a central business district. We can also see that the districts on the border of Tower Hamlets and Hackney are not always identified as part of a UAOI, which suggests that the degree of popularity of these districts is influenced by different factors and may vary seasonally.

The temporal nature of UAOIs is explored in Figure 3.8, which shows how their extent changes during a single year (i.e., 2013). We can see that some UAOIs emerged and

disappear suddenly in the span of one or two months, which indicates that there is a high probability that large-scale but temporary events took place in these areas. For example, the UAOI extracted in the north of Camden existed only in January and February and then disappeared during the following months. This is likely caused by the first snowfall in London in January 2013, as Hampstead Heath is known as a good place for people to enjoy snow by sledging, activities that are usually recorded in photographs. This event was reported in multiple media (Emms, 2013; Pettitt, 2013).

Figure 3.8 The spatiotemporal evolution of Urban Areas of Interest in 2013

Although useful, it is difficult to scale the spatiotemporal variation in Figure 3.8. Every additional month involves a full map, and comparing a large number of maps at the same time carries a large cognitive load. To be able to extend the analysis and consider the entire period of three years at a fine temporal resolution, we created area profiles for stable geographical entities. We designed this approach to avoid directly examining and comparing the shape of each UAOI over time, as it is difficult and unintuitive to track and follow change with such an approach. Because of their organic and rapidly evolving nature, their shape and extent may vary significantly over time. This makes consistent temporal analysis complicated if the original shapes are to be used. For this reason, we returned to the MSOAs. Area profiles are a series of time plots that display, for every MSOA, the percentage of the area that is considered part of a UAOI in a given month. These figures are able to intuitively summarise the degree of participation of a given MSOA in UAOIs, as well as their evolution over the period considered, jointly capturing space and time in a single figure. To put this profile into context, the time plot is complemented with a map that shows the location of the area considered.

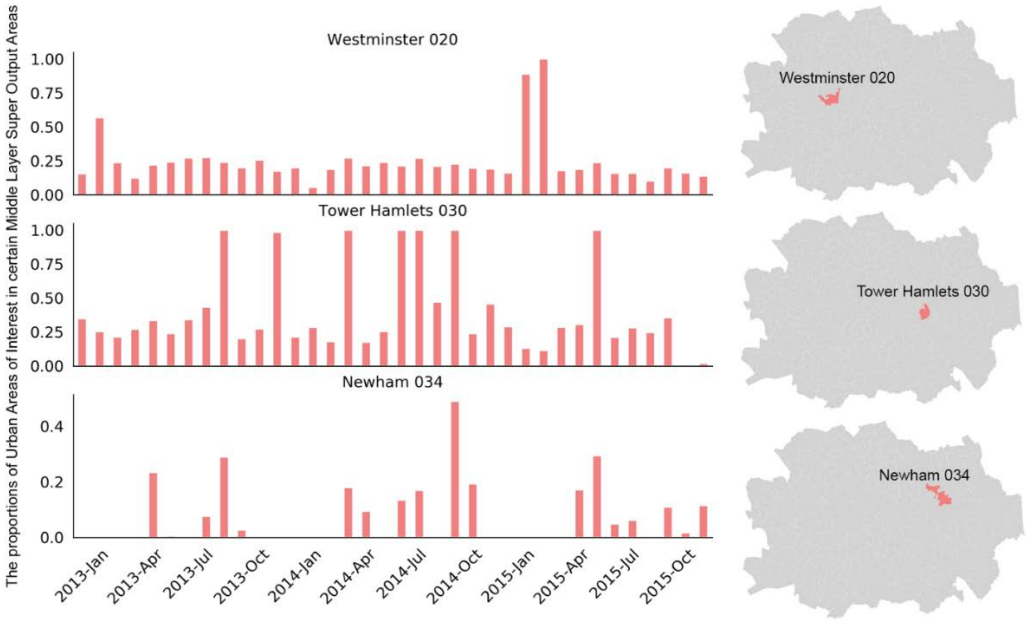


Figure 3.9 Spatiotemporal profiles for Urban Areas of Interest based on Middle Layer Super Output Layer geographic areas

Figure 3.9 shows the UAOI profiles of three MSOAs with distinct characteristics throughout the three years from January 2013 to December 2015. These spatiotemporal profiles can thus help stakeholders better understand the dynamic characteristics of these districts when, for example, allocating resources more effectively, or enhance their understanding of the seasonal interest in specific geographic areas of the city.

The first profile corresponds to an area in Westminster. The profile clearly shows a seasonal evolution, oscillating around 15-20%, with higher percentages in warmer months (June, July and August), and lower participation in UAOIs in colder months. In addition, there are also three outliers corresponding to February 2013, and January and February 2015, which display a larger share of the area being part of a UAOI. In particular, the 2015 outliers reach the full extent of the MSOA. It is hard to tell why these occurred, and in-depth exploration of each of these warrants further research (e.g., semantic analysis or image recognition), which is beyond the scope of this chapter. However, what they help to highlight is the ability of the profile to make these patterns explicit and alert the analyst about their existence in a way that traditional maps do not. The ability is even clearer if we consider the profile of the area in Tower Hamlets. In this case, the seasonal variation is more pronounced, moving from about 20% to the entire coverage of the MSOA. These spikes are not necessarily outliers, as they occur in each of the three years considered during the warmer months. The only one that could be considered an anomaly is that of March 2014, which took place at a time outside the summer period. Equally, the MSOA was not part of any UAOI during November and December of 2015 which, compared to the previous years, was expected. Again, these patterns warrant further research to explore the drivers behind them, but the role of the profile in highlighting them is clear. Finally, the third panel in Figure 3.9 shows a different type of area. The Newham example displays several months in which the area is not part of any UAOI. However, the spring and summer months see it consistently having around a third of its extension within an identified UAOI. This pattern implies that the popularity of this district is significantly

influenced by season and its role in the overall hierarchy is less prominent than that of the other two areas considered here.

3.6 Discussion and Conclusions

This chapter provides insight into several questions relevant for research concerned with VGI as a means of better understanding urban environments. We propose a framework to extract UAOI boundaries from geotagged image data, and use them to build spatiotemporal profiles of areas. When compared to existing literature, our approach is distinct in two key dimensions. First, we introduce the use of the recent HDBSCAN clustering algorithm, which we show improves on the results of other commonly used density-based algorithms employed in previous studies (Kisilevich et al. 2010; Hollenstein and Purves, 2010; Li, Goodchild, and Xu, 2013; Lee 2013; Gao et al., 2017). Second, our approach is significantly more detailed in terms of temporal resolution, which allows us to characterize areas based on their seasonal profiles. This again brings a new perspective to previous approaches (e.g., Andrienko and Andrienko, 2013; Hu et al., 2015), which focus on coarser temporal scales.

The results on the spatial dimension of the analysis suggest that the urban environment influences human activity, shaping the attention of people and attracting them to areas where many unique buildings and important landmarks are located. Conversely, the temporal aspect of the results reflects how human activity changes and shapes the use of the urban environment. Putting these two together, our spatiotemporal profiles visualise how the popularity of certain regions is influenced by factors such as time of the year and season, and also make visually explicit how popularity levels differ across areas. This approach is distinct from related works that use VGI to study UAOIs, such as Hu et al (2015), in that our perspective is more granular and thus allows us to uncover qualitatively different types of dynamics. Spatially, we are focused on the internal dynamics of urban environments and comparing areas within the same city. Temporally, we use the higher resolution to consider seasonal changes, rather than longer-term evolution.

The methods and results presented in this chapter are of interest for several fields and domains. For example, it can help urban planners to develop better strategies related to tourism planning. If certain tourist attractions showed a seasonal pattern according to the spatiotemporal profiles produced in this study, urban planners could allocate resources for tourism more efficiently. Local authorities may also be interested in those UAOIs that are the most stable and have a larger area throughout the year for purposes such as police patrol and traffic monitoring. The results can also be used by researchers and practitioners as an additional geographic layer to understand the use of the urban built environment. Furthermore, part of the relevance of our contribution lies in the fact that it can be deployed using data that are available in near real-time. Unlike more traditional data sources, geotagged images are constantly added to services such as Flickr, thus providing an opportunity to study the evolution of UAOIs not only retrospectively but as they evolve over time. This holds distinct value for practitioners such as urban planners and policy makers.

There are several avenues towards which the work presented in this chapter could be extended. Although the data set used here is extracted from Flickr, geotagged images from other websites could be used. Different platforms provide slightly different services that attract different populations (Lazer & Radford, 2017). Incorporating different sources would thus likely improve the coverage of the analysis and provide a novel comparison of the inherent biases of each platform. Furthermore, the algorithm used to cluster UAOI could be improved by using Spatio-temporal DBSCAN to visualise clusters in a space-time cube rather than just take into consideration of spatial information. Besides, a few approaches could be used to evaluate the clustering results to increase the reliability. Additionally, our current focus has been on the spatial and temporal aspects of the images. A promising further avenue for research is to include information in the analysis other than spatiotemporal stamps such as, for example, the text included in tags, or the images themselves. The former would expand existing work on semantic ontologies (Kisilevich et al. 2010; Lee et al., 2014), while the latter would complement recent advances on deep learning that aim at extracting features

from images (Krizhevsky et al., 2017; Zhou et al., 2017; Redmon and Farhadi, 2017). Finally, this analysis could also be further extended by considering the socioeconomic characteristics of Flickr users, seeking to establish a link between e.g., Flickr metadata and census data. These applications, although beyond the scope of this present chapter, warrant future attention by researchers.

4. Quantifying the Characteristics of the Local Urban Environment through Geotagged Flickr Photographs and Image Recognition

N.B. The research presented in this chapter is an adapted version of the publication: Chen, M., Arribas-Bel, D., & Singleton, A. (2020). Quantifying the characteristics of the local urban environment through geotagged Flickr photographs and image recognition. ISPRS International Journal of Geo-Information, 9(4). https://doi.org/10.3390/ijgi9040264

Abstract: Urban environments play a crucial role in the design, planning, and management of cities. Recently, as urban population expands, the ways in which humans interact with their surroundings has evolved, presenting a dynamic distribution in space and time locally and frequently. Therefore, how to better understand the local urban environment and differentiate varying preferences for urban areas have been big challenges for policymakers. This chapter leverages geotagged Flickr photographs to quantify characteristics of varying urban areas and exploit the dynamics of areas where more people assembled. An advanced image recognition model is used to extract features from large numbers of images in Inner London within a period of 2013-2015. After the integration of characteristics, a series of visualisation techniques are utilised to explore the characteristic differences and their dynamics. We find that urban areas with higher population density cover more iconic landmarks and leisure zones, while others are more related to daily life scenes. The dynamic results demonstrate that season determines human preferences for travel modes and activity modes. Our study expands the previous literature on the integration of image recognition methods and urban perception analytics and provides new insights for stakeholders, who can use these findings as vital evidence for decision making.

4.1 Introduction

Urban environments play a crucial role in decision making in terms of the design, planning, and management of cities, which are closely linked with urban functions and their ecosystems. From a social perspective, understanding how humans experience these environments is important for improving urban functions. For example, areas with a large population density and exposure require more attention and in-depth strategies. In recent years, as the urban population has expanded, how humans interact with their surroundings has evolved (Singleton et al., 2018). The distribution of the population has changed over space and time locally and frequently.

Traditional approaches to understanding the urban environment have relied on survey data. These approaches can be used to characterise urban morphology, but they can generate gaps between data collection and data quality that are costly and problematic (Stubbings et al., 2019). Although recently emerging street-level imagery data can overcome these gaps, these data are mostly from Google's street view fleets, which rarely capture human perceptions of the urban environment. Therefore, challenges remain for policymakers to plan and manage urban environments. In the past few decades, improvements in location technology, such as the global positioning system (GPS), have produced plenty of georeferenced urban data sources (Arribas-Bel, 2014), such as social media data and mobile data. In addition to geographic information, many of these new forms of data also have other attributes, such as time, user profiles, user evaluation, or user photographs, providing great opportunities for research in social and urban domains (Hollenstein & Purves, 2010). Among these attributes, photographs offer a wealth of information on the environment that can be analysed to determine why and how humans interact with urban areas (Dorwart et al., 2010). However, previous research on the content analysis of photographs is relatively rare (Chen et al., 2019; Crandall et al., 2009; Y. Hu et al., 2015a; Kisilevich et al., 2010).

Recently, thanks to advances in computer vision and deep learning techniques, especially improvements in convolutional neural network (CNN) performance, images

have gradually been proven to be powerful for investigating the visual perception of our environment (Naik et al., 2017; Seresinhe et al., 2018; F. Zhang, Zhang, et al., 2018). Since the early 2000s, CNNs have been applied to image recognition but were neglected until a big success during an ImageNet competition in 2012 (Lecun et al., 2015). CNNs have since become the dominant method for all image recognition tasks.

Drawing on the limited research of dynamic urban perceptions and the ongoing improvements in image recognition performance, this chapter focuses on urban areas of interest (UAOIs) and their outer urban environments. An UAOI is a perceptual space captured by the social morphology of the city, which reflects the real interests of large numbers of people and may emerge and disappear at different times (Chen et al., 2019; Crooks et al., 2016). A UAOI is not only a perceived region of a place but an outcome of human interactions with the environment. More importantly, many geotagged photographs that represent the physical appearances of UAOIs are available. As such, research on UAOIs offers a way to explore the connections between human cognition and digitally and visually represented geographies.

The objective of this study is to quantitively formalise and understand urban areas through geotagged images. Not only do we analyse photographic metadata, but we also exploit information from the images themselves. Also, dynamic analysis is considered, which bridges a research gap. The research questions are proposed as follows: 1) Why do people gather at certain areas all year or at certain times? 2) Is there any difference between UAOIs and other areas? 3) What are the visual characteristics of UAOIs over time? We first extract the UAOIs in Inner London through a method framework proposed by (Chen et al., 2019); then, an advanced and novel CNN model called Places365-CNN is utilised to extract features inside and outside the UAOIs. These features are then integrated to explore the regular characteristics of the urban environment. Finally, a finer temporal scale is applied to understand the dynamic characteristics of the UAOIs through a heatmap based on a z-score.

The structure of this chapter is organised as follows. The next section discusses the past and recent work using geotagged images in urban analytics, as well as common techniques of image recognition in this domain. Section 4.3 introduces the methods used to characterise UAOIs, including data description, the CNN model, characteristic integration, z-score standardisation, and heatmap analysis. This is followed by an interpretation and discussion of the results of the overall and dynamic characteristics of UAOIs. Finally, section 4.6 concludes the chapter and suggests future extensions to this research.

4.2 Literature Review

4.2.1 Previous Studies on Geotagged Images from Social Media

In earlier research, geotagged images from photo sharing social media websites like Flickr, Instagram, and Picasa have been widely utilised to address a series of urban issues. Previous research includes proving the utility of Flickr data in mapping the urban environment (Crandall et al., 2009; Dunkel, 2015), analysing user behaviour (Antoniou et al., 2010; Miah et al., 2017), facilitating event detection (Kisilevich et al., 2010; Papadopoulos et al., 2011; Rattenbury et al., 2007), travel route recommendations (Sun et al., 2015; Zheng et al., 2012), places/areas of interest identification Chen et al., 2019; Lee et al., 2014; Li et al., 2013), and cultural ecosystem analysis (Hristova et al., 2018). However, certain information in geotagged photographs is currently underused, such as the content of photographs that were taken in urban areas. The density of photographs can only reflect the popularity of a place or an area but cannot demonstrate the reasons behind those patterns. It is thus necessary to understand if the photographs are relevant to the built environment and what aspects of the city are of greatest interest to people in a specific area (Richards & Friess, 2015). Many studies have used the ‘tags’ attribute of photographs to estimate public interest or capture large-scale events (Crandall et al., 2009; Kisilevich et al., 2010; Lansley & Longley, 2016; Luo et al., 2016; Papadopoulos et al., 2011; Rattenbury et al., 2007;

Steiger et al., 2015). However, these studies have ignored the key attributes (i.e., photographs) of geotagged Flickr photographs. Furthermore, these tags may not be related to the photographs themselves due to their heterogeneity (Goodchild, 2007), while several users add no tags at all.

4.2.2 Image Recognition and Urban Analytics

Due to the great improvements in computer vision and deep learning techniques in recent years, a growing number of works have attempted to apply image recognition techniques to understand urban environments, mostly relying on Google Street View (GSV) images. Some harnessed GSV images to measure the perception of safety, class, and uniqueness, thus creating reproducible quantitative measures of urban perceptions and characterising the inequality of different cities (Salesses et al., 2013). Law and his colleagues combined GSV images with 3D models generated from the GSV images and used a CNN to classify the street frontages of a front-facing street image in Greater London (Law et al., 2017). Similarly, Liu et al., (2016) exploited GSV images to predict the visual quality of the urban environment by comparing ratings based on a survey to train an image classification ConvNet model to predict a façade's quality scale. Some studies have combined GSV images with other imagery datasets to extract parcel features for urban land use classification (Kang et al., 2018; Zhang et al., 2018). Naik and his colleagues used an image segmentation approach and support vector regression to monitor neighbourhood changes and correlate socioeconomic characteristics to uncover predictors for the improvement of physical appearance (Naik et al., 2017). More recent research developed a deep CNN model, a hierarchical urban forest index, to quantify the amount of vegetation visible based on street-level imagery (Stubbings et al., 2019).

However, GSV is not the only image source that can be used to explore the urban environment. Alternatives have also appeared in recent urban studies. For example, images from Flickr, the most prevalent online photograph sharing website, were proven to be usable by (Antoniou et al., 2016; Xing et al., 2018) for land cover

classification and validation, and the 3D reconstruction of the city. Flickr was also exploited in the work of (Richards et al., 2018), who developed a novel framework for ecosystem service assessment using Google Cloud Vision and hierarchical clustering to analyse the contents of Flickr photographs automatically. Apart from Flickr, 'Place Pulse 1.0', a crowdsourced image dataset created by Salesses and colleagues (2013), was used to predict the human judgement of a streetscape's safety (Naik et al., 2014). The results showed that geotagged imagery combined with neural networks can quantify urban perceptions on a global scale. Other novel image datasets, such as 'Scenic-or-not', an online game that crowdsources the ratings of the beauty of geotagged outdoor images, was used to quantify the beauty of outdoor places in the UK through Places365-CNN models (Seresinhe et al., 2017).

All these works demonstrate that geotagged images in collaboration with image recognition techniques in computer vision can provide a deeper understanding of our built environments. Meanwhile, a variety of challenges have emerged in these applications. Most studies are based on the global urban environment, while finer urban areas are rarely involved. More importantly, few efforts have associated image recognition with urban change (Ilic et al., 2019; Naik et al., 2017). Nevertheless, urban dynamics play an important role in understanding cities, especially for the perceived urban spaces that reflect human interactions with the built environment. Therefore, this study will bridge this research gap to quantify the characteristics of local urban built environments (i.e., UAOIs) and explore their dynamic patterns.

4.2.3 Recent Approaches to Image Recognition

For about a decade, there have been improvements in the techniques used for image recognition. Some of the most notable techniques include image classification, object detection, and image segmentation. Image classification refers to labelling a photograph based on its content from a fixed set of categories (Karpathy, 2016). Image classification gained significant attention when the 'AlexNet' model became the winner of the ImageNet Large Scale Visual Recognition Challenge 2012 (ILSVRC-

2012), which was a breakthrough that significantly reduced the error rate of images to 15.3% (Krizhevsky et al., 2012). ILSVRC is an annual contest that aims to automatically estimate the content of photographs from a subset of a large hand-labelled ImageNet dataset (1000 object categories for training and validation). Since then, an increasing number of pre-trained CNN architectures/models have been proposed for the contest, such as GoogleNet, ResNet-152, Inception-v4, etc., which have constantly improved the accuracy of image classification (He et al., 2016; Szegedy et al., 2015, 2017). Several studies in recent years have used image classification to solve empirical problems—for example, to retrain one's own image dataset based on pre-trained architecture for prediction (Law et al., 2017; Liu et al., 2016) or to extract features from images through a pre-trained model (Richards et al., 2018; Seresinhe et al., 2017; Xing et al., 2018). By manually labelling data or using ready-made training data, an image can be identified by a single attribute/label or by multiple features.

More sophisticated techniques include object detection and image segmentation. Compared to image classification, these two methods are able to recognise and locate multiple objects from an image. The former method identifies different sub-images, drawing a bounding box around a recognized object, while the latter partitions an image into regions or parts present with accurate boundaries (Gandhi, 2018; Murali, 2018). Recent approaches that have gained wide popularity include Faster R-CNN (Region Convolutional Neural Network; Ren et al., 2017) and YOLO (You Only Look Once; Redmon & Farhadi, 2017) for object detection and Mask R-CNN (He et al., 2016) for image segmentation. Unlike image classification tasks that primarily using the ImageNet dataset for training, most object detection and image segmentation tasks are trained on COCO. COCO is a large-scale image dataset, with 80 categories used for object detection and segmentation (COCO, 2018). These categories mainly include everyday objects, such as vehicles, people, and a few animals. These data have been widely applied in pose estimation (Papandreou et al., 2017), medical imaging

(Johnson, 2018), real-time video surveillance (Shaifee et al., 2017), and so on (Naik et al., 2017).

Considering the suitability and availability of these approaches, a recently introduced and scene-related image classification model, Places365 CNN (Zhou et al., 2018), is used in our study. Places365 CNN worked as a classifier to identify 365 scenes from the built environment, which was mainly trained on millions of Flickr images over the world. Compared to other pre-trained CNN models, this model corresponds to our motivation to identify scene attributes from a built environment, while other object detections or segmentation models are office furniture, vehicle and animal-related. More importantly, this model is freely available and well documented (Zhou et al., 2018) but has been rarely used in previous urban analytics (Seresinhe et al., 2017).

4.3 Methods

In the following section, we introduce the Flickr data, study area, and UAOI extraction and subsequently characterise the features of the UAOIs and the outer areas through an image classification model. In addition, a finer time dimension is included to further explore the dynamic characteristics of UAOIs.

4.3.1 Data and UAOI Extraction

Data were collected from Greater London, as Greater London is the capital of, and the largest city in, the United Kingdom, with a population of over 8 million, according to the latest 2011 census. Furthermore, the raw data show that Greater London has a larger volume of geotagged Flickr photographs than many other cities. In particular, Inner London (London City Hall, 2019b), the interior part of Greater London, is used for characterisation, as a large volume of Flickr photographs is available from Inner London over a variety of years. Figure 4.1 demonstrates the spatial density of the photographs in Inner London and Greater London visualised by KDE (O'Sullivan & Unwin, 2010).

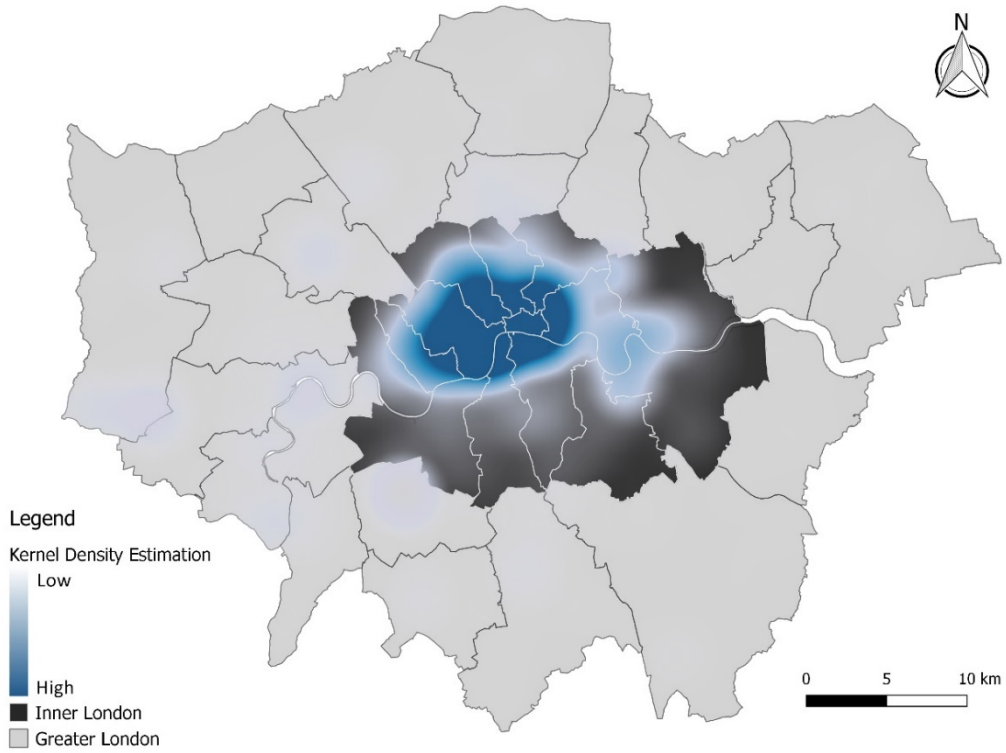


Figure 4.1 Spatial distribution of the geotagged Flickr photographs in Inner London and Greater London.

Flickr is an online photograph management and sharing website, where public photographs uploaded by users can be requested and downloaded from its public API, https://www.flickr.com/services/api/). The scale of Flickr is extensive, with 122 million users and over 10 billion photographs as of 2016, with a large degree of penetration (Smith, 2021). Unlike commonly used geotagged GSV images that are not real-time (Google Maps Street View, 2019), Flickr image data are accessible at any time and have been available since 2004, making it feasible to investigate the dynamic characteristics of UAOIs in a finer time dimension (Chen et al., 2019). Furthermore, the locations of Flickr images result from human choices and are a representation of human interactions with the built environment. However, photographs are captured in a biased way, as the aspects of the urban environment rely on how populations interact with that environment. As such, the representation of Flickr images is skewed and not necessarily realistic. These warrants caution when drawing conclusions. Nevertheless, we argue that Flickr image data are still meaningful for our study due to their

embodiment of human perceptions of the built environment and flexibility in the time dimension.

The first two stages of data pre-processing and UAOI extraction are based on the framework of (Chen et al., 2019). All geotagged Flickr metadata uploaded within Inner London has been collected through a bounding box, with a time span from the first day of 2013 to the last day of 2015. The attributes of each data record include geographic coordinates, the capture times of the photographs, user IDs, and download URLs for each photograph. This three-year time span has more Flickr photographs than others, since the site was launched in 2004. It also allows us to explore the dynamic characteristics of images within UAOIs by subdividing the time by month. To decrease the influence of a few active users who will dominate the analysis outcomes, we retained only one photograph for each user based on the tags used and the time when the photograph was taken (Chen et al., 2019). It is because some active users may take many similar photographs in a high-density area which will influence the extraction of UAOI. Specifically, if a user took several photographs in a minute but with the same tags, only one photograph was retained. The rationale for this approach was to remove photographs within a limited spatial extent based on the hypothesis that a person's average walking speed is 5 km/h (Onaverage, 2017). On this basis, the maximum walking distance within a minute is approximately 83 m. Within this short distance, only a single user's photograph with the same text is retained.

For UAOI extraction we rely on the methodology from previous chapter which combines HDBSCAN (hierarchical density-based spatial clustering for application with noise; McInnes et al., 2017) and alpha shapes (Akkiraju et al., 1995). We identified UAOI every month by HDBSCAN and constructed the corresponding boundary for each UAOI via Alpha shapes. Figure 4.2 shows the spatial distribution of all extracted UAOIs from 2013 to 2015 in Inner London in a light coral colour. We subsequently downloaded all photographs within Inner London through the URL links embedded in the Flickr metadata. Since spatial information is available for the UAOIs, in other words, images that are grouped as UAOI are available, we subsequently

divided them into two image subsets: UAOI and NON-UAOI images, with a total number of 187,064 and 816,058 photographs, respectively.

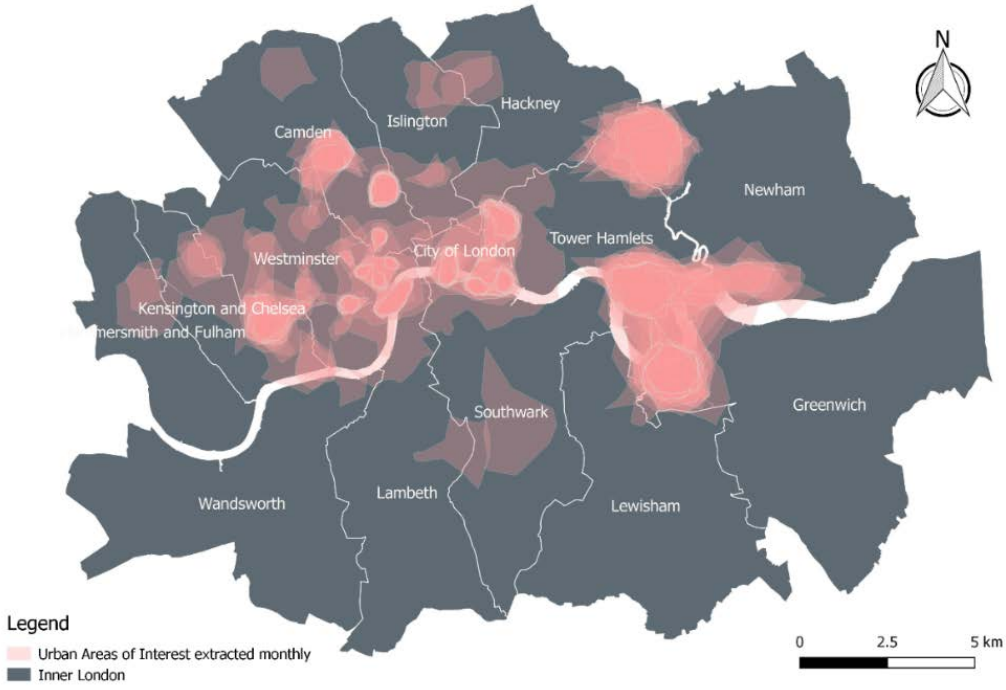


Figure 4.2 The spatial distribution of all urban areas of interest extracted per month for three years from (same with Figure 3.6).

4.3.2 Extracting the Characteristics from UAOIs and Outer Areas

To uncover the potential driving factors that influence the formation of UAOIs, an image recognition technique is conducted to identify the objects in each Flickr image. CNN models are generally designed to process data in the form of multiple arrays, such as colourful image data consisting of three 2D arrays presented as pixel values in the three colour channels.

In this work, an image classification model Places365 CNN is used to extract the characteristics of UAOIs. The reason for using this classification model instead of object detection is primarily because we are interested in the characteristics of places. Places365 CNN can work as a classifier to identify scenes from the built environment. Alternatively, other image recognition models could be used as well, while we deemed

Places365 CNN as the most productive model in this context, Places365 is the latest subset of the Places2 Database, which is trained by 1.8 million images from 365 scene categories, where there are, at most, 5,000 images per category (Zhou et al., 2018). We specifically use the Places365-ResNet model, fine-tuned on the ResNet152 (152-layer Residual Network) architecture. This CNN model has the best performance; its top 5 classification accuracy reaches 85.08%, whereas the top 5 classification accuracy for other popular CNNs, such as Places365-AlexNet, Places365-GoogleNet, and Places365-VGG is 82.89%, 83.88%, and 84.91%, respectively (Zhou et al., 2018).

All photographs within and outside the UAOIs are fed to the Places365-Resnet model[3], with the aim of exploring if there are any unique characteristics at UAOIs compared to other areas. As each photograph may contain more than one scene class, the model is set to return the maximum top five labels based on the probability for each photograph of our dataset. Further, the top five labels' classification accuracy (85.08%) is far beyond that of the top one label (54.74%), which was validated in the work of (Zhou et al., 2018). Then, we integrate the probability of all identical labels together and divide by the total number of photographs for UAOIs and other areas separately. This step helps us acquire the mean regular probability of each label in different areas. Table 4.1 features a numeric illustration of how the results are interpreted and visualised in section 4.4.1. It displays portions of the extraction from the 365 categories/labels, where the higher probability of a label represents more significant characteristics in that area, and vice versa.

[3] For high-efficiency implementation, the recognition process of all photographs (approximately 100 GB) was undertaken using a single Nvidia Quadro M5000 GPU with 8 GB memory.

Table 4.1 The mean probability of partial labels quantified inside and outside urban areas of interest.

	Bus Station	Street	Stage	Skyscraper	Downtown	Tower	Museum	Train Station	Music Studio
UAOI	0.0223	0.0253	0.0032	0.0291	0.0191	0.0385	0.0084	0.0071	0.0020
Non-UAOI	0.0448	0.0301	0.0169	0.0133	0.0115	0.0104	0.0096	0.0096	0.0094

Considering the temporal nature of UAOIs, certain UAOIs emerged and disappeared within just a few months (see examples in Figure 4.3). The UAOI in the northwest of Newham appears in July and August but disappears in September 2013, and the UAOI emerges in the middle of Southwark in August but vanishes in the next month. However, the regular characteristics recognised at the UAOIs over three years are unable to capture these minor seasonal changes. As a result, it remains challenging to explain why people would gather at certain UAOIs at specific times without identifying the dynamic patterns underlying these images.

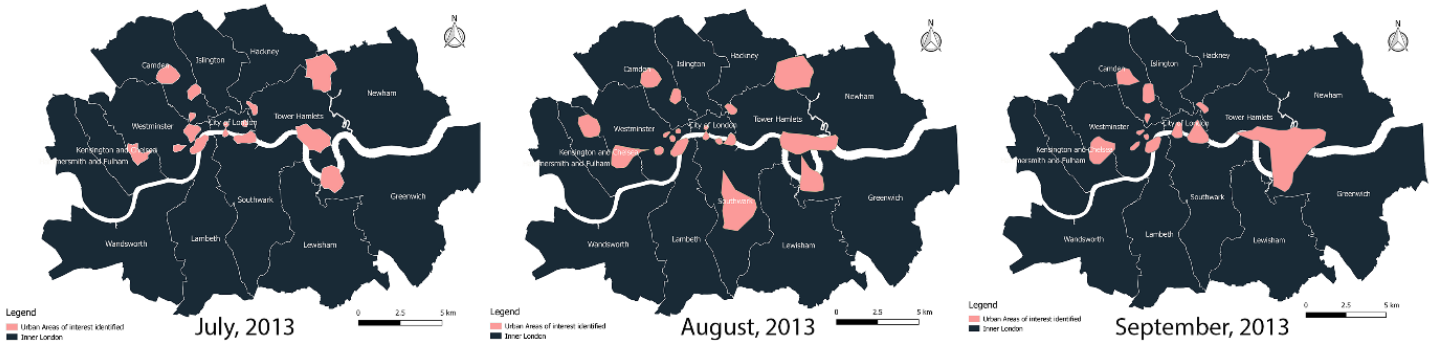


Figure 4.3 A few urban areas of interest emerged and disappeared at certain months.

To understand the factors that contributed to the dynamic changes of UAOIs, we subdivided photographs into a finer temporal resolution (i.e., we grouped photographs by month). Similarly, the maximum top five probabilities of labels were returned, and the mean probability of each label for UAOIs and Non-UAOIs in a month was calculated. Next, 36 tables similar to Table 4.1 were acquired at different months. Then, we concatenated them into a single table and determined the label probability of the UAOIs, where the row and column represent 365 features and 36 different months separately. We finally calculated the average values of the label probability for

identical months but different years, as shown in Table 4.2, which includes a small sample from the 365 labels and a numeric illustration for section 4.4.2. By doing this, the significant characteristics for the UAOIs in different months are identified, thereby allowing us to capture several interesting dynamic patterns.

Table 4.2 The mean probability of the partial labels quantified in urban areas of interest per month.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
tower	0.038	0.032	0.036	0.039	0.032	0.041	0.042	0.044	0.041	0.042	0.036	0.041
skyscraper	0.034	0.034	0.035	0.028	0.025	0.026	0.028	0.026	0.028	0.027	0.026	0.033
bridge	0.026	0.021	0.022	0.026	0.023	0.029	0.027	0.024	0.027	0.028	0.027	0.029
street	0.026	0.024	0.023	0.026	0.025	0.038	0.024	0.026	0.021	0.023	0.025	0.022
hospital	0.002	0.003	0.003	0.002	0.002	0.003	0.002	0.002	0.002	0.002	0.003	0.002
outdoor library	0.002	0.002	0.002	0.002	0.003	0.003	0.004	0.002	0.003	0.003	0.001	0.001
jewellery shop	0.002	0.003	0.002	0.002	0.003	0.001	0.003	0.002	0.002	0.002	0.003	0.003
carousel	0.002	0.002	0.001	0.001	0.001	0.001	0.002	0.002	0.001	0.001	0.003	0.010

The probability values from Table 4.2 vary greatly among individual labels. For example, the values of the label ‘tower’ are about 20 times higher than the values for the label ‘carousel’. The disparity of scales created a large challenge in simultaneously comparing the variety of all characteristics. To handle this, we calculated the z-score to standardise all label probability values by row; these values can be used to compare the results to the sample mean of the label probability for every row. This method returns a normalised value (z-score) based on its mean and standard deviation. The basic Z-Score can be calculated by the formula below:

$$\mathrm{Z} = \frac{x - \bar{\mathrm{x}}}{s}$$

where x represents the value of the data point, and $\bar{\mathrm{x}}$ and s represent the sample mean and sample standard deviation. This process ensures that the values in each row in Table 4.2 are on the same scale, thus laying the foundation for the subsequent heatmap analysis. A heatmap is a graphical presentation of data where the values contained in

a matrix are represented as colours; the darker the colour is, the higher the value or the density. We performed heatmap analysis on the z-score of the probability of a label because it returns an instant visual pattern of the labels in a timeline, offering better insight into the dynamic characteristics of UAOIs.

4.4. Results and Discussion

Based on the mean regular probabilities of the 365 categories for UAOIs and outside areas, we visualised the top 50 categories for both in an inverted pyramid graph (see Figure 4.4). The labels for the left and right y-axes were organised hierarchically, representing the significance of the characteristics from most to least within and outside the UAOIs. The top three characteristics for UAOIs are ‘tower’, ’skyscraper’, and ‘bridge’, suggesting that the Tower of London, skyscrapers, and a variety of bridges, such as Millennium Bridge and Tower Bridge, are the most significant representations of UAOIs and the primary reasons for why people gathered in these places. The overall composition of the UAOIs includes iconic landmarks, historic and famous buildings, entertainment places, and museums and galleries, as the most high-frequency appearances of these characteristics include the tags ‘canal’, ‘harbour’, ‘church’, ‘amusement park,’ ‘museum’, ‘gallery’, and so on. The components of areas outside the UAOIs are more strongly related to buses or train stations, as well as several indoor venues, such as ‘music studio’, ‘beauty salon’, ‘coffee shops’, ‘bakery shops’ and ‘bars’. These are ordinary scenes from daily life, which are less attractive to large numbers of people.

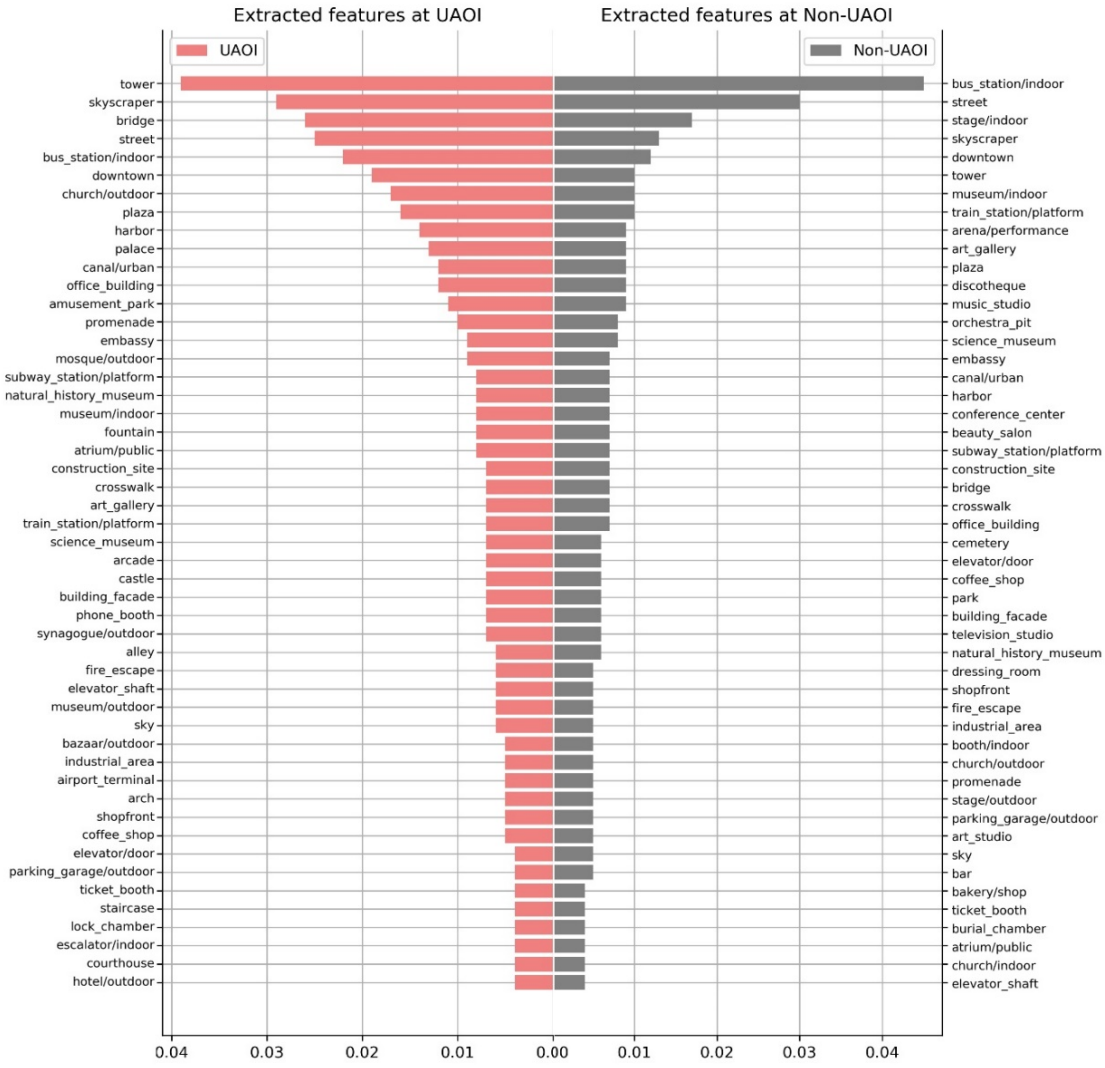


Figure 4.4 Top 50 feature probabilities extracted at urban area of interests and other areas.

4.4.1 Regular Characteristics of UAOIs and Non-UAOIs

There are a few repetitive characteristics in the top 50 for both categories, making it difficult to determine the differences between UAOIs and Non-UAOIs. For example, the labels ‘tower’, ‘street’, ‘bus_station’, ‘skyscraper’, and ‘downtown’ are identified in the top 10 for both. We then distinguished the most significant characteristics for both areas by calculating the different values of the mean regular probability of all labels in the UAOIs and Non-UAOIs. Figure 4.5 shows the differences in features between UAOIs and Non-UAOIs. By plotting this, features that are common in both

would cancel out if their probabilities are the same and thus not features both higher in the figure. The bars in light coral and grey, respectively, represent more significant features for UAOIs and Non-UAOIs. A total number of 28 labels have a higher probability in UAOIs, while more labels are identifiable in Non-UAOIs. This can be attributed to the huge and manifold areas of Non-UAOIs, where larger numbers of photographs were taken. Although the significant levels of characteristics in UAOIs and Non-UAOIs are slightly different from those in Figure 4.4, the overall pattern conforms to the features shown above. UAOIs involve more scenic spots and places of entertainment, such as ‘tower’, ’church’, ‘canal’, ’fountain’, ’amusement park’, and ’shopping mall’, while the areas of less interest are more strongly related to daily life, including labels like ‘bus station’, ’street’, ‘bar’, ‘conference centre’, and ’railroad track’.

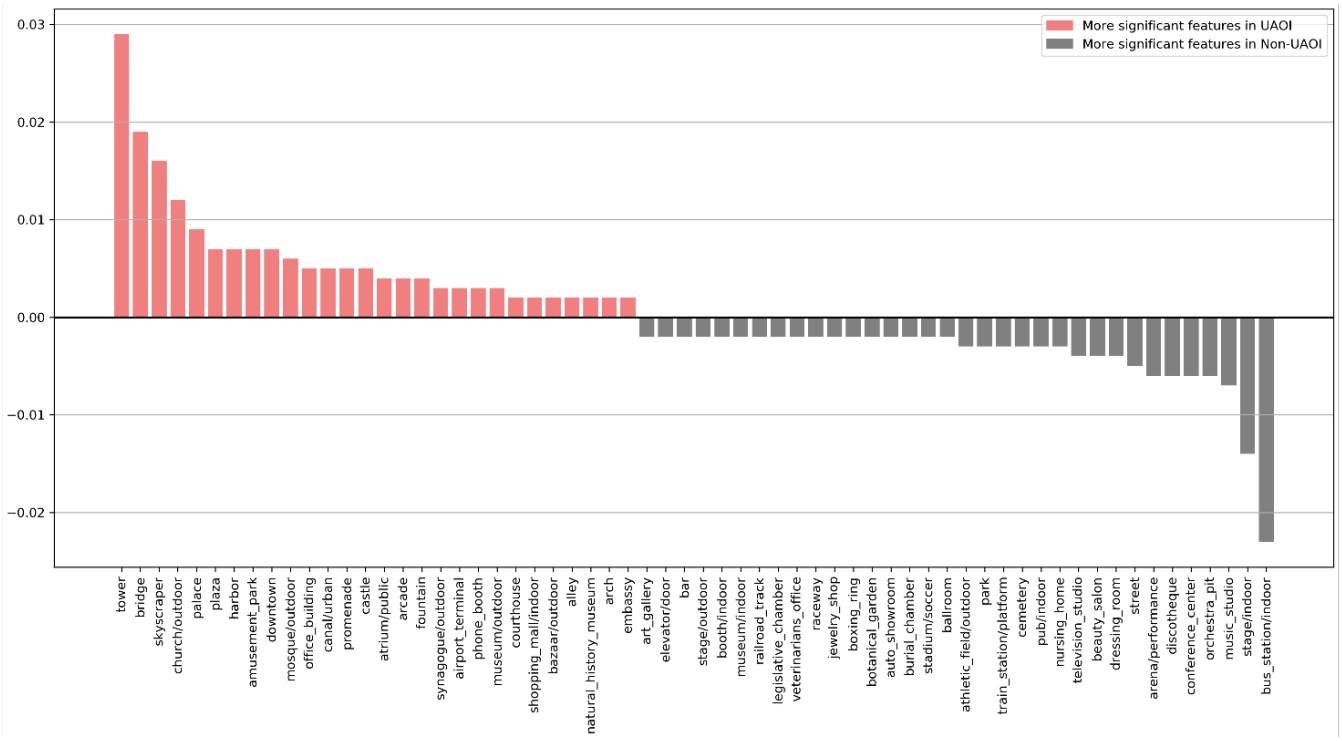


Figure 4.5 Significant features in the urban areas of interest and outer areas separately.

These regular characteristics quantitatively suggest why people would gather at UAOIs regularly over several years, as well as the characteristic differences between UAOIs and other areas. A large number of world-famous landmarks, modern skyscrapers, large-scale shopping malls, plazas, and places of entertainment are

located at UAOIs. The uniqueness of these elements has attracted thousands of people (both travellers and residents in Inner London) to take photographs of them. Conversely, the characteristics of photographs taken outside UAOIs are relatively common and anonymous and are primarily associated with daily-life scenes. We would like to highlight that the features like music studio and pub display a small lean over Non-UAOI but do not feature as a clear signifier of the class (in other words, they can be found in the middle of the figure). Subjectively, this could correspond to people taking photos with no specific purpose in these areas compared to the more purposeful photographs taken within UAOIs, such as recording certain tourist attractions like the Tower Bridge.

More importantly, the results demonstrate that geotagged Flickr images can be used to quantify the characteristics of the urban environment instead of tags. This has been rarely explored in past research, where quite a few works have instead used tags of Flickr to understand the urban environment and people's perceptions of it (Hu et al., 2015; Kisilevich et al., 2010; Li et al., 2013). Moreover, these results will help familiarize us with the perception features of large communities on a local scale, whereas previous attempts were primarily focused on global urban appearance features.

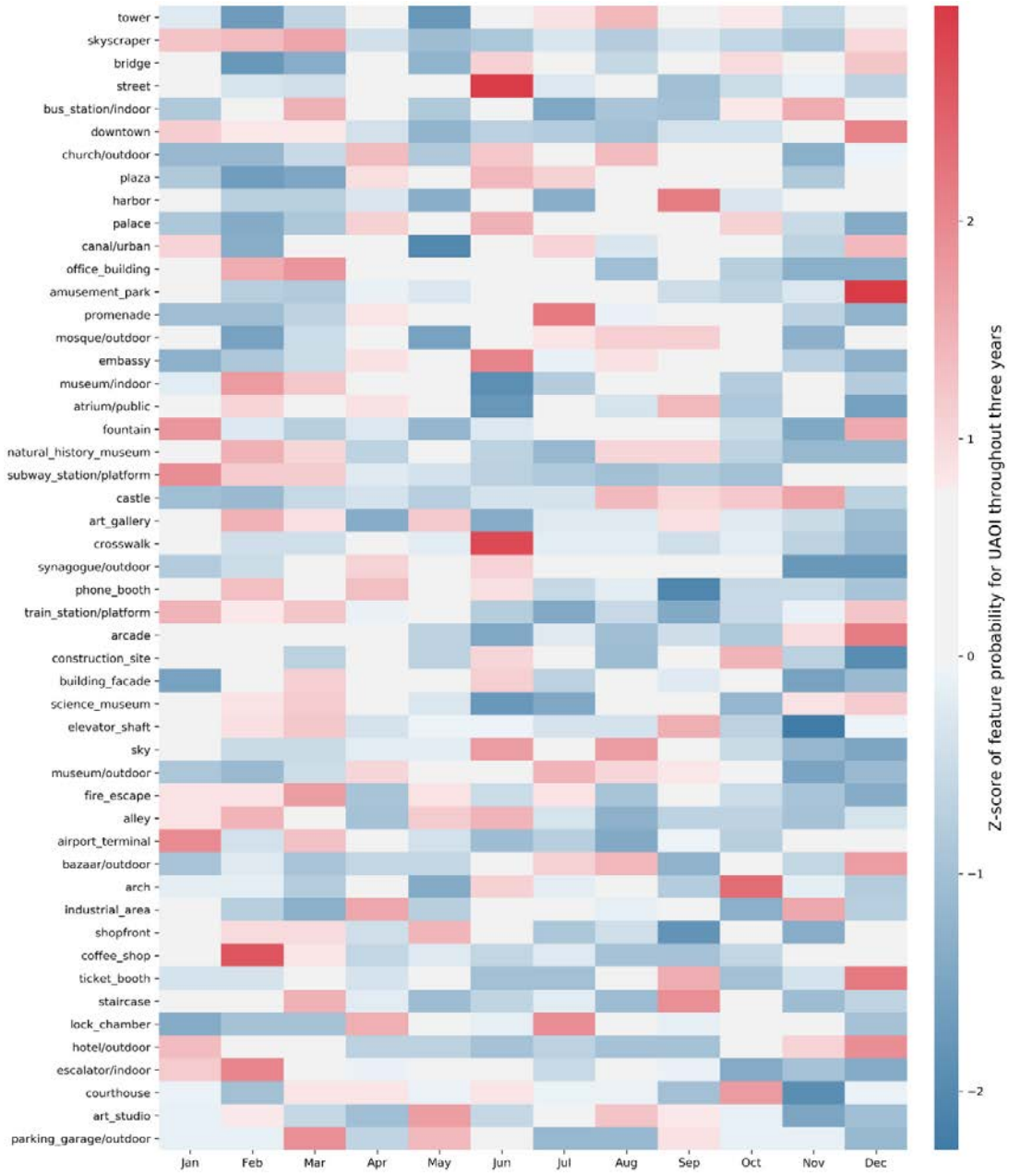


Figure 4.6 Seasonal variations in the dynamic characteristics of UAOIs based on the z-score.

4.4.2 Dynamic Characteristics of UAOIs

Based on the z-score conversion, Figure 4.6 displays a heatmap with the top 50 labels in terms of probability of occurrence. This representation uncovers the underlying characteristics of UAOIs at certain time periods, where darker red or darker blue represent the standard deviation above or below the mean of a label over the period,

respectively. The top three characteristics of the UAOIs ‘tower’, ‘skyscraper’, and ‘bridge’ primarily present an intermediate colour between red and blue, with z scores ranging from -1 to 1, implying that these three characteristics remain attractive to people all year round. The colour for several transport-related labels, such as ‘subway_station’, ‘train_station’, and ‘airport_teminal’ was slightly red from January to March but was blue for the rest of the year, suggesting that more photographs with these travel modes were taken during these months. Conversely, people’s travel mode priorities might differ when the weather becomes warmer, possibly including more walking and fewer vehicles. This manifests in the ‘street’, ‘promenade’, and ‘crosswalk’ labels, whose z-scores of probability peak in June or July but remain at an average probability during the other months. We also uncovered various seasonal patterns of indoor and outdoor activities for UAOIs. For example, a series of indoor museums and galleries labelled as ‘museum/indoor’, ‘natural_history_museum’, ’science museum’, and ’art_gallery’ were more prevalent during relatively cold months (February and March) compared with the others, while a number of magnificent buildings, as well as outdoor leisure places with the labels like ‘church’, ‘palace’, ‘mosque’, ‘castle’, and ‘plaza’, ‘bazaar’, and ‘sky’ were more likely to be identified in relatively warm seasons.

These dynamic patterns demonstrate that the season has an important impact on human activity and considerably changes the travel modes and activity modes of people, leading to the different scene characteristics of UAOIs over the year. UAOI features tend to contain more vehicles and indoor buildings in winter, as people prefer to take photographs of vehicles and indoor activities during the cold season. Correspondingly, the UAOI features consist of more crosswalks, magnificent buildings, and recreational areas in warmer months, as more photographs related to these features were taken during this period.

These results also illustrate how urban perception changes over time, showing that dynamic analytics are important for the urban environment. These bridge the identified research gap on the dynamic features of cities (Ilic et al., 2019; Naik et al., 2017).

Meanwhile, the practical implications of the dynamic characteristics of UAOIs can be reflected in the actions of retailers and local authorities. For example, a few retailers within UAOIs could expand their opening hours or deliver targeted advertising to potential customers in the summer, as people were more active during this period.

4.4.3 Capacity and Bias of Using Places365-CNN within This Context

In addition, the above heatmap also suggests that certain patterns deserve special attention. It is obvious that some characteristics are highly popular (i.e., reddest) over just a single month, such as coffee shops, streets, crosswalks, and amusement parks. To investigate what happened during these months with the corresponding characteristics, the 'amusement_park' label was selected as an example for inspection. Specifically, we extracted the photographs that were classified as 'amusement_park' in December for three years, setting a classification probability of 0.5 to filter photographs less than the threshold. A total of 175 photographs were kept after filtering, the majority of which (54.7%) were distributed at UAOIs, where Hyde Park, Trafalgar Square, London Bridge, and North Greenwich are located. Figure 4.7[4] displays a handful of samples from the 175 photographs we extracted, which were taken by various photographers in various years. Here we can see a ferris wheel, street food markets, roller coaster rides, ice skating, and carousels; these kinds of scene attributes are located in from the upper half of the images that were taken at Hyde Park. This seems related to Hyde Park's Winter Wonderland, a Christmas extravaganza that is open to the public for 6 weeks every year from mid-November to the end of December (Wonderland, 2019). This is one of the reasons why 'amusement_park' peaked in December, in agreement with our common knowledge.

[4] Due to the different shapes of the photographs, some images have been rescaled and cropped to aid visualisation in this figure. Photographers (Flickr user IDs) of images in Figure 4.7: ©17576427@N00, ©89333651@N00, ©91832335@N04, ©42230049@N03, ©16483105@N02, ©87076514@N02, ©64882892@N08, ©24605992@N06, ©75209620@N00, ©42112515@N06, ©42230049@N03, ©29558445@N00, ©36054481@N00, ©74264857@N00. Copyright of the images is retained by the photographers.

However, this does not relate exactly to the installation of an actual amusement park when examining the photographs shown in the rest of Figure 4.7. These photographs were taken at Trafalgar Square instead of Hyde Park, where a sculpture of a giant blue chicken, a Christmas tree, and a fountain with a red light was captured by multiple photographers. These scenes are not parts of an amusement park in the strictest sense, but their integration at a specific place and time can be considered a provisional amusement park, as the blue sculptures, green trees, and red fountains are similar to the colourful characteristics of an amusement park. The probable reason for this phenomenon is that groups of people gathered around Trafalgar Square in December because the Christmas tree appeared here in early December, and manifold events, such as a lighting ceremony and carol singing, happened during this period (London City Hall, 2019a). Therefore, ‘amusement_park’ became extremely prevalent in December because many seasonal landmarks appeared, and spectacular events happened in a few UAOIs due to Christmas.

Figure 4.7 Representative photographs taken in December; identified as an amusement park.

Additionally, one of the other distinct classes shown in Figure 4.6 was identified as ‘crosswalk’, which peaked in June over the year. Follow the same procedure, the photographs labelled as ‘crosswalk’ larger than the threshold 0.5 were selected to further manually validate the outputs of the Places365-CNN. 56 photographs were remained after filtering and a few representative ones were selected as samples (see Figure 4.8). Generally, the large proportions of photographs identified as ‘crosswalk’ contained a person or a few people gathered on a road. Although crosswalk lines were limitedly captured from these photographs and people were not necessarily crossing the roads, the model was able to identify the pedestrians on a street or a road to a certain degree.

Figure 4.8 Representative photographs taken in June, identified as the crosswalk

These patterns demonstrate that the pre-trained Places365-CNN model may not fit Flickr images very well, as several images can be identified based on biased characteristics. Nevertheless, the capacity of this CNN model to unpack the characteristics of the local built environment cannot be underestimated, which other models can rarely have. Although the model cannot accurately identify features as labelled, it was able to extract similar scene features and thus still helpful to this study, which can be used as a reference for policymakers and stakeholders.

4.5 Conclusions

In this study, a recent and rarely used image recognition method, Places365 CNN, was used to extract and quantify features of the local urban environment from Flickr photographs. We first compared the differences of the regular characteristics within and outside UAOIs over three years. Then, we explored the dynamic characteristics of UAOIs over that period. The results help explain why people become interested in certain urban areas more than others, what characteristics these areas possess, and if these characteristics can change over time. We found that the UAOIs were mainly identified in areas where iconic landmarks, tourist attractions, magnificent buildings, and leisure zones are located, such as towers, bridges, skyscrapers, churches, plazas, and shopping malls, which are different from the characteristics of Non-UAOIs, where more daily life-related areas are captured, such as stations, shops, and indoor venues. In terms of the dynamic characteristics of the UAOIs, UAOIs extracted in winter contained more vehicles and indoor buildings, while UAOIs extracted in other seasons consisted of more crosswalks, magnificent buildings, and recreational areas. These patterns demonstrate that the season has an important impact on human preferences for travel and activity modes. People tend to travel by various vehicles and conduct indoor activities on cold winter days but walk and engage in outdoor activities when the weather gets warmer.

This study contributes to both the theoretical and practical domains. We demonstrated that Flickr photographs themselves can be used to understand the perceived features of cities, instead of traditional methods, by using Flickr tags and other image sources like GSV images. More importantly, this work provides a potential way to bridge the research gap between image recognition techniques and urban perception analytics. Local scales and dynamic characteristics play important roles in recognising the features of the urban environment. In terms of practical significance, the regular and dynamic characteristics of the urban environment provide new insights for policymakers, who can use these findings as vital evidence for decision making. The

regular characteristics of UAOIs would be informative for urban planners to give them a macroscopic understanding of urban areas and aid them in formulating relevant policies, such as investing more funds in certain UAOIs to stimulate consumption for economic growth. The dynamic characteristics of UAOIs can help transport planners regulate trip frequency in various seasons, with a greater trip frequency in winter than in summer. Furthermore, a few retailers may also be inspired by the dynamic characteristics of UAOI, helping them better design personalised advertisements at specific places and time or expand their opening hours in summer.

However, the limitations of this study warrant further attention in future work. Flickr offers only one type of geotagged image data. Future work should incorporate multiple image sources together, which would make the results more persuasive and improve the coverage of the analysis. In addition, although the Places365 CNN model that we used to extract the urban features has a relatively high classification accuracy compared to others, the model is trained on the Flickr dataset globally instead of certain local cities or areas, implying that the model could be biased for this study area. This could lead to several features identified by Places365-CNN being incompatible with the real features of images. This issue can be addressed by manually labelling the features for a certain number of images and then retraining them by fine-tuning the parameters in the max-pooling layer of the Places365-CNN. Furthermore, except for comparing the characteristics variations between UAOI and non-UAOI at the aggregated level, the work could also be extended through identifying characteristics of each UAOI. By creating a unique profile for each UAOI, a few distinct events or characteristics can be extracted from specific urban areas. Finally, the study area we selected was located at the local level of Inner London; more interesting patterns could be uncovered at a smaller scale by including more cities in future work.

5. Using Geotagged Images and Machine Learning to Unpack the Impacts of Housing Prices

Abstract: The characteristics of housing and neighbourhood can be viewed as measurements of human progress and the quality of life, which play an important role in urban planning process. This chapter identifies the relationships between urban perceptual features and the surrounding housing market. The analysis is based on the image features recognised in the previous chapter and property transaction records. By Combining with ancillary datasets and built around a traditional housing price model (i.e., HPM), structural, neighbourhood, and perceived scene characteristics are identified to uncover their impacts on housing prices. Two machine learning algorithms – random forest and gradient boosting machines – are utilised to compare their performance and interpretability with the baseline model. The results demonstrate the usefulness of volunteered geographic image information in housing market studies. This could capture impacts of how people interacted with the built environment rather than traditional neighbourhood features extracted from Point of Interest data. Furthermore, machine learning algorithms are shown to be comparable to traditional HPM in terms of their performance and interpretability. This study could help the restructuring and optimisation of residential areas in future regional construction, planning and development.

5.1 Introduction

The desire for a good life is easier within a built environment that is equipped with facilities enabling connections to other communities and encouraging a healthy lifestyle (Molinsky & Forsyth, 2018). Within this context, housing and neighbourhood environments are particularly important since they are the locations where people spend most of their time, making a significant impact on physical, psychological, and social health. From the perspective of urban planning, the characteristics of housing and neighbourhoods can be viewed as measurements of human progress and the quality of life, and also as an alternative to shape the economy (Molinsky & Forsyth, 2018). A key indicator of these characteristics is housing price, which is an outcome of the interaction of several parties (Law et al., 2019). The price of a property is an integrated reflection of housing characteristics such as age, the property type and geographical location, as well as neighbourhood features such as accessibility of transportation and facilities. Accordingly, these features jointly influence the housing market and people's willingness to purchase, bringing challenges to urban planners, urban designers, and practitioners for regulation, construction, and evaluation.

When people perceive a city, their experience with the surrounding environment could be considered as different mental images (Lynch, 1960). Perceived urban scenes represent distinctive place characteristics or physical attributes of the city evaluated and identified by individuals (Fu et al., 2019; Zhou et al., 2014). The spatial distributions of these perceived scenes show diversity, complexity and heterogeneity, affecting the recognition and understanding of urban citizens (Dubey et al., 2016; Haney & Knowles, 1978; Zhang et al., 2018). As such, elements of perceived urban scenes play an important role in people's quality of life, public health and urban design. For instance, urban greenery has multiple functions including decreasing the pressure and negative emotions of pedestrians (Maller et al., 2006), increasing residents' probability of walking (Lu, 2018), and helping local government in road design (Zhang & Dong, 2018); scenes like open plazas allow residents and pedestrians to hold

multiple outdoor activities which can improve their physical health (Gubbels et al., 2016; Jackson, 2003); urban areas with rubbish and graffiti are more likely to be unsafe for pedestrians that decrease people's willingness to live in (L. He et al., 2017); and public and social scenes could be utilised to create design prototypes to make cities more vibrant, equitable, and resilient over time (Barkham et al., 2018).

In recent years, the wider availability of new urban data, growing computational power, and advancements in machine learning and computer vision methods have made it easier for more urban features to be identified from urban photographs. An image could contain abundant information related to human activity in the environment and the cities, providing opportunities in addressing complex questions in the cities. Its applicability has been confirmed in crime surveillance, greenery coverage detection, and natural landscape aesthetics (Collins et al., 2000; Seresinhe et al., 2017; Stubbings et al., 2019). However, the quantifiable measurements of the impacts of perceived urban scenes on real estate have been rarely discussed (Fu et al., 2019; Law et al., 2020; Zhang & Dong, 2018), which could help multi-stakeholders. For instance, the research could be meaningful for urban studies to recognise what scenes make a city more attractive, or facilitative for government and urban planners to revitalise the residential neighbourhood, or inspirational for urban design to beautify the city appearance (Barkham et al., 2018).

This study explores whether perceived urban scenes correlate to housing prices and how they influence real estate values. To achieve this, we use convolutional neural networks (CNN) to identify perceived urban scenes from user-generated image data (i.e., geotagged Flickr), which allow us to combine them with common housing price indicators. In addition to the traditional hedonic price model, two machine learning algorithms (random forest and gradient boosting) are also deployed to be compared in performance and interpretation. Our work differs from previous studies in two main aspects. Perceived urban scenes could capture elements of human perception that have impacts on housing prices. This is an aspect that remains largely unexplored in the

literature that extracting neighbourhood features of housing prices from physical POI data rather than actual perception from human perspective. Moreover, the confirmed usability and interpretability of random forest and gradient boosting methods demonstrate the bias of using hedonic price modelling (e.g., linear regression models) as a baseline in housing price studies.

The remainder of the chapter is structured as follows. Section 5.2 reviews measurements and models related to housing price estimation, and the potential of geotagged images employed in this application field. The next section describes three primary datasets collected and pre-processed to obtain housing structural, neighbourhood and image characteristics for estimation models. Section 5.4 presents the methods used and the process implemented in this study. The experimental results and discussion are reported in Section 5.5. Finally, we conclude the main contributions and limitations of this study in Section 5.6.

5.2 Literature Review

5.2.1 Hedonic Models of House Prices

The characteristics that affect housing prices vary on different scales. Within the macro scale, housing price is generally influenced by economic bases (Wang et al., 2017), such as population, household income and building cost at the administrative or city level (Baker et al., 2016; Cai & Lu, 2015). While at the microscale or intra-urban dimension, urban residents are affected by common macroeconomic variables, surrounding environmental and social characteristics, therefore, become dominant factors (Hu et al., 2019). There have been many studies to investigate the influence of housing prices within intra-urban settings, such as the age of houses, urban greenery, landscape features, distance to the city centre, air, water, or noise quality, etc. (Chen & Jim, 2010; Jim & Chen, 2006; Yao et al., 2018; Zhang & Dong, 2018; Morancho, 2003). These are a measurement of social and environmental characteristics that significantly affect the value of the house as buyers tend to purchase houses with better amenities

(Chen & Jim, 2010). The characteristics generally can be summarised as three types: structural features, location features and neighbourhood features (Xiao et al., 2017). Structural features refer to the features related to the property itself such as the type of property or the number of bathrooms; location features reflect characteristics of the geographic location of the property, such as the distance to the city centre or suburban areas; and neighbourhood features can be viewed as the availability and accessibility of several important urban amenities or landscapes, such as educational facilities, urban parks, and healthcare services.

A typical and frequently used theoretical model to analyse characteristics that influence housing value is the hedonic price model, which has been used in various studies (Chen & Jim, 2010; Zhang & Dong, 2018; Hamilton & Morgan, 2010; Wen & Tao, 2015). This model measures how each of the potential characteristics affects housing prices, playing a role in uncovering the intrinsic value of a single attribute based on the estimation of the marginal changes in observed prices (Rosen, 1974). For instance, Hamilton & Morgan (2010) integrated Lidar data and GIS into a hedonic price model to estimate the household's desire to purchase for beach access and view. Wen & Tao (2015) employed a hedonic price model to examine the polycentric urban structure in determining housing prices. However, the hedonic price model has been criticised for its strong assumptions on the linear relation between characteristics and prices and its inability to handle spatial heterogeneity (Anglin & Gençay, 1996; Dubé & Legros, 2014). Although alternative methods such as spatial econometrics and geographically weighted regression (GWR) have been proposed to incorporate spatial effects (Choumert et al., 2014; Z. Huang et al., 2017), they require prior knowledge, the assumption of linear relationships between attributes and housing prices as well, and cannot address multiscale effects well (Hu et al., 2019). As a result, to overcome the above issues, more recent studies have turned to machine learning techniques in housing research. Some have compared the model performance among multiple regression approaches to determine better models for real estate prices estimation (Chen et al., 2016; Hu et al., 2019; Park & Bae, 2015), and others have proposed

improvements of original models or a combination of two models in housing studies (Wang et al., 2014; Yao et al., 2018; Hu et al., 2019). These works have proven the usability and advantage of machine learning methods in the field of forecasting housing prices due to their fit for non-linear relationships and better prediction accuracy over traditional hedonic prices. However, the interpretation and visualisation of machine learning results remain limited, granting further investigation.

5.2.2 The Potential of Social Media Image Data in Housing Studies

Traditional research on the determinants of housing prices primarily relies on data collected from official statistical databases, proprietary listings, and questionnaire surveys (Granziera & Kozicki, 2015). These traditional data sources are labour intensive in collection and management and may not be freely available to the public. Nowadays, a growing number of studies have employed new forms of data since there is growing evidence of their potential in the analysis of regional and urban research (Arribas-Bel, 2014). Regarding the nature of easy access and available spatiotemporal attributes, data derived from social media platforms has become prevalent in housing studies in recent years (Chen et al., 2016; Hu et al., 2019; Liu & Long, 2016; Rae & Sener, 2016). For instance, housing data obtained from Anjuke, a real estate platform, was used to map spatial patterns of housing rental prices in Guangzhou, China (Chen et al.,2016). Similar research was implemented in London, UK, where Rae & Sener (2016) explored the spatial patterns of housing search using data generated from Rightmove, the UK's leading housing market portal. Social media data have been largely used in previous works for measuring landscape (Neuhaus, 2012), land uses (Shen & Karimi, 2016), urban areas (Chen et al., 2019), and the public's perception of the built environment (Chen et al., 2020), all of which are social and environmental characteristics that have significant impacts on housing prices (Bowes & Ihlanfeldt, 2001; Kong et al., 2007). However, fewer studies have captured how these characteristics generated from social media data affect housing prices (Soo, 2013; Wu

et al., 2016), which is a promising data source to measure people's perception of a place that may affect the neighbourhood housing value.

In terms of human perception of a city, it has been claimed by scientists that an image physically or mentally is an intuitive and direct perspective to capture this type of information (Ittelson, 1978; Lynch, 1960). However, as previous studies primarily relied on qualitative analysis such as visual surveys and interviews (Nasar, 1990; Scott, 1998), quantitative measurements remained limited until the technological advances in computer vision that revolutionised the field a decade ago. Since then, an increasing number of researchers have utilised images in urban perception studies. Some are interested in the identification of visual representations in the city (Chen et al., 2020; Comber et al., 2020; Doersch et al., 2012; Zhang, et al., 2018), some focused on quantifying perceptual characteristics of the city or their further relationships with non-visual socioeconomic attributes, such as population density and crime rate (Arietta et al., 2014; Dubey et al., 2016; Khosla et al., 2014; Naik et al., 2014; Salesses et al., 2013). These works measured the perception of the places through varying image recognition techniques. Many of them are based on street view imagery mostly captured by street view fleets rather than geotagged imagery (i.e., photographs) derived from human users (Google Maps Street View, 2020). As a response, the perceptual characteristics identified in previous works focus more on urban perception directly captured from physical appearance instead of perception based on how people experience the environment.

Taking the above into consideration, this study seeks to exploit the potential of image-based social media (i.e., Flickr photographs) to housing price studies, aiming to uncover if social media images can be used to explore the environmental impacts on the property nearby and how these perceptual characteristics affect the housing values.

5.3 Data

Three primary datasets are collected and used in this study. The first two represent traditional housing attributes which include housing structural characteristics and neighbourhood characteristics, the third includes geotagged images collected from the social media platform Flickr, capturing scenes around properties. Our data focuses on London as a case study, as it is the most populated city with over 8 million people in the United Kingdom and provides a good degree of Flickr usage. To obtain a higher density of images, Inner London is selected to explore our house price model, which has a larger volume of Flickr images (73%) than Outer London. In terms of the time dimension, datasets are collected within the years from 2013 to 2015, as Flickr users were the most dynamic, based on the number of photographs uploaded, during these three years. Figure 5.1(a) shows the distribution of housing price transactions in Inner London collected from the UK HM Land Registry and its geographical extent of London, where the prices display a decreasing pattern from western borough Hammersmith and Fulham to eastern Newham; Figure 5.1(b) creates a hexagonal aggregation of geotagged Flickr images within the three years, calculating the density of images at the grid size of 50 and 30 (i.e., the number of hexagons in the x-direction and y-direction) to show the spatial density.

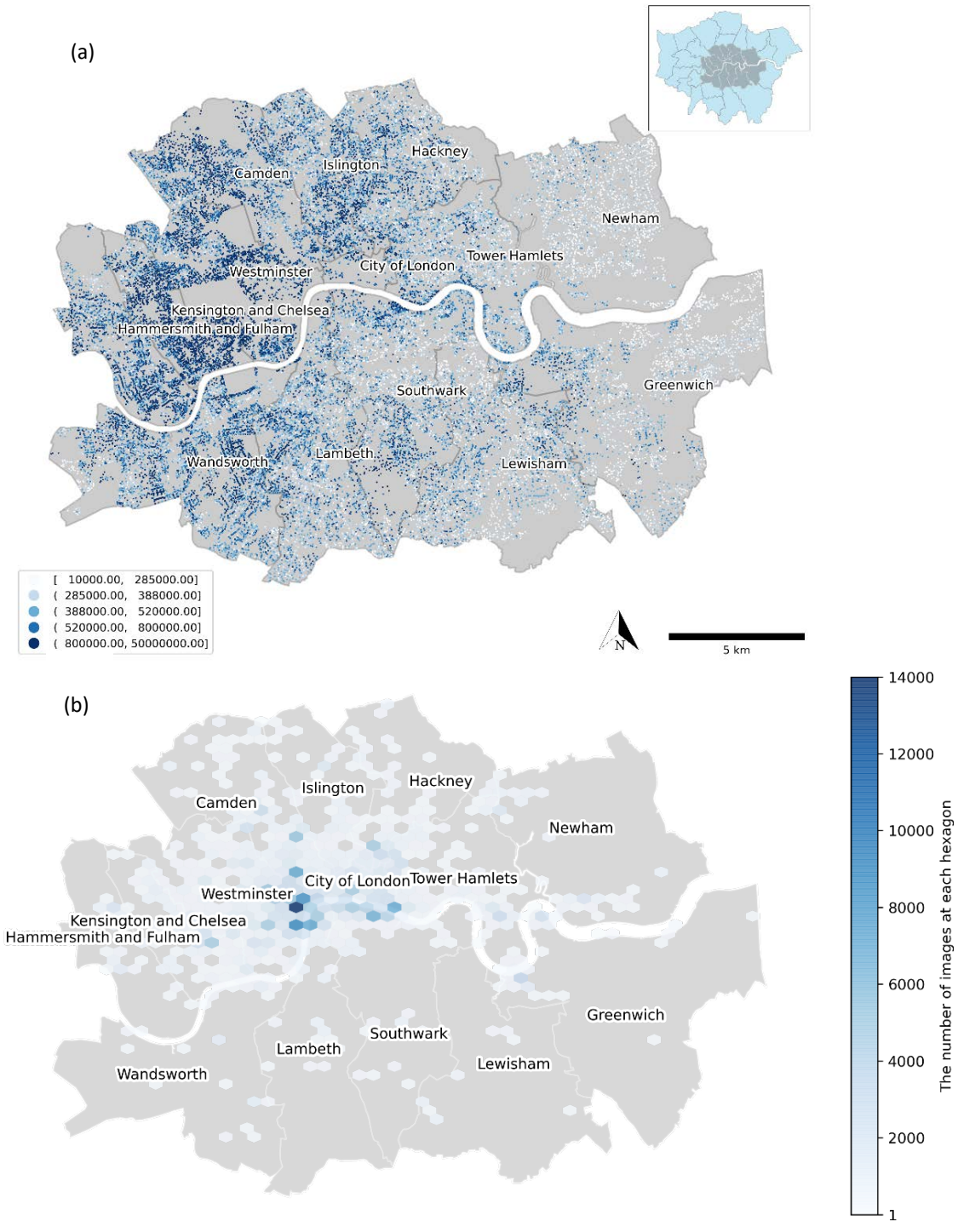


Figure 5.1 The spatial distribution of housing price and Flickr imagery datasets. (a) Choropleth of property transaction prices, (b) hexagonal aggregation of the density of Flickr images in Inner London

5.3.1 Traditional Housing Characteristics

Structural characteristics are obtained from house price paid data published by the UK HM Land Registry (HM Land Registry, 2019), which tracks property sales in England and Wales monthly since 1995. The original dataset we collected for London includes a total of 226,332 property transactions within the time range from 2013 to 2015. The dataset is subsequently cleaned based on the work by Dong et al. (2019), which only keeps properties sold for full market values, as repossessions or buy-to-lets are not a reflection of real estate market values. Since all housing structural characteristics are categorical data, we converted them into indicator variables for each category. Additionally, to identify neighbourhood and perceived scene characteristics, a geocoding process is necessary to assign spatial coordinates to all postcode addresses of houses. Furthermore, as we only focus on Inner London, the data outside our study area is removed. After this pre-processing steps, 137,132 property transaction records remained, with each containing property transaction price, the postcode address, the spatial coordinates, the date of transfer, the property type (flats, semi-detached, and terraced), whether the property is new or old, and the tenure type (freehold or leasehold).

Neighbourhood characteristics are collected from Point of Interest data and Open Greenspace data published by Ordnance Survey (Ordnance Survey, 2020), which produces detailed location information for Great Britain. As demonstrated in the literature reviewed above, buyers tend to purchase houses with perfect amenities that relate to the area where it is located, such as convenient transportation, easy access to social infrastructure and access to open spaces (Hu et al., 2019). We, therefore, measured the number of buses, underground stations, schools, medical care, and entertainment centres within an area of 800 m of each house and the distance from one house to the nearest amenities. Furthermore, areas of green space within 800 m distance of houses are also calculated following evidence of their relevance in determining housing values (Hu et al., 2019; Sirmans et al., 2005). Particularly, 800 m

Euclidean distance is used as a threshold because it was argued as a pedestrian and cycling-friendly distance for residents that lived in this neighbourhood (Liu et al., 2020).

5.3.2 Scene (Image) Characteristics

Scene characteristics are identified from geotagged social media images, which are collected from Flickr, an online photo-sharing community with over 90 million monthly users (Smith, 2020). Unlike street view imagery, Flickr image data can reflect people's perception of the built environment. On the one hand, the image contents are collected, derived, and shared by different individuals, reflecting their preferences and, in aggregate, suggesting how the city is collectively perceived. On the other hand, Flickr data can distinguish the most iconic landmarks in scenes such as towers, bridges and skyscrapers, from those including daily-life scenes such as bars, stages and conference centres (Chen et al., 2020). Despite that Flickr data has biased aspects such as possible image distortion, possible GPS bias of image geolocation, and most importantly self-selection, its usability in identifying urban representative characteristics is powerful and has been demonstrated by many studies (Chen et al., 2019; Kisilevich et al., 2010; Seresinhe et al., 2018; Zhou et al., 2014).

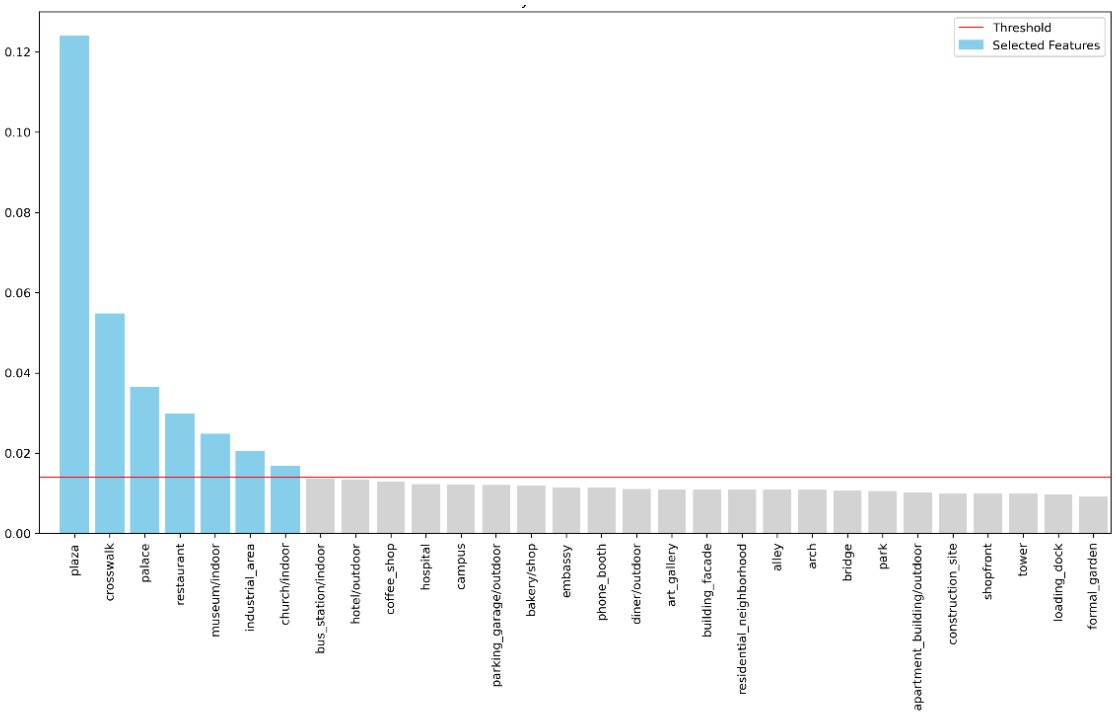


Figure 5.2 7 more relevant features selected through the feature selection process

A total of 501,098 geotagged images are collected from the official Flickr API (Flickr, 2021) and used for image recognition, with each image including latitude, longitude and specific date taken. Scene characteristics are extracted by a pre-trained Places365 convolutional neural network model, an image recognition technique designed for identifying 365 scene related categories or places (Zhou et al., 2018). This model is used due to its high performance, the recognised accuracy of the top five categories approaching 85.08% (Zhou et al., 2018). Furthermore, its capability of identifying scenario-based places from the built environment has gone beyond many other image recognition models such as YOLO (You Only Look Once, Shaifee et al., 2017), which were trained on 20 objects that include office furniture, animals and food-related categories. Through this procedure, each image is assigned to return only five scenario-based categories with identified probabilities from high to low, where a category with higher percentage values implies more significant characteristics of that image.

To investigate the possible impacts of features encoded in images on houses, we select images georeferenced within the same 800 meters distance of houses, subsequently quantifying the probability of each scene characteristic based on corresponding space and time. However, several scene characteristics we identified may have little impact on housing prices, leading to higher computational cost and lower model performance. Thus, feature selection is implemented to select a subset of 365 scene features that are important and relevant to housing prices so they can be included in the traditional modelling framework. Given limited computational capacity and possible multicollinearity among features (Li et al., 2017), we employ the feature importance of random forest to select features by its built-in mean decrease impurity (MDI) function. MDI refers to the total decrease in node impurity averaged over all trees of the ensemble, where the impurity represents a function that is weighted by the proportion of samples reaching that node (Pedregosa et al., 2011). The impurity measures the goodness of any node of decision trees (i.e., variance for regression). The smaller impurity, the purer the node and the better the prediction accuracy (Louppe, 2014). The logic of this feature selection mechanism is, when training a tree, the more a feature decreases the impurity, the more important the feature is. For many decision trees in random forest, the impurity decrease from each feature can be averaged across trees to compute the final importance of the variable (Breiman, 2001). To select more robust and important features, the standard deviation of variance is set as a threshold to drop features that are lower than the value. Figure 5.2 displays more important scene features selected (sky blue bars) after feature selection using Random Forest feature importance. Finally, we have a total of 23 independent variables to explore their impacts on housing prices, the overall descriptions and statistics of three types of characteristics are displayed in Table 5.1.

Table 5.1 Descriptions and statistics of three types of variables for housing prices

Categories	Variables	Descriptions	Mean
Structural characteristics	type_F	Dummy variables, 1 if the property type is flat	0.785
	type_S	Dummy variables, 1 if the property type is semi-detached	0.025
	type_T	Dummy variables, 1 if the property type is terraced	0.183
	new_Y	Dummy variables, 1 if the property is newly built	0.128
	tenure_L	Dummy variables, 1 if the tenure is Leasehold	0.795
Neighbourhood characteristics	bus_num	Number of bus or coach stations within 0.8km distance	0.031
	sub_num	Number of underground stations within 0.8km distance	0.219
	lei_num	Number of leisure or sports centres within 0.8km distance	0.157
	med_num	Number of medical care centres within 0.8km distance	0.187
	sch_num	Number of primary schools within 0.8km distance	2.165
	bus_dis	Distance to the nearest bus and coach station	2.174
	sub_dis	Distance to the nearest underground station	1.477
	lei_dis	Distance to the nearest leisure or sports centre	0.875
	med_dis	Distance to the nearest medical care centre	0.918
	sch_dis	Distance to the nearest primary school	0.240
	park_area	Coverage of parks and gardens within 0.8km distance	0.048
Scene characteristics (within 0.8 km distance of houses)	church	Mean probability of images classified as church	0.002
	crosswalk	Mean probability of images classified as crosswalk	0.008
	plaza	Mean probability of images classified as plaza	0.013
	restaurant	Mean probability of images classified as restaurant	0.005
	ind_area	Mean probability of images classified as industrial area	0.008
	museum	Mean probability of images classified as museum	0.006
	palace	Mean probability of images classified as palace	0.002

5.4 Methods

We propose a method framework (see the flowchart in Figure 5.3) to explore two aspects of housing price estimation: (1) whether geotagged images are an efficient data source to unpack impacts on housing price; (2) whether novel machine learning methods are more promising tools in terms of performance and interpretation compared with a traditional hedonic price model.

To explore the first aspect, models will be trained on two sets of variables and their performance compared. One includes 16 basic housing structural and neighbourhood characteristics and another is the entire 24 characteristics summarised in Table 5.1. To

consider the second aspect, three models are employed, including one hedonic price model that serves as a baseline, and two machine learning models: The Random Forest and Gradient Boosting Machines. The estimations will then be analysed and evaluated through model performance and model interpretation. Model performance evaluates how well the constructed models fit the observations and model interpretation unpacks the relationships between all independent variables and housing prices. The best model will be recognised based on prediction performance and interpretation. The remainder of this section provides a summary of each technique and procedure involved in our method framework.

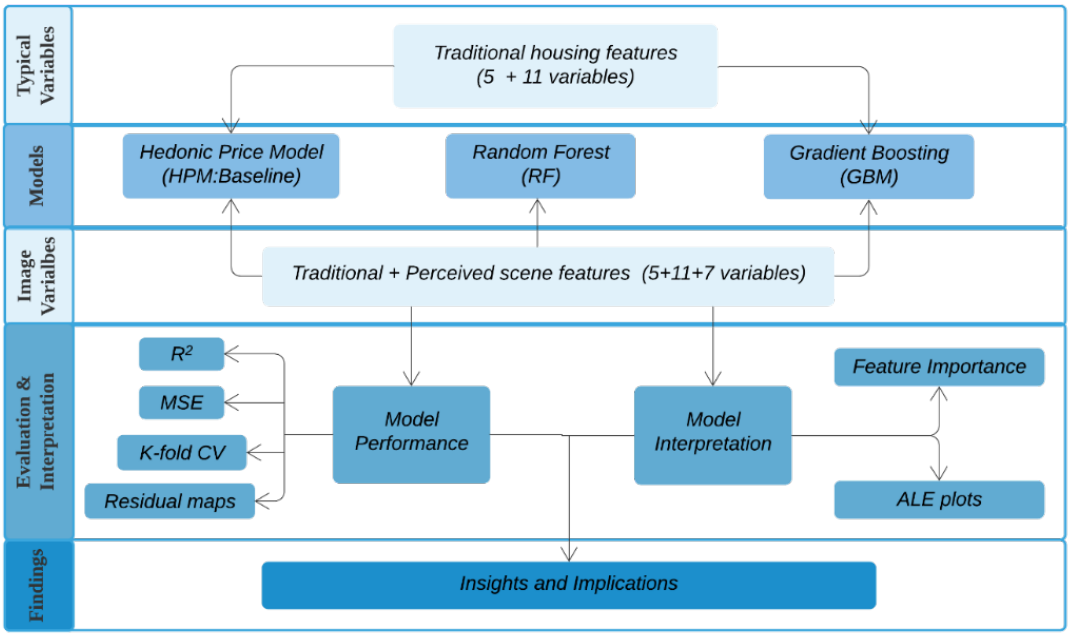


Figure 5.3 Overall methodological framework

5.4.1 Baseline Hedonic Price Model

A parametric hedonic price model (HPM) is firstly used as a benchmark in our approach. The HPM assumes a linear functional form described by a group of parameters, the coefficients of independent variables (Horowitz & Lee, 2002). Although non-parametric approaches such as Kernal estimates avoids the strong linear assumptions underlying the parametric methods, it has been criticised for its “curse of high dimensionality” and computational burden. Therefore, a semi- logarithm HPM is

selected regarding its intuitive interpretation, ease of use and variation consideration (Sirmans et al., 2005; Zhang & Dong, 2018). Particularly, the property transaction price is viewed as the dependent variable and all characteristics are independent variables. Before fitting the models, the input variables are standardized to express each variable in the same units (i.e., between 0 and 1) and thus ease interpretation. The mathematical formula of semi-log hedonic model is displayed in Equation (1):

$$LnP = \alpha + \sum \beta k Ck + \varepsilon \qquad (1)$$

LnP refers to the logarithmic form of transaction price at the postcode level, βk is the coefficient of one housing characteristic where k represents the number of independent variables C, α is a constant term, and ε is the random error term.

The mechanism of this model is to find the optimal coefficients for all the variables that minimize the error. Its interpretation is relatively straightforward: the estimated coefficients represent that the marginal change of the dependent variable when a unit increase in one of the independent variables. However, this approach is highly sensitive to multicollinearity and outliers, limited to capture non-linear relationships and large numbers of variables (Sirmans et al., 2005). As a result, the Variance Inflation Factor (VIF) is used to check for multicollinearity in our model trained with all 23 independent variables. VIF is a measure of collinearity and correlation among predictor variables within a multiple regression model. A rule of thumb is that if the VIF is larger than the threshold of 10, then the variable is considered highly colinear and correlated with the other variables (Kutner et al., 2004). The calculated VIFs of the property type of flat (35.40), property type of terraced (21.47) and tenure type of leasehold (12.59) are greater than 10, representing these three housing variables are highly correlated with each other. Multicollinearity is not a problem for non-parametric tree-based methods. Hence, two machine learning methods are compared with the baseline semi-log HPM.

5.4.2 Machine Learning Methods

Both RF and GBM are ensemble machine learning methods, which combine the predictions of several base estimators on a given algorithm to gain better robustness than a single estimator (Pedregosa et al., 2011). These two have been commonly used due to their ability to handle larger features, high accuracy performance, and robustness to skewed distributions, multicollinearity, outliers and missing values (Pal, 2017). On the other hand, these two models have been stated usability in housing price studies and higher interpretability compared to other machine learning models, such as neural networks (Arribas-Bel et al., 2017; Hu et al., 2019). Their joint pitfalls are computationally expensive and may overfit particularly noisy datasets.

RF generates a multitude of uncorrelated decision trees based on averaging random selection of predictor variables from the training set (Breiman, 2001). According to Pal, (2017), “it is a form of nonlinear regression model where samples are partitioned at each node of a binary tree based on the value of one selected input feature”. The bootstrap sampling (bagging) for each decision tree generation and the random selection of features at each node de-correlate the trees and thus reduce the variance of the prediction error when trees are averaged. The predictor variables for RF can be of any type: numerical, categorical, continuous, or discrete. The method implicitly includes interaction among the predictor variables in the model due to the hierarchical structure.

GBM trains a series of models in a stage-wise, additive, and sequential manner: it allows the optimization of arbitrary differentiable loss functions (Friedman, 2001). Unlike RF where each tree can be trained independently, each tree in GBM is determined by previous outputs. Specifically, decision trees are constructed greedily, choosing the best split nodes in each phase based on purity scores. A gradient descent procedure is used to fit and minimise the residuals (errors) in the predictions when adding subsequent trees one at a time. The training process stops once loss reaches an acceptable level or is no longer being improved (Brownlee, 2016).

A distinct benefit of using GBM is that it allows users to select any differentiable loss function or define their own, which offers more control and increases the robustness to the effects of the outliers (Arribas-Bel et al., 2017). There are several popular loss criteria, each aligned with various real-world contexts, such as least squares regression generally used for regression and logarithmic loss used for classification. More detailed information about these two approaches is referred to, for example, Breiman (2001) and Mason et al.(2000) for the interested readers.

A common characteristic in machine learning methods is that they are parameterised by a range of hyperparameters, which are required to be tuned and optimised to yield an optimal model that minimises some predefined loss function (Claesen & Moor, 2015). Manual and grid search are the most frequently used hyperparameter optimisation methods, however, they have difficulties reproducing results and suffer from too many trails to dimension exploration. (Bergstra & Bengio, 2012). Hence, random search, where each parameter setting is sampled independently from a specified distribution over the cross-validated search, is implemented due to mostly high efficiency and less computational time. To obtain a reasonably decent set of values of the hyperparameters, either a distribution over possible and random values or a list of discrete choices can be specified for each parameter. The important parameters to adjust for RF are the number of trees, the minimum number of samples at a leaf node and the number of features for a split. For GBM are the number of boosting stages, learning rate, the minimum number of samples at a leaf node and to split the node, the maximum depth to limit the number of nodes (Pedregosa et al., 2011)[5]. 5-fold cross-validation, a typical split-train-test strategy that minimises the estimator error is used in our random search. More details of cross-validation are explained in section 5.4.3. The optimised hyperparameters of RF and GBM are shown in the footnote.

[5] The values for hyper parameters that we use include:
RF: n_estimators = 200, min_samples_leaf = 2, max_features= ‘auto’, max_depth = 30;
GBM: n_estimators = 350, learning_rate= 0.1, min_samples_split= 25, min_samples_leaf= 50, max_depth = 10.

5.4.3 Model Performance and Interpretability

In any modelling context, validation and performance are crucial to evaluate how accurate and reliable the constructed models are (Hastie et al., 2009). We use a set of visualisation tools to validate the prediction, and two popular statistical metrics, mean squared estimation (MSE) and the coefficient of determination (R^2), to evaluate the model performance. The combination of these two metrics can imply the predictive power of the model and also what variation of observed variable is described by independent variables. Alternative metrics such as mean absolute error and mean absolute percentage error is also feasible in this case. MSE computes the average squared error or loss between the predicted and the actual values, which is always positive and represents better predictions the smaller its value. R^2 is an index that represents the percentage of the variance in the output that is explained by predictors (i.e., independent variables) in a regression model, which ranges from 0 to 1 and where larger values represent more explanatory models.

To avoid overfitting and unreliable results, we use cross-validation (CV) to evaluate all model performances on our limited data sample. The basic approach is called k-fold CV, which divides the dataset into the number of k non-overlapping partitions (James et al., 2013). For each of the k groups or folds, a model is trained on k-1 of the folds and the remaining part of the data is treated as testing data to measure the model performance. The resulting measure is often summarised with the average of the values computed in the k loop. Considering the data size of this study and the computational cost, a commonly used k=5 (Arribas-Bel et al., 2017; James et al., 2013) is configured to calculate cross-validated MSE and R^2, then the model with better performance will be recognised for interpretation.

Since RF and GBM cannot be interpreted by examining regression coefficients and significance due to their non-parametric nature, we, therefore, rely on permutation importance and accumulated local effects (ALE) plots to explore the relationships between variables and the observations. These two methods primarily assist us to gain

insights into how input variables relate to the target. Permutation importance is calculated by two steps: firstly, a baseline metric of the estimator is evaluated on the training dataset; secondly, a single feature column from the validation set is permuted and the metric is recomputed (Breiman, 2001; Pedregosa et al., 2011). The importance is the difference between the baseline and the drop in overall metric by permuting the column. In addition to being more reliable, permutation importance can also overcome the misleading of many unique values compared to the traditional feature importance method of several ensemble methods. MSE is the metric used in this study to measure feature importance. ALE plots visually reflect how features affect the prediction of a machine learning model on average (Apley & Zhu, 2016). To estimate local effects, the feature is divided into many intervals defined by the quantiles of the feature distribution to measure their differences in the predictions. The value of the ALE represents the key effect of the feature at a given value compared to the average prediction. Unlike the more popular partial dependence plots (PDPs), which display the marginal effect of one or two features on a machine learning prediction model, ALE plots are faster, unbiased and a more interpretable tool (Molnar, 2019). This is because PDPs can greatly bias the estimated feature effect if features are correlated, which is the case in our study.

5.5 Results and Discussions

5.5.1 Model Performance

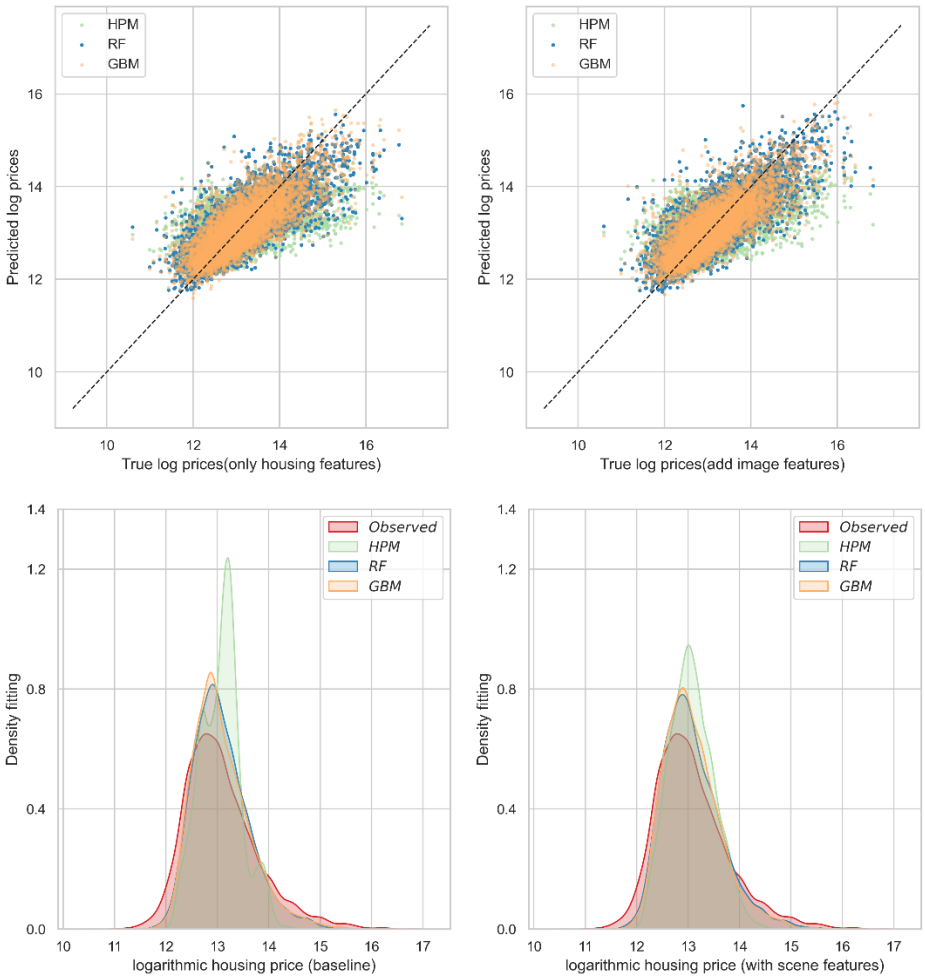


Figure 5.4 Visualisation of actual and predicted values of all the models

We first visualise in Figure 5.4 the actual and predicted values (upper graph) and their density distribution fittings (lower graph) of different models. A KDE plot is used for visualizing the density distribution of observations and predicted values of varying models, using a continuous probability density curve in more dimensions. The left side and right side represent models trained on data with housing characteristics only and with additional scene characteristics, respectively. Overall, we can see that the predicted values derived from two HPMs (green dots) deviate from actual values and only a small fraction of predicted values distribute at the same density fitting of actual

logarithmic housing prices. The larger gaps of HPMs are most likely due to the influence of multicollinearity on the models, which have been discussed in section 5.4.1. The predictions of machine learning models appear to fit the observations better than HPM. The density distribution of RF and GBM is approaching actual log prices more when scene characteristics are added into independent variables. The figure indicates that models trained with two ensemble machine learning methods fit our data well and have better performance than both predictions of HPMs no matter if images are added into the variables.

Next, the cross-validated MSE and R^2 are calculated to reflect generalization performance, as shown in Table 5.2. The results demonstrate that the overall performance of the three models improved significantly, with higher R^2 and lower MSE, when image attributes are considered. The performance of HPM is inferior to RF and GBM, as shown by larger MSE and smaller R^2; on the contrary, RF model shows better accuracy and robustness, with the highest R^2 and smallest MSE, where 66.5% of the variance in the observation could be explained by the entire 24 input variables. This table illustrates the superiority and flexibility of the two machine learning models due to their smaller uncertainty (lower MSE) and higher accuracy (higher R^2). Furthermore, the improvements in prediction with additional perceived scene characteristics demonstrate that geotagged images can be viewed as a useful data source in housing price estimation.

Table 5.2 Accuracy and error score for various models with various attributes

Metrics	Housing attributes only			Housing attributes + Image attributes		
	HPM	RF	GBM	HPM	RF	GBM
R^2	0.305	0.619	0.626	0.356	0.665	0.635
MSE	0.355	0.193	0.189	0.363	0.169	0.185

Since RF with the entire 24 variables performs better than others, we further create a spatial residual map shown in Figure 5.5 to visualise the difference of logarithmic house prices between observations and estimations to gain an intuitive insight into the spatial distribution of the errors. The dots with blue and red colours represent the estimations higher and lower than the actual log house prices. The overall residuals of housing prices in Inner London fluctuate around 0 (i.e., white dots), while properties distributed at western areas of Inner London (Borough of Kensington and Chelsea, Westminster as well as Camden) have lower estimations than the actual prices, implying that the errors located at these areas are difficult to be explained by our regression model. A possible factor of this error is the average house prices of these three boroughs are far higher than the other areas in Inner London as shown in Figure 5.1(a).

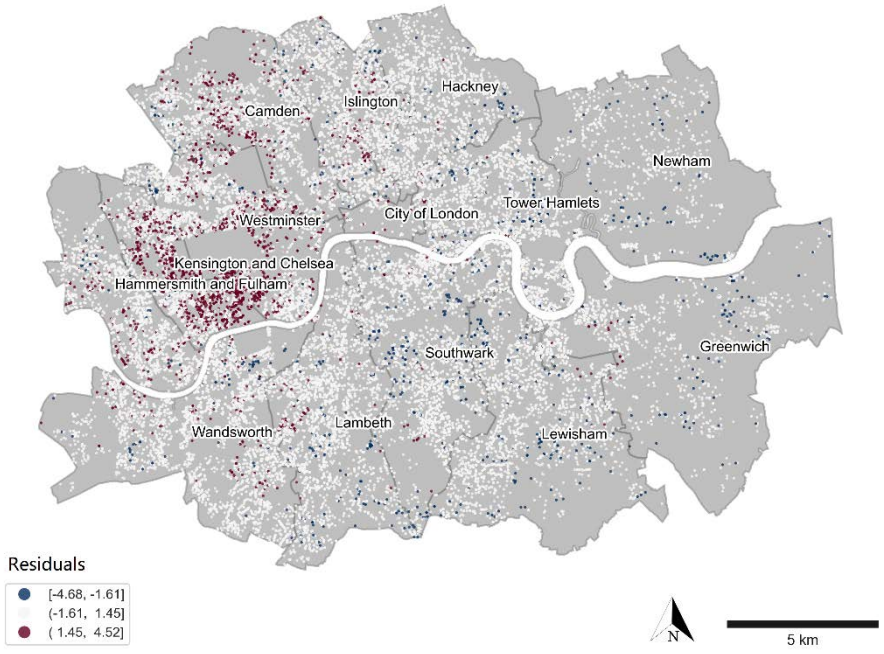


Figure 5.5 Spatial distribution of residuals of actual and predicted log house prices

5.5.2 Model Interpretation

Before looking into the interpretability of RF which has the best performance, it would be good to see how the baseline linear HPM behaves, in other words, the magnitude

of all 24 independent variables. This procedure could help us enhance the reliability of the interpretation of RF. The effect sizes (coefficient) and significance of estimated variables (p-values) within 95% confidence interval are displayed in Table 5.3, the larger coefficient values the more marginal changes in the outcome related to a unit increase in each determinant.

Table 5.3 Standardised coefficients of the baseline HPM model with different number of variables

Features	traditional housing features (16)		with additional perceived scene features (24)	
	coefficient	P>|t|	coefficient	P>|t|
intercept	13.08	0.00	13.08	0.00
type_F	-0.31	0.00	-0.34	0.00
type_S	-0.03	0.00	-0.04	0.00
type_T	-0.17	0.00	-0.18	0.00
new_Y	0.02	0.00	0.02	0.00
tenure_L	-0.14	0.00	-0.14	0.00
sch_dis	0.05	0.00	0.05	0.00
med_dis	-0.07	0.00	-0.03	0.00
lei_dis	0.04	0.00	0.04	0.00
bus_dis	-0.03	0.00	-0.01	0.00
sub_dis	-0.36	0.00	-0.32	0.00
num_sch	0.00	0.27	0.00	0.45
num_medi	-0.02	0.00	-0.01	0.00
num_lei	0.01	0.01	0.01	0.01
num_bus	0.00	0.87	0.00	0.32
num_sub	-0.07	0.00	-0.09	0.00
park_area%	-0.02	0.00	-0.02	0.00
plaza	-	-	0.06	0.00
crosswalk	-	-	0.05	0.00
palace	-	-	0.06	0.00
restaurant	-	-	0.05	0.00
museum	-	-	0.04	0.00
industrial_area	-	-	-0.01	0.07
church	-	-	0.05	0.00

Overall, traditional housing features such as the type of house (i.e., flat or terraced), the distance to the nearest subway station and whether the tenure type is leasehold or freehold (type_F, type_T, sub_dis, tenure_L) affect the housing prices much more than others and are also statistically significant. Most coefficients and significance remain

stable regardless of whether perceived scene features are considered or not. Only a minor reduction of impacts on the location-type variables by introducing scene variables. In addition to the number of subway stations, the number of other POIs including schools, health centres, retail centres and bus stations has little impact on the house prices within 0.8 km distance. On the contrary, scene features including plaza, crosswalk, palace, restaurant, museum and church have more influence and significant explanatory power in predicting the housing prices. Particularly, the type of flat and terraced, the nearest distance to subway station and health centres have negative relationships with housing prices, the higher values of these variables, the more decrease in housing prices. Conversely, most of the scene variables have positive effects on the estimation, which conform to prior knowledge that the more attractive scenery and robust infrastructure around a house, the higher the price.

Based on the understanding of the baseline model, we then turn to the interpretation of RF trained on all features. Figure 5.6 plots the importance scores computed, where the light blue bar represents more important features to the prediction that larger than the median of importance. It is obvious that the distance to the nearest subway stations within 0.8 km distance contributes the most to the predictive power of the model, the other four accessibility variables (i.e., the nearest distances to a property) and the coverage of parks also have important effects on the estimations. The type of flat and terraced and tenure type of leasehold is far more important than the other housing structural features. Significant perceived scene characteristics are palace, plaza and crosswalk, conforming to common knowledge that attractiveness and accessibility have clear impacts on house prices. However, whether the property is new or old and the type of semi-detached shows almost no association to its housing price. Besides, the number of different POIs and infrastructures within 0.8 km distance of the property proves less relevant to the estimation. The possible reason for the above scenarios is a very small fraction of transactions have records for these features during the period in this study, such as the valid values for the degree of new or old and the type of semi-

detached property only accounting for 10% and 3% of the data, consequently, hardly contribute much to the predictive power of the model.

The overall interpretation of RF model is similar to that of benchmark HPM except that RF model captures more significance in terms of service accessibility and the coverage of green parks. The more important input variables are associated with convenient transportation, accessibility of essential social infrastructure, the property type of flat, the property type of terrace, tenure type of leasehold and a few perceived scenes on housing prices. The results display that in addition to conventional influence characteristics of housing prices, how people interacted with the surrounding environment of the properties also have impacts on housing markets. Compared with the neighbourhood characteristics identified through POI data, image-based perceived scene characteristics highlight the dynamic significance of attractiveness of certain local amenities and places to housing prices. This is the core merit of considering perceived scene characteristics into housing prices as well, helping restructuring and optimisation of residential areas in future regional construction, planning and development.

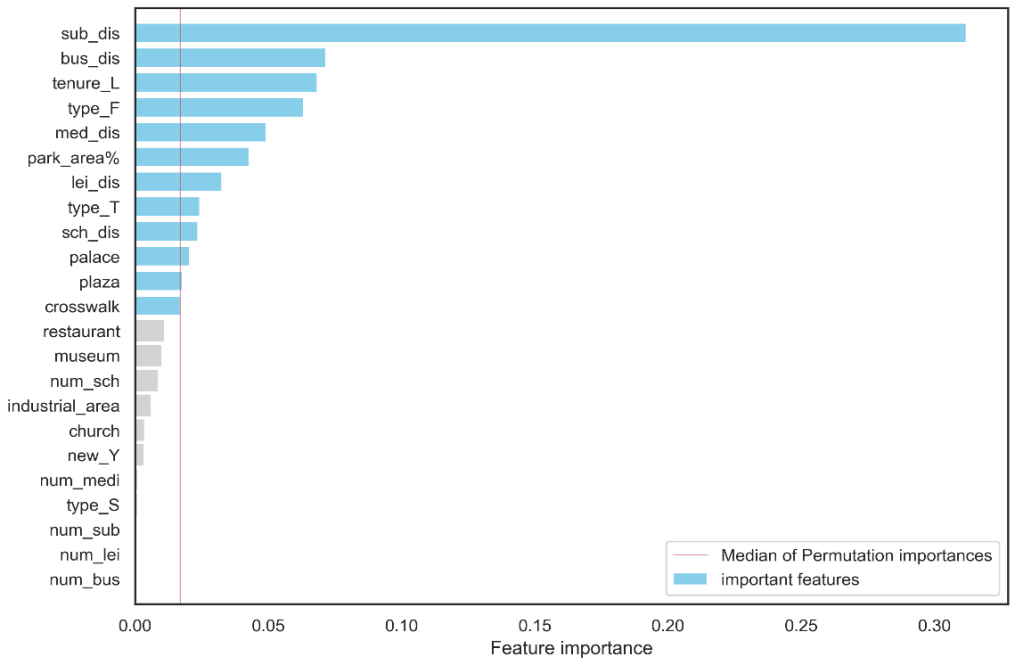


Figure 5.6 Feature importance of Random Forest based on different input variables

To further explore the relationships between variables and estimations, Figure 5.7 displays the ALE plots of the most relevant variables from three types of characteristics as shown in Figure 5.6, which are the type of flat, the nearest distance to the subway station and perceived palace scene. To help the interpretability of the results, we also include ALE plots of baseline HPM for comparison. The horizontal and vertical axes represent the range of variables and accumulated local effect values, respectively. We can see that the overall patterns of ALE plots for both baseline HPM and RF are consistent, the features property type of flat and the nearest distance to subway station have negative relationships with the observation, while perceived scene feature palace is positively associated. The differences between the two models are firstly linear and non-linear relations, and secondly, the average prediction of HPM changed more than RF with the same increasing values of features. Specifically, the average prediction decreases with the increasing value of property type of flat, but it flattens out until 0.5 for the rest. The higher values of distance to the nearest subway station, the lower the prediction; conversely, the perceived scene feature palace has a strong positive effect on the prediction.

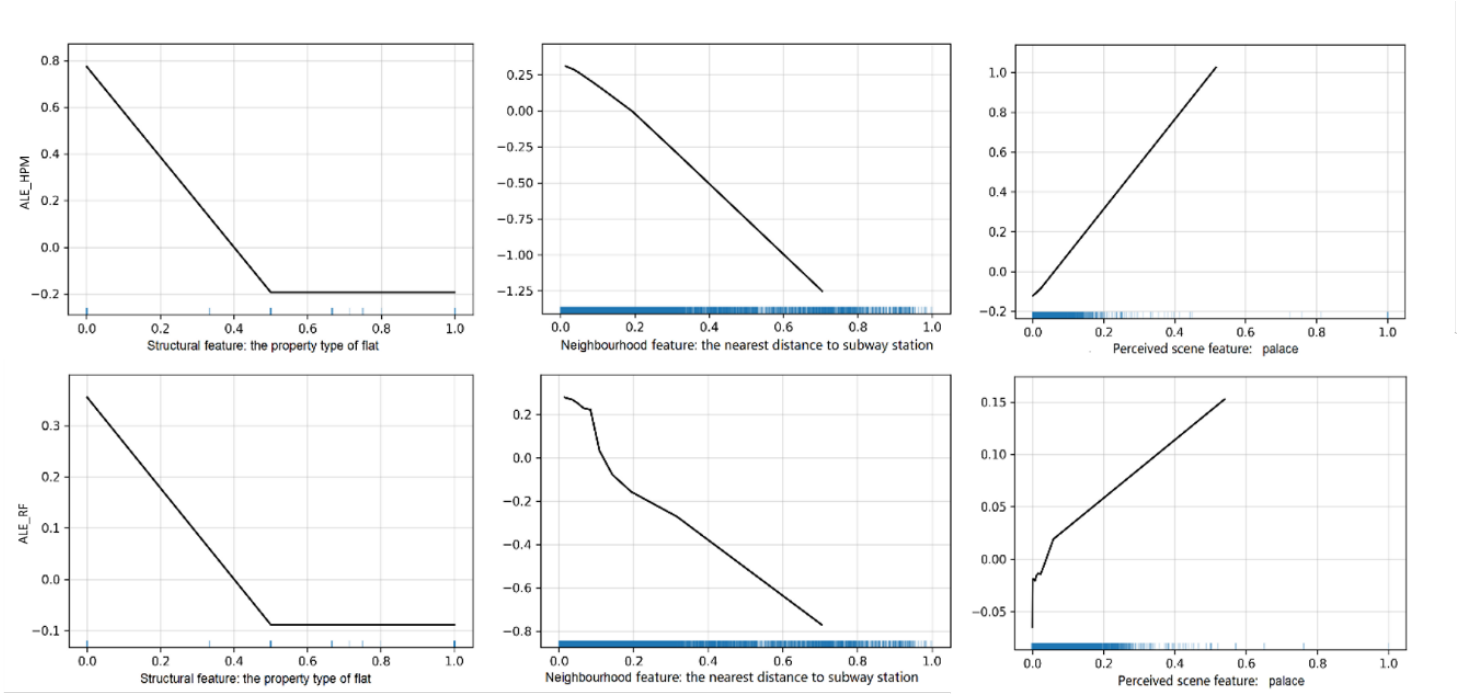


Figure 5.7 Accumulated local effects plots for partial representative characteristics

The results suggest that a closer distance to subway stations, less opportunity to live in the property type of flat, and more potential to live next to the attractive scene such as palace, can significantly increase the local real estate prices. The findings could be informative for policymakers to formulate equitable housing policies and help urban governance not only based on physical environment, but also relied on its popularity through how people perceived and interacted with. Additionally, dynamic changes in housing prices are of significance for local planners. For example, developing a healthy housing market through neighbourhood public services configuration and distribution and more affordable homes to a wider population.

5.6 Conclusions

This chapter explores the utility of geotagged social media images for monitoring housing transaction prices and the superiority and flexibility of using machine learning methods to understand the impacts of various characteristics on the housing market. We employ multiple datasets to extract three types of characteristics: structural, neighbourhood and scene attributes. With these, we check if scene characteristics can be considered as a potential data source for the understanding of housing market. Two machine learning methods, random forest and gradient boosting machines, are used in comparison with the traditional linear hedonic price model. Our results illustrate random forest proved to be the best model, based on performance, and it is also interpretable through a series of visualisation to HPM. In summary, the empirical results indicate that scene characteristics extracted from geotagged social media images have clear impacts on housing prices in Inner London. Properties surrounded by well-equipped amenities and natural scenes tend to have higher attraction and value.

Our main contributions are two-fold. On the one hand, we proved the potential of geotagged social media images as an additional dataset to incorporate in housing price estimation and monitoring. This filled in current research gaps that neighbourhood characteristics identified from POI data was unable to capture how people experienced and interacted with the physical environment. Through the inclusion of urban scene

characteristics extracted by various citizens, the impacts of popularity of scenes are uncovered and convinced on the housing market. On the other hand, our findings demonstrate that HPM is not an appropriate baseline to investigate the impacts of housing prices when multicollinearity was presented in the data. This is also applicable to other traditional empirical-statistical methods that considering spatial heterogeneity such as GWR, which has a high dependence on prior knowledge and poor capability of addressing multiscale effects (Hu et al., 2019). However, the machine learning models are not only proven to be superior performance-wise and more flexible than the traditional hedonic price model, but also interpretable in similar ways to linear models, avoiding the common black-box problems attributed to these algorithms.

In addition to traditional datasets produced by official departments, the stakeholders may also consider user-generated images as an additional dataset when assessing the housing transaction market. This data source can capture the interactions with the urban environment from residents, reflecting their interests and perceptions on urban scenes. The patterns would be informative to real estate developers for early-stage site selection of the residential buildings. Living environments with good amenities such as convenient transportation, accessibility of green space, recreational places, as well as distinctive scenes such as plaza, palace and crosswalk have important impacts on housing values. Furthermore, the government should pay more attention to the adjustment and design of housing development based on various facilities and surrounding urban features. It can assist to improve the vitality of the area surrounding a property, which subsequently influences people's willingness to buy that property.

Our works could be extended and improved in a few ways. First, other image recognition methods such as image segmentation can be used to extract more precise scene characteristics for our housing price estimation model. Second, additional datasets could be used to capture possible impact factors on real estate prices, such as the data includes more housing structural characteristics like the size or the number of bedrooms for a single house, and the visual images from property inside (Ahmed & Moustafa, 2016) and street view (Law et al., 2020). Furthermore, more cities could be

included to compare the relationships and differences. For example, how identical impact factors influence local housing prices and are there any distinctive scene characteristics for each city. Moreover, the time dimension could be further considered to unpack dynamic impacts of housing prices, helping monitor the changes of the housing market and regulate housing prices in time.

6. Conclusions

6.1 Summary and Discussions

This thesis aims to answer the general research question “How can the perception of the city be better understood by volunteered geographic image information?”. Through a theoretical argument and a data-driven methodological framework, answers to the question are unpacked from varying perspectives. From a theoretical perspective, this thesis enriches urban analytics in two aspects. First, it brings geotagged Flickr images into urban analytics to emphasise the importance of human cognition relating to urban perception. It fills the gaps that relevant research often only focuses on perception captured from a single built environment rather than the perception from human-environment interactions and it also highlights how much the representativeness of using this data. Second, a clear connection between physical scene attributes and neighbourhood non-visual attributes of the environment is developed, suggesting new avenues for the application of urban perception. In terms of technical contributions, this thesis introduces and develops a detailed methodological framework that involves exploratory spatiotemporal analysis, dynamic change profiling, image recognition techniques, and a range of machine learning algorithms. This framework provides a coherent workflow to study human perception of a city, starting from data collection and pre-processing, all the way to the integration of perceptual features in housing studies.

The objectives proposed in the introduction section are achieved and answered in different chapters throughout the document. Chapter 3 engages with Objective 1, identifying UAOIs at finer spatiotemporal granularity and profiling their dynamic patterns through the HDBSCAN machine learning algorithm, the computational technique alpha shapes, and a statistical measure based on the geographical area of UAOI. This analysis approach is distinct from other works that explore POIs or UAOIs

through VGI which have been reviewed in Section 2.2.3 and Section 3.2. It is more granular and allows varying types of dynamics in both spatial and temporal dimensions. The findings suggest how the built environment influence human activity and how human activity could potentially shape the use of the built environment. The areas where people taken photos suggest that people are attracted by areas where many unique buildings and important landmarks are located. Furthermore, it also reflects how human activity evolves and shapes the use of the urban environment. People are likely to be attracted by iconic landmarks and unique buildings, and meanwhile, the functions of these areas may be affected by human activity due to varying months of the year or seasons. The findings of this chapter would be especially beneficial for urban planners and policymakers in local authorities or city councils. For instance, urban planners could manage resource allocation more efficient in tourism if certain urban areas showed clear seasonal dynamics based on the spatiotemporal analysis of this chapter.

Chapter 4 engages with Objectives 2 and 3. A recently introduced image recognition technology is utilised to identify scene features from urban areas with varying popularity levels (i.e., UAOIs and non-UAOIs). By comparing the difference of the general characteristics within and outside UAOIs, the driving factors for why certain areas are more attractive for people are uncovered. In terms of the reason for the formation of UAOIs, the findings further confirm the implications obtained in Chapter 3 but provide richer evidence. Urban areas with more popularity (i.e., high-density population flow) are more likely to be tourist attractions, iconic landmarks, and leisure zones, which are attractive to both residents and tourists in Inner London. Conversely, urban areas with less popularity are closely associated with people's daily lives, which may only be relevant for residents rather than tourists. Moreover, by analysing the dynamic perceived characteristics of UAOIs, the seasonal nature of these is further demonstrated as an important index to influence human activity in the built environment, particularly affecting travel modes and activity modes: Cold winter days contribute to a higher frequency of use of various vehicles and more indoor activities

while warmer seasons lead to more walking and outdoor activities. These findings on urban dynamics are of significance to transport-related stakeholders, helping them to monitor and maybe adjust travel trips by seasons in variation.

Chapter 5 accomplishes Objectives 4 and 5 by analysing relationships between identified perceived scenes and the housing market to illustrate the usability and interpretability of geotagged social media images for housing studies. Compared to a traditional housing price model, more modern and interpretable machine learning approaches are used to imply the bias of using HPM. Specifically, it combines perceived features with features extracted from ancillary datasets and builds around a traditional HPM to analyse their influence on housing prices. Given perceived features have a significant impact on housing prices, volunteered geographic image information is, therefore, deemed a useful additional data source for the understanding of the housing market and real estate appraisal. Furthermore, two machine learning algorithms -RF and GBM- are developed and compared with the baseline HPM in terms of performance and interpretability. The higher performance and lower error indicate their flexibility and superiority, and a series of available visualisation techniques including feature importance and ALE plots illustrate how they can be interpreted in comparable ways to an HPM. The findings of this chapter are mainly of interest to stakeholders in housing including real estate developers, housing policymakers and housing price assessors to take geotagged social media images into consideration within their work progress. An example of impacts could be configuring and distributing neighbourhood infrastructure better to develop a healthy housing market and more affordable homes to a wider population.

The overall significance and impacts of the thesis are beneficial to urban planning that forked into tourism planning, transport planning, housing planning and design. First, UAOIs can be used to manage tourism dynamically through more flexible regional resource allocation that has more popularity at specific months over the year. Furthermore, transport departments could regulate trip frequency in various seasons, with greater trip frequency in the winter than in the summer, which relied on dynamic

perceived features of UAOIs. Additionally, regular characteristics could help urban planners to have a macroscopic understanding of urban areas and thus formulating relevant policies. For example, allocating more attention and resources to attractive urban areas with more tourism resources and infrastructure development. In terms of housing planning, results in this thesis suggest that policymakers could also pay attention to geotagged images perceived by different individuals. This enables human perception is also captured which did have impacts on housing but is not achievable through traditional POI data. In doing so, the housing market could be better understood, and more appropriate policies could be made to help reduce wealth inequalities.

6.2 Limitations and Further Works

Despite these contributions, we note several general data issues that remain unresolved in the literature. Firstly, new forms of data are often generated as a by-product without specific design, implying that the volunteered geographic image information (i.e., geotagged Flickr data) harnessed in this thesis is likely to be anonymous and unstructured, which is less targeted and reliable than traditional planned data sources. A possible solution is to select specific groups that relied on keywords in a bounded area, e.g., all Flickr images grouped as parks in London. Nevertheless, this may also generate a drawback that the number of available geotagged images within the areas is relatively small, which cannot be used to gain general patterns.

Secondly, we should note that the data has a limitation in representation. On the one hand, the use of the platform is a kind of self-selection process, indicating that the number of users cannot represent the population at all age groups and genders. The latest survey about the demographics of social media demonstrates that the dominant age groups for many social media users are teenagers or middle-aged males (Barnhart, 2021). Hence, the urban perception we captured were mainly derived from certain population groups, which were not representative enough. Additionally, since more photographs were associated with tourist attractions and important buildings, implying

that representation skewed for tourists more than for residents. On the other hand, as stated in the literature, the content of VGI is provided by varying volunteers without any reference, citation, or restrictions. Different Flickr users display uneven volume and quality of contributions. For example, active users may contribute the largest share of images while predominant users only share a few (see Figure 3.1b). As such, the data bias is still included even if a series of data preprocessing works have been conducted in Section 3.1.

To reduce the above-mentioned data biases, one possible way in further works is to combining other VGI as input, such as text-based Twitter and image-based Google Photos (https://www.google.com/photos/about/). By linking them through aggregated spatial and temporal scales, varying datasets could jointly improve the coverage of the analysis (e.g., include semantic analysis as well to identify the emotional perception of the city). The combination and comparison of various results of analysis making it more possible to gain reliable and general patterns. Furthermore, administrative data (e.g., census data) could also be used to blend with geotagged Flickr images. This will enable us to understand the percentage of population groups of users and their socioeconomic characteristics, mitigating data bias of Flickr images.

In addition to conceptual issues, a few empirically driven limitations should also be outlined. Most techniques utilised to build the methodological framework require careful parameter tuning, such as the minimal size of point for HDBSCAN to identify UAOIs, the α value for alpha shape to delineate the area boundary, and the number of estimators or the minimal number of samples at a leaf node in the RF or GBM machine learning models. Although a series of processes have been conducted in section 3.4 and section 5.4 to find a reasonable value for each core parameter, the outputs of these approaches cannot be assured of the best results due to empirically pre-defined computational thresholds. For example, the random search for hyperparameter optimisation of machine learning models in Chapter 5 is implemented within a set of thresholds based on general empirical values, and different thresholds probably will generate different results for the parameter values and thus influence the final results

and interpretation.

Furthermore, a few other issues remain unresolved that could be improved in future works. First, the semantics of Flickr images are not considered in this thesis in that we argue that text (i.e., title and tags of Flickr images) may not necessarily be related to the contents of images and many images do not have any text. However, the text could still be used as an additional supplement to develop features based on the images. For this, careful data processing and modelling need to be applied to reduce the issues. Second, HDBSCAN used in this thesis could be replaced by a more advanced algorithm such as Spatiotemporal DBSCAN that can generate clusters with the consideration of both spatial and temporal dimensions instead of only spatial information. Furthermore, hundreds of scene categories were generated from Flickr images to explain how people perceived the urban areas, some of which were quite similar and could be grouped further to improve the interpretability of the results in Chapter 4. Last but not the least, this thesis uncovers the influence of volunteered geographic (image) information on housing prices, building a connection between urban perception and housing studies. A similar approach could be used in a variety of other applications. For example, areas such as population mobility or social status could prove fertile soil for the use of features based on perceived scenes.

6.3 Concluding Remarks

In summary, this thesis provides a methodological framework for obtaining insights using volunteered geographic image information and advanced analytical techniques. Through a combination of multiple approaches such as spatiotemporal dynamic analysis, image recognition algorithms, statistical analysis, and various machine learning approaches, this thesis contributes to our understanding of how a city is perceived by the people who experience it, and how such perception interacts with the environment that makes up the urban fabric. The overall research and findings also add to our existing knowledge of how volunteered geographic image information not only reflects but also shapes the city. This is particularly significant considering the

increasing popularity of volunteered geographic knowledge and the recent rise of urban analytics. The findings outlined in this study provide a richer understanding that seeks to aid in the advancement of urban planning processes and enable policymakers to make more informed decisions about urban governance.

Bibliography

Ahlfeldt, G. (2013). Urbanity. In *LSE-SERC*. http://eprints.lse.ac.uk/59244/1/__lse.ac.uk_storage_LIBRARY_Secondary_libfile_shared_repository_Content_LSE Spatial Economic Research Centre_sercdp0136.pdf

Ahlqvist, T. (2008). *Social Media Roadmaps, VTT Research Notes*. VTT Research Notes. http://www.vtt.fi/inf/pdf/tiedotteet/2008/T2454.pdf?q=sociable-media

Ahmed, E. H., & Moustafa, M. (2016). House price estimation from visual and textual features. *IJCCI 2016 - Proceedings of the 8th International Joint Conference on Computational Intelligence*, *3*(Ijcci), 62–68. https://doi.org/10.5220/0006040700620068

Akdag, F., Eick, C. F., & Chen, G. (2014). Creating polygon models for spatial clusters. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/978-3-319-08326-1_50

Akkiraju, N., Edelsbrunner, H., Facello, M., Fu, P., Mücke, E. P., & Varela, C. (1995). Alpha shapes: definition and software. In *Proceedings of the 1st International Computational Geometry Software Workshop*.

Anglin, P. M., & Gençay, R. (1996). Semiparametric estimation of a hedonic price function. *Journal of Applied Econometrics*, *11*(6), 633–648. https://doi.org/10.1002/(SICI)1099-1255(199611)11:6<633::AID-JAE414>3.0.CO;2-T

Antoniou, V., & Skopeliti, A. (2015). Measures and indicators of vgi quality: An overview. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. https://doi.org/10.5194/isprsannals-II-3-W5-345-2015

Antoniou, Vyron, Fonte, C. C., See, L., Estima, J., Arsanjani, J. J., Lupia, F., Minghini, M., Foody, G., & Fritz, S. (2016). Investigating the feasibility of geo-Tagged photographs as sources of land cover input data. *ISPRS International Journal of Geo-Information*. https://doi.org/10.3390/ijgi5050064

Antoniou, Vyron, Morley, J., & Haklay, M. (2010). Web 2.0 geotagged photos: Assessing the spatial dimension of the phenomenon. *Geomatica*.

Apley, D. W., & Zhu, J. (2016). *Visualizing the Effects of Predictor Variables in Black Box Supervised Learning Models*. http://arxiv.org/abs/1612.08468

Arietta, S. M., Efros, A. A., Ramamoorthi, R., & Agrawala, M. (2014). City forensics: Using visual elements to predict non-visual city attributes. *IEEE Transactions on Visualization and Computer Graphics*. https://doi.org/10.1109/TVCG.2014.2346446

Arribas-Bel, D. (2014). Accidental, open and everywhere: Emerging data sources for the understanding of cities. *Applied Geography*. https://doi.org/10.1016/j.apgeog.2013.09.012

Arribas-Bel, D., Patino, J. E., & Duque, J. C. (2017). Remote sensing-based measurement of Living Environment Deprivation: Improving classical approaches with machine learning. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0176684

Baker, E., Bentley, R., Lester, L., & Beer, A. (2016). Housing affordability and residential mobility as drivers of locational inequality. *Applied Geography*. https://doi.org/10.1016/j.apgeog.2016.05.007

Barkham, R., Bokhari, S., & Saiz, A. (2018). *Urban Big Data: City Management and Real Estate Markets*.

Barnhart, B. (2021). *Social media demographics to inform your brand's strategy in 2021*. https://sproutsocial.com/insights/new-social-media-demographics/

Batty, M. (2005). Agents, cells, and cities: New representational models for simulating multiscale urban dynamics. *Environment and Planning A*. https://doi.org/10.1068/a3784

Batty, M. (2016). Big data and the city. *Built Environment*. https://doi.org/10.2148/benv.42.3.321

Batty, M. (2019). Urban analytics defined. In *Environment and Planning B: Urban Analytics and City Science*. https://doi.org/10.1177/2399808319839494

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*.

Blaikie, N. (2003). Analyzing quantitative data: from description to explanation. In *SAGE Publications Ltd*. https://doi.org/10.5860/choice.41-0975

Boarnet, M. G. (2003). The Built Environment and Physical Activity: Empirical Methods and Data Resources. In *TRB Special Report 282*. http://reconnectingamerica.org/assets/Uploads/Build-Enviro-and-Physical-Activity.pdf

Boulding, K. E. (1957). The Image: Knowledge in Life and Society. In *University of Michigan Press*. John Wiley & Sons, Ltd. https://books.google.co.uk/books?hl=en&lr=&id=w11X66GwvNIC&oi=fnd&pg=PA3&dq=The+Image:+Knowledge+in+Life+and+Society&ots=PGYHHC7tOK&sig=yMJwOSLrYqENycqem1sLAI928LQ&redir_esc=y#v=onepage&q=The Image%3A Knowledge in Life and Society&f=false

Bowes, D. R., & Ihlanfeldt, K. R. (2001). Identifying the impacts of rail transit stations on residential property values. *Journal of Urban Economics*. https://doi.org/10.1006/juec.2001.2214

Breiman, L. (2001). Random forests. *Machine Learning*. https://doi.org/10.1023/A:1010933404324

Brown, M., Sharples, S., Harding, J., Parker, C. J., Bearman, N., Maguire, M., Forrest, D., Haklay, M., & Jackson, M. (2013). Usability of Geographic Information: Current challenges and future directions. *Applied Ergonomics*. https://doi.org/10.1016/j.apergo.2012.10.013

Brownson, R. C., Hoehner, C. M., Day, K., Forsyth, A., & Sallis, J. F. (2009). Measuring the Built Environment for Physical Activity. State of the Science. *American Journal of Preventive Medicine*, *36*(4 SUPPL.), S99-S123.e12. https://doi.org/10.1016/j.amepre.2009.01.005

Burke, J., Estrin, D., & Hansen, M. (2006). *Participatory sensing*. http://eprints.cdlib.org/uc/item/19h777qd.pdf

Cai, W., & Lu, X. (2015). Housing affordability: Beyond the income and price terms, using China as a case study. *Habitat International*. https://doi.org/10.1016/j.habitatint.2015.01.021

Cao, Z., Zheng, X., Liu, Y., Li, Y., & Chen, Y. (2018). Exploring the changing patterns of China's migration and its determinants using census data of 2000 and 2010. *Habitat International*. https://doi.org/10.1016/j.habitatint.2018.09.006

Caspersen, C. J., Powell, K. E., & Christenson, G. M. (1985). Physical Activity, Exercise and Physical Fitness Definitions for Health-Related Research. *Public Health Reports*.

Catt, R. D. (2009). *100,000,000 geotagged photos (plus)*. Code.Flickr.Com. https://code.flickr.net/2009/02/04/100000000-geotagged-photos-plus/

Chen, M., Arribas-Bel, D., & Singleton, A. (2019). Understanding the dynamics of urban areas of interest through volunteered geographic information. *Journal of Geographical Systems*. https://doi.org/10.1007/s10109-018-0284-3

Chen, M., Arribas-Bel, D., & Singleton, A. (2020). Quantifying the characteristics of the local urban environment through geotagged flickr photographs and image recognition. *ISPRS International Journal of Geo-Information*, *9*(4). https://doi.org/10.3390/ijgi9040264

Chen, W. Y., & Jim, C. Y. (2010). Amenities and disamenities: A hedonic analysis of the heterogeneous urban landscape in Shenzhen (China). *Geographical Journal*. https://doi.org/10.1111/j.1475-4959.2010.00358.x

Chen, W. Y., & Li, X. (2018). Impacts of urban stream pollution: A comparative spatial hedonic study of high-rise residential buildings in Guangzhou, south China. *Geographical Journal*. https://doi.org/10.1111/geoj.12246

Chen, Y., Liu, X., Li, X., Liu, Y., & Xu, X. (2016). Mapping the fine-scale spatial pattern of housing rent in the metropolitan area by using online rental listings and ensemble learning. *Applied Geography*. https://doi.org/10.1016/j.apgeog.2016.08.011

Choumert, J., Stage, J., & Uwera, C. (2014). Access to water as determinant of rental values: A housing hedonic analysis in Rwanda. *Journal of Housing Economics*. https://doi.org/10.1016/j.jhe.2014.08.001

Claesen, M., & De Moor, B. (2015). *Hyperparameter Search in Machine Learning*. https://arxiv.org/abs/1502.02127

ClockBackward. (2019). *Ordinary Least Squares Linear Regression: Flaws, Problems and Pitfalls*.

COCO. (2018). *COCO - Common Objects in Context*.

Collins, R. T., Lipton, A. J., & Kanade, T. (2000). Introduction to the special section on video surveillance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(8), 745–746. https://doi.org/10.1109/34.865192

Comber, S., Arribas-Bel, D., Singleton, A., & Dolega, L. (2020). Using convolutional autoencoders to extract visual features of leisure and retail environments. *Landscape and Urban Planning*. https://doi.org/10.1016/j.landurbplan.2020.103887

Craik, K. H. ., & Zube, E. (1976). Perceiving Environmental Quality. In *Perceiving Environmental Quality*. https://doi.org/10.1007/978-1-4684-2865-0

Crandall, D. J., Backstrom, L., Huttenlocher, D., & Kleinberg, J. (2009). Mapping the world's photos. *Proceedings of the 18th International Conference on World Wide Web - WWW '09*, 761. https://doi.org/10.1145/1526709.1526812

Crooks, A. T., Croitoru, A., Jenkins, A., Mahabir, R., Agouris, P., & Stefanidis, A. (2016). User-Generated Big Data and Urban Morphology. *Built Environment*, *42*(3), 396–414. https://doi.org/10.2148/benv.42.3.396

Daniel, T. C., & Boster, R. S. (1976). Measuring landscape esthetics: the scenic beauty estimation method. In *USDA Forest Service Research Paper*. https://doi.org/10.1017/CBO9781107415324.004

Doersch, C., Singh, S., Gupta, A., Sivic, J., & Efros, A. A. (2012). What makes paris look like Paris? *ACM Transactions on Graphics*. https://doi.org/10.1145/2185520.2185597

Dong, G., Wolf, L., Alexiou, A., & Arribas-Bel, D. (2019). Inferring neighbourhood quality with property transaction records by using a locally adaptive spatial multi-level model. *Computers, Environment and Urban Systems*. https://doi.org/10.1016/j.compenvurbsys.2018.09.003

Dorwart, C. E., Moore, R. L., & Leung, Y. F. (2010). Visitors' perceptions of a trail environment and effects on experiences: A model for nature-based recreation experiences. *Leisure Sciences*. https://doi.org/10.1080/01490400903430863

Dubé, J., & Legros, D. (2014). Spatial econometrics and the hedonic pricing model: what about the temporal dimension? *Journal of Property Research*. https://doi.org/10.1080/09599916.2014.913655

Dubey, A., Naik, N., Parikh, D., Raskar, R., & Hidalgo, C. A. (2016). Deep learning the city: Quantifying urban perception at a global scale. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9905 LNCS*, 196–212. https://doi.org/10.1007/978-3-319-46448-0_12

Duncan, M. J., Spence, J. C., & Mummery, W. K. (2005). Perceived environment and physical activity: A meta-analysis of selected environmental characteristics. In *International Journal of Behavioral Nutrition and Physical Activity* (pp. 2–11). https://doi.org/10.1186/1479-5868-2-11

Dunkel, A. (2015). Visualizing the perceived environment using crowdsourced photo geodata. *Landscape and Urban Planning*, *142*, 173–186. https://doi.org/10.1016/j.landurbplan.2015.02.022

El-Assi, W., Salah Mahmoud, M., & Nurul Habib, K. (2017). Effects of built environment and weather on bike sharing demand: a station level analysis of commercial bike sharing in Toronto. *Transportation*, *44*(3), 589–613. https://doi.org/10.1007/s11116-015-9669-z

Elwood, S., Goodchild, M. F., & Sui, D. Z. (2012). Researching Volunteered Geographic Information: Spatial Data, Geographic Research, and New Social Practice. *Annals of the Association of American Geographers*. https://doi.org/10.1080/00045608.2011.595657

Feng, C., Li, W., & Zhao, F. (2011). Influence of rail transit on nearby commodity housing prices: A case study of Beijing Subway Line Five. *Dili Xuebao/Acta Geographica Sinica*.

Ferrare, R. A., Fraser, R. S., & Kaufman, Y. J. (1990). Satellite measurements of large-scale air pollution: Measurements of forest fire smoke. *Journal of Geophysical Research: Atmospheres*, *95*(D7), 9911–9925. https://doi.org/10.1029/JD095iD07p09911

Flanagin, A. J., & Metzger, M. J. (2008). The credibility of volunteered geographic information. In *GeoJournal*. https://doi.org/10.1007/s10708-008-9188-y

flickr. (2021). *App Garden*. https://www.flickr.com/services/api/

Forsyth, A., Schmitz, K. H., Oakes, M., Zimmerman, J., & Koepp, J. (2016). Standards for Environmental Measurement Using GIS: Toward a Protocol for Protocols. *Journal of Physical Activity and Health*. https://doi.org/10.1123/jpah.3.s1.s241

Frank, L. D., & Engelke, P. O. (2001). The built environment and human activity patterns: Exploring the impacts of urban form on public health. *Journal of Planning Literature*. https://doi.org/10.1177/08854120122093339

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*. https://doi.org/10.2307/2699986

Fu, X., Jia, T., Zhang, X., Li, S., & Zhang, Y. (2019). Do street-level scene perceptions affect housing prices in Chinese megacities? An analysis using open access datasets and deep learning. *PLoS ONE*, *14*(5), 1–18. https://doi.org/10.1371/journal.pone.0217505

Gandhi, R. (2018). *R-CNN, Fast R-CNN, Faster R-CNN, YOLO — Object Detection Algorithms*.

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). *Rich feature hierarchies for accurate object detection and semantic segmentation Tech report (v5)*. http://www.cs.berkeley.edu/˜rbg/rcnn.

González, M. C., Hidalgo, C. A., & Barabási, A. L. (2008). Understanding individual human mobility patterns. *Nature*. https://doi.org/10.1038/nature06958

Goodchild, M. F. (2007). Citizens as sensors: The world of volunteered geography. In *GeoJournal*. https://doi.org/10.1007/s10708-007-9111-y

Goodchild, M. F., & Glennon, J. A. (2010). Crowdsourcing geographic information for disaster response: a research frontier. *International Journal of Digital Earth*, *3*(3), 231–241. https://doi.org/10.1080/17538941003759255

Goodchild, M. F., & Li, L. (2012). Assuring the quality of volunteered geographic information. *Spatial Statistics*. https://doi.org/10.1016/j.spasta.2012.03.002

Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning*. https://doi.org/10.4258/hir.2016.22.4.351

Google Maps Street View. (2019). *Google-Contributed Street View Imagery Policy*.

Google Maps Street View. (2020). *Google Maps Street View Google-Contributed Street View Imagery Policy*.

Granziera, E., & Kozicki, S. (2015). House price dynamics: Fundamentals and expectations. *Journal of Economic Dynamics and Control*. https://doi.org/10.1016/j.jedc.2015.09.003

Gubbels, J. S., Kremers, S. P. J., Droomers, M., Hoefnagels, C., Stronks, K., Hosman, C., & de Vries, S. (2016). The impact of greenery on physical activity and mental health of adolescent and adult residents of deprived neighborhoods: A longitudinal study. *Health and Place*. https://doi.org/10.1016/j.healthplace.2016.06.002

Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., & Lew, M. S. (2016). Deep learning for visual understanding: A review. *Neurocomputing*, *187*, 27–48. https://doi.org/10.1016/j.neucom.2015.09.116

Hamilton, S. E., & Morgan, A. (2010). Integrating lidar, GIS and hedonic price modeling to measure amenity values in urban beach residential property markets. *Computers, Environment and Urban Systems*. https://doi.org/10.1016/j.compenvurbsys.2009.10.007

Haney, W. G., & Knowles, E. S. (1978). Perception of neighborhoods by city and suburban residents. *Human Ecology*. https://doi.org/10.1007/BF00889095

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning. In *Springer Series in Statistics*. https://doi.org/10.1007/b94608

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). *Mask R-CNN*. https://arxiv.org/abs/1703.06870

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. https://doi.org/10.1109/CVPR.2016.90

He, L., Páez, A., & Liu, D. (2017). Built environment and violent crime: An environmental audit approach using Google Street View. *Computers, Environment and Urban Systems*. https://doi.org/10.1016/j.compenvurbsys.2017.08.001

Herzog, T. R., Kaplan, S., & Kaplan, R. (1976). The prediction of preference for familiar urban places. *Environment and Behavior*, *8*, 627–645. https://doi.org/10.1177/001391657684008

HM Land Registry. (2019). *HM Land Registry Price Paid Data*. https://data.gov.uk/dataset/4c9b7641-cf73-4fd9-869a-4bfeed6d440e/hm-land-registry-price-paid-data

Holcomb, B., & Saarinen, T. F. (1977). Environmental Planning: Perception and Behavior. *Geographical Review*. https://doi.org/10.2307/214027

Hollenstein, L., & Purves, R. S. (2010). Exploring place through user-generated content: Using Flickr tags to describe city cores. *Journal of Spatial Information Science*. https://doi.org/10.5311/JOSIS.2010.1.3

Horowitz, J., & Lee, S. (2002). Semiparametric methods in applied econometrics: Do the models fit the data? *Statistical Modeling*. https://doi.org/10.1191/1471082x02st024oa

Hristova, D., Aiello, L. M., & Quercia, D. (2018). The new urban success: How culture pays. *Frontiers in Physics*, *6*(APR). https://doi.org/10.3389/fphy.2018.00027

Hu, L., He, S., Han, Z., Xiao, H., Su, S., Weng, M., & Cai, Z. (2019). Monitoring housing rental prices based on social media:An integrated approach of machine-learning algorithms and hedonic modeling to inform equitable housing policies. *Land Use Policy*. https://doi.org/10.1016/j.landusepol.2018.12.030

Hu, T., Yang, J., Li, X., & Gong, P. (2016). Mapping urban land use by using landsat images and open social data. *Remote Sensing*. https://doi.org/10.3390/rs8020151

Hu, Y., Gao, S., Janowicz, K., Yu, B., Li, W., & Prasad, S. (2015a). Extracting and understanding urban areas of interest using geotagged photos. *Computers, Environment and Urban Systems*. https://doi.org/10.1016/j.compenvurbsys.2015.09.001

Hu, Y., Gao, S., Janowicz, K., Yu, B., Li, W., & Prasad, S. (2015b). Extracting and understanding urban areas of interest using geotagged photos. *Computers, Environment and Urban Systems*, *54*, 240–254. https://doi.org/10.1016/j.compenvurbsys.2015.09.001

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely Connected Convolutional Networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2261–2269. https://doi.org/10.1109/CVPR.2017.243

Huang, Z., Chen, R., Xu, D., & Zhou, W. (2017). Spatial and hedonic analysis of housing prices in Shanghai. *Habitat International*. https://doi.org/10.1016/j.habitatint.2017.07.002

Humpel, N., Owen, N., & Leslie, E. (2002). Environmental factors associated with adults' participation in physical activity. A review. In *American Journal of Preventive Medicine*. https://doi.org/10.1016/S0749-3797(01)00426-3

Ibrahim, M. R., Haworth, J., & Cheng, T. (2020). Understanding cities with machine eyes: A review of deep computer vision in urban analytics. *Cities*, *96*(October 2019). https://doi.org/10.1016/j.cities.2019.102481

Ilic, L., Sawada, M., & Zarzelli, A. (2019). Deep mapping gentrification in a large Canadian city using deep learning and Google Street View. *PLOS ONE*, *14*(3), e0212814. https://doi.org/10.1371/journal.pone.0212814

Ittelson, W. H. (1978). Environmental Perception and Urban Experience. *Environment and Behavior*. https://doi.org/10.1177/0013916578102004

Jackson, L. E. (2003). The relationship of urban design to human health and condition. *Landscape and Urban Planning*. https://doi.org/10.1016/S0169-2046(02)00230-X

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

James, G., Witten, D., Hastie, T., & Tibishirani, R. (2013). An Introduction to Statistical Learning with Applications in R (older version). In *Springer Texts in Statistics*.

Jason Brownlee. (2016). *A Gentle Introduction to the Gradient Boosting Algorithm for Machine Learning*. https://machinelearningmastery.com/gentle-introduction-gradient-boosting-algorithm-machine-learning/

Jedlovec, G. (2013). Transitioning research satellite data to the operational weather community: The SPoRT Paradigm [Organization Profiles]. *IEEE Geoscience and Remote Sensing Magazine*, *1*(1), 62–66. https://doi.org/10.1109/mgrs.2013.2244704

Jim, C. Y., & Chen, W. Y. (2006). Impacts of urban environmental elements on residential housing prices in Guangzhou (China). *Landscape and Urban Planning*. https://doi.org/10.1016/j.landurbplan.2005.12.003

Johnson, J. W. (2018). Adapting Mask-RCNN for Automatic Nucleus Segmentation. *2019 Computer Vision Conference, Vol. 2*. https://doi.org/10.1007/978-3-030-17798-0

Kandt, J., & Batty, M. (2021). Smart cities, big data and urban policy: Towards urban analytics for the long run. *Cities*, *109*(October 2020), 102992. https://doi.org/10.1016/j.cities.2020.102992

Kang, C., Liu, Y., Ma, X., & Wu, L. (2012). Towards Estimating Urban Population Distributions from Mobile Call Data. *Journal of Urban Technology*. https://doi.org/10.1080/10630732.2012.715479

Kang, J., Körner, M., Wang, Y., Taubenböck, H., & Zhu, X. X. (2018). Building instance classification using street view images. *ISPRS Journal of Photogrammetry and Remote Sensing*, *145*, 44–59. https://doi.org/10.1016/j.isprsjprs.2018.02.006

Karpathy, A. (2016). CS231n Convolutional Neural Networks for Visual Recognition. *Stanford University*.

Kennedy, L., & Naaman, M. (2008). Generating diverse and representative image search results for landmarks. *Proceeding of the 17th International Conference on World Wide Web 2008, WWW'08*. https://doi.org/10.1145/1367497.1367539

Khosla, A., An, B., Lim, J. J., & Torralba, A. (2014). Looking beyond the visible scene. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. https://doi.org/10.1109/CVPR.2014.474

Kisilevich, S., Krstajic, M., Keim, D., Andrienko, N., & Andrienko, G. (2010). Event-based analysis of people's activities and behavior using Flickr and Panoramio geotagged photo collections. *Proceedings of the International Conference on Information Visualisation*. https://doi.org/10.1109/IV.2010.94

Kong, F., Yin, H., & Nakagoshi, N. (2007). Using GIS and landscape metrics in the hedonic price modeling of the amenity value of urban green space: A case study in Jinan City, China. *Landscape and Urban Planning*. https://doi.org/10.1016/j.landurbplan.2006.02.013

Kou, Z., & Cai, H. (2019). Understanding bike sharing travel patterns: An analysis of trip data from eight cities. *Physica A: Statistical Mechanics and Its Applications*. https://doi.org/10.1016/j.physa.2018.09.123

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *ImageNet Classification with Deep Convolutional Neural Networks*.

Kutner, M., Nachtsheim, C., & Neter, J. (2004). Applied Linear Regression Models- 4th Edition. *McGraw-Hill Irwin*.

Lansley, G., & Longley, P. A. (2016). The geography of Twitter topics in London. *Computers, Environment and Urban Systems*, *58*, 85–96. https://doi.org/10.1016/j.compenvurbsys.2016.04.002

Law, S., Paige, B., & Russell, C. (2019). Take a look around: Using street view and satellite images to estimate house prices. *ACM Transactions on Intelligent Systems and Technology*, *10*(5). https://doi.org/10.1145/3342240

Law, S., Seresinhe, C. I., Shen, Y., & Gutierrez-Roig, M. (2020). Street-Frontage-Net: urban image classification using deep convolutional neural networks. *International Journal of Geographical Information Science*, *34*(4), 681–707. https://doi.org/10.1080/13658816.2018.1555832

Law, S., Shen, Y., & Seresinhe, C. (2017). An application of convolutional neural network in street image classification: The case study of London. *Proceedings of the 1st Workshop on GeoAI: AI and Deep Learning for Geographic Knowledge Discovery, GeoAI 2017, February 2018*, 5–9. https://doi.org/10.1145/3149808.3149810

Lawson, B. R., & Ittelson, W. H. (1977). An Introduction to Environmental Psychology. *Leonardo*. https://doi.org/10.2307/1573445

Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Lee, I., Cai, G., & Lee, K. (2014). Exploration of geo-tagged photos through data mining approaches. *Expert Systems with Applications*. https://doi.org/10.1016/j.eswa.2013.07.065

Lee, S., Maisonneuve, N., Crandall, D., Efros, A. A., & Sivic, J. (2015). *Linking Past to Present: Discovering Style in Two Centuries of Architecture*. https://doi.org/10.1109/iccphot.2015.7168368

Lee, Y. J., Efros, A. A., & Hebert, M. (2013). Style-aware mid-level representation for discovering visual connections in space and time. *Proceedings of the IEEE International Conference on Computer Vision*. https://doi.org/10.1109/ICCV.2013.233

Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R. P., Tang, J., & Liu, H. (2017). Feature selection: A data perspective. In *ACM Computing Surveys*. https://doi.org/10.1145/3136625

Li, L., & Goodchild, M. F. (2012). Constructing places from spatial footprints. *GEOCROWD 2012 - Proceedings of the 1st ACM SIGSPATIAL International Workshop on Crowdsourced and Volunteered Geographic Information*. https://doi.org/10.1145/2442952.2442956

Li, L., Goodchild, M. F., & Xu, B. (2013). Spatial, temporal, and socioeconomic patterns in the use of twitter and flickr. *Cartography and Geographic Information Science*. https://doi.org/10.1080/15230406.2013.777139

Li, X., Zhang, C., & Li, W. (2015). Does the visibility of greenery increase perceived safety in urban areas? Evidence from the place pulse 1.0 dataset. *ISPRS International Journal of Geo-Information*. https://doi.org/10.3390/ijgi4031166

Liu, L., Wang, H., & Wu, C. (2016). *A machine learning method for the large-scale evaluation of urban visual environment*. *Harvey 2014*. http://arxiv.org/abs/1608.03396

Liu, X., & Long, Y. (2016). Automated identification and characterization of parcels with OpenStreetMap and points of interest. *Environment and Planning B: Planning and Design*. https://doi.org/10.1177/0265813515604767

Liu, Yu, Liu, X., Gao, S., Gong, L., Kang, C., Zhi, Y., Chi, G., & Shi, L. (2015). Social Sensing: A New Approach to Understanding Our Socioeconomic Environments. *Annals of the Association of American Geographers*, *105*(3), 512–530. https://doi.org/10.1080/00045608.2015.1018773

Liu, Yunzhe, & Cheng, T. (2020). Understanding public transit patterns with open geodemographics to facilitate public transport planning. *Transportmetrica A: Transport Science*. https://doi.org/10.1080/23249935.2018.1493549

Liu, Yunzhe, Singleton, A., & Arribas-Bel, D. (2020). Considering context and dynamics: A classification of transit-orientated development for New York City. *Journal of Transport Geography*. https://doi.org/10.1016/j.jtrangeo.2020.102711

Lloyd, A., & Cheshire, J. (2017). Deriving retail centre locations and catchments from geo-tagged Twitter data. *Computers, Environment and Urban Systems*, *61*, 108–118. https://doi.org/10.1016/j.compenvurbsys.2016.09.006

London City Hall. (2019a). *Christmas at Trafalgar Square*.

London City Hall. (2019b). *Policy 2.9 Inner London*.

Louppe, G. (2014). Understanding Random Forests from theory to practice [University of Liège]. In *Cornell University Library*. https://arxiv.org/pdf/1407.7502.pdf

Lowenthal, D. (1968). The American Scene. *Geographical Review*, *58*(1), 61–88. https://doi.org/10.2307/212832

Lu, Y. (2018). The association of urban greenness and walking behavior: Using google street view and deep learning techniques to estimate residents' exposure to urban greenness. *International Journal of Environmental Research and Public Health*. https://doi.org/10.3390/ijerph15081576

Luo, F., Cao, G., Mulligan, K., & Li, X. (2016). Explore spatiotemporal and demographic characteristics of human mobility via Twitter: A case study of Chicago. *Applied Geography*, *70*, 11–25. https://doi.org/10.1016/j.apgeog.2016.03.001

Lynch, K. (1960). *The Image of the City*. The MIT Press.

Maller, C., Townsend, M., Pryor, A., Brown, P., & St Leger, L. (2006). Healthy nature healthy people: "contact with nature" as an upstream health promotion intervention for populations. In *Health Promotion International*. https://doi.org/10.1093/heapro/dai032

Mason, L., Baxter, J., Bartlett, P., & Frean, M. (2000). Boosting algorithms as gradient descent. *Advances in Neural Information Processing Systems*.

McCollum, D., Ernsten-Birns, A., Feng, Z., & Everington, D. (2020). Mobile no more? The innovative use of administrative data linked to a census-based longitudinal study to investigate migration within Scotland. *Population, Space and Place*, e2312. https://doi.org/10.1002/psp.2312

McCullough, M. (2005). Digital Ground: Architecture, Pervasive Computing, and Environmental Knowing. In *Digital Ground*.

McInnes, L., Healy, J., & Astels, S. (2017). hdbscan: Hierarchical density based clustering. *The Journal of Open Source Software*. https://doi.org/10.21105/joss.00205

McKenzie, G., Janowicz, K., Gao, S., Yang, J. A., & Hu, Y. (2014). POI Pulse: A multi-granular, semantic signature-based information observatory for the interactive visualization of big geosocial data. *Cartographica*. https://doi.org/10.3138/cart.50.2.2662

Miah, S. J., Vu, H. Q., Gammack, J., & McGrath, M. (2017). A Big Data Analytics Method for Tourist Behaviour Analysis. *Information and Management*. https://doi.org/10.1016/j.im.2016.11.011

Milgram, S. (1976). Psychological Maps of Paris. In *Environmental psychology: people and their physical settings.* (pp. 104–124). Holt, Rinehart, and Winston.

Molinsky, J., & Forsyth, A. (2018). Housing, the Built Environment, and the Good Life. *Hastings Center Report*. https://doi.org/10.1002/hast.914

Molnar, C. (2019). 5.3 Accumulated Local Effects (ALE) Plot. In *Interpretable Machine Learning- A Guide for Making Black Box Models Explainable.* https://christophm.github.io/interpretable-ml-book/ale.html

Monson, M. (2009). Valuation Using Hedonic Pricing Models. *Cornell Real Estate Review*.

Moosavi, V. (2017). Urban morphology meets deep learning: Exploring urban forms in one million cities, town and villages across the planet. *ArXiv*. http://arxiv.org/abs/1709.02939

Murali, S. (2018). *An analysis on computer vision problems*.

Naik, N., Kominers, S. D., Raskar, R., Glaeser, E. L., & Hidalgo, C. A. (2017). Computer vision uncovers predictors of physical urban change. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(29), 7571–7576. https://doi.org/10.1073/pnas.1619003114

Naik, N., Philipoom, J., Raskar, R., & Hidalgo, C. (2014). Streetscore-predicting the perceived safety of one million streetscapes. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. https://doi.org/10.1109/CVPRW.2014.121

Nasar, J. L. (1990). The Evaluative Image of the City. *Journal of the American Planning Association*. https://doi.org/10.1080/01944369008975742

Neuhaus, F. (2012). New city landscape – Mapping urban Twitter usage. *Technoetic Arts*. https://doi.org/10.1386/tear.9.1.31_1

Ng, M., Fleming, T., Robinson, M., Thomson, B., Graetz, N., Margono, C., Mullany, E. C., Biryukov, S., Abbafati, C., Abera, S. F., Abraham, J. P., Abu-Rmeileh, N. M. E., Achoki, T., Albuhairan, F. S., Alemu, Z. A., Alfonso, R., Ali, M. K., Ali, R., Guzman, N. A., … Gakidou, E. (2014). Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: A systematic analysis for the Global Burden of Disease Study 2013. *The Lancet*. https://doi.org/10.1016/S0140-6736(14)60460-8

O'Brien, O., Cheshire, J., & Batty, M. (2014). Mining bicycle sharing data for generating insights into sustainable transport systems. *Journal of Transport Geography*, *34*, 262–273. https://doi.org/10.1016/j.jtrangeo.2013.06.007

O'Sullivan, D., & Unwin, D. J. (2010). Geographic Information Analysis: Second Edition. In *Geographic Information Analysis: Second Edition*. https://doi.org/10.1002/9780470549094

Onaverage. (2017). *Average walking speed.*

Ordnance Survey. (2020). *Order OS OpenData*. https://www.ordnancesurvey.co.uk/opendatadownload/products.html

Ordonez, V., & Berg, T. L. (2014). Learning high-level judgments of urban perception. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/978-3-319-10599-4_32

Pal, R. (2017). Overview of predictive modeling based on genomic characterizations. In *Predictive Modeling of Drug Sensitivity*. https://doi.org/10.1016/b978-0-12-805274-7.00006-3

Palmquist, R. B. (1984). Estimating the Demand for the Characteristics of Housing. *The Review of Economics and Statistics*. https://doi.org/10.2307/1924995

Papadopoulos, S., Zigkolis, C., Kompatsiaris, Y., & Vakali, A. (2011). Cluster-based landmark and event detection for tagged photo collections. *IEEE Multimedia*. https://doi.org/10.1109/MMUL.2010.68

Papandreou, G., Zhu, T., Kanazawa, N., Toshev, A., Tompson, J., Bregler, C., & Murphy, K. (2017). Towards accurate multi-person pose estimation in the wild. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*. https://doi.org/10.1109/CVPR.2017.395

Park, B., & Bae, J. K. (2015). Using machine learning algorithms for housing price prediction: The case of Fairfax County, Virginia housing data. *Expert Systems with Applications*. https://doi.org/10.1016/j.eswa.2014.11.040

Parker, C. J., May, A., & Mitchell, V. (2013). The role of VGI and PGI in supporting outdoor activities. *Applied Ergonomics*. https://doi.org/10.1016/j.apergo.2012.04.013

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*.

Peterson, S., & Flanagan, A. B. (2009). Neural network hedonic pricing models in mass real estate appraisal. *Journal of Real Estate Research*.

Piazzesi, M., Schneider, M., & Stroebel, J. (2014). Segmented Housing Search. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2543267

Pollakowski, H. O., & Wachter, S. M. (1990). The effects of land-use constraints on housing prices. *Land Economics*. https://doi.org/10.2307/3146732

Powe, N. A., Garrod, G. D., & Willis, K. G. (1995). Valuation of urban amenities using an hedonic price model. *Journal of Property Research*. https://doi.org/10.1080/09599919508724137

Rae, A., & Sener, E. (2016). How website users segment a city: The geography of housing search in London. *Cities*. https://doi.org/10.1016/j.cities.2015.12.002

Rapoport, A., & Hawkes, R. (1970). The perception of urban complexity. *Journal of the American Planning Association*. https://doi.org/10.1080/01944367008977291

Rattenbury, T., Good, N., & Naaman, M. (2007). Towards automatic extraction of event and place semantics from flickr tags. *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR'07*. https://doi.org/10.1145/1277741.1277762

Reades, J., Calabrese, F., & Ratti, C. (2009). Eigenplaces: analysing cities using the space – time structure of the mobile phone network. *Environment and Planning B: Planning and Design*, *36*(5), 824–836. https://doi.org/10.1068/b34133t

Redmon, J., & Farhadi, A. (2017). YOLO9000: Better, faster, stronger. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, *2017-Janua*, 6517–6525. https://doi.org/10.1109/CVPR.2017.690

Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. https://doi.org/10.1109/TPAMI.2016.2577031

Richards, D. R., & Friess, D. A. (2015). A rapid indicator of cultural ecosystem service usage at a fine spatial scale: Content analysis of social media photographs. *Ecological Indicators*. https://doi.org/10.1016/j.ecolind.2015.01.034

Richards, D. R., Tunçer, B., & Tunçer, B. (2018a). Using image recognition to automate assessment of cultural ecosystem services from social media photographs. *Ecosystem Services*, *31*, 318–325. https://doi.org/10.1016/j.ecoser.2017.09.004

Richards, D. R., Tunçer, B., & Tunçer, B. (2018b). Using image recognition to automate assessment of cultural ecosystem services from social media photographs. *Ecosystem Services*, *31*, 318–325. https://doi.org/10.1016/j.ecoser.2017.09.004

Roick, O., & Heuser, S. (2013). Location based social networks - definition, current state of the art and research agenda. In *Transactions in GIS*. https://doi.org/10.1111/tgis.12032

Roof, K., & Oleru, N. (2008). Public Health: Seattle and King County's Push for the Built Environment. *Journal of Environmental Health*.

Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition Sherwin Rosen The. *Tetrahedron Letters*. https://doi.org/10.1016/S0040-4039(00)85403-9

Salesses, P., Schechtner, K., & Hidalgo, C. A. (2013). The Collaborative Image of The City: Mapping the Inequality of Urban Perception. *PLoS ONE*, *8*(7). https://doi.org/10.1371/journal.pone.0068400

Schroeder, H. W., & Anderson, L. M. (1984). Perception of Personal Safety in Urban Recreation Sites. *Journal of Leisure Research*. https://doi.org/10.1080/00222216.1984.11969584

Scott, J. C. (1998). Seeing like a state: how certain schemes to improve the human condition have failed. In *Yale University Press*. https://doi.org/10.5860/choice.36-1224

Secord, J., & Zakhor, A. (2007). Tree detection in urban regions using aerial lidar and image data. *IEEE Geoscience and Remote Sensing Letters*. https://doi.org/10.1109/LGRS.2006.888107

Seresinhe, C. I., Moat, H. S., & Preis, T. (2018a). Quantifying scenic areas using crowdsourced data. *Environment and Planning B: Urban Analytics and City Science*, *45*(3), 567–582. https://doi.org/10.1177/0265813516687302

Seresinhe, C. I., Moat, H. S., & Preis, T. (2018b). Quantifying scenic areas using crowdsourced data. *Environment and Planning B: Urban Analytics and City Science*, *45*(3), 567–582. https://doi.org/10.1177/0265813516687302

Seresinhe, C. I., Preis, T., & Moat, H. S. (2017). Using deep learning to quantify the beauty of outdoor places. *Royal Society Open Science*, *4*(7). https://doi.org/10.1098/rsos.170170

Shaifee, M. J., Chywl, B., Li, F., & Wong, A. (2017). Fast YOLO: A Fast You Only Look Once System for Real-time Embedded Object Detection in Video. *Journal of Computational Vision and Imaging Systems*. https://doi.org/10.15353/vsnl.v3i1.171

Shen, Y., & Karimi, K. (2016). Urban function connectivity: Characterisation of functional urban streets with social media check-in data. *Cities*. https://doi.org/10.1016/j.cities.2016.03.013

Silva, T. H., Vaz De Melo, P. O. S., Almeida, J. M., & Loureiro, A. A. F. (2013). Social media as a source of sensing to study city dynamics and urban social behavior: Approaches, models, and opportunities. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/978-3-642-45392-2_4

Simonyan, K., & Zisserman, A. (2015, September 4). Very deep convolutional networks for large-scale image recognition. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*. http://www.robots.ox.ac.uk/

Singleton, Alex D., Spielman, S. E., & Folch, D. C. (2018). Urban Analytics. In *SAGE Publications Ltd*. https://doi.org/10.1177/2399808318759193

Singleton, Alexander D;, & Longley, P. A. (2009). Geodemographics, visualisation, and social networks in applied geography. *Applied Geography*, *29*, 289–298. https://doi.org/10.1016/j.apgeog.2008.10.006

Sirmans, G. S., Macpherson, D. A., & Zietz, E. N. (2005). The composition of hedonic pricing models. In *Journal of Real Estate Literature*.

Skogan, W. (1990). Disorder and decline: Crime and the spiral of decay in American neighborhoods. In *University of California Press*. https://doi.org/10.5860/choice.28-3588

Smith, C. (2019). *20 Interesting Flickr Stats and Facts (2019) | By the Numebrs*.

Smith, C. (2021). *20 Interesting flickr States and Facts (2021)| By the Numbers*. https://expandedramblings.com/index.php/flickr-stats/

Soo, C. K. (2013). Quantifying Animal Spirits: News Media and Sentiment in the Housing Market. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2330392

Steiger, E., Westerholt, R., Resch, B., & Zipf, A. (2015). Twitter as an indicator for whereabouts of people? Correlating Twitter with UK census data. *Computers, Environment and Urban Systems*, *54*, 255–265. https://doi.org/10.1016/j.compenvurbsys.2015.09.007

Stokols, D., & Moos, R. H. (1979). The Human Context: Environmental Determinants of Behavior. *Contemporary Sociology*. https://doi.org/10.2307/2064994

Strauss, A. L. (1988). Qualitative Analysis for Social Scientists. In *Cambridge University Press*. https://doi.org/10.2307/2069712

Stubbings, P., Peskett, J., Rowe, F., & Arribas-Bel, D. (2019). A hierarchical Urban forest index using street-level imagery and deep learning. *Remote Sensing*. https://doi.org/10.3390/rs11121395

Sui, D., & Goodchild, M. (2011). The convergence of GIS and social media: Challenges for GIScience. In *International Journal of Geographical Information Science*. https://doi.org/10.1080/13658816.2011.604636

Sulis, P., Manley, E., Zhong, C., & Batty, M. (2018). Using mobility data as proxy for measuring urban vitality. *Journal of Spatial Information Science*, *16*(16), 137–162. https://doi.org/10.5311/JOSIS.2018.16.384

Sun, J. B., Yuan, J., Wang, Y., Si, H. B., & Shan, X. M. (2011). Exploring spacetime structure of human mobility in urban space. *Physica A: Statistical Mechanics and Its Applications*, *390*(5), 929–942. https://doi.org/10.1016/j.physa.2010.10.033

Sun, Y., Fan, H., Bakillah, M., & Zipf, A. (2015). Road-based travel recommendation using geo-tagged images. *Computers, Environment and Urban Systems*. https://doi.org/10.1016/j.compenvurbsys.2013.07.006

Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. (2017). Inception-v4, inception-ResNet and the impact of residual connections on learning. *31st AAAI Conference on Artificial Intelligence, AAAI 2017*.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. *Proceedings of the IEEE*

Computer Society Conference on Computer Vision and Pattern Recognition. https://doi.org/10.1109/CVPR.2015.7298594

Toole, J. L., Ulm, M., González, M. C., & Bauer, D. (2012). Inferring land use from mobile phone activity. *Proceedings of the ACM SIGKDD International Workshop on Urban Computing - UrbComp '12*, 1. https://doi.org/10.1145/2346496.2346498

Tuan, Y.-F., Downs, R. M., & Stea, D. (1975). Image and Environment: Cognitive Mapping and Spatial Behavior. *Geographical Review*. https://doi.org/10.2307/213853

Wang, X., Wen, J., Zhang, Y., & Wang, Y. (2014). Real estate price forecasting based on SVM optimized by PSO. *Optik*. https://doi.org/10.1016/j.ijleo.2013.09.017

Wang, Y., Wang, S., Li, G., Zhang, H., Jin, L., Su, Y., & Wu, K. (2017). Identifying the determinants of housing prices in China using spatial regression and the geographical detector technique. *Applied Geography*. https://doi.org/10.1016/j.apgeog.2016.12.003

Wen, H., & Tao, Y. (2015). Polycentric urban structure and housing price in the transitional China: Evidence from Hangzhou. *Habitat International*. https://doi.org/10.1016/j.habitatint.2014.11.006

Wherrett, J. R. (2000). Creating landscape preference models using internet survey techniques. *Landscape Research*. https://doi.org/10.1080/014263900113181

Wilhelmsson, M. (2009). Construction and updating of property price index series: The case of segmented markets in Stockholm. *Property Management*. https://doi.org/10.1108/02637470910946426

Wohlwill, J. F. (1976). Environmental Aesthetics: The Environment as a Source of Affect. In *Human Behavior and Environment*. https://doi.org/10.1007/978-1-4684-2550-5_2

Wonderland, H. P. W. (2019). *Visit London's Christmas Extravaganza!*

Wu, C., Ye, X., Ren, F., Wan, Y., Ning, P., & Du, Q. (2016). Spatial and social media data analytics of housing prices in Shenzhen, China. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0164553

Wu, H., Jiao, H., Yu, Y., Li, Z., Peng, Z., Liu, L., & Zeng, Z. (2018). Influence factors and regression model of urban housing prices based on internet open access data. *Sustainability (Switzerland)*. https://doi.org/10.3390/su10051676

Xiao, Y., Chen, X., Li, Q., Yu, X., Chen, J., & Guo, J. (2017). Exploring determinants of housing prices in Beijing: An enhanced hedonic regression with open access POI data. *ISPRS International Journal of Geo-Information*. https://doi.org/10.3390/ijgi6110358

Xing, H., Meng, Y., Wang, Z., Fan, K., & Hou, D. (2018). Exploring geo-tagged photos for land cover validation with deep learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, *141*(May), 237–251. https://doi.org/10.1016/j.isprsjprs.2018.04.025

Yao, Y., Zhang, J., Hong, Y., Liang, H., & He, J. (2018). Mapping fine-scale urban housing prices by fusing remotely sensed imagery and social media data. *Transactions in GIS*, *22*(2), 561–581. https://doi.org/10.1111/tgis.12330

Ye, M., Yin, P., Lee, W. C., & Lee, D. L. (2011). Exploiting geographical influence for collaborative point-of-interest recommendation. *SIGIR'11 - Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*. https://doi.org/10.1145/2009916.2009962

Yi, C., & Huang, Y. (2014). Housing Consumption and Housing Inequality in Chinese Cities During the First Decade of the Twenty-First Century. *Housing Studies*. https://doi.org/10.1080/02673037.2014.851179

You, Q., Pang, R., Cao, L., & Luo, J. (2017). Image-Based Appraisal of Real Estate Properties. *IEEE Transactions on Multimedia*. https://doi.org/10.1109/TMM.2017.2710804

Zabel, J. (2015). The hedonic model and the housing cycle. *Regional Science and Urban Economics*. https://doi.org/10.1016/j.regsciurbeco.2015.07.005

Zhang, F., Zhang, D., Liu, Y., & Lin, H. (2018). Representing place locales using scene elements. *Computers, Environment and Urban Systems*, *71*(May), 153–164. https://doi.org/10.1016/j.compenvurbsys.2018.05.005

Zhang, F., Zhou, B., Liu, L., Liu, Y., Fung, H. H., Lin, H., & Ratti, C. (2018). Measuring human perceptions of a large-scale urban region using machine learning. *Landscape and Urban Planning*. https://doi.org/10.1016/j.landurbplan.2018.08.020

Zhang, S., Lee, D., Singh, P. V., & Srinivasan, K. (2017). How Much Is an Image Worth? Airbnb Property Demand Estimation Leveraging Large Scale Image Analytics. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2976021

Zhang, W., Li, W., Zhang, C., Hanink, D. M., Li, X., & Wang, W. (2017a). Parcel-based urban land use classification in megacity using airborne LiDAR, high resolution orthoimagery, and Google Street View. *Computers, Environment and Urban Systems*, *64*, 215–228. https://doi.org/10.1016/j.compenvurbsys.2017.03.001

Zhang, W., Li, W., Zhang, C., Hanink, D. M., Li, X., & Wang, W. (2017b). Parcel-based urban land use classification in megacity using airborne LiDAR, high resolution orthoimagery, and Google Street View. *Computers, Environment and Urban Systems*, *64*, 215–228. https://doi.org/10.1016/j.compenvurbsys.2017.03.001

Zhang, Y., & Dong, R. (2018). Impacts of street-visible greenery on housing prices: Evidence from a hedonic price model and a massive street view image dataset in Beijing. *ISPRS International Journal of Geo-Information*. https://doi.org/10.3390/ijgi7030104

Zhao, K., Kang, J., Jung, J., & Sohn, G. (2018). Building extraction from satellite images using mask R-CNN with building boundary regularization. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2018-June*, 242–246. https://doi.org/10.1109/CVPRW.2018.00045

Zheng, Y., Capra, L., Wolfson, O., & Yang, H. (2014). Urban computing: Concepts, methodologies, and applications. *ACM Transactions on Intelligent Systems and Technology*. https://doi.org/10.1145/2629592

Zheng, Y. T., Zha, Z. J., & Chua, T. S. (2012). Mining travel patterns from geotagged photos. *ACM Transactions on Intelligent Systems and Technology*. https://doi.org/10.1145/2168752.2168770

Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., & Torralba, A. (2018). Places: A 10 Million Image Database for Scene Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 40*(6), 1452–1464. https://doi.org/10.1109/TPAMI.2017.2723009

Zhou, B., Liu, L., Oliva, A., & Torralba, A. (2014). Recognizing city identity via attribute analysis of geo-tagged images. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/978-3-319-10578-9_34

Zube, E. H., Sell, J. L., & Taylor, J. G. (1982). Landscape perception: Research, application and theory. *Landscape Planning*. https://doi.org/10.1016/0304-3924(82)90009-0