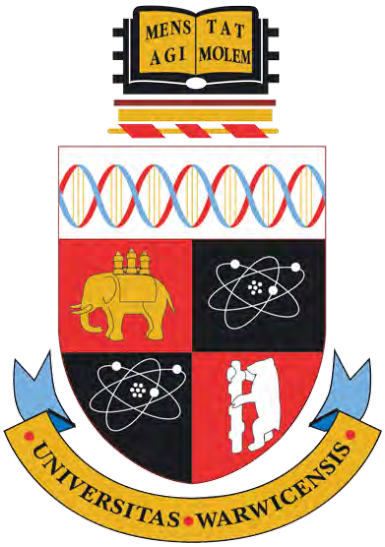

Using deep learning to infer house prices from street view imagery

by

Bhavan Kaur Chahal

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy in

Mathematics of Systems

Mathematics for Real-World Systems CDT

September 2022

Contents

List of Tables

List of Figures

Acknowledgments

Firstly, I would like to extend my deepest gratitude to my supervisors, Suzy Moat and Tobias Preis. Their wisdom and guidance have been invaluable, and I am thankful for their endless patience and support during challenging times. I would also like to extend my gratitude to Mario Gutiérrez-Roig and Federico Botta for sharing their expert advice and helping to shape my research.

I would like to extend my sincere thanks to all of the staff and students at the MathSys CDT for their guidance and great company. In particular, I would like to thank my D2.17 office family for interesting conversations, and I would like to thank Kutlwano Bashe, Kieran Kalair, and Christopher Norman for providing me with mathematical discussions, entertainment, and life-long friendships.

Thanks should also go to my closest friends for their unwavering support and for ensuring that I have a work-life balance, namely Lorna Payne, Charlotte Heath, Simi Kang, Rimi Kang, and Ruman Hayre. I am particularly indebted to my best friend, Frances Cadd, for everything. I could write another thesis discussing everything she has done for me but I will save that for another day.

I cannot begin to express my thanks to my family, without whom I would not be where I am in life. Thanks to my Bibi for continually supporting me, regardless of my decisions; my Baba for always believing in me; Ranjodh and Dilshere for challenging me; my father for opening up my mind to the world's possibilities; and my mother, for her daily (sometimes more) phone calls and always having my best interests at heart.

Last but not least, I would like to express my deepest appreciation to Ian Atkinson. I am extremely grateful every day that I met him at the MathSys CDT. The impact he has had on my life cannot be described by words, but I would like to say thank you for everything.

The research described in this thesis was supported by the EPSRC/MRC Centre for Doctoral Training for Mathematics for Real-World Systems under grant number EP/L015374/1. High-performance computing facilities were provided by the Scientific Computing Research Technology Platform of the University of Warwick.

Declarations

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy in Mathematics of Systems. It has been composed by myself and has not been submitted in any previous application for any degree.

Abstract

House prices are a key economic indicator. However, there is usually a delay of two months between the sale of a house and the availability of price data through the Land Registry. Over the past decade, vast quantities of imagery of streets and neighbourhoods have become available online, and the speed at which such imagery is being updated is growing. Simultaneously, advances in deep learning have enhanced our ability to determine the contents of an image automatically. In this thesis, we therefore ask: can we use deep learning to automatically infer house prices from large volumes of photos of houses and neighbourhoods?

Focusing on London as a case study, we analyse millions of photos from three sources, covering photographs of streets (*Google Street View*); crowdsourced photographs of outdoor neighbourhood locations (*Geograph*); and marketing images of house exteriors and interiors (*Zoopla*). For each image, we use a convolutional neural network to extract data on the presence of visual features. We investigate whether these features can be used in conjunction with an elastic net regression model to estimate the median house price during 2015 and 2016 for London neighbourhoods.

Our results demonstrate that it is possible to automatically infer local house prices from all three sources of imagery. The best estimates are made using *Zoopla* photographs. This may be due to the larger number of *Zoopla* images available in our sample (3,727,890 photographs, compared to 519,295 images for *Google Street View* and 273,999 images for *Geograph*). However, we find that increasing the number of images available to the model by combining images from different sources has limited benefits.

As the speed of collection of such imagery increases, this approach could equip national and local policymakers with more timely information on house prices, enabling them to make more informed decisions about economic policy.

Chapter 1

Introduction

Why are house prices so relevant to society? At an individual level, shelter is vital for a multitude of reasons such as privacy, stability, and safety. Moreover, owning a property can increase the chance of obtaining financial security. However, the escalating lack of affordable housing can make this a challenging endeavour. At a governmental level, policymakers may be looking to set policies that would promote or protect economic growth, and house prices are a key economic indicator.

Currently, it is difficult to obtain accurate and up-to-date information concerning house prices in England. Not only are they constantly changing, but there often exists a time delay of two months between a house transaction and the availability of data on the transaction from the Land Registry. In turn, the ability to make informed decisions based on house prices is hindered. Fortunately, alternative data sources of relevance are now available, and the rate at which they are updated is rising. Therefore, exploiting them to make inferences regarding house prices could allow individuals and organisations alike to make better-informed and more timely judgements.

Instinctively, the appearance of a property and its environment give us an indication of its value, as well as the value of properties in the general area. The volume of street view imagery containing these visuals is growing swiftly, with Google Street View being a large provider of such data. Simultaneously, improvements within the field of computer vision have paved the way for new methods that can analyse images rapidly, and extract information about what the images themselves contain. Therefore, in this thesis, we investigate whether we can use millions of images of houses and their neighbourhoods in conjunction with machine learning tools to automatically infer local house prices, as well as changes in local house prices. The photographs we analyse consist of images from three frequently updated data

sources: street-level imagery from Google Street View; images of local outdoor areas from Geograph; and interior and exterior images from the property website Zoopla. Note that the images in our dataset incorporate a natural, human perspective of neighbourhoods and properties, rather than satellite images, for instance, which show a bird's eye view. This gives us the opportunity to develop models that make inferences about neighbourhoods with the same information that humans have access to when walking through a new area.

Researchers and practitioners that analyse house prices rarely incorporate visual information in their models. However in this thesis, we develop a methodology that could make it easier for them to gain useful information about house prices from relevant image data and make use of it. This could potentially improve their understanding of house prices. Furthermore, relevant imagery could reveal insights into other topics in social science. As such, our methodology could be valuable for researchers in the field of computational social science, as the methods could be applied to other topics.

In this thesis, we start by taking an extensive look at how research has evolved in the field of computational social science in Chapter 2. We explore literature that focuses on the benefits of large datasets and the ways in which data science has advanced social science. We also discuss the increase in availability of image data, as well as the challenges surrounding the efficient collection of house price data. We then consider improvements in the performance of specific machine learning tools relevant to this thesis, namely convolutional neural networks and elastic net models.

In Chapter 3, we take the first step towards understanding how photos of houses can be utilised to estimate local house prices. Focusing on London, we apply a computer vision algorithm to street-level images of house facades from Google Street View. The visual features that we automatically extract from the images, such as 'skyscraper' and 'fountain', are then used to estimate median house prices for London neighbourhoods at a small area granularity. We discover that house prices can be estimated well from these images alone. We also identify visual features that the model associates with higher prices and lower prices. Furthermore, we explore whether we can capture relative changes in house prices over a decade.

Since we can successfully infer local house prices in London from Google Street View imagery, can we also infer local house prices in London from an alternative street-level image source? We utilise images from Geograph in Chapter 4 to answer this question. Geograph images capture outdoor scenes but are not restricted to the street network. They also contain images of environments such as parks and canals. The analyses we conduct in this chapter are similar to those in Chapter 3

but with Geograph images, rather than Google Street View images. Our findings demonstrate that we can estimate median house prices in London neighbourhoods reasonably well using Geograph imagery too. We also find that the more photos we use in the models, the better the performance is. However, the models we build with Geograph imagery do not perform as well as the models we build with Google Street View imagery, potentially due to fewer Geograph images being available.

The fact that we can infer median house prices of London neighbourhoods from outdoor images naturally leads us to ask whether we can estimate local house prices from indoor images too. Therefore in Chapter 5, we analyse indoor photographs of properties in London from Zoopla. Zoopla is a real estate website that is constantly updated with properties that are advertised via its platform. Our results show that we can also use Zoopla imagery to estimate local house prices in London and changes in local house prices very well. Furthermore, we find that if we build a model that infers house prices from both indoor and outdoor Zoopla imagery, performance only slightly improves in comparison to a model that uses indoor imagery alone. However, we find that using either indoor or indoor and outdoor Zoopla images delivers a substantial improvement on the performance of models we build with Google Street View or Geograph images, again perhaps because we analyse a much greater number of Zoopla images.

Each of the datasets discussed thus far can therefore be used to infer higher and lower median house prices in local areas of London. However, the nature of the imagery differs. Furthermore, we find that we achieve varying levels of success when estimating median house prices in London from Google Street View, Geograph, and Zoopla images independently. Therefore, in Chapter 6, we explore what makes a house price estimation model based on image data perform at its best. We conduct this investigation by combining the information we extract from our data sources and using this combined information to estimate local house prices and changes in local house prices in London. Our findings show that performance is strongly influenced by the inclusion of Zoopla data, regardless of whether we generate estimates of median house prices or changes in median house prices. We also provide further evidence suggesting that incorporating data from a larger volume of images in a house price estimation model leads to better results up to a limit, after which the improvements are marginal.

Our findings confirm that we can use large-scale, frequently updated image data with deep learning techniques to estimate local house prices successfully. In Chapter 7, we discuss our results, their consequences for policymakers, and possible steps for future research.

Chapter 2

Background

At the beginning of the twenty-first century, when the availability of data on the behaviour of large groups of people skyrocketed, interest grew in carrying out social science research using more computational approaches to solving problems. As a result, there is now an entire sub-discipline of the social sciences dedicated to computational social science (Conte et al., 2012; Lazer et al., 2009).

Over the past few years, the field of data science has evolved and now slightly overlaps with computational social science, leading to a new sub-discipline of data science known as social data science. Similarly to computational social science, social data science involves using computational methods to understand social phenomenon, but they differ in terms of the foundations for research. Whereas computational social science starts with a theoretical framework to be tested with data, social data science starts with data and data science techniques that lead to findings that inform theory. This relatively new paradigm can be beneficial, as data can reveal information that would not have been discovered otherwise, which could lead to the formulation of new yet realistic theories surrounding human behaviour.

The research in this thesis follows the practical framework of social data science. We start with data, as we have access to an abundance of them at not extra cost, and intend to understand house prices using new sources. We then compare our findings to what we already know to be true, as well as discovering new information. The field of social data science is very new relative to computational social science, so very little research has been conducted regarding social data science. Additionally, since the two disciplines somewhat overlap, computational social science is relevant to this thesis. We therefore review existing research in the well-established field of computational social science in this chapter.

In particular, we focus on the insights that have been gained from sizeable

datasets on human behaviour. Since our research explores the use of street view imagery with deep learning tools to infer house prices, we take a closer look at research using both online and offline data sources, with a keen focus on online image data and offline house price data. The use of machine learning tools is vital for our research, so we also evaluate advances in deep learning techniques for extracting information regarding the contents of images, and relevant statistical learning techniques for using this information to build accurate estimation models.

2.1 Computational social science

Using a computational approach to answer questions in the social sciences has led to a wealth of knowledge that furthers understanding of human behaviours, from modelling the relationship between attention and choice to the impact of bots on politics (Bhavnani & Choi, 2012; Mullett & Stewart, 2016; Tesfatsion, 2002; Woolley, 2016). In particular, using computational techniques with large datasets has paved the way for research that would not have been possible otherwise. Here, we review research using both large online datasets, generated through our interactions with the internet, and large offline datasets, produced through administrative processes such as house sales.

2.1.1 Studies using online data

Massive volumes of data on human behaviour come from online platforms, whose importance in daily life has grown considerably over the past decade. With greater use of various platforms, a greater variety of data are generated and available for analysis, providing insight into human behaviour in the real world (Letchford, Preis, & Moat, 2016b; Maruf, Meshkat, Ali, & Mahmud, 2015; Moat, Preis, Olivola, Liu, & Chater, 2014).

An example of a popular social media platform is Twitter, which is accessed by millions of users daily (Twitter, 2022). It is a microblogging site where users can post messages called tweets, but the messages uniquely have a character limit. The limit was 140 up until 2017, after which it became 280. Rapidly available Twitter data have been used for a number of applications (Bollen, Mao, & Zeng, 2011; Borge-Holthoefer, Rivero, & Moreno, 2012; De Domenico, Lima, Mougel, & Musolesi, 2013; González-Bailón, Borge-Holthoefer, Rivero, & Moreno, 2011; Sloan & Morgan, 2015; Takhteyev, Gruzd, & Wellman, 2012): for example, estimating crowd size (Botta, Moat, & Preis, 2015), predicting elections based on political sentiment (Tumasjan, Sprenger, Sandner, & Welpe, 2010), and generating quicker indicators of

the prevalence of infectious diseases, such as dengue (Marques-Toledo et al., 2017; Mizzi et al., 2021). Alis et al. (2015) take advantage of this novel data source and study whether the length of these Twitter messages is different across regions of the UK. This analysis is inspired by the discovery that the median length of Twitter messages varies across the US (Alis & Lim, 2013), and that the median message length is impacted by demographic grouping.

The microblogging site Sina Weibo is also a popular social media platform with millions of active users (Weibo, 2020). Like Twitter, the messages had a character limit of 140, but Sina Weibo removed the limit in 2016. If researching human behaviour in China, these data can be useful, as shown by Jendryke, Balz, and Liao (2017). They make use of the data by accumulating text messages and location details to identify associations between Sina Weibo data and census information. For example, they find that the messages sent by Sina Weibo users in a particular location are not necessarily from residents, but could also be from tourists, highlighting popular tourist spots.

An additional study by Jendryke, Balz, Mcclure, and Liao (2017) looks at the relationship between urbanisation in Shanghai and social media messages from Sina Weibo, concluding that social media data alone can contribute towards a more granular analysis of urbanisation. In other words, the evolution of a rural area to an urban area can be traced at a smaller scale using social media messages, rather than measuring urbanisation from aggregated data only. Furthermore, Sina Weibo data have been used to understand the relationship between air pollution and happiness (Zheng, Wang, Sun, Zhang, & Kahn, 2019), humour reactions in crisis (Beeston et al., 2014), and the relationship between Chinese tourist mobility and consumption in London (Ye, Newing, & Clarke, 2020). For the latter, the activities of Chinese tourists in London can be tracked when they 'check in' to places. This includes places of consumption, such as restaurants and shops.

Another popular social media platform offering a variety of data is Facebook, which currently has billions of monthly users (Meta, 2022). An abundance of data are collected by the multimedia sharing platform, such as interactions between users and their network of friends and associates. These interactions are explored by Bakshy, Messing, and Adamic (2015), who try to understand how people interact via social networks. By quantifying the diversity in information received by users due to sharing, they find that individual networks play a large role in exposing content. More specifically, they compare the political alignment of millions of individual Facebook users with their friends network, and discover that individuals are more likely to see information that concurs with their ideologies in their news feed, rather than

opposing viewpoints, as the friends network often has similar alignments.

Many other research questions have been investigated with the data available from Facebook (Bond et al., 2012; Ferrara, 2012; Shen, Brdiczka, & Liu, 2015), including understanding the spreading of misinformation (Del Vicario et al., 2016), providing evidence of emotional contagion outside of physical human interaction (Kramer, Guillory, & Hancock, 2014), and predicting private and personal information (Kosinski, Stillwell, & Graepel, 2013). The latter two studies, in particular, inadvertently reveal the threat to user privacy on social media, as Facebook users unintentionally reveal personal traits and are seemingly unaware of doing so. Flick (2015) further expands on the poor ethical practices in research projects that use large datasets by using the (Kramer et al., 2014) paper as a case study. They specifically focus on the lack of informed consent and suggest practical recommendations for obtaining consent ethically.

In two studies, Chetty et al. (2022a, 2022b) also explore the way in which people interact via Facebook, and the impact of the strength of an individual's social network (or social capital) on their economic prospects. They define measures of social capital in Chetty et al. (2022b) and determine the factors that lead to social interaction between people with high socioeconomic status and people with low socioeconomic status in Chetty et al. (2022a). Amongst other conclusions, they reveal the importance of social capital for escaping poverty, and discover that exposure to economically diverse individuals increases the chances of friendships to occur amongst people with different backgrounds, which in turn increases the chances of individuals to improve their economic status.

Similarly to social media platforms, the usage of online search engines has risen rapidly over time and a large volume of data are generated due to the wealth of information available at our fingertips. The most favoured search engine, by far, is Google, with billions of searches conducted everyday (Nayak, 2019; StatCounter, 2022). They have also developed the Google Trends platform, providing public access to data on the volume of Google searches for different keywords, phrases and topics from 2004 onwards. All data accessible are aggregated (and hence anonymised), and Google state that the sample used to produce the statistics is representative of all Google searches (Google, n.d.-c).

A number of researchers have demonstrated how Google Trends data can provide insights into human behaviour (Botta, Preis, & Moat, 2020; Caputi, Leas, Dredze, Cohen, & Ayers, 2017; Kristoufek, Moat, & Preis, 2016; Letchford et al., 2016b; Moat, Olivola, Chater, & Preis, 2016; Noguchi, Stewart, Olivola, Moat, & Preis, 2014; Olivola, Moat, & Preis, 2019; Preis, Moat, Stanley, & Bishop, 2012;

Schootman et al., 2015). For example, Choi and Varian (2012) make use of Google Trends data to nowcast economic indicators. Nowcasting is similar to forecasting, but specifically involves predicting the present or very near future, rather than the distant future. This can be beneficial because information regarding the present tense is vital for decision making in policy and business. Unfortunately, measurements of the current state of society are frequently delayed. For example, gross domestic product data, financial asset prices, and statistics associated with inflation are often subject to a publication lag. Therefore, being able to nowcast such measurements would allow for better decision-making in the present, before the official publication of measurements.

Choi and Varian (2012) specifically build simple seasonal auto-regressive models using time-series data regarding different topics of relevance to society and compare them to models that also include relevant search terms related to these topics. They can then determine whether it is possible to nowcast their chosen measurements. One measurement they attempt to nowcast is the number of initial claims for unemployment benefits using the Google Trends categories *Jobs* and *Welfare & Unemployment*. They find that all of the models they build with Google Trends terms, including this one, tend to outperform those that exclude the terms.

Another method of utilising search data is explored by Ginsberg et al. (2009), who find that Google Trends data can be used to track influenza-like illnesses in areas with a large population of internet users. Official data on the number of current flu cases tends to arrive with one or two weeks' delay, due to the time required for patients to speak to a doctor and for the doctor to report the case to the country's central health authorities, such as the Centers for Disease Control and Prevention (CDC) in the US, or the UK Health Security Agency (previously Public Health England) in the UK. Quicker insights into the current number of flu cases are very helpful for medical professionals, as they can help them manage the resources required to fight against a disease that causes many deaths every year: for example, by rescheduling non-urgent operations to free up hospital beds for flu patients. The researchers hypothesise that users search for their symptoms as they appear and demonstrate that the popularity of certain search terms can be used to estimate the level of influenza before official figures on the number of influenza cases become available. The resulting approach is known as 'Google Flu Trends' and was long heralded as a major success story in computational social science.

Unfortunately, in the winter of 2012–2013, the Google Flu Trends estimate suggested that at the peak of the US flu season, over 10% of the US population was suffering from influenza-like illness. When the official CDC data eventually arrived,

the true figure was closer to 6% (Butler, 2013). One hypothesis as to the source of the error was that there was a lot of media coverage of the danger of catching flu in that winter season, such that people had started searching for flu-related terms to gather information about the disease without having a current flu infection.

Preis and Moat (2014) provide evidence that such problems can be overcome by augmenting the original Google Flu Trends approach in two key ways. First, they show how a model can be constantly recalibrated with the most recent data, so that estimates are based on an up-to-date picture of the relationship between the volume of flu-related searches and the true number of flu cases. Second, they observe that recent official data on the number of flu cases are likely to provide some insight into the current number of flu cases, and show how old, delayed data on the true number of flu cases can be used to inform estimates alongside search volume data. They call their approach 'adaptive nowcasting', as it produces estimates of the current state of affairs whilst constantly adapting the model used to generate the estimates. Their results demonstrate that such an approach can be used to successfully avoid the problems witnessed in the winter of 2012–2013.

A different approach is required if official data do not arrive with a simple delay of one or two weeks, and instead trickle in over multiple months. This is the case for data on many infectious diseases worldwide, including the mosquito-borne diseases dengue and chikungunya in Brazil. Mizzi et al. (2021) and Miller et al. (2022) show how data on the volume of Google searches for the diseases can help in these circumstances too, when coupled with a Bayesian approach for estimating how many cases will be reported after varying lengths of delay.

Previous work has also revealed how data on Google searches can be used to anticipate future human behaviour. For example, Preis, Moat, and Stanley (2013) show that increases in searches for more financially-related terms on Google tend to be followed by stock market falls. Preis and Moat (2015) show that increases in searches for company names or company stock ticker symbols similarly tend to be followed by falls in the price of the stock. Finally, Curme, Preis, Stanley, and Moat (2014) present evidence that increases in searches for politically-related terms on Google also provide early warning signs of stock market falls. Bollen et al. (2011) and Bordino et al. (2012) also use online data to predict the stock market. The former uses mood states identified in Twitter feeds to estimate changes in the value of the Dow Jones Industrial Average, whereas the latter uses Yahoo! search engine query data related to specific stocks to infer trading volumes of those stocks. Both studies exhibit some degree of success.

A further popular source of online information due to its accessible nature

and high frequency of updates is Wikipedia, a free online encyclopedia with millions of articles. The metadata associated with Wikipedia articles in particular have been useful for a number of studies (Capocci et al., 2006; Kristoufek, 2013; Mestyán, Yasseri, & Kertész, 2013; Yasseri, Kornai, & Kertész, 2012; Yasseri, Sumi, & Kertész, 2012; Yasseri, Sumi, Rung, Kornai, & Kertész, 2012). Moat et al. (2013), for instance, find that the number of views of certain Wikipedia pages can be used to anticipate stock market movements that occur shortly after. Specifically, an increase in views of pages relating to finance tends to precede falls in the stock market.

When studying human behaviour, it is also possible to collect huge volumes of data from large groups of people who volunteer information by crowdsourcing data online. Crowdsourced data are essentially large-scale survey and experiment data, and can be collected and analysed in numerous ways for a range of topics (Mastrandrea, Fournet, & Barrat, 2015; Quercia, Schifanella, & Aiello, 2014; Zhang, Li, & Hong, 2016). For example, MacKerron and Mourato (2013) develop a smartphone app called Mappiness that asks users to report on their wellbeing twice a day and records their location to find links between the wellbeing and environment of an individual. They discover that their app users are substantially happier in natural outdoor areas than urban areas. Solymosi, Bowers, and Fujiyama (2017) explore the subjective bias involved when crowdsourcing data in relation to criminological research. They introduce a new approach for measuring the perception of crime and safety using crowdsourced data by extrapolating from the biases, rather than discarding them as a limitation.

2.1.1.1 Image data

Beyond text-based online information, increasing volumes of images are now also shared widely on the internet, offering new insights into human behaviour around the globe. For example, the image and video hosting site Flickr has a large historic corpus of images, with considerable metadata attached to each photo, such as tags, comments, and location data (Flickr, n.d.; Preis, Moat, Bishop, Treleaven, & Stanley, 2013; Seresinhe, Moat, & Preis, 2018; Seresinhe, Preis, & Moat, 2016). Alanyali, Preis, and Moat (2016) track protests using Flickr photographs that are tagged with location data. They discover that the proportion of photos in a country tagged with the word 'protest' is higher when there are greater numbers of reports of protests in that country in the newspaper The Guardian. This highlights the potential for using image metadata as a rapid measurement of human behaviour.

An alternative use of location data from Flickr photographs is illustrated by Barchiesi, Preis, Bishop, and Moat (2015). They devise an algorithm to infer

the probability that someone is in a certain location, as well as the probability of movement between two locations. Their results show that estimates from their algorithm match up with official UK figures on travel flows. Barchiesi, Moat, Alis, Bishop, and Preis (2015) demonstrate that data on the location of Flickr photographs can be used to estimate the annual number of visitors to the UK from countries around the world, and Preis, Botta, and Moat (2020) show that this approach can be extended to all G7 countries. In a similar vein, Wood, Guerry, Silver, and Lacayo (2013) test whether Flickr metadata can be used to quantify the popularity of nature attractions. They find that their estimates are comparable to official survey data on the number of visits to nature parks around the world.

Instagram is a another popular social media platform with a focus on photographs, and regularly attracts millions of users (Rodriguez, 2021). A number of researchers have demonstrated how data associated with Instagram photographs can provide insight into behaviour in the real world (Hassanpour, Tomita, DeLise, Crosier, & Marsch, 2018; Hausmann et al., 2017; Heikinheimo et al., 2017; Paül i Agustí, 2021; Rossi, Boscaro, & Torsello, 2018; Tenkanen et al., 2017; Trevisan et al., 2019). For example, Ferwerda, Schedl, and Tkalcic (2016) show that they are able to infer personality traits from features identified in Instagram photos. Botta, Moat, and Preis (2020) illustrate how data from Instagram photos can be used to estimate the size of a crowd. Le Moignan, Lawson, Rowland, Mahoney, and Briggs (2017) explore how Instagram has changed the concept of 'family snapshots'.

Other sources of online imagery, such as the crowdsourcing project Geograph (Geograph, n.d.-a), seek to capture the appearance of the outdoor environment. Seresinhe, Preis, and Moat (2015) draw on hundreds of thousands of Geograph images to demonstrate that people who live in more scenic environments report their health to be better, even when considering the amount of local greenspace, socioeconomic metrics and data on pollution. Their analyses use scenic ratings from the online game Scenic-or-Not (Data Science Lab, n.d.), which has gathered over 1.5 million votes on Geograph photos of over 200,000 locations across Great Britain. By combining Scenic-or-Not data with three years of happiness data from over 15,000 people from the mobile app Mappiness, Seresinhe, Preis, MacKerron, and Moat (2019) also demonstrate that people in more scenic locations report themselves to be happier, even in cities. Using deep learning tools, Seresinhe, Preis, and Moat (2017) show that scenic locations are not just green and natural locations: buildings with character can boost the value of a scene too. McKenna et al. (2021) also use data from Scenic-or-Not to investigate the impact on British onshore wind energy capacity and cost if wind farms are not permitted in the most scenic locations.

A further source of image data on the outdoor environment is Google Street View (Google, n.d.-d). Street View is a collection of images from all across the world depicting outdoor spaces, collected by Google and their contributors. Google originally collected images at a satellite level only with Google Maps (Reid, 2020), but like Geograph, Street View takes this one step further and introduces a more human perspective to the data collected. Vandeviver (2014) exploits this vast data source and explores how online mapping applications can be useful for criminological research. Further studies also use Google Street View data to improve our understanding of neighbourhoods, from identifying accessibility issues on the streets to determining whether safer-looking neighbourhoods are more lively (De Nadai et al., 2016; Doersch, Singh, Gupta, Sivic, & Efros, 2012; Griew et al., 2013; Hara, Le, & Froehlich, 2013; Odgers, Caspi, Bates, Sampson, & Moffitt, 2012; Rundle, Bader, Richards, Neckerman, & Teitler, 2011). For example, Arietta, Efros, Ramamoorthi, and Agrawala (2014) provide an initial assessment of whether Google Street View data can be used to automatically infer relationships between the visuals of US neighbourhoods and their non-visual characteristics, such as crime statistics and house prices.

The examples we have discussed here only cover a small number of the various different kinds of image data available for analysis, from images of animals to brain scans. Furthermore, the quantity of images available is rapidly increasing and a large amount of visual information exists to be exploited. This presents a huge opportunity in relation to housing because visual information can be quite indicative of house prices in a given area, and forms the first step in the practical framework we implement.

2.1.1.2 Issues with online data

Throughout this section, we have demonstrated that online data can be very useful for understanding human behaviour and furthering research in the social sciences. However, no data concerning humans and their environments can ever be perfect, since humans themselves are subjective creatures. Issues can occur when using online data, from accessing it to drawing conclusions from it. We therefore need to be aware of the disadvantages of using online data in order to accurately interpret results based on it.

Specifically in the field of social science, surveys are commonly used to gain information about individuals, and online surveys can be a more cost-effective and less time-consuming way of gathering information relative to in-person surveys. Lefever, Dal, and Matthíasdóttir (2007) and Duffy, Smith, Terhanian, and Bremer (2005) discuss the use of online survey data, including the disadvantages. For example, online

surveys can be easier to complete and can be more transparent. This could allow participants to see the results as they are updated, and manipulate their answers to change the results. In turn, the results of the survey may not be reliable. It is important to access data from trustworthy sources and be aware of how reliable a data source is.

Privacy and ethical concerns should also be considered when using online data. Israel (2015) discuss ethical approaches that can be taken in social science that also maintain privacy, for example, informed consent. Since online data often contain personal information about people, those involved should be informed of how their data are being used, allowing them to choose what they are comfortable with sharing. If privacy and ethical concerns are ignored, data could potentially be misused. Kröger, Miceli, and Müller (2021) outline numerous ways in which data can be used against people, such as blackmail or social media burglary, which is when information about someone's whereabouts is exploited to break into their home.

Additionally, online data can be biased for many reasons, including the contents of the data and the methodology used to collect it. Therefore, when drawing conclusions based on online data, it is crucial that biases are taken into consideration. Mehrabi, Morstatter, Saxena, Lerman, and Galstyan (2022) conduct a survey of the many types of biases that could occur in machine learning algorithms and data, including population bias and behavioural bias. Population bias occurs when a sample of data does not accurately represent the population being studied, an example of which would be surveying males about female health. Behavioural bias occurs when user behaviour differs between data sources, an example of which would be using emojis on Facebook but not Instagram. In the case of online surveys, one type of bias that could occur is order bias, which means that the order of the questions influences a participants' responses. Some of the ways in which this bias could be reduced include randomising questions and grouping survey items based on topic.

Some issues with online data are issues that occur with all data in general. If the volume of data available is insufficient, it will be difficult to draw accurate conclusions from them. Interpretation would also be hindered if the quality of data is poor, for instance, if a large image data source consisted of pixelated or blurry imagery, it would be difficult to extract useful information from it. Furthermore, messy datasets, such as those containing duplicate entries or missing values, could reduce the volume of information available to use for analysis, and introduce biases in the dataset.

In this section, we briefly discuss a few common issues surrounding online data, but many more exist. Provided that a user of online data is aware of the

limitations of their data, online data can be very useful for research, whether it is within the field of social science or not.

2.1.2 Studies using other data sources

A large number of studies have focused on data that captures human behaviour at large-scale, but is not solely generated by online services or our use of the internet. A rich source of data on the border between online and offline data sources comes from mobile phone companies, who collect usage data from their customers (Barlacchi et al., 2015; Blondel, Decuyper, & Krings, 2015; Deville et al., 2014; Douglass, Meyer, Ram, Rideout, & Song, 2015; González, Hidalgo, & Barabási, 2008; Krings, Calabrese, Ratti, & Blondel, 2009). Cheng, Fang, Hong, and Yang (2017) discuss the type of data collected, including GPS monitoring and call transmissions, and potential applications of the data. For example, the time and location information of a user could contribute towards relevant recommendations at that point in time. Mobile phone data could also enhance traffic monitoring procedures and improve congestion (Steenbruggen, Tranos, & Nijkamp, 2015), as well as help estimate crowd size (Botta, Moat, & Preis, 2015).

Stock market and financial data are also widely accessible for analysis (Distinguin, Rous, & Tarazi, 2006; Münnix et al., 2012; Preis, Kenett, Stanley, Helbing, & Ben-Jacob, 2012; Preis, Schneider, & Stanley, 2011; Sornette, Woodard, & Zhou, 2009). For instance, Botta, Moat, Stanley, and Preis (2015) quantify the probability of large price changes in stock markets. They discover that the tails of the return distributions obey a power law for time scales ranging from 300 seconds to 3600 seconds.

As well as data on online information gathering, large-scale data on stories in the news have been shown to provide insights into stock market moves, perhaps reflecting how traders rely on news developments to make trading decisions (Atkins, Niranjan, & Gerding, 2018; Curme, Stanley, & Vodenska, 2015; Hagenau, Liebmann, & Neumann, 2013; Narayan & Bannigidadmath, 2017; Piškorec et al., 2014; Souma, Vodenska, & Aoyama, 2019; Tetlock, 2010). Using six years of data on stories published in the Financial Times, Alanyali, Moat, and Preis (2013) find that on days on which a company is mentioned more frequently in the news, transaction volume for that company's stock increases both the day before and on the same day. Curme, Zhuo, Moat, and Preis (2017) find that a measure of the diversity of stories covered by the Financial Times can improve forecasts of trading volume. They also report that stock market falls tend to be followed by the news focusing in on a smaller number of topics.

Other researchers have focused on the volumes of data produced by our everyday travels (Manley, Zhong, & Batty, 2016; O'Brien, Cheshire, & Batty, 2014; Sulis & Manley, 2018; Zhong, Arisona, Huang, Batty, & Schmitt, 2014; Zhong, Manley, Müller Arisona, Batty, & Schmitt, 2015). For example, Miller, Moat, and Preis (2020) demonstrate that billions of real-time messages on the location of aeroplanes undertaking millions of flights can help generate quicker estimates of elements of GDP. Additionally, Goddard (1970) conducted a study that uses taxi flow data from 1962 to extract functional regions of Central London, where a functional region refers to an area in London where taxis are frequently used. However, due to technological limitations, the analysis was relatively small. Therefore, Demšar, Reades, Manley, and Batty (2018) attempt to replicate the study using modern technology and a modern taxi flow dataset from 2011 and conduct a comparable analysis. Their initial goal was to gain insights into the temporal changes in structure in the city centre, but they were unable to fully replicate the study. Therefore, they extract functional regions using a slightly different method and reflect on the issues surrounding the replication of the study. From their analysis, they are able to make some meaningful comparisons between the original and replicated studies, such as learning about the monetary impact of pre-booking taxis in 2011, which was not possible when making taxi journeys in 1962.

Many studies have explored the use of large crime datasets to understand criminal behaviour and its after-effects on society (Bowers, Johnson, & Pease, 2004; N. Johnson et al., 2011; N. F. Johnson et al., 2016; Mohler, Short, Brantingham, Schoenberg, & Tita, 2011; Skarlatidou, Ludwig, Solymosi, & Bradford, 2021; Solymosi, Bowers, & Fujiyama, 2015; Solymosi, Cella, & Newton, 2017). For instance, Davies and Johnson (2014) hypothesise that streets with a higher potential for usage have a higher risk of burglary. Note that potential usage refers to how likely a street is to be used by pedestrians, and the street could be used more because of its close proximity to a train station or it leads to a shopping centre, for example. Davies and Johnson (2014) use crime data from the city of Birmingham in England, in addition to street network data from the Ordnance Survey, to test and provide evidence in favour of their hypothesis. Solymosi, Ashby, Cohen, and Sidebottom (2017) explore the use of open data to estimate the risk of violent crime on the London Underground. Open data refers to data that can be used and redistributed freely by anyone, possibly subject to an attribution. Trendl, Stewart, and Mullett (2021) use 10 years' worth of crime and incidence data from West Midlands Police and historical sports data to investigate the link between football, alcohol, and domestic abuse in England. Their key finding is that a victory for England leads to an increase in the reported number

of domestic abuse incidents that involve alcohol.

Computational social science research has also had a strong focus on health, drawing on official health data, transport data, and more (Chinazzi et al., 2020; Colizza, Barrat, Barthélemy, & Vespignani, 2006; Eetemadi et al., 2020; Grunerbl et al., 2015; Kraemer et al., 2020; Pavel, Jimison, Korhonen, Gordon, & Saranummi, 2015). Colizza, Barrat, Barthelemy, Valleron, and Vespignani (2007) model the spread of the avian influenza virus at a global scale, and compare various strategies for containment of the virus. They use data on worldwide air travel and census data from urban areas near airports for their investigation, and are able to determine the impact of different levels of antiviral drug usage on containing and mitigating the spread of viruses with varying reproductive rates.

Scientists have also turned the microscope of computational social science on themselves, and considered the volumes of data produced through scientific publishing (Fortunato et al., 2018; Frenken, Hardeman, & Hoekman, 2009; Mingers & Leydesdorff, 2015; Penner, Pan, Petersen, Kaski, & Fortunato, 2013; Petersen, Riccaboni, Stanley, & Pammolli, 2012; Radicchi, Fortunato, & Castellano, 2008; Radicchi, Fortunato, Markines, & Vespignani, 2009). Letchford, Moat, and Preis (2015) show that journals that publish papers with shorter titles receive more citations per paper. Similarly, Letchford, Preis, and Moat (2016a) demonstrate that papers with abstracts that contain more frequently used words receive more citations.

2.1.2.1 House price data

Of most relevance to this thesis amongst offline data, many researchers have carried out extensive analyses using house price datasets. Shelter is a basic human need, which is why data on house prices are so important both to individuals and policy makers. Unfortunately, there is often a time lag between the sale of a house and the availability of the transaction data, meaning that decision-makers are left with out-of-date data.

In the UK, the HM Land Registry is the official source for land and property data (HM Land Registry, 2016). They collect house price data as properties are registered, but there is typically a delay of two months between the purchase of a house and its registration. The data collected are then shared with the Office for National Statistics (ONS), who use it to conduct analyses and calculate national statistics (HM Land Registry, 2020). Researchers have taken advantage of this resource to attempt to both predict and better understand influences on house prices in the UK for some time (Badarinza & Ramadorai, 2018; Begiazi & Katsiampa, 2019; Brown, Song, & McGillivray, 1997; Cook, 2005; Feng & Jones, 2015; Pain &

Westaway, 1997). For instance, Yao and Fotheringham (2016) use a mixed modelling approach that considers the impact of both spatial and temporal components on house prices in Fife, Scotland.

Over in the United States, the methodology of house price data collection is not as straight forward because there is no official publicly available dataset of house prices. Despite the absence of a publicly available dataset of all house prices in the US, numerous housing-related data sources are publicly available for analysis, as discussed by Weinberg (2014). However, the data often come from surveys conducted by the United States Census Bureau (Weinberg, 2014). A particularly useful dataset is available from the Federal Housing Finance Agency (FHFA), who have a complete collection of open access house price indices with data from all of the states in the US (FHFA, n.d.).

Many researchers have used these data regarding house price indices to further understand measurements relating to house prices, such as changes in house prices, as well as house prices themselves (Bogin & Doerner, 2019; Bourassa, Hoesli, & Oikarinen, 2016; Brady, 2014; Kim & Rous, 2012). One such example is given by Doerner and Leventis (2013), who analyse the effects of distressed sales on house price changes over several years. In this instance, distressed sales refer to property transactions that must occur quickly, and these properties are often sold below market value. Sellers may accept this financial loss for a number of reasons, including urgent debts or divorce.

In general, there is a wealth of literature surrounding the impact of different variables on house prices and they often use a hedonic modelling approach (Gibbons, 2004, 2015; Gibbons, Heblich, Lho, & Timmins, 2016; Gibbons & Machin, 2003; Gibbons, Peng, & Tang, 2021; Higgins, Rezaei, & Wood, 2019; Morancho, 2003). Gibbons and Machin (2005) explore the effects of distance from stations in London on house valuations, whereas Sims, Dent, and Oskrochi (2008) model the impact of wind farms on house prices around a wind farm in Cornwall. Furthermore, Sirmans, MacDonald, Macpherson, and Zietz (2006) study housing characteristics, including the number of bedrooms and whether there exists air conditioning, that frequently appear in hedonic pricing models for single-family housing. They determine whether their importance varies based on external factors, including geographical location and model specification, i.e. the number of variables included in the model.

House price data and related statistics are often published with a delay, so it could be useful to nowcast house prices, especially since house prices are a key economic indicator. As mentioned previously, nowcasting could improve decision making in policy and business. Online data may be suitable for this purpose, as

many online data sources are frequently updated and may be able to indicate house prices movements in advance of publication. Indeed, online data have previously been used to successfully nowcast house prices and housing-related measures, such as house price indices, across the globe (Askitas, 2016; Mitra, Sanyal, & Choudhuri, 2017; Oestmann & Bennöhr, 2015; Widgrén, 2016).

A key study on this topic was carried out by McLaren and Shanbhogue (2011), who use internet search data to nowcast house prices in the UK. More specifically, they consider generic search terms related to house prices, such as 'estate agents' and 'mortgage', and use the corresponding Google Trends search volumes to detect relationships between the search volumes and monthly house price growth. They provide evidence suggesting that utilising the information available from online search data can help generate quicker indicators of house price growth.

House price data is generally not as easy to access for researchers as image data is, but there still exists an enormous amount of accurate house price data for the UK. Traditional models estimating house prices often include variables that are known to be relevant and associated with housing, indicating that the models tend to be based on theories. However, we are looking to gain new insights about house prices in this thesis, and do not know what these insights will be in advance.

2.1.3 Summary

In this section, we examine key literature in the field of computational social science that involve using large volumes of data to understand human behaviour. In particular, we focus on studies that use online data from sources such as Google Trends and Twitter, and studies that use offline data sources such as data from the stock market and data on crime. Since this thesis focuses on the use of street view imagery to infer house prices, we also take a closer look at research that analyses image data and house price data, whilst also considering the issues that could occur when using online data.

While plenty of interesting work has been done with both image data and house price data, very few house price models currently consider the visual environment in which a house is situated. Traditionally, incorporating image-based information in a computational model has been challenging, due to the complexity of such data. In the next section, we explore machine learning techniques for extracting information from images and investigating how it relates to house prices.

2.2 Computational methods for image analysis

In recent years, advances in technology have paved the way for novel methods, known as machine learning techniques, that are especially useful when working with large datasets. Well-known machine learning techniques include topic modelling, hierarchical clustering, and random forest regression. In this section, we focus on the machine learning techniques that are particularly helpful for our research. To make use of huge amounts of image data for our analyses, we need to use techniques that are able to extract useful information from the images in order to investigate the relationship between them and house prices. To model this relationship, we further need to use a method that is capable of handling a large quantity of variables without sacrificing accuracy. Therefore, we review research associated with convolutional neural networks and elastic net models.

2.2.1 Convolutional neural networks

Convolutional neural networks are machine learning algorithms that have transformed our ability to extract information from images. However, to understand how convolutional neural networks operate and why they are favoured for image recognition tasks, we must first discuss neural networks.

In machine learning, neural networks refer to a series of algorithms that work together to learn patterns and classify data (Anthony & Bartlett, 1999; Hagan, Demuth, Beale, & De Jesús, 2014; LeCun, Bengio, & Hinton, 2015). The idea of neural networks was first introduced by McCulloch and Pitts (1943), who characterised neural activity in the brain as logical expressions that follow set conditions. Hebb (1949) built upon this idea, suggesting that as neural activity grows, neural pathways strengthen and information can be learned.

A neural network consists of layers, where each layer is a collection of nodes. The layers are stacked one after the other, and the nodes from one layer are connected to nodes of its nearby layers. The selection of nodes, layers, and connections is called the architecture of the neural network. An example architecture is visually depicted in Fig. 2.1. The circles represent nodes and the arrows represent connections between nodes. The network contains an input layer, an output layer, and two hidden layers.

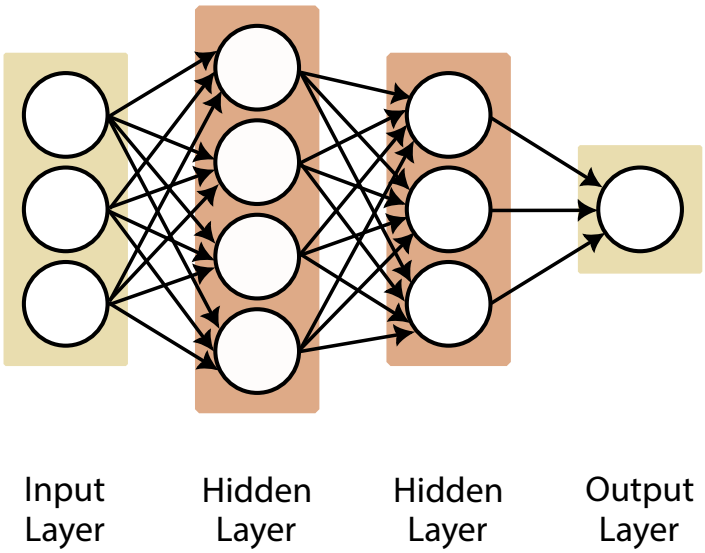

Figure 2.1: A visual depiction of a neural network. The circles represent nodes and the arrows represent connections between nodes. This example network contains an input layer, an output layer and two hidden layers. Each layer is fully connected to the subsequent layer.

There are many types of neural networks with varying architectures and different applications, but for the purpose of this discussion, we will focus on a simple feedforward neural network, as depicted in Fig. 2.1. Feedforward refers to the fact that the information only flows in one direction, rather than forwards and backwards between layers. To train such a neural network using a standard algorithm called backpropagation (Rumelhart, Hinton, & Williams, 1986), pre-classified training data are required, and at a minimum, the network needs an input layer and an output layer. The nodes of the input layer represent the variables from the training data, and the nodes of the output layer represent the classifications of the data. Between these layers can be any number of hidden layers, nodes, and connections, and different architectures are implemented for different classification tasks. This type of network connects each subsequent layer, so for each layer, the output from the current layer is input to the subsequent layer. Furthermore, the connections between nodes from the current layer to the subsequent layer are weighted. For each node n in the subsequent layer, a function of the weights and nodes from the current layer connected to n

determines the value of n.

A neural network often aims to solve the optimisation problem of minimising a selected loss function, where the loss function measures how well the algorithm performs. If the errors between the values of the observed dependent variable and the values estimated by the model are large, then the loss function would also output a large value. Therefore, a neural network often aims to reduce these errors based on the output from the loss function. In our example, the classification from the neural network is compared to the observed classification of the data at the final layer of the neural network, and the learning process known as backpropagation is applied to improve the classification. Specifically, the neural network works backwards to automatically adjust each weight and then tries to classify the training data once more. It continues this process until the loss function is minimised, and the final output from the network represents the final classifications. Once the neural network is trained and the weights are fixed, it can be deployed on unseen and unclassified datasets to obtain estimates of classifications of the data.

Configuring the architecture of a neural network in a specific style can considerably improve the performance on different tasks. As such, there are several types of neural networks (Alzubaidi et al., 2021; Liu et al., 2017; Sharkawy, 2020; Wilamowski, 2009). One type is the convolutional neural network (CNN), which has led to substantial improvements in image processing. Prior to CNNs, image processing would involve using algorithms to detect distinct shapes and patterns, or using image matching techniques that compare the pixels of a target image to the pixels of images whose contents are already known (Bhatt et al., 2021; Dufaux, 2021; Ma, Jiang, Fan, Jiang, & Yan, 2020). However, through the advent of CNNs, it is now possible to automatically identify multiple complex shapes and patterns in images that were more difficult to recognise before. In fact, CNNs are often used for pattern detection in image recognition tasks (Goodfellow, Bengio, & Courville, 2016; LeCun, 1989; LeCun et al., 1989).

The architecture of a CNN typically includes a convolution layer followed by a pooling layer. Non-linear transformations can also be incorporated between these layers to introduce non-linearity and noise to what would otherwise be a linear model. The transformations are known as activation functions, and they enable the CNN to learn and perform more complex tasks, as well as make the CNN more robust. Example activation functions that are often used in CNNs include softmax, which maps values between 0 and 1, and ReLU (Rectified Linear Unit), which only outputs positive values and sets all negative values to 0.

When building a CNN, the convolution and pooling layers can be incorporated

any number of times, but what do these layers do? A simple visual example of a CNN is shown in Fig. 2.2. Consider a coloured image represented by a matrix of pixels on a computer. Each pixel is a node in the input layer of the neural network. In this example network, the image is then 'convolved', or combined, in the convolutional layer with filters, which are collections of weights that look for features in the image. Each filter corresponds to a specific feature, so multiple filters are used if multiple patterns are to be identified.

The pooling layer then summarises the features extracted and reduces the size of the input for the next layer, effectively reducing the number of nodes from the convolution layer to the pooling layer. This is highly beneficial, as it both avoids the CNN overfitting and reduces the time of computation. Once all of the convolution and pooling layers have been computed, the penultimate layer of the example CNN is a fully connected layer, so each node in the fully connected layer receives input from each node in the previous layer. In our example, this layer flattens all of the pooled inputs into one feature vector instead of a matrix, and this feature vector then provides a classification in the output layer. Note that deep learning usually refers to neural networks that use multiple layers to extract multiple features from input data. Therefore, since a CNN consists of multiple layers that do just that, it is a deep learning technique.

One of the first key studies highlighting the success of using a CNN for image recognition tasks was conducted by LeCun, Bottou, Bengio, and Haffner (1998), who deploy a CNN and train it to classify handwritten numbers. Since this study, larger CNNs with more complex architectures have been constructed. These have led to pre-trained CNNs that are designed for distinct image recognition tasks, saving many researchers the time and data collection required to train an entirely new CNN (Carneiro, Nascimento, & Bradley, 2015; Redmon & Farhadi, 2016).

For example, Russakovsky et al. (2015) create a dataset consisting of over 15 million images of objects belonging to roughly 22,000 categories, and make a subset of the dataset publicly available in preparation for object recognition tasks. Specifically, the researchers introduce the ImageNet large-scale Visual Recognition Challenge (ILSVRC) and task other researchers with developing computer vision algorithms that could detect objects in images. They provide approximately 1,000 images for 1,000 different object categories. Some participants of the challenge achieved considerable success, including Krizhevsky, Sutskever, and Hinton (2012), who developed a very influential network architecture for image classification, known as AlexNet. Once it has been pre-trained on ImageNet data, AlexNet is able to detect objects, such as 'mushroom' and 'leopard', in unseen images.

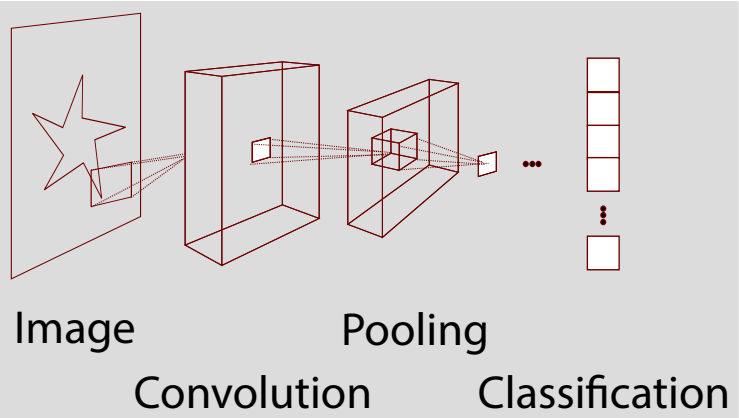

Figure 2.2: A visual depiction of a simple convolutional neural network. An image is fed to the convolutional neural network, where any number of convolution and pooling layers extract information from it. The final output represents the classification.

In addition to object recognition, scene recognition is also possible using a CNN, as shown by Zhou, Lapedriza, Xiao, Torralba, and Oliva (2014). To train their Places CNN for scene recognition tasks, they first create a dataset of 7 million images of scenes, called the Places database. These images are then labelled with semantic categories related to the kind of scene they depict, such as 'tower', 'highway' and 'forest'. Finally, Places CNN is trained on 2.5 million images from this dataset, and is capable of identifying 205 different kinds of scenes. Note that the architecture of Places CNN is almost identical to that proposed by Krizhevsky et al. (2012). The Places CNN network as pre-trained by Zhou et al. (2014) has been successfully exploited for various scene-related tasks (Hou, Zhang, & Zhou, 2015; Sachin, Sowmya, Govind, & Soman, 2017; Seresinhe et al., 2017).

Moreover, the Places database has since been updated and now consists of 10 million images of scenes, known as the Places2 database (Zhou, Lapedriza, Khosla, Oliva, & Torralba, 2018). Zhou et al. (2018) train a new CNN on 1.8 million images from this dataset and can successfully identify 365 different kinds of scenes. In general, CNNs have widespread usage for computational tasks, from translating sign language to identifying the style of an image (Ameen & Vadera, 2017; Feichtenhofer, Pinz, & Zisserman, 2016; Jaderberg, Simonyan, Vedaldi, & Zisserman, 2015; Nguyen et al., 2017; Parkhi, Vedaldi, & Zisserman, 2015; Simonyan & Zisserman, 2014; Sun,

Wang, Yang, & Hu, 2017), as well as classifying YouTube videos (Karpathy et al., 2014).

In summary, CNNs are capable of extracting vast quantities of features from images. For example, 205 features can be extracted by Places CNN, and 365 features can be extracted by the updated version of the CNN. Could these features be used to understand other measures from relevant images, for example, house prices from photos of houses? A problem that could arise when investigating this question is multicollinearity, which is when independent variables in a regression model are correlated. Modelling a relationship using a huge volume of features could lead to this, especially in linear models, resulting in less reliable statistical inferences. In the next section, we examine the problem further and introduce a possible solution: the use of an elastic net (Seresinhe et al., 2017).

2.2.2 Elastic net models

An elastic net model is a regularised linear regression model that is particularly useful when a large volume of independent variables are involved in a model, as it is able to mitigate the impact of multicollinearity and overfitting. To understand how it works, we first look at the ordinary least squares (OLS) regression, which the elastic net model is based on.

OLS regression is used to measure the statistical relationship between a dependent variable and one or more independent explanatory variables. The method involves estimating the parameters of a linear model subject to a loss function, and is useful because it provides an unbiased mean estimation with a low variance. Recall that the loss function measures how well the algorithm performs, so the model aims to minimise the loss function, as a smaller output from the loss function indicates smaller model errors and greater model accuracy. The fitted OLS regression model takes the form

$$\hat{y}_i = \hat{\beta}_0 + \sum_{j=1}^{N} \hat{\beta}_j x_{i,j},\ i = 1, \cdots, n \quad \text{or} \quad \hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$$

where

- n = the sample size and $i = 1, \cdots, n$ represents a sample
- $\hat{y}_i$ = the estimated values of the dependent variable for sample i
- $\hat{\mathbf{y}} = \begin{bmatrix} \hat{y}_1 & \hat{y}_2 & \cdots & \hat{y}_n \end{bmatrix}^T$ = the vector of estimated values of the dependent variable

- N = the number of independent variables and $j = 1, \cdots, N$ represents a variable
- $\hat{\beta}_0$ = the estimated value of the intercept of the linear model and $\hat{\beta}_j$ represents the estimated coefficient for variable j
- $\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_0 & \hat{\beta}_1 & \hat{\beta}_2 & \cdots & \hat{\beta}_N \end{bmatrix}^T$ = the vector of estimated values of the independent variable coefficients
- $x_{i,j}$ = the observed values of the independent variable for sample i and variable j
- $\mathbf{X} = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} & \cdots & x_{1,j} \\ 1 & x_{2,1} & x_{2,2} & \cdots & x_{2,j} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{i,1} & x_{i,2} & \cdots & x_{i,j} \end{bmatrix}$ = the matrix of observed values of all independent variables

and minimises

$$L_{OLS}(\hat{\boldsymbol{\beta}}) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = ||\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}||_2^2$$

where

- y_i = the observed value of the dependent variable for sample i
- $\mathbf{y} = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix}^T$ = the vector of observed values of the dependent variable

According to Stigler (1981), the first published mention of OLS regression was by Adrien Marie Legendre in 1805, although the pioneer is unknown. To use OLS effectively, its associated assumptions need to be met. As outlined by Fox (2015), Poole and O'Farrell (1971), and Frost (2018), the assumptions are:

1. The regression model is linear in the coefficients and the error term, so the relationship between the independent and dependent variables is linear
2. The error term has a population mean of zero, so is random
3. All independent variables are uncorrelated with the error term
4. Observations of the error term are uncorrelated with each other
5. The error term has constant variance

6. No independent variables are perfect linear functions of other variables

7. The error is normally distributed

Datasets may not necessarily meet all of the requirements of OLS, and this would impact the accuracy of the model. In terms of our research problem, using the large volume of features extracted from an image by a CNN, such as Places CNN, in a model based on OLS could lead to multicollinearity in the data. As a consequence, the OLS residuals would tend to be larger, since the predictions vary greatly from the observed value, and assumption 6 is violated (no independent variables are perfect linear functions of other variables). Hence, there are extensions and variations of OLS to mitigate these consequences (Hastie, Tibshirani, & Friedman, 2009; James, Witten, Hastie, & Tibshirani, 2013).

In regards to multicollinearity, it would be useful to shrink the resulting errors, even if the estimator becomes slightly biased. In 1970, Hoerl and Kennard (1970) introduced the ridge regression to achieve this reduction in errors. They extend the OLS regression model by augmenting the loss function with an L_2 penalty, also known as L2 regularisation. In other words, they reduce the impact of larger model coefficients and bias the model towards smaller coefficients by squaring the magnitude of the coefficients in the loss function. Therefore, the loss function becomes

$$L_{ridge}(\hat{\boldsymbol{\beta}}) = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{N} \hat{\beta}_j^2 = ||\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}||_2^2 + \lambda ||\hat{\boldsymbol{\beta}}||_2^2$$

where $\lambda \geq 0$ is a tuning parameter that represents the regularisation penalty. As λ increases, the OLS coefficient estimates decrease and tend towards zero.

Another potential issue with OLS is overfitting. Adding explanatory variables to a regression increases the number of parameters estimated, and the model tends to improve simply because there is more information to learn from. This leads to a model that works well in certain circumstances only. The ridge regression helps negate this effect, as the coefficients are shrunk, but they are all shrunk by the same factor. Furthermore, there is no variable selection method, as no variables are set to zero. A more sparse model would improve interpretability of the model, as the most relevant and impactful variables give the strongest indication of the relationship between the independent and dependent variables.

This issue is tackled by Tibshirani (1996), who developed the lasso regression in. Tibshirani (1996) builds upon the work of Breiman (1995), who proposes a new method for subset regression called the non-negative garrote. The non-negative garrote method focuses on shrinking coefficients and setting them to 0, rather than a small value like the ridge regression. However, compared to the stable ridge

regression, the non-negative garrote is unstable. Stability refers to how slight changes in the model can affect model coefficients and their significance, so a stable model is relatively unaffected by changes in the model, whereas an unstable model is highly sensitive to changes in the model. The more unstable a method is, the harder it is to accurately obtain parameter estimates. Therefore, Tibshirani (1996) developed the lasso regression, which is similar to the non-negative garrote method but more stable like the ridge regression.

Tibshirani (1996) achieved this balance by extending the OLS methodology and augmenting the loss function with an L_1 penalty, also known as L1 regularisation. This regularisation adds the absolute value of the magnitude of the coefficients to the loss function. I.e. the loss function becomes

$$L_{lasso}(\hat{\boldsymbol{\beta}}) = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{N} |\hat{\beta_j}| = ||\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}||_2^2 + \lambda||\hat{\boldsymbol{\beta}}||_1$$

Again, $\lambda \geq 0$ is a tuning parameter that represents the regularisation penalty, but it can actually set unimportant coefficients to zero, rather than simply making them smaller. This makes the lasso regression appropriate for variable selection tasks.

However, the lasso regression has also been critiqued by academics. The variable selection is fairly dependent on the data provided, so can still be quite unstable (Nogueira, Sechidis, & Brown, 2018). Combining the qualities of both the ridge and lasso regressions could lead to stable models that conduct variable selection but also negate the effects of multicollinearity and overfitting. Hence, Zou and Hastie (2005) developed the elastic net in 2005. It is also an extension of the OLS regression and combines the regularisation penalties of the lasso and ridge techniques. The loss function is then defined as

$$L_{enet}(\hat{\boldsymbol{\beta}}) = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{2n} + \lambda(\frac{1-\alpha}{2}\sum_{j=1}^{N}\hat{\beta}_j^2 + \alpha\sum_{j=1}^{N}|\hat{\beta_j}|) =$$
$$||\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}||_2^2 + \lambda(\frac{1-\alpha}{2}||\hat{\boldsymbol{\beta}}||_2^2 + \alpha||\hat{\boldsymbol{\beta}}||_1)$$

where $0 \leq \alpha \leq 1$ represents the mixing parameter between a ridge and lasso regression. If $\alpha = 0$, then the model is a purely ridge regression and if $\alpha = 1$, then the model is a purely lasso regression. This machine learning technique can be especially useful when working with considerable quantities of features extracted from images because the impact of multicollinearity and overfitting can be mitigated, and a balance between sparsity and important information can be reached.

Widespread usage of the elastic net is facilitated by machine learning libraries, which are well maintained for multiple programming languages and simplify the process of using the model. For example, the 'glmnet' library is commonly used for

elastic net regression in the programming language R (Friedman, Hastie, & Tibshirani, 2010; Friedman et al., 2022), and the 'scikit-learn' library is popular for elastic net regression in Python (Pedregosa et al., 2011; scikit-learn, n.d.). Unfortunately, the notation and terminology between libraries is inconsistent. For example, in 'glmnet', the parameters α and λ are defined in the same way as in Zou and Hastie (2005), so α is the mixing parameter between the L_1 and L_2 penalties, and λ is the regularisation penalty. Yet, in 'scikit-learn', *l1_ratio* is the mixing parameter between the L_1 and L_2 penalties, and α is the regularisation penalty. This makes it slightly confusing to switch between programming languages. For our research, we use the Python library 'scikit-learn'.

2.2.3 Summary

In this section, we focus on recent advances in machine learning that enable us to extract information from images. We look at how convolutional neural networks operate and are able to extract high volumes of features from images. Since modelling with such a huge quantity of features could lead to multicollinearity, we further discuss machine learning models that are capable of dealing with large datasets with considerable multicollinearity.

The aim of our research is to extract information from large volumes of street view images to infer house prices. The machine learning techniques discussed in this section therefore provide a foundation for the analyses in the rest of the thesis.

2.3 Outlook

Computational social science has progressed rapidly over the past few decades, predominantly due to the abundance of accessible large datasets on human behaviour. Numerous studies have furthered our understanding of human behaviour by exploring a variety of different data sources, from Twitter data to data from mobile phone networks, as well as online imagery and house price data. At the same time, recent advances in machine learning have made it easier to extract information from images and build models with large numbers of potentially correlated predictors.

However, studies of house prices rarely consider any visual information, even though the appearance of a property and its surroundings often provide an indication of its price. Furthermore, many sources of imagery of streets and houses are updated frequently (Biljecki & Ito, 2021; Columbro, Eudave, Ferreira, Lourenço, & Fabbrocino, 2022), including the crowdsourced services Mapillary (Mapillary, n.d.; Neuhold, Ollmann, Rota Bulo, & Kontschieder, 2017) and KartaView (KartaView, n.d.).

Anyone can contribute images to both Mapillary and KartaView, so images on these platforms cover locations from across the world and are theoretically updated as soon as someone uploads an image.

As such, there exists the potential to mitigate the impact of delays on official house price data at a fraction of the cost. If house prices can be estimated accurately using imagery that is updated more frequently than official house price data, then members of society could have access to this information at a faster rate than previously possible. This would allow them to make better-informed decisions that are based on house prices. For example, if an individual was looking to buy a house in the UK, they could have a better understanding of how much it would cost to buy a property in a particular location at a particular time, as the information would be available sooner.

Furthermore, if the methodology we propose is able to successfully lead to new insights about house prices, this could allow researchers of house prices to develop more accurate house price estimation models that incorporate both visual and non-visual characteristics related to house prices. The methodology could also be of interest to social scientists who may be looking for novel techniques to use to extract new information about their topic of research, as our propose d method can be implemented with different data.

Therefore, in Chapter 3, we follow the principles of social data science and explore the opportunity to combine street view imagery with deep learning techniques to understand and estimate house prices in the UK. More specifically, we use the Places CNN for scene recognition to extract visual information from Google Street View images of London and use this visual information to estimate house prices in London using an elastic net.

Chapter 3

Inferring house prices from Google Street View images

3.1 Introduction

The most famous source of street-level imagery is Google Street View, one of the first projects to start collecting such images. Google Street View panoramas mostly consist of images of building facades and outdoor street environments from a human perspective. The project continues to grow over time, with new imagery of existing locations being captured, and new locations being added. At the same time, as discussed in detail in Chapter 2, advances in deep learning have greatly increased our ability to determine the contents of an image automatically. Humans are often able to make inferences about house prices in a neighbourhood from the appearance of the streets and houses. This chapter therefore asks the following two questions:

1. Can we automatically infer house prices from current Google Street View images?

2. Can we automatically identify areas which have recently seen relatively high increases in house prices from Google Street View images?

We build on recent analyses suggesting that median income in small urban areas can be inferred from Google Street View data using deep learning techniques (Gutiérrez-Roig, Preis, Seresinhe, Letchford, & Moat, 2017). For this case study, we focus on house prices in London, the capital city of the United Kingdom.

3.2 Data and methods

3.2.1 Google Street View images

Figure 3.1 depicts our methodology. The initial step is to acquire large scale data on visual features of areas across London. We accomplished this by accessing images from Google Street View from 2016 onwards. Google mainly collects these images by driving around areas in a Google car (Google, n.d.-b). However, in locations where a car cannot pass, they use other means of transport such as sailing or walking. The photos themselves are captured with special cameras that can take photos from multiple directions at once. For each individual point where photos are taken, the photos are stitched together to create a complete 360-degree picture of the area.

We build a corpus of visual features from Google Street View images of London in the same manner as that for a previous analysis by Gutiérrez-Roig et al. (2017). To access these images, we define a grid of squares over London of 100 metres by 100 metres and at the corner of each square, we take the latitude and longitude co-ordinates. Using the 'streetview' library for Python (Letchford, 2018), we identify Google Street View panorama IDs for images near each latitude and longitude co-ordinate pair and the date that the image was taken if it was available. Using these data, we access four images at each latitude and longitude co-ordinate pair, looking north, east, south, and west, via the Google Street View API. In early 2017, exploratory analyses of the availability of Google Street View imagery in London revealed that the best coverage with recent imagery could be achieved by accessing images captured in 2014 and 2015. In total, we access 518,808 images from 2014 and 2015.

To allow us to combine house sale price data with data from Google Street View, we carry out our analysis at a small area level, where we associate multiple house sales and multiple Google Street View images with each small area. To give us a good number of house sale and images in each small area, we use middle layer super output areas (MSOAs), of which there are 983 in London. MSOAs are geographic areas that either contain a population of between 5,000 and 15,000 inhabitants, or between 2,000 and 6,000 households (Stokes, 2012). Note that for each MSOA in London, the number of images we access is proportional to the area of that MSOA. As such, where MSOAs are geographically larger and there is more opportunity for the appearance of the neighbourhood to vary across its full area, we access more images to capture this additional variation. The number of images per MSOA ranges from 88 to 2,164, with a median of 496.

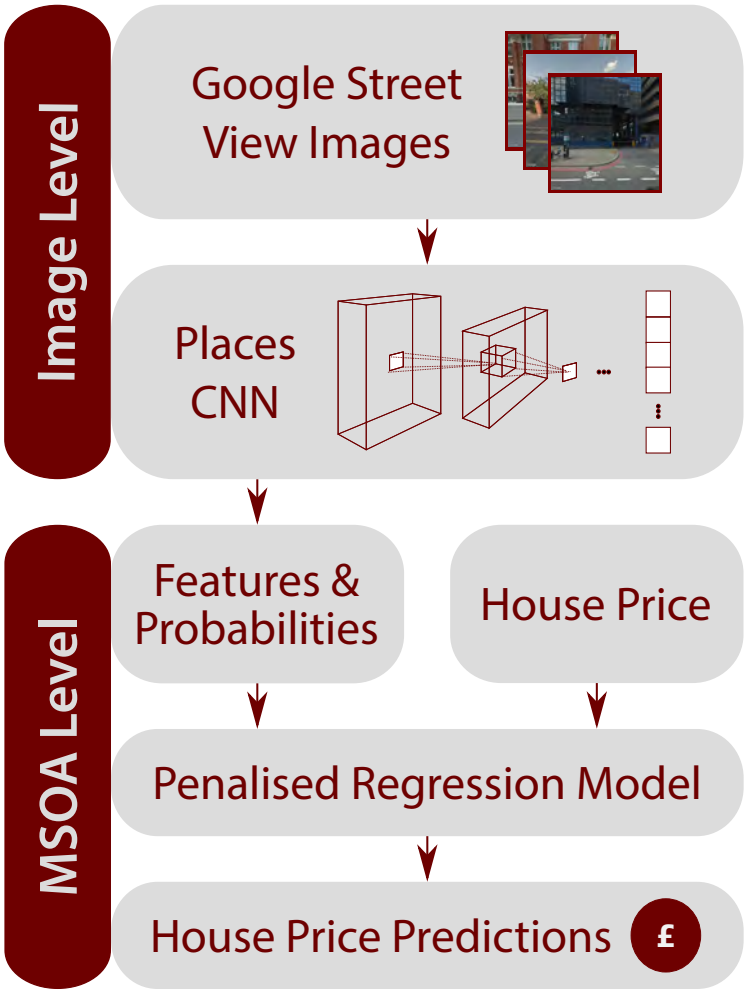

Figure 3.1: Deep learning framework. We sample Google Street View images at a 100-metre resolution throughout London and assign each image to its middle layer super output area (MSOA). MSOAs are geographic areas designed to aid the reporting and analysis of statistics, with 983 MSOAs in London. We feed each image individually into the Places convolutional neural network (CNN) developed by Zhou et al. (2014), which we will refer to as Places205. For each individual image, we extract 205 Places features and 102 SUN attributes, and store the feature values assigned by the CNN in a vector of length 307. The feature values relate to the likelihood of a given feature appearing in a given image. Example features include 'highway' and 'tower'. Next, we calculate the mean feature vector for each MSOA and use it to train a penalised regression model known as an elastic net. *(continues on the following page)*

Figure 3.1: *(continues from previous page)* The elastic net attempts to learn the relationship between the feature vector for an MSOA and the median house price for that area. The predictions of the model are tested using a 'leave-one-out' approach. In other words, for each MSOA m, the elastic net is trained on data from all other MSOAs. The feature vector for MSOA m is then passed to the elastic net, which generates a prediction of the house price for MSOA m. Finally, we compare the resulting predictions to the actual median house price for each MSOA. Street View images © 2017 Google.

3.2.2 Places CNN

To extract features from Google Street View images, we run a computer vision algorithm on each of our collected images (Section 2.2.1). Specifically, we use the Places205-AlexNet convolutional neural network (CNN) developed by Zhou et al. (2014) at MIT, which is trained on 2.5 million images from the Places database of 7 million images. We use Places205 to extract the values of 205 Places scene categories and 102 scene understanding (SUN) attributes (Patterson, Xu, Su, & Hays, 2014) from each individual image. The features themselves represent scenic elements, such as 'yard' and 'skyscraper', and environmental attributes, such as 'soothing' and 'concrete'. Since Places205 is trained to classify scenes, each feature value can be interpreted as relating to the likelihood of a given feature appearing in a given image.

We first extract the Places scene categories from Places205. To do so, we simply use the output vector from the CNN which contains the 205 feature values associated with the scenic features. The feature values are normalised using the softmax activation function and as a result, all feature values range from 0 to 1 and the vector sums to 1.

To extract the SUN attributes from Places205, we then extract the vector produced at the penultimate layer of the CNN and transform the vector into attributes. The SUN attributes are then normalised, giving us a vector of 102 feature values associated with environmental features that sums to 1. Thus, for each image in our dataset, we have a feature vector of 307 feature values combining 205 Places scene categories and 102 SUN attributes, and the feature vector subsequently sums to 2.

Since Google Street View images are of the outdoors, it does not make sense to include indoor features in the model, such as 'kitchen' or 'bedroom'. For this reason,

we remove 69 indoor features from the model. This leaves 136 outdoor features attached to each image, plus the 102 SUN attributes. As there are multiple images associated with each MSOA, we calculate the feature vector for each MSOA as the mean value of the feature vectors for all images in that MSOA. In other words, each feature value in the feature vector of each MSOA represents the mean feature value across all images in that MSOA. These mean feature vectors then act as input to our house price estimation model.

To demonstrate how Places205 operates, we choose a selection of features and top-scoring Google Street View images for those features, and display them in Fig. 3.2. Specifically, for three features, we show three images that generate high feature values, which are provided underneath each image. As we can see, the CNN is able to recognise elements in images well, and assigns labels to images appropriately. For example, multiple variations of a bridge are assigned a high feature value for the label 'bridge' (Fig. 3.2a). This provides sufficient evidence that we are able to automatically extract patterns in and gain an understanding of the visual environments of areas at MSOA granularity.

Additionally, we can explore the features further and ask whether they are distributed across London as we would expect from intuition. To do so, we inspect the spatial distributions of two example features in Fig. 3.3. Specifically, we inspect the mean feature value distribution maps for the features 'highway' (Fig. 3.3a) and 'tower' (Fig. 3.3b) at MSOA granularity across London. This helps us understand whether there is any value in the the visual characteristic measurements provided by the CNN when processing Google Street View imagery to infer the distribution of visual characteristics across London.

We would expect to find more highways in the outskirts of London, as the centre of London is more pedestrianised and there is substantial reliance on public transport, including the London Underground. Moreover, the centre is much older and as such the roads are generally narrower. Visual inspection of the map we produce confirms that mean feature values for 'highway' are much higher for areas in the suburbs. Conversely, we would expect to find more towers in the centre of London, as there are more high rise buildings in the centre than in the suburbs. Indeed, the maps reflect that mean feature values for 'tower' are much higher in the centre. In both cases, these spatial distributions of feature values produced by applying Places CNN to Google Street View imagery accord with intuition. This provides further reason to believe that these data can help us understand the distribution of different visual features across London.

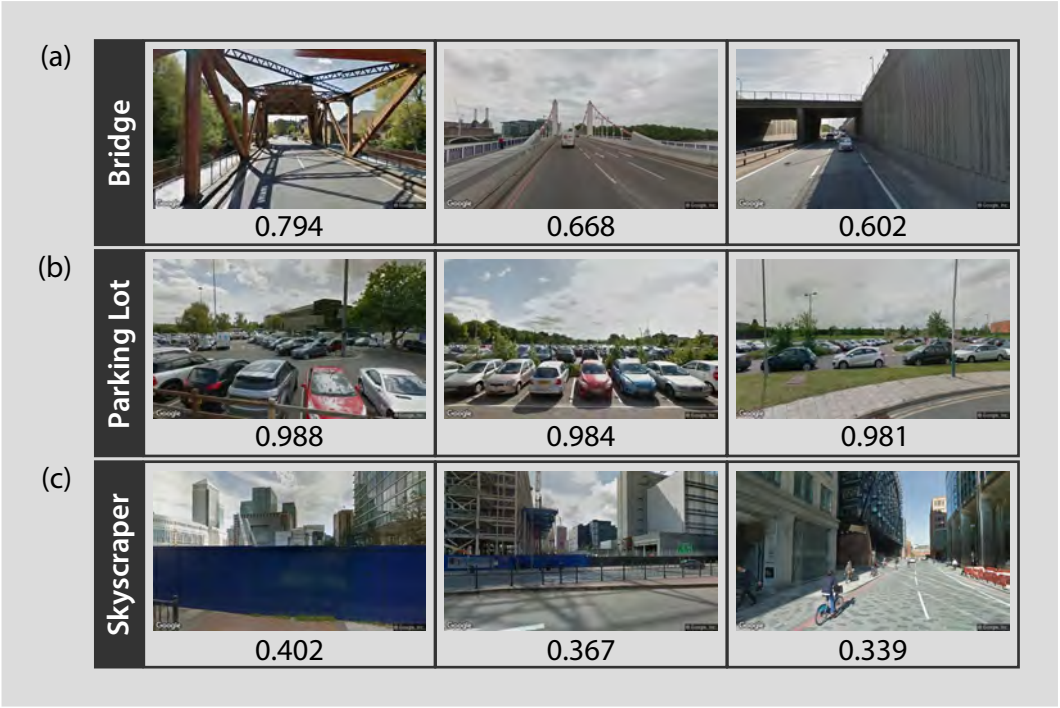

Figure 3.2: Labelling the scene shown in Google Street View images using Places205. To demonstrate that Places205 generally performs well at identifying the scene shown in Google Street View images, we first choose three features. For each feature, we display three images from the top ten images with the highest feature values, as they reflect what the CNN sees when assigning labels and feature values to the highest degree. The feature values can be interpreted as relating to the likelihood of a given feature appearing in a given image, and are assigned using the methodology described in Section 3.2.2. For each image, the associated feature is its highest-scoring Places label. **(a)** Top images for the feature 'bridge'. The CNN correctly identifies bridges with varying architectures. **(b)** Top images for the feature 'parking lot'. The CNN correctly identifies car parks filled with cars. **(c)** Top images for the feature 'skyscraper'. The CNN can identify skyscrapers correctly, regardless of whether there are structures blocking the view. Generally, the CNN identifies the scene shown in Google Street View images correctly. Street View images © 2017 Google.

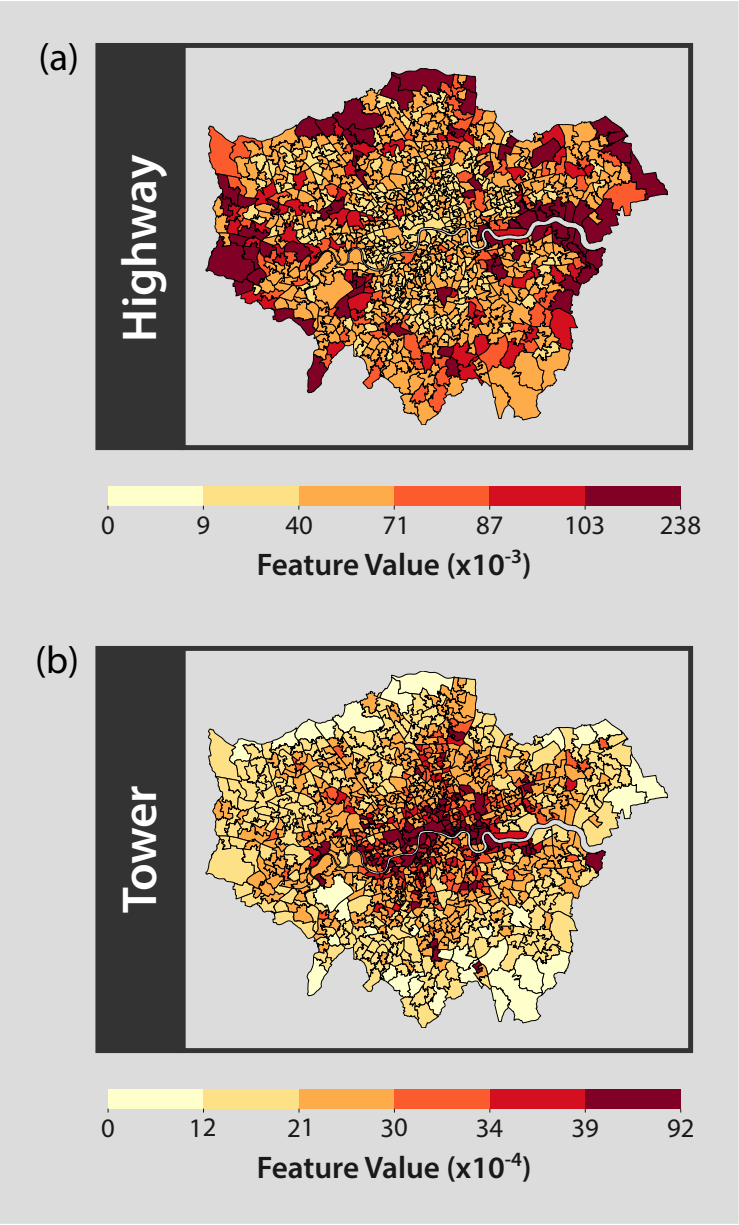

Figure 3.3: Prevalence of features in London from Google Street View images. Do the spatial distributions of the feature values of visual features identified by the CNN follow intuition? We examine the maps of two example features, where we analyse Google Street View images of London from our dataset. For a given feature, the feature values for each MSOA can be interpreted as the prevalence of the feature in that area, and the mean feature value for each MSOA is generated using the methodology described in Section 3.2.2. *(continues on the following page)*

Figure 3.3: *(continues from previous page)* For both maps, a darker red indicates a higher feature value. **(a)** Map of mean feature values for the feature 'highway'. The CNN identifies more highways in the outskirts of London than in the centre of London. This can be reasonably expected, as the centre of London is heavily pedestrianised relative to the outskirts of London. As such, public transport is frequently relied upon for travelling around the centre. **(b)** Map of mean feature values for the feature 'tower'. The CNN identifies more towers in the centre of London than the remainder of London. Since the centre of London contains more high rise buildings than the suburbs of London, we would intuitively expect to see more tall towers around the centre of London too. Hence, the CNN operates intuitively. Overall, we can confirm that the spatial distributions of features broadly follow intuition. For all maps, standard deviation breaks are implemented with six bins. Note that the maps therefore use different scales. Contains National Statistics data © Crown copyright and database right 2011. Contains Ordnance Survey data © Crown copyright and database right 2011. Contains Street View data © 2017 Google.

3.2.3 House price data

We analyse sale price data from house transactions completed during 2015 and 2016, the two most recent complete years of data at the time of analysis. Data regarding house transactions are freely available from the HM Land Registry (2016), where the full postcode is given for each transaction. For each postcode, we determine which MSOA the postcode falls in using data accessed from the London Datastore (Greater London Authority, n.d.). Of the 250,171 transactions that took place during 2015 and 2016, the postcodes of 691 transactions could not be matched with an MSOA, as the postcodes were not present in the London Datastore data at the time of analysis.

The house price data themselves are skewed with a long positive tail, representing a small number of properties with very high house prices. For this reason, we work with median house prices per MSOA rather than mean prices, so that the small number of expensive properties in an area does not disproportionately affect the average measure of prices in an area.

3.2.4 Elastic net model

For each MSOA, we calculate a mean visual feature vector of 307 features (Section 3.2.2). However, including too many features in a simple linear regression could

lead to problems such as overfitting and multicollinearity. In turn, overfitting and multicollinearity result in poor accuracy of model predictions.

One approach to dealing with these problems is to use a penalised regression model. These types of models build on a simple linear regression by including a loss function, known as the penalty. The coefficients of the parameters are then constrained and penalised to prevent problems such as overfitting from occurring.

Due to the sheer volume of features we extract from our images, we require a penalised regression model with a robust variable selection method. Therefore, we use an elastic net model, following a methodology introduced by Seresinhe et al. (2017). As discussed in Section 2.2.2, the benefits of using an elastic net model include its ability to automatically pick the most important features and set the remaining features to zero.

With the features as predictors and the house prices as the outcome variable, we build an elastic net model. We use the 'scikit-learn' library in Python and employ the elastic net with various values of *l1_ratio*, where $0 \leqslant$ *l1_ratio* $\leqslant 1$ is the regularisation penalty such that *l1_ratio* $= 0$ is a lasso regression and *l1_ratio* $= 1$ is a ridge regression. We start by trying *l1_ratio* values from $0.1, 0.2, \cdots, 0.9$ and narrow down to the *l1_ratio* values $0.9, 0.99, 0.999, 0.9999, 0.99999, 0.999999, 0.9999999, 0.99999999$. Each and every small difference in the *l1_ratio* impacts the model performance, so we report it fully. We then obtain the *l1_ratio* and corresponding α that generates the best model. Recall that the α refers to the shrinkage parameter, which allows the model to determine which features are important.

We check our optimal parameters via cross-validation, which is a technique that determines how well a model generalises by analysing the errors. This approach is useful when trying to avoid overfitting, as it involves splitting a dataset into a 'training' set that the model is trained on and a 'testing' set which is validated. For the model, we use 3-fold cross-validation, to find the model with the highest R^2, and hence best fit. More precisely, 3-fold cross-validation involves splitting the dataset into three sections and using two of the sections as the training set and one section as the testing set.

After we find our optimal values of *l1_ratio* and α, we implement a separate leave-one-out cross-validation on the entire dataset to generate predictions of house prices. Specifically, we leave out one MSOA, m, from the dataset as our 'testing' data and train a new elastic net for estimating house prices, with our optimal values of *l1_ratio* and α, on our 'training' data consisting of the remaining 982 MSOAs.

During training, the features of the remaining 982 MSOAs, as obtained from Google Street View images and the CNN, are the predictors. Their corresponding

median house prices are the outcome variable. Following training, we test the model and use the visual features of the excluded MSOA, m, in order to obtain a prediction of the median house price in MSOA m.

3.2.5 Measuring change in house prices

We also investigate whether it is possible to automatically identify areas in which house prices have recently risen or fallen from Google Street View imagery. A simple approach to modelling changes in house prices is to look at absolute changes in price for each MSOA in London. However, we find that the areas with the largest absolute increases usually have the highest original prices, reflecting the phenomenon that the rich get richer. As a result, identifying areas with the highest absolute increases in house price may not be a greatly different task to identifying the areas with the highest median house prices.

An alternative approach to modelling changes in house prices would be to look at how the areas of London rank against each other in terms of house prices, and then examine the change in rank over time. Ranking areas by house price brings the focus more on the change between areas, irrespective of the exact differences in house prices. A weaker ordering of areas, where we ignore the exact values of the house prices but instead base the ordering on their relative rankings alone, allows us to focus on the qualitative differences between house prices but in a quantitative manner. This forms the basis of our second analysis, where we aim to develop a model that can infer changes in house price rank.

We consider house price changes over the past ten years, from 2005 and 2006 to 2015 and 2016. As in Section 3.2.3, we obtain house price data from the HM Land Registry for the years 2005 and 2006. When matching MSOAs from the London Datastore to postcodes for that time period, only 240 of the 312,344 transactions could not be matched, as the postcodes for those transactions were not present in the London Datastore data at the time of analysis.

To rank, we set the highest rank of 983 to the MSOA with the highest house price, and the lowest rank of 1 to the MSOA with the lowest house price. If two areas have the same median house price, we give them both the mean rank. The change in rank is then given by the rank during 2015 and 2016 minus the rank during 2005 and 2006, as this is the simplest to interpret. For example, let us calculate the change in rank for an MSOA in Barking and Dagenham (E02000012) from our dataset. During 2005 and 2006, this MSOA has a ranking of 532, and during 2015 and 2016, it has a ranking of 426.5. This rank is not a whole number because there are 10 MSOAs with the same median house price, including E020000012, and they would normally have

ranked from 422 to 431. However, since we take the mean of the ranks and assign the same mean rank to all MSOAs with the same median house price, they are all ranked 426.5. Given that the 2015 to 2016 rank is 426.5 and the 2005 to 2006 rank is 532, the change in rank is calculated as 426.5 - 532 = -105.5. A change in rank of -105.5 indicates a decrease in ranked price over the decade, hence a relative reduction in median house price.

Our analysis aims to determine whether we can identify the areas that had the biggest changes in house price rank from Google Street View imagery from 2014 and 2015. We therefore employ an elastic net model using the same methodology described in Section 3.2.4 with the same features as input to predict the change in house price rank instead of the median house price.

3.3 Results

3.3.1 Estimating local house prices in 2015 and 2016 from Google Street View images

How well does our house price estimation model perform? We first examine the house price data from 2015 and 2016, which reveal that median house prices for MSOAs across London range from £69,460 to £1,270,000. Figure 3.4a shows the observed median house prices, where a darker area represents a higher house price. The observed median house price refers to the true median house price for each MSOA. A number of patterns in house prices across London can be observed from this map. For example, houses in the centre of London are consistently priced at a high level, whereas parts of the east of London are priced at relatively low levels (Fig. 3.4a).

Undoubtedly, there are differences in visual characteristics of different areas. These visual fingerprints might reflect socioeconomic characteristics of the area, like house prices. Therefore, our goal is to investigate whether it is possible to infer house prices across London from Google Street View data. Employing the methodology summarised in Fig.3.1 results in positive findings. In the elastic net, we find the optimal value of *l1_ratio* to be 0.999999 with an α of 4.73 (to 2 decimal places). We also conduct a regression analysis comparing the observed median house price for each MSOA with the estimates produced by the elastic net using a leave-one-out analysis. This analysis indicates that the model performs reasonably well ($R^2 = 0.56$, $N = 983$, $p < 0.001$).

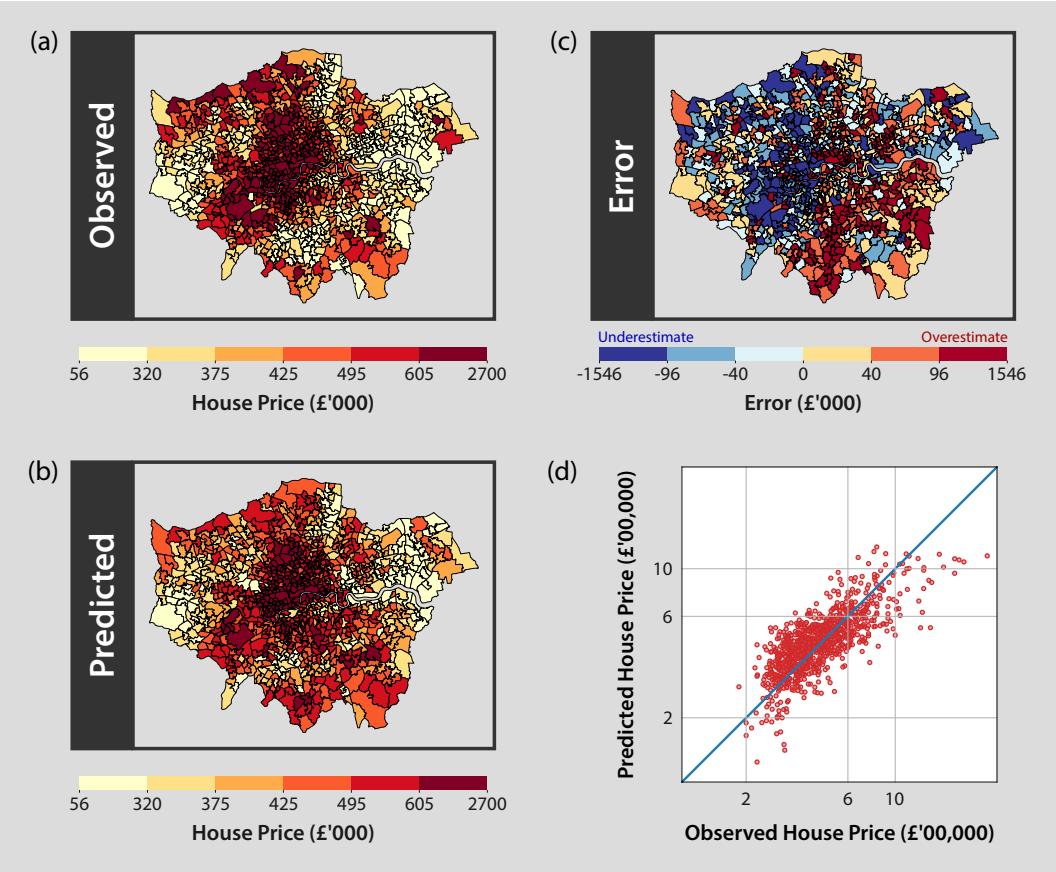

Figure 3.4: Evaluation of estimates of London house prices during 2015 and 2016 from Google Street View images. (a) Map of median house prices for each MSOA ('Observed'). Darker areas correspond to higher house prices. **(b)** Map of median house prices for each MSOA estimated from Google Street View images ('Predicted'). Again, darker areas correspond to higher house prices. We generate estimates using the methodology outlined in Fig. 3.1. Visual inspection reveals that median house prices estimated from Google Street View images correspond well to the true median house prices for a given area. For example, the east of London is both observed and predicted to have cheaper housing than the centre of London. A regression analysis confirms that the elastic net model generates reasonable estimates of house price ($R^2 = 0.56$, $N = 983$, $p < 0.001$). **(c)** Map of the errors in median house price estimates for each MSOA ('Error'). Blue denotes an underestimate and red denotes an overestimate. *(continues on the following page)*

Figure 3.4: *(continues from previous page)* Median house prices in the south-east tend to be overestimated, whereas median house prices in the north-west tend to be underestimated. **(d)** Log-log plot of observed and predicted median house prices. The red dots represent the MSOAs of London and the blue line represents where perfect predictions should fall. We find that median house prices under £1,000,000 are generally estimated well, but estimates for areas with higher median house prices are generally too low. This is possibly because the majority of median house prices are between £300,000 and £600,000, such that the model is more accurate for MSOAs with a median house price in that range. For all maps, quantile breaks are implemented with six bins. Contains National Statistics data © Crown copyright and database right 2011. Contains Ordnance Survey data © Crown copyright and database right 2011. Contains Street View data © 2017 Google.

To determine whether a model is successful or not based on the R^2, we use the recommendations outlined by Chin (1998) throughout this thesis. Specifically, Chin (1998) suggests that models achieving an R^2 value of 0.67 indicate substantial success, an R^2 value of 0.33 indicates moderate success, and an R^2 value of 0.19 indicates weak success. We take this into consideration when we determining how well our models perform.

Figure 3.4b shows a map of the predicted median house prices, or in other words, the estimate of median house prices produced from Google Street View imagery from 2014 and 2015. We find that there is a good match between the geographic distribution of the observed house prices shown in Fig. 3.4a and the predicted house prices in Fig. 3.4b.

To compute the error for each MSOA, we subtract the observed median house price from the predicted median house price. In Fig. 3.4c, we display the errors, where blue denotes an underestimate of median house prices, and red denotes an overestimate of median house prices. We find that the mean absolute error is £97,710 and the mean absolute percentage error is 20.24% (both to 4 significant figures). We compare these errors with those achieved by a baseline mean model, to see if our model is an improvement upon the baseline model. The mean model predicts all median house prices in all MSOAs to be the median house price of £479,800, and achieves a mean absolute error of £152,800 and mean absolute percentage errors of 32.13% (all to 4 significant figures). Indeed, our model outperforms the baseline model. Mapping the prediction errors from our model reveals that house prices in the north-west and south-west of London are generally underestimated, and house

prices in the south-east of London are generally overestimated.

Why does the model tend to overestimate house prices in the south-east, but underestimate house prices in the north-west and south-west? One possible explanation could be the role of transport connections. The presence of good transport connections is likely to have a strong effect on house prices (Efthymiou & Antoniou, 2013; Gibbons & Machin, 2005), as better transport connections allow people to make quicker journeys to places of work, amenities, and other locations. The south-east of London in particular has poorer transport connections than the north-west of London, as it lacks a dense network of underground stations. As the model only uses visual information to estimate house prices, it is unable to account for the fact that journey times from the south-east to the centre of London tend to be longer than from areas in the north-west with similar appearances, which impacts property prices in those areas. Hence, it seems likely that the algorithm concludes that areas in the south-east have visual characteristics that resemble those of areas in London suburbs with higher house prices.

Figure 3.4d shows a scatter plot of the observed house prices against the predicted house prices, where red dots represent MSOAs of London and the blue line represents where perfect predictions should fall. The model has particular difficulty with areas where the median house price is higher than £1,000,000, which are underestimated more frequently than they are overestimated. It seems likely that this is due to the distribution of house prices having a long tail of particularly high house prices. The median house price of all areas is £416,875, whereas the maximum house price is £27,000,000. Only 30 MSOAs have house prices above £1,000,000 (3%) and 677 of 983 MSOAs have a median house price between £300,000 and £600,000 (69% to 2 significant figures). A lack of a large number of examples of areas with high median house prices may have caused the model difficulty in learning to recognise the visual characteristics of such areas. Alternatively, it is possible that the model needs some recalibration to better deal with this long tail of high house prices, a point we return to in Section 3.4.

Each non-zero elastic net coefficient extracted from the model in Fig. 3.5 indicates the impact that the corresponding visual feature has on inferring median house prices. For each coefficient, the coefficient value indicates the size of the impact of that feature on house price estimation, and whether the impact is positive or negative. In other words, it shows us whether the feature is associated with higher house prices (positive coefficient) or lower house prices (negative coefficient) and to what degree. This allows us to gain insight into how the model generates estimates of median house prices across London.

As the model performs well, we would therefore reasonably expect positive model coefficients to represent characteristics of areas with higher median house prices, and negative model coefficients to represent characteristics of areas with lower median house prices. From the positive coefficients in Fig. 3.5, we find that some of the features that the model associates with areas with higher median house prices are 'palace', 'office building', and 'courthouse'. From the negative coefficients, we find that some of the features that the model associates with lower median house prices are 'slum', 'playground', and 'hospital'. At least some of these associations follow expectations. For instance, office buildings are more likely to be found in the centre of the city, where prices are higher.

However, some of the associations that the model makes are unexpected. For example, 'rice paddy' is included as a positive coefficient, but we would not expect to find a 'rice paddy' in London given the UK climate. Similarly, the model identifies 'rainforest' as a negative coefficient, but no rainforests exist in London.

Given that at least some of the connections made by the elastic net between features and house prices make sense, we want to understand why the elastic net is using unexpected features to estimate median house prices in London. We investigate whether some of the unexpected features are not very prevalent in the images of London. We find that while only 1.4% of our Google Street View images have 'rainforest' as one of their top five features, 22% of images have 'hospital' as one of the top five highest-scoring features (to 2 significant figures). We also note that some of the most unexpected features, such as 'rainforest' and 'rice paddy', have low coefficients in the model.

Although some of these unexpected features do not appear that frequently in our images, we decide nevertheless to examine examples of images receiving these labels to gain a full picture of the operation of the methodology. To begin our examination, we choose four unexpected features from Fig. 3.5, and for each feature, we sample three images that generate high feature values. We display three high-scoring images for each feature because they most reflect what the CNN sees when assigning labels and generating feature values. Note that for each image, the associated feature is its highest-scoring Places label. By inspecting these images, we should then be able to understand why such features are associated with house prices. Figure 3.6 displays our chosen features and their corresponding images. Feature values are given underneath each image.

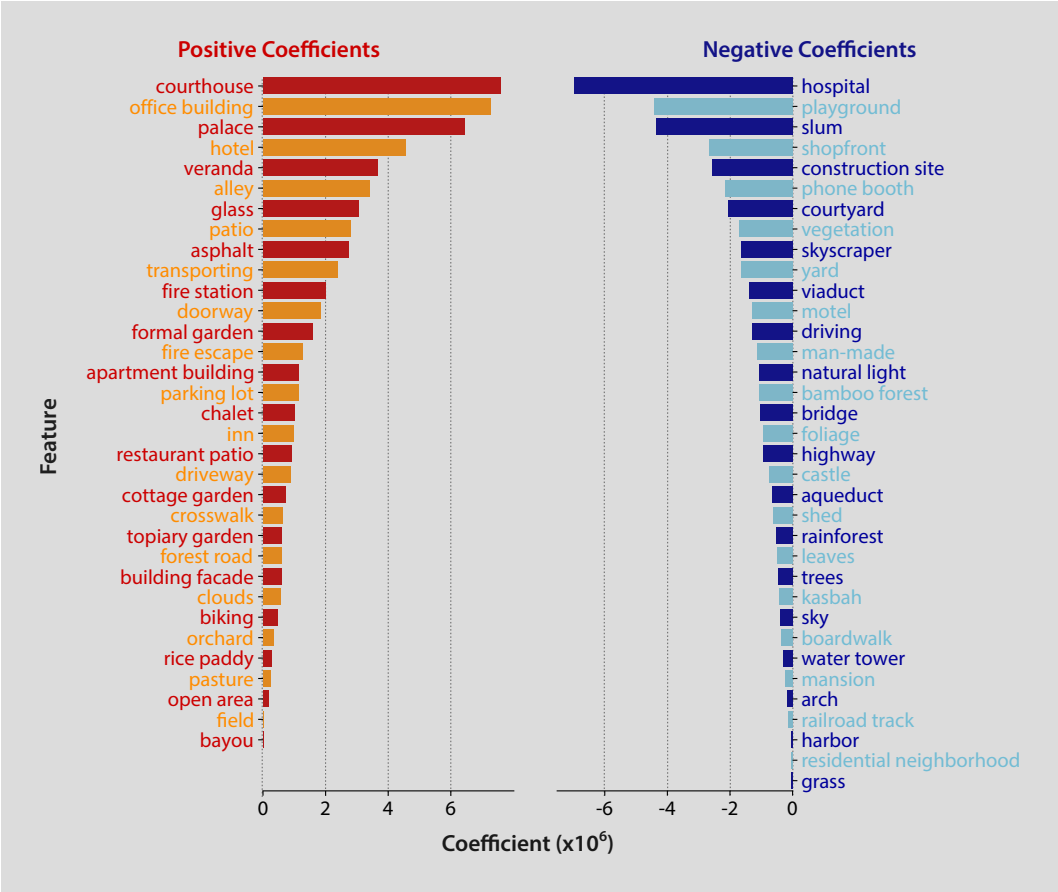

Figure 3.5: Elastic net coefficients for estimates of London house prices during 2015 and 2016 from Google Street View images. Positive coefficients relate to visual features that are associated with higher median house prices, and negative coefficients relate to visual features that are associated with lower median house prices. We build an elastic net model using the methodology outlined in Fig. 3.1, and depict the resulting coefficients for all features where the coefficient is not zero. The model has learned to associate elements of images that resemble various places and features with higher and lower median house prices. For example, the model has learned to associate the elements 'palace', 'office building', and 'courthouse' with areas that have higher median house prices, and 'slum', 'playground', and 'hospital' with areas that have lower median house prices. At least some of these associations make intuitive sense. For example, a 'palace' is a grand, ornate property that is expensive to both buy and maintain, so we would expect it to be in an affluent area. However, the model also associates some surprising features with areas in London. For example, neither a 'rice paddy' nor a 'rainforest' would be expected in London.

Figure 3.6: Understanding unexpected features. We want to understand why some features that the elastic net is using to generate estimates of median house prices are not features that one would expect to find in London on a widespread basis. For example, what pictures are being labelled as a rainforest? To do so, we start by choosing four unexpected features from Fig. 3.5 and for each feature, we select three images from the ten images with the highest feature values. We showcase a selection of top images for each feature because those images reflect, to the highest degree, what the CNN sees when assigning the feature label to those images. Furthermore, the associated feature for each image is its highest-scoring Places label. All images are from the set of Google Street View images accessed for this analysis, all feature values are determined by the CNN, and the feature values can be interpreted as relating to the likelihood of a given feature appearing in a given image. By inspecting these images, we hope to understand why such unexpected features are detected in images of London. *(continues on the following page)*

Figure 3.6: *(continues from previous page)* **(a)** Top images for the feature 'palace'. The CNN labels large buildings with a historical aesthetic and numerous windows as palaces. **(b)** Top images for the feature 'rice paddy'. The CNN labels large fields with light coloured barriers as rice paddies. **(c)** Top images for the feature 'mansion'. The CNN labels housing that is connected, such as terraced and semi-detached housing, as mansions. **(d)** Top images for the feature 'rainforest'. The CNN labels large clusters of lush, green trees as rainforests. Overall, the CNN seems to give images that share shapes and patterns the same label but the labels themselves do not always reflect the actual contents of the images accurately. This may be because the CNN was trained on diverse images from all across the world where the prior probability of scenes such as rainforests and rice paddies would be higher, whereas in our case study, we only analyse images from London. Street View images © 2017 Google.

Notice that the images that are given high scores for a particular label tend to look very similar. For example, all the images labelled as 'rainforest' in Fig. 3.6d are images that are nearly exclusively filled with greenery. Similarly, all the images labelled as 'mansion' in Fig. 3.6c contain housing that is sufficiently broad to nearly fill the frame from left to right, with repeating patterns in the architecture. However, the labels do not always reflect the content of the images precisely. For example, the images labelled 'rainforest' are not images of rainforests but other greenery, and the images labelled 'mansion' are, in some cases, images of impressive housing, but not necessarily houses that many British citizens would refer to as mansions.

This could be due to the fact that the CNN was trained on images from all across the world, whereas we are only analysing images from the UK in this case study. For example, we do have large trees with lush greenery in the UK but we know that they are not part of a rainforest, and simply luscious trees. However, there are many locations over the world that do have the natural environment for rainforests that consist of large, leafy trees, such that the prior probability of a 'rainforest' score would be higher. This may lead to the CNN associating leafy tree visuals with rainforests more frequently than is appropriate in the UK. Architecture and landscapes vary from country to country, so the generalisations made by the CNN do not always translate to our images. However, images given a certain label are usually similar. This means that a label will generally reflect some sort of consistent visual characteristic in a scene that could be plugged into further modelling, for example with an elastic net. However, care needs to be taken in interpreting the labels too literally, as features can attain high scores without exactly reflecting the content of

the image.

3.3.2 Estimating changes in local house price rank from 2005 and 2006 to 2015 and 2016 using Google Street View images

Which areas have recently seen relatively high increases in house prices? The next step of our analysis is to evaluate whether Google Street View data can provide any insight into this question. Before we do this, we wish to further understand our house price data for the two periods we will compare. Specifically, we want to confirm that we have sufficient data coverage across local areas of London to draw accurate conclusions. Figure 3.7 contains maps of the distribution of house transactions for 2015 and 2016, as well as 2005 and 2006 across all areas in London. For both maps, a darker red indicates a higher number of transactions in a given area.

During 2015 and 2016, the number of houses sold across MSOAs in London ranges from 52 to 1,654. The median number of transactions is 227. This highlights that the distribution of property transactions across MSOAs in London during this time period has a long tail of greater numbers of property transactions. During 2005 and 2006, the number of houses sold across MSOAs in London ranges from 37 to 1,198, and the median number of transactions is 373. This suggests that there were somewhat more sales across London during 2005 and 2006 than during 2015 and 2016. Indeed, 250,171 transactions took place between 2015 and 2016, whereas 312,344 transactions took place between 2005 and 2006. In line with this observation, only 150 MSOAs (15% to 2 significant figures) had fewer than 200 transactions in 2005 and 2006 compared to 373 MSOAs (37% to 2 significant figures) in 2015 and 2016. Broadly speaking, we see that more transactions occur in areas towards the centre of London in both time periods. However, there is good data coverage across all areas in London. At least 37 transactions take place in each MSOA in both time periods, so there are no areas with zero transactions.

As described in Section 3.2.5, we measure the changes in median house prices from the median house price rankings of all MSOAs in London from two time periods. This allows us to understand how areas evolve over the decade relative to each other. Preliminary analysis of the changes in ranked price reveals that areas evolve at different paces, since some areas do not change in ranked price over the decade, whereas other change substantially. To help us better understand the data we are modelling, we inspect the rank distribution for 2015 to 2016 and 2005 to 2006 visually. In Fig. 3.8, we map the rank distributions of MSOAs across London during 2015 and 2016, as well as 2005 and 2006.

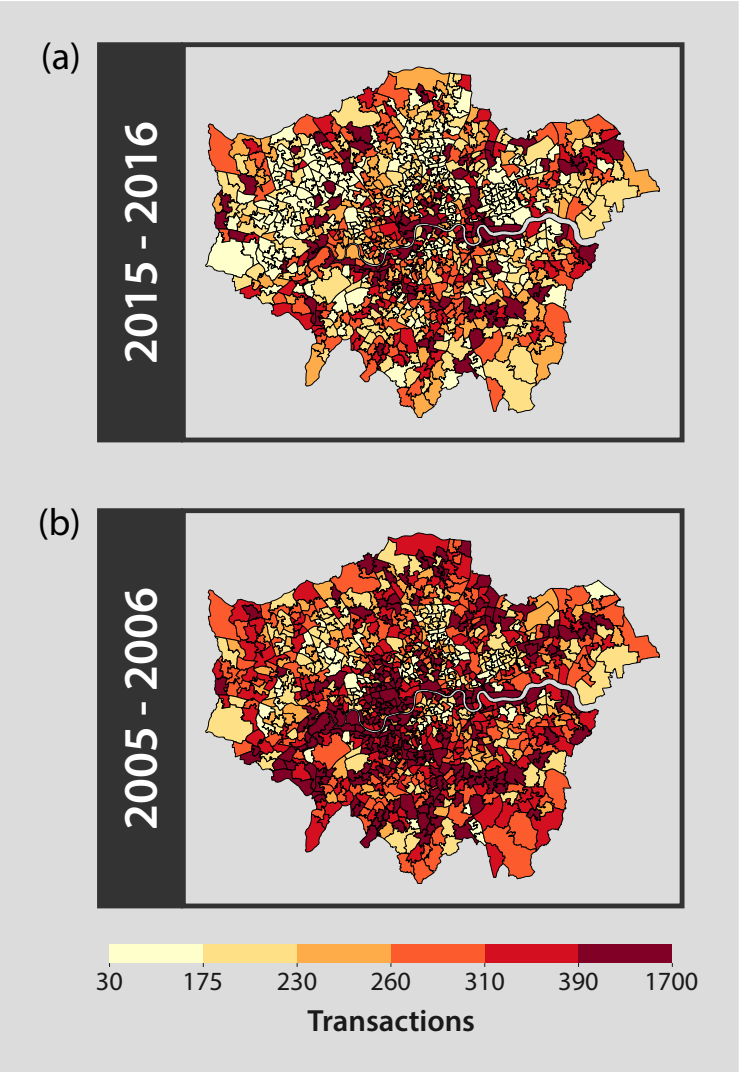

Figure 3.7: The distribution of house transactions across London from 2015 and 2016 to 2005 and 2006. Before attempting to estimate changes in median house prices across local areas of London, we want to confirm that we have sufficient data coverage across all areas of London for the two time periods we compare, as we can draw more accurate conclusions about changes in house prices from more representative data. Therefore, we investigate how the number of transactions across local areas of London is distributed. For both of the following maps, a darker red represents a greater number of house transactions. *(continues on the following page)*

Figure 3.7: *(continues from previous page)* **(a)** Map of the number of transactions in each MSOA during 2015 and 2016. The number of transactions ranges from 52 to 1,654, and the median number of transactions in each MSOA is 227. **(b)** Map of the number of transactions in each MSOA during 2005 and 2006. The number of transactions ranges from 37 to 1,198, and the median number of transactions in each MSOA is 297. During both time periods, there is quite a large range of transactions, and the transactions exhibit a positively skewed distribution with a long tail of greater numbers of property transactions. Visual inspection of the maps shows that more transactions occur in and around the centre of London during both time periods. However, there is good coverage across all MSOAs of London. For all maps, quantile breaks are implemented with six bins. Contains National Statistics data © Crown copyright and database right 2011. Contains Ordnance Survey data © Crown copyright and database right 2011.

A darker red indicates a higher ranked house price. From visual inspection, the rankings do not appear to change drastically over the decade. Areas in the centre of London tend to exhibit a large improvement over the decade, especially towards the west. In contrast, areas towards the east and south of London exhibit a lesser improvement over the decade.

We investigate whether areas with greater changes in ranked median house price can be identified from Google Street View imagery. Figure 3.9a displays the actual change in rank across areas in London. Blue indicates a decrease in ranked price and red indicates an increase in ranked price over the decade. The darker the colour, the larger the change.

Figure 3.9a reveals that some areas in the eastern part of London have risen in the ranks, especially those closest to the centre from the north and the south. We can also see that there is a clear divide between the suburbs and the centre of London. Namely, areas in the centre of London tend to increase in ranked price over the decade, whereas the suburbs appear to decrease in ranked price. Some of the changes in ranked price may seem large, but we note that a change in median house price from £235,000 to £340,000 would be sufficient to make an MSOA jump a quarter of the table, as was the case for an MSOA in the south-west of London (E02000847).

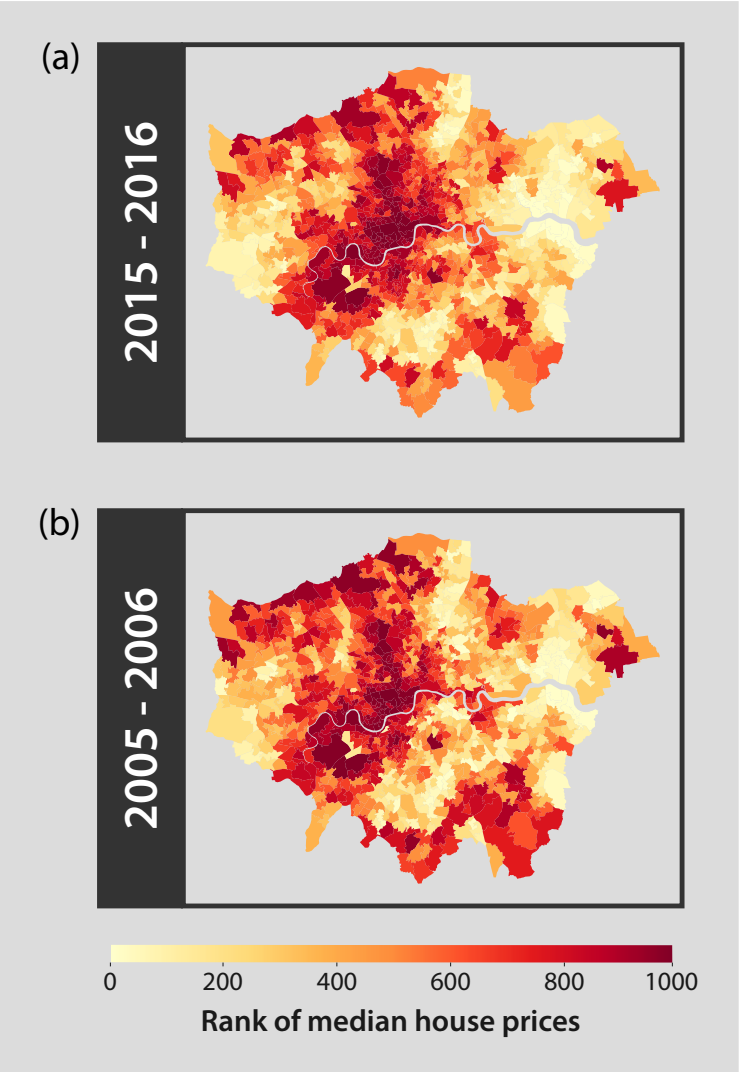

Figure 3.8: The distribution of rankings of median house prices in MSOAs across London during 2015 and 2016, and 2005 and 2006. We visually inspect the median house price rank distribution for 2015 to 2016 and 2005 to 2006 to help us better understand the data we are modelling. During each time period, we rank MSOAs in London from 983 (highest median house price) to 1 (lowest median house price). **(a)** Map of median house price rankings during 2015 and 2016. **(b)** Map of median house price rankings during 2005 and 2006. For both maps, a darker red represents a higher ranked median house price. We find that the ranks have not changed dramatically over the decade. The greatest changes can be observed in the centre and east of London. Contains National Statistics data © Crown copyright and database right 2011. Contains Ordnance Survey data © Crown copyright and database right 2011.

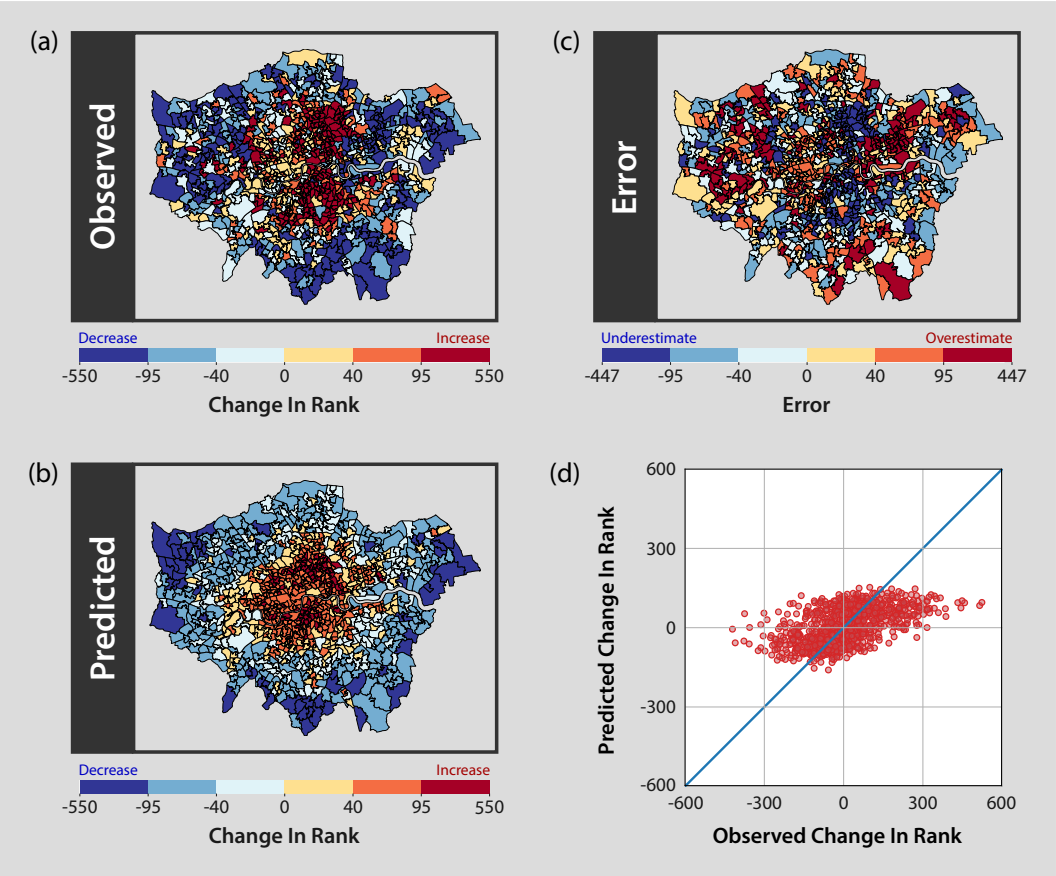

Figure 3.9: Inferring changes in ranked house prices across local areas in London using Google Street View images. Can house price changes be inferred from Google Street View images? **(a)** Map of change in median house prices for each MSOA ('Observed'). We rank MSOAs in London from 983 (highest price) to 1 (lowest price) using house price data from 2015 and 2016. We then rank the MSOAs in the same fashion using house price data from 2005 and 2006. The change in each MSOA is then given by the rank during 2015 and 2016 minus the rank during 2005 and 2006. In the map, blue denotes a decrease in ranked price over the past decade and red denotes an increase in ranked price over the decade. Broadly speaking, we see that areas in the centre of London have continued to rise in relative house price. However, some areas in the eastern parts of London have also risen in the ranks, particularly those closest to the centre in both the north and the south. **(b)** Map of change in median house prices for each MSOA estimated from Google Street View images ('Predicted'). Again, blue denotes a decrease in ranked price over the past decade and red denotes an increase in ranked price over the decade. *(continues on the following page)*

Figure 3.9: *(continues from previous page)* We estimate the change in ranked price using the methodology discussed in Section 3.2.5. Visual inspection of both maps shows that the elastic net model can capture some patterns, but overgeneralises somewhat, mostly predicting relative increases in price in the centre of London. A regression analysis confirms that the model captures some of the observed pattern in median house price changes, but not all ($R^2 = 0.27$, $N = 983$, $p < 0.001$). **(c)** Map of the errors in estimates of change in median house price rank for each MSOA ('Error'). Blue denotes an underestimate and red denotes an overestimate. The map again highlights that the model has underestimated rank increases in parts of the east of London, close to the centre. **(d)** Scatter plot of observed and predicted change in rank. The red dots represent MSOAs of London and the blue line represents where perfect predictions should fall. We find that the change in rank estimates are capped between -200 and 200, whereas the true changes in rank range between -424.5 and 522. For all maps, quantile breaks are implemented with six bins.

When employing the elastic net model to produce estimates of changes in ranked median house prices, we find the optimal value of *l1_ratio* to be 0.99 with a corresponding α of 0.0036 (to 2 significant figures). As with our analysis of median house prices in Section 3.3.1, we conduct a regression analysis using leave-one-out cross-validation to understand how well the model performs. We find that the model does pick up some patterns in the changes in ranked house price, but not all ($R^2 = 0.27$, $N = 983$, $p < 0.001$).

Figure 3.9b displays the estimated change in rank across areas in London. We generate the estimates using the methodology described in Section 3.2. Upon inspection, we can confirm that the elastic net model captures some of the aforementioned patterns across London. However, it has a tendency to overgeneralise the difference between the centre and the suburbs, estimating increases in ranked price in the centre of London and decreases in ranked price in the suburbs, whilst missing further nuances in the pattern. A plausible cause for this tendency is that areas that have seen higher recent changes in house price may not generally be visually distinctive in recent imagery, whereas visual differences between the centre and the suburbs are easy to pick up on. To improve these estimates, it may be necessary to examine how images change over time instead.

From the maps, we can see that the majority of areas that see a large increase in ranked price over the decade are indeed those near the centre of London, particularly towards the north-east and south-east. Consequently, it is likely that the model has simply learnt to associate features that are common in the centre of London during 2014 and 2015 with rank increases over the decade.

From the prediction errors in Fig. 3.9c, we notice that the model struggles to capture relatively extreme changes in ranked price. This is clearly visible in the scatter plot (Fig. 3.9d), since the model always produces estimates of the change in rank between -200 and 200. The rank changes range from -424.5 to 522 with a median change in rank of -3.5. Of the 14% of areas whose ranking changes by more than 200 places, 84 areas increase in ranked price and 55 areas decrease in ranked price from 2005 and 2006 to 2015 and 2016. We may be seeing these errors because the distribution of ranked median house price changes is heavy-tailed, so it violates the assumption of normality that is required by the model.

As with the house price estimation analysis (Section 3.3.1), we extract elastic net coefficients from the model and present all non-zero coefficients in Fig. 3.10. Positive coefficients are connected to a rise in ranked price over the decade, and negative coefficients are connected to a reduction in ranked price over the decade. We would expect that some coefficients associated with an increase in ranked price might also be associated with higher median house prices, since some areas that exhibit high median house prices correspond to areas that have seen increases in rank prices over the decade (e.g. some areas of the centre of London). In other words, we would expect the positive and negative coefficients extracted from the model to be somewhat similar to those extracted during the house price estimation analysis (Fig. 3.5), even if the exact coefficient values are different.

Some of the coefficients we extract from the model are the same, like the positive coefficient 'apartment building' and negative coefficient 'motel'. However, some coefficients are new, like the positive coefficient 'fairway' and the negative coefficient 'shrubbery'. In fact, some coefficients are inverted such that they are positively associated with median house prices but negatively associated with changes in house price rank. Examples include 'courthouse' and 'palace'. Similarly, some coefficients that are negatively associated with median house prices are positively associated with changes in house price rank, such as 'hospital' and 'playground'.

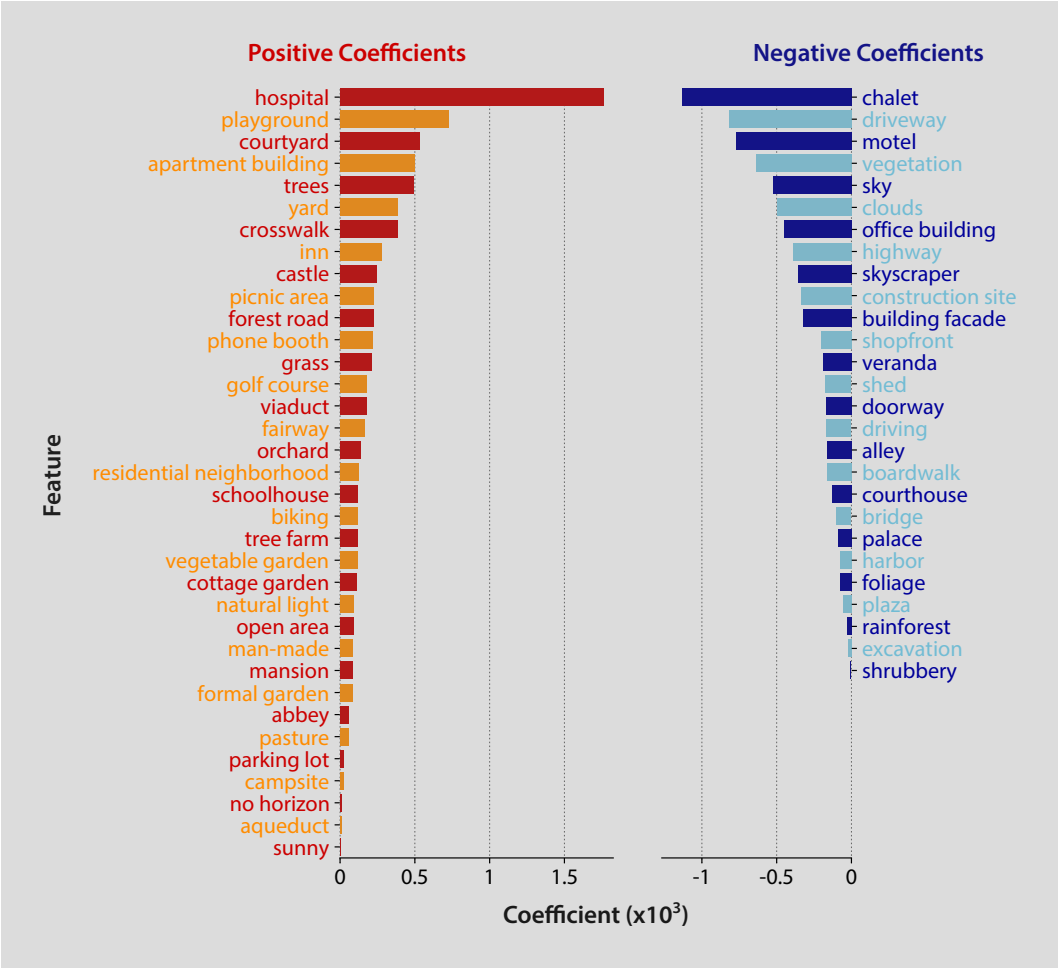

Figure 3.10: Elastic net coefficients for estimated changes in ranked median house prices of MSOAs across London from Google Street View images. We rank MSOAs in London and estimate the change in rank over a decade using the methodology described in Section 3.2.5. The resulting coefficients from our elastic net are depicted for all features where the coefficient is not zero. Positive coefficients correspond to features associated with increases in ranked price over the decade, and negative coefficients correspond to features associated with decreases in ranked price over the decade. It would be reasonable to expect some of the features associated with higher median house prices in Fig. 3.5 to also be associated with increases in ranked prices. However, this is not the case for many of the coefficients. For example, the features 'hospital' and 'playground' are strongly positive coefficients in this analysis, instead of negative coefficients as in Fig. 3.5. Likewise, 'courthouse' and 'palace' are strongly negative coefficients in this analysis, instead of positive coefficients as in Fig. 3.5.

Examination of the map of observed median house prices (Fig. 3.4a) and the map of observed changes in median house price rank (Fig. 3.9a) shows that some areas have high median house prices but did not climb as highly in the ranks, whereas other areas have low median house prices but did climb high in the ranks. For example, one MSOA in the centre of London (E02000978) has a large median house price of £2,700,000 but only a very modest increase in rank of 2. The swap in some of the coefficients may relate to some of these areas.

3.4 Discussion

In this chapter, we take initial steps towards understanding whether we can gain information about local house prices across London from street-level imagery on Google Street View. Our initial steps consist of two main analyses. We attempt to infer median house prices in London during 2015 and 2016, and we also attempt to model changes in ranked median house prices in London from 2005 and 2006 to 2015 and 2016.

To carry out these analyses, we exploit a sample of Google Street View images from London during 2014 and 2015, and feed them into a CNN to extract visual features and their quantitative values. We then aggregate the feature values of all images at MSOA granularity to obtain mean feature values for each feature in each MSOA. Finally, the mean features are used as input in an elastic net regression to estimate median house prices and changes in ranked median house prices. We discover that median house prices can be inferred well from Google Street View images alone, but the model struggles with more extreme high prices. This may be due to the positively skewed distribution of our median house price data, where there are only a few examples of areas with high median house prices. A possible solution to this would be to change the configuration of the model to better take this skew into account. For example, we could estimate the logarithm of median house prices. This may lead to a model that estimates median house prices with greater accuracy.

The model particularly tends to overestimate prices in the north-west of London, and underestimate prices in the south-east of London. We believe this could be due to the impact of external factors on house prices, such as the availability of transport links. In future, we could incorporate such external details into the model, or even combine image data with pre-existing models that estimate house prices based on more traditional variables, such as the number of beds in a property, the size of the house, and when a property was built (Case, Clapp, Dubin, & Rodriguez, 2004; Lu, Charlton, Harris, & Fotheringham, 2014; Osland, 2010). This could pave

the way for house price estimation models that incorporate differing yet relevant variables and perform with high accuracy.

In addition to estimating median house prices well, we also find that the model has some success in estimating changes in ranked median house prices. As with our median house price data though, our ranked median house price changes are not normally distributed, so the model estimating changes in median house prices struggles with extreme values. More precisely, the distribution of changes in median house price rank are heavy-tailed. Changing the model in future may alleviate this issue and improve the accuracy of the model. For example, a neural network could be used to estimate changes in ranked price. An architecture could be developed such that heavy-tailed distributions are the focus, and the leave-one-out cross-validation approach could be used to generate estimates of the changes in median house price rank.

Moreover, the model struggles to capture some of the variations in median house prices over the decade from the data we hold. This may be because we lack visual information on changes over time, as we only estimate the price changes from images taken towards the end of the period in question. In future, we could improve the accuracy of the model by using images from two separate time periods to estimate the changes between these two time periods. While Google Street View launched in 2007, such that image data representative of 2005 and 2006 are not available, this approach would be feasible for many other time periods. We could also make use of more images in the model overall. Currently, we make estimates at MSOA level and sample images across London using a 100 metre by 100 metre grid. However, if we increased the number of images in our sample by reducing the size of the grid squares, we may be able to improve the accuracy of estimates at MSOA level, and make estimates at a higher spatial granularity too.

To summarise, our findings confirm that it is possible to infer house prices from large scale street-level image data combined with deep learning techniques. As we find that we can successfully infer median house prices from Google Street View imagery, we ask whether this success is due to the specific data source. Could an alternative street-level image source allow us to construct house price estimation models of equal or greater accuracy? We investigate this question in Chapter 4.

Chapter 4

Inferring house prices from Geograph images

4.1 Introduction

In Chapter 3, we found that we can use Google Street View images of neighbourhoods to infer local house prices in London. We hypothesise that this approach works because the appearance of a neighbourhood is likely to both impact and be affected by local house prices. However, other sources of street-level imagery are also available, for example from crowdsourcing projects. Would this approach work with other street-level imagery sources too? How might performance compare?

An example of an alternative street-level image source would be Geograph. The Geograph project aims to collect images of every square kilometre of Great Britain and Ireland. Members of the public upload photos to Geograph and are encouraged to provide images that would give a viewer a good sense of what a local area looks like. As such, the Geograph image dataset does not only consist of photos of streets, but contains images of off-road areas such as parks, canals, and mountains too.

Furthermore, Geograph images are non-proprietary and shared under a Creative Commons license. The website is updated frequently, so large quantities of data are available for all areas across Britain and Ireland, including London, from March 2005 onwards when the project began. Photos are not restricted to this time frame though, as users are able to upload photos from any time period in the past. Geograph therefore offers a rich source of open data imagery, potentially offering fuller insights into local neighbourhood environments, with a relatively long history available for analysis.

Thus, we aim to determine whether the combination of Geograph images and deep learning can be utilised to gain insights into the development of house prices in a given area. To do this, we ask the following questions:

1. Can we automatically infer house prices from current Geograph images?
2. Can we automatically identify areas which have seen relatively high recent increases in house prices from Geograph images?
3. How do the insights gained from Geograph images differ from those gained from Google Street View images?
4. How does performance of the house price estimation methodology compare for these two data sources?

4.2 Data and methods

4.2.1 House price data

Figure 4.1 depicts our methodology. To help us compare the Geograph analyses with the Google Street View analyses (Section 3.3.1), we focus on house transactions from 2015 and 2016, and use the same house price data as for the Google Street View analysis. Again, as the house price data are skewed with a long positive tail, we calculate the median house price for each MSOA.

As in the Google Street View analysis, we analyse images and house prices at the Middle Layer Super Output Area (MSOA) level, where an MSOA contains a population of between 5,000 and 15,000 inhabitants, or between 2,000 and 6,000 households, with 983 MSOAs in London (Stokes, 2012).

4.2.2 Geograph images

We retrieved Geograph data in November 2019 by using the data request form at Geograph (n.d.-b). We accessed all images of London that had been uploaded to Geograph since its inception in March 2005 until November 2019, as well as metadata associated with the images, such as the Geograph website link to each image and the date that each image was originally taken. We find that 328,014 images of London were uploaded to Geograph's website between March 2005 and November 2019. The quantity of Geograph images per MSOA ranges from 8 to 14,768 with a median of 151 images per MSOA.

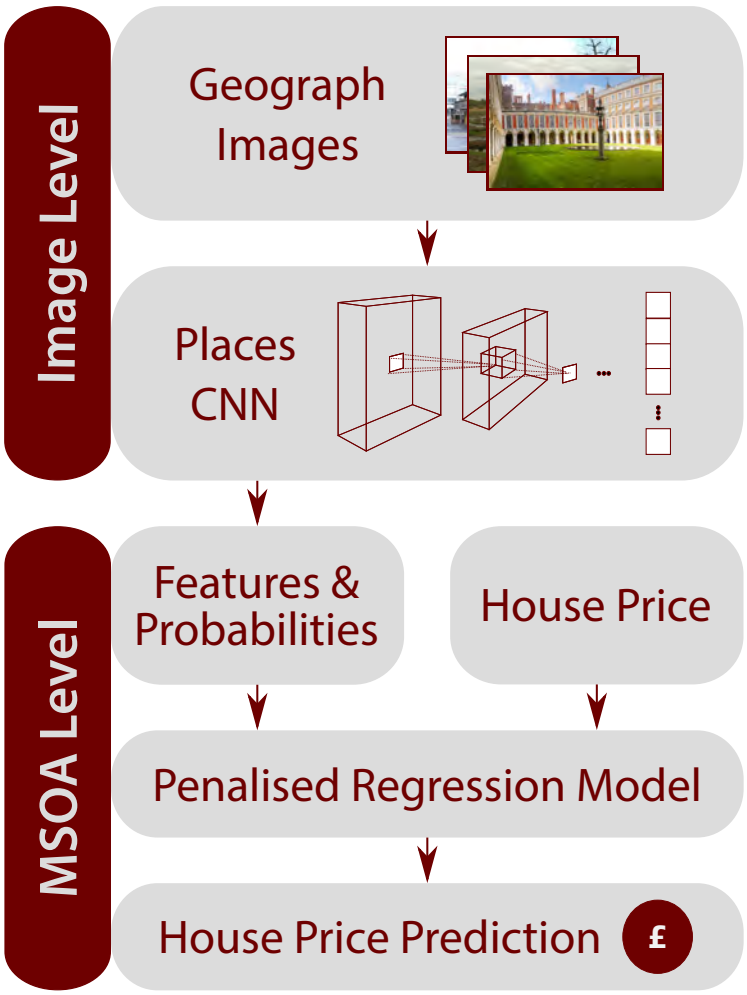

Figure 4.1: Deep learning framework. We adapt the deep learning framework from the Google Street View analysis described in Chapter 3 to infer house prices from Geograph imagery. We obtain all Geograph images available for the Greater London area, instead of Google Street View images, and designate each image to an MSOA. To process each image, we use an updated version of the Places CNN developed by Zhou et al. (2018), which we refer to as Places365. Examples of the 365 Places features identified by Places365 include 'motel' and 'park'. Given an input of an image, Places365 produces a vector of 365 feature values, rather than the 205 features produced by Places205. As in the Google Street View Places analysis, the feature values relate to the likelihood of a given feature appearing in a given image. We calculate the mean feature vector for each MSOA and then train an elastic net to learn the relationship between the mean feature vector for an MSOA and the median house price for that area. Finally, we test the estimates from the model using a leave-one-out cross-validation approach and compare the resulting predictions to the actual median house price for each MSOA. *(continues on the following page)*

Figure 4.1: *(continues from previous page)* Note that we do not include SUN attributes in this analysis. All Geograph photos are licensed for reuse under creativecommons.org/licenses/by-sa/2.0. Front photo © David Dixon; middle photo © Robert Lamb; back photo © Jim Osley.

As we are analysing house prices from 2015 to 2016, we wish to analyse images taken during 2015 and 2016 too, as they will represent the environment during the time period being analysed. The metadata attached to Geograph images make it easy to select images taken in 2015 and 2016 only. We find that 51,918 photos of London were taken and uploaded to Geograph between 1st January 2015 and 31st December 2016. For this time frame, the quantity of Geograph images per MSOA ranges from 0 to 2,821 with a median of 21 images per MSOA. As there are 30 MSOAs in London with no images available from Geograph for this time period, we exclude these MSOAs from our analyses and focus on the remaining 953 MSOAs.

There are approximately ten times as many Google Street View photos than there are Geograph photos available in our 2015–2016 Geograph dataset. We use 518,808 Google Street View images to estimate house prices in 2015 and 2016, and the number of images per MSOA ranges from 88 to 2,164 with a median of 496 photos. This raises the question of whether the quantity of estimates from Geograph imagery will be adversely impacted simply by the dramatic fall in imagery volume. Therefore, in addition to estimating house prices in 2015 and 2016 from Geograph images from the same time frame, we also estimate house prices in 2015 and 2016 without restricting the date on which the Geograph images were taken. This will help us investigate whether performance improves when more images are available, regardless of the date on which the photo was taken.

4.2.3 Places CNN

Similarly to our approach for the Google Street View analysis, we utilise Places CNN to extract features from Geograph images. In this analysis, we apply an updated version of Places CNN to the Geograph images (Zhou et al., 2018), specifically the Places365-AlexNet CNN. Places365 was trained on 1.8 million images from the Places2 database of 10 million images (Zhou et al., 2018), in contrast to the Places205 network used in the Google Street View analysis, which was trained on 2.5 million images from the original Places database of 7 million images (Zhou et al., 2014).

Places365 allows for 365 features to be extracted, whereas Places205 allows for 205 features to be extracted. As such, we process each of the Geograph images with Places365 and extract the values of all 365 features into a feature vector. To process images efficiently by processing them in parallel, we use high-performance computing clusters managed by the Scientific Computing Research Technology Platform at the University of Warwick. As in the Google Street View analysis, the feature values from Places365 relate to the likelihood of a given feature appearing in a given image. Since the feature values are normalised using a softmax activation function, the feature values range from 0 to 1 and each feature vector sums to 1.

As Geograph imagery is outdoor imagery, we remove all 160 indoor features, such as 'bedroom' and 'staircase'. This leaves 205 outdoor features, whereas in the Google Street View analysis, 136 outdoor features remained, excluding SUN attributes. For each MSOA, we again calculate the mean value of the feature vectors for all images in that MSOA.

We note that use of the newer Places network represents a change in methodology in comparison to Chapter 3. When comparing the results between the Geograph and Google Street View analyses, it will be necessary to consider how an update to the methodology may have impacted performance.

4.2.4 Elastic net model

The visual characteristics of each MSOA are therefore represented by a 205 feature vector. This vector then acts as input to the house price estimation regression model. As in the Google Street View analysis, we use an elastic net to help us manage correlations between the features. We train and test the elastic net model using the leave-one-out methodology described in the Google Street View analysis (Section 3.2.4).

4.2.5 Measuring change in house prices

We also investigate whether Geograph imagery reveals which areas have recently seen the largest jumps in house prices. We again consider house price changes over a decade from 2005 and 2006 to 2015 and 2016. When analysing Geograph photos from 2015 and 2016 only, 953 MSOAs have street-level imagery available. As in the Google Street View analysis, we rank MSOAs by median house price by setting the highest rank of 953 to the MSOA with the highest median house price, and the lowest rank of 1 to the MSOA with the lowest median house price. If two areas have the same median house price, we give them both the mean rank. When analysing all

Geograph images for London without restricting the dataset to photos taken in 2015 and 2016, we set the highest rank to 983, as images are then available for all 983 MSOAs. The change in rank is again calculated as the 2015–2016 rank minus the 2005–2006 rank. To interpret the change in ranked price, consider a change of -100. This would indicate a decrease in ranked price by 100 places, hence a relative rank reduction in median house price over the decade.

To model the change in rank, we again employ an elastic net, taking the 205 outdoor Places CNN features as input, and predicting the change in rank.

4.3 Results

4.3.1 Estimating local house prices in 2015 and 2016 from Geograph images taken in 2015 and 2016

Can we build an accurate model that estimates house prices from Geograph images? Excluding the 30 MSOAs with no Geograph images, examination of the house price data from 2015 and 2016 reveals that median house prices for the remaining 953 MSOAs across London range from £185,000 to £1,270,000. The minimum median house price is £115,540 higher than the minimum of £69,460 in the corresponding Google Street View analysis, where all 983 MSOAs were considered. This difference suggests that Geograph users are less likely to visit and photograph MSOAs with lower house prices. This could be due to a lack of amenities in these areas, or because Geograph photographers are less likely to live in those areas themselves.

Figures 4.2a and 4.2b display the observed and predicted median house prices from 2015 to 2016, where a darker area represents a higher median house price. The observed median house price is the actual median house price for each MSOA (Fig. 4.2a), and the predicted median house price is the estimate of median house price from Geograph photos taken in 2015 and 2016 (Fig. 4.2b). We generate estimates using the methodology outlined in Fig. 4.1. As in the Google Street View analysis, the observed median house price map in Fig. 4.2a shows that areas in the centre of London tend to exhibit high house prices, whereas areas in East London tend to exhibit relatively low house prices.

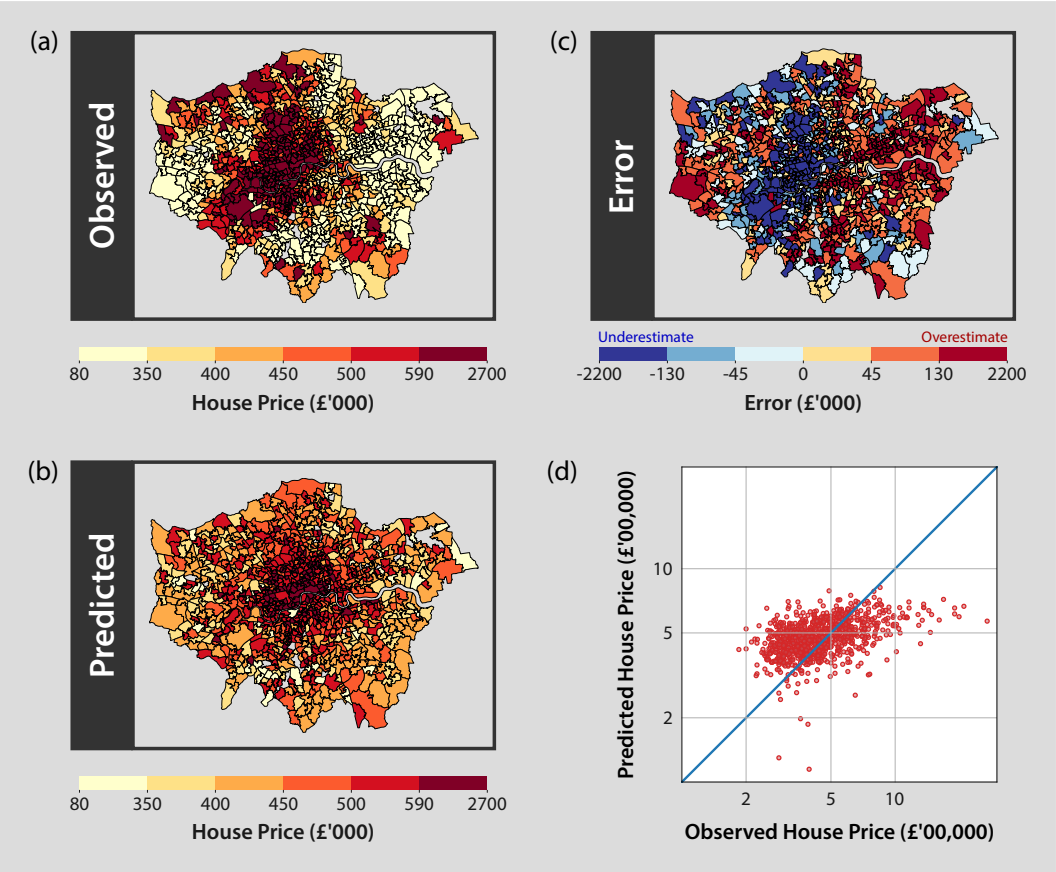

Figure 4.2: Evaluation of estimates of London house prices during 2015 and 2016 from Geograph images. **(a)** Map of median house prices for each MSOA ('Observed'). Darker areas correspond to higher house prices. **(b)** Map of median house prices for each MSOA estimated from Geograph images ('Predicted'). Darker areas again correspond to higher house prices. We generate estimates using the methodology outlined in Fig. 4.1. A regression analysis shows that the elastic net model has only limited success in estimating house prices ($R^2 = 0.18$, $N = 953$, $p < 0.001$). The model captures higher median house prices in the centre of London but tends to overestimate areas outside of the centre. **(c)** Map of the errors in median house price estimates for each MSOA ('Error'). Blue denotes an underestimate and red denotes an overestimate. Median house prices in the east of London tend to be overestimated, whereas house prices in the west of London towards the centre tend to be underestimated. *(continues on the following page)*

Figure 4.2: *(continues from previous page)* **(d)** Log-log plot of observed and predicted median house prices. The red dots represent MSOAs of London and the blue line represents where perfect predictions should fall. For areas with observed median house prices over £1,000,000, we find that the estimates tend to be too low. Some areas with observed median house prices under £500,000 are also strongly underestimated. Overall, these results suggest that the features extracted from Geograph images from 2015 and 2016 do not provide enough information to make particularly accurate estimates of local house prices. For all maps, quantile breaks are implemented with six bins. Contains National Statistics data © Crown copyright and database right 2011. Contains Ordnance Survey data © Crown copyright and database right 2011. Contains Geograph data © Geograph Project Limited and licensed for reuse under creativecommons.org/licenses/by-sa/2.0.

We investigate whether these differences in median house price across London can be inferred from Geograph image data. When fitting the elastic net, we find the optimal value of *l1_ratio* to be 0.9999 with an α of 2.56 (to 2 decimal places). A further regression analysis comparing the observed median house price for each MSOA with the estimates produced by the elastic net using a leave-one-out analysis indicates that the model has only limited success in estimating house prices ($R^2 = 0.18$, $N = 953$, $p < 0.001$). This is possibly because the median number of images in each MSOA is lower for this dataset of Geograph photos taken in 2015 and 2016 compared to the Google Street View dataset. Hence, there is a lower amount of information for each MSOA and a lack of representation in this analysis. The lack of representation is the most likely cause for low model performance. In contrast, the comparable Google Street View analysis (Section 3.3.1) exhibits much higher performance ($R^2 = 0.56$, $N = 983$, $p < 0.001$).

Furthermore, from visual inspection of the predicted median house price map in Fig. 4.2b, we see that the model mostly captures the tendency for higher house prices in the centre of the city and the north-west, but few nuances beyond this. This is in contrast to the results from the corresponding Google Street View analysis, where more detailed patterns are captured, such as lower median house prices in the west of London.

Figure 4.2c presents the prediction errors of the model, where blue denotes an underestimate and red denotes an overestimate of median house price. Mapping the prediction errors reveals that median house prices in lower house price areas in the east of London and some areas in the west of London are generally overestimated.

Conversely, median house prices in higher house price areas in the west of London, closest to the centre, are typically underestimated. This may be because the model is struggling to identify anything other than the distinctive city centre architecture of the centre of London and is therefore drifting back to a mean estimate outside of this central area.

The model achieves a mean absolute error of £136,500 and a mean absolute percentage error of a considerable 29.18% (both to 4 significant figures). In comparison, the analogous Google Street View analysis achieves a mean absolute error is £97,710 and a mean absolute percentage error is 20.24% (both to 4 significant figures). Note that this model is still an improvement over the baseline mean model, which achieves a mean absolute error of £152,800 and mean absolute percentage errors of 32.13%. The scatter plot of observed against predicted median house prices further illustrates the difficulty the model experiences in capturing the variability in house prices (Fig. 4.2d). The red dots represent MSOAs and the blue line shows where perfect predictions should fall. The model tends to produce underestimates for areas where higher prices are observed, and overestimates for areas where lower prices are observed. In other words, the model does not sufficiently diverge from the mean, leading to these large prediction errors.

Such large errors suggest that insufficient information is available from the features extracted from the Geograph images to make accurate estimates of local house prices. Again, this could be due to the relatively low number of images in our dataset. The median number of images for each MSOA is 21, which is fairly low compared to the median of 496 images in each MSOA for the Google Street View dataset. Furthermore, 122 of 953 MSOAs have fewer than five images attached to them and many of the areas simply do not have enough images to fully represent them. Therefore, the features extracted would not provide enough information about the areas, leading to large prediction errors.

Alternatively, visual information contained in Geograph imagery may not be useful in house price estimation. Geograph images contain information on the local neighbourhoods of houses, but perhaps such visuals are not as indicative of house prices as other street-level imagery. The reduced number of visual features extracted from the images may also impact the model. Although we extract more Places features, we do not extract any SUN attributes like we did with the Google Street View images. Their inclusion could make it easier for the model to estimate local house prices accurately.

Figure 4.3 depicts the top 50 positive and negative elastic net coefficients from the model. For each coefficient, the absolute value indicates how big an effect

the feature has on house price estimates, and the sign indicates whether the effect is positive or negative.

Since the model has some success in estimating median house prices, we would expect positive model coefficients to represent visual features of areas with higher median house prices and negative model coefficients to represent visual features of areas with lower median house prices. From Fig. 4.3, we see that positive coefficients include 'mansion', 'embassy', and 'courtyard', and negative coefficients include 'industrial area'. Mansions, for example, are luxurious houses and are expensive by definition, so are indeed likely to be found in areas with higher house prices. On the other hand, industrial areas are usually loud and noisy and such attributes are undesirable near residential areas, such that lower house prices in such areas might be expected.

The model also includes a greater proportion of features than the corresponding Google Street View model. Of the 205 outdoor features available to the Geograph model from Places365, 191 are included (93%; Fig. 4.3). However, of the 239 outdoor features and attributes available to the Google model from Places205 and SUN respectively, only 68 are included (28%; Fig. 3.5).

This may be partially because the type of images available from Geograph resemble those used to train Places CNN more closely than images from Google Street View. The Places365-standard dataset that the Places365 CNN was trained on includes many scenes of differing types (Zhou et al., 2018), as does the Places205-standard dataset that the Places205 CNN was trained on (Zhou et al., 2014). Likewise, Geograph imagery includes a variety of elements from the environment, whereas Google Street View imagery is often only of streets and facades. Therefore, the CNN may identify shapes and patterns more easily in Geograph imagery, and the elastic net may consequently deem more features important. Places365 is also an updated version of Places205, so an improvement in scene detection is to be expected.

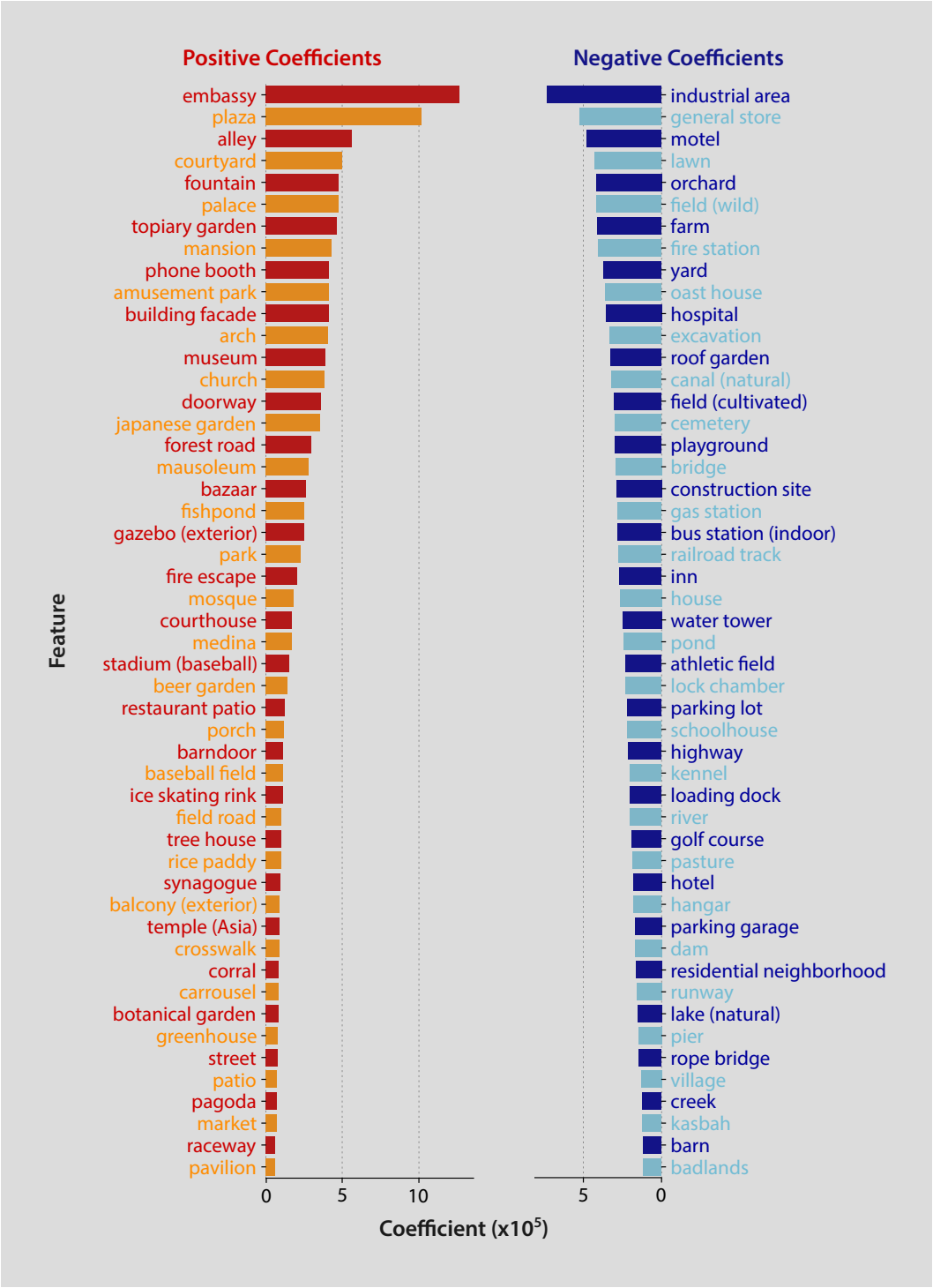

Figure 4.3: Top 50 elastic net coefficients for estimates of London house prices during 2015 and 2016 from Geograph images. *(continues on the following page)*

Figure 4.3: *(continues from previous page)* Positive coefficients relate to visual features that are associated with higher median house prices, and negative coefficients relate to visual features that are associated with lower median house prices. As in the Google Street View analysis, we train an elastic net model using the methodology outlined in Fig. 4.1 and depict the top 50 resulting positive and negative coefficients from the elastic net model. We find that the model has learnt to associate at least some features in images with higher or lower median house prices in a fairly intuitive way. For example, the features 'embassy', 'courtyard' and 'arch' are linked with areas with higher median house prices, and the feature 'industrial area' is linked with lower median house prices. A courtyard for instance requires a large amount of space, so would be present in larger properties. Larger properties are often more expensive than smaller properties, so are associated with higher house prices. Conversely, houses in industrial areas, or on streets that resemble industrial areas, might be expected to have lower house prices. There are also a few unexpected features that appear in the top 50 positive and negative coefficients. A 'desert' is a feature that would not be expected in London due to an incompatible climate. We discuss this further in Section 4.3.1.

Another reason why more features are included in the Geograph model than the Google Street View model might be that the Geograph model has had greater difficulty in distinguishing relevant visual characteristics from irrelevant visual characteristics, as reflected by the poorer performance of the model. Moreover, some coefficients from the model are unexpected. For example, a 'desert' would not be expected in London and neither would a 'bamboo forest', due to the climate in the UK being too cold. It seems that Places365 identifies patterns and shapes resembling these features in some Geograph images of London. To understand why Places365 attaches labels to the images in the manner it does, we explore a selection of top scoring images for positive and negative coefficients. Figure 4.4 displays three images that generate very high feature values for three strongly positive and three strongly negative coefficients. The feature values underneath each image are those extracted from the CNN, as described in the CNN methodology (Section 4.2.3).

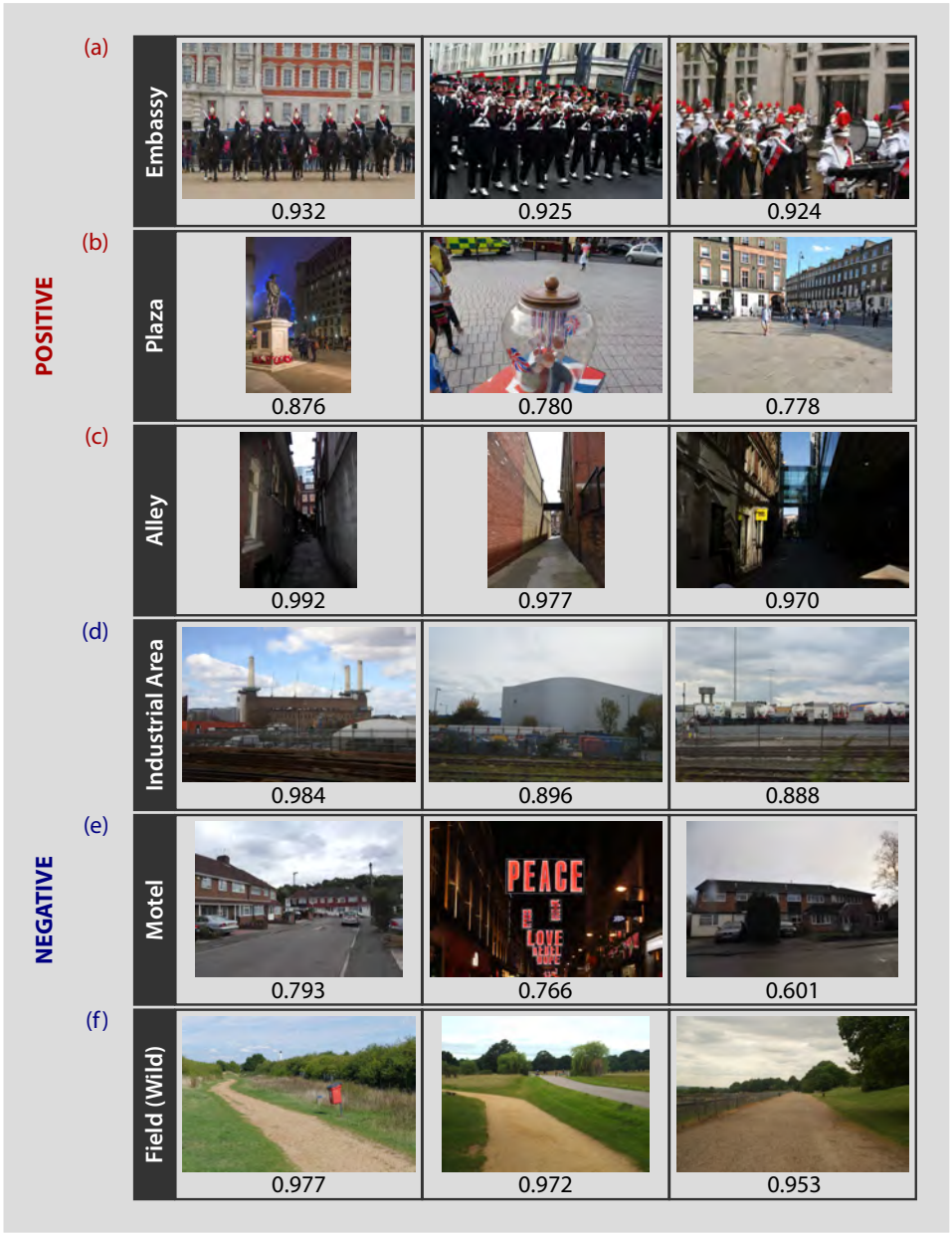

Figure 4.4: Top Geograph images of London during 2015 and 2016 for top positive and negative features. We obtain Geograph images and extract corresponding feature vectors using the methodology outlined in Fig. 4.1. *(continues on the following page)*

Figure 4.4: *(continues from previous page)* From the feature vectors, we identify three images with high feature values for the features that are most positively and negatively related to median house prices in London according to the elastic net model. This should help us understand what the CNN sees for each feature. Note that for each image, the associated feature is its highest-scoring Places label. **(a)** The images with a high score for the feature 'embassy' often contain a group of people dressed as a marching band, and always contain buildings that look historical. Although the images do not contain embassies, the pictured scenes do resemble scenes that might be seen outside embassies. All photos © Robert Lamb. **(b)** The images with a high score for the feature 'plaza' contain plazas. The label accurately reflects the content of the images. Photos from left to right © Christine Matthews, Robert Lamb, and Martin Bodman. **(c)** The images with a high score for the feature 'alley' contain alleys. Once again, the label is representative of the content in the images. Photos from left to right © Basher Eyre, Basher Eyre, and Robert Lamb. **(d)** The images with a high score for the feature 'industrial area' contain an industrial building surrounded by a fence. Therefore, the label for the images is appropriate. Photos from left to right © Mike Pennington, N Chadwick, and N Chadwick. **(e)** The images with a high score for the feature 'motel' do not appear to contain motels, but do contain visual elements that are reminiscent of a motel. For example, some images contain long groups of buildings, such as terraced housing, with red materials, and the middle image is focused on neon lighting, which is often found at motels. Photos from left to right © James Emmans, Robert Lamb, and David Howard. **(f)** The images with a high score for the feature 'field (wild)' contain sandy paths along fields. It is not clear that the fields are particularly wild, but the label is certainly reflective of the essence of the images. Photos from left to right © Alan Hunt, Malc McDonald, and Robert Lamb. Generally, the images that have a high score for a given feature have similar characteristics to each other, and are either labelled correctly or have labels with a clear relationship to the image content. All Geograph photos are licensed for reuse under creativecommons.org/licenses/by-sa/2.0.

Of the six example features portrayed in Fig. 4.4, four appear to be allocated to imagery appropriately ('plaza', 'alley', 'industrial area', and 'field (wild)'). For 'embassy' and 'motel', it is less clear that the label is appropriate. However, it is very clear that there are similarities between images with the same label (e.g. 'embassy') or visual similarities between the images and the concept in the label (e.g. 'motel').

To help us better understand how the elastic net operates, we examine which areas in London exhibit high mean feature values for positive coefficient features, and which areas in London exhibit high mean feature values for negative coefficients

features. This is an extended version of our examination of the features extracted from Google Street View imagery by Places205 (Fig. 3.3), but for features extracted from Geograph imagery using Places365. For highly positive coefficients, we would expect the distribution of features across London to coincide with areas of higher median house prices.

To investigate, we use the features from Fig. 4.4 as examples. Figure 4.5 shows a map of median house prices across London (Fig. 4.5a, as also shown in Fig. 4.2a), as well as maps of the prevalence of features across London areas for three top positive coefficients from the Geograph model, 'embassy' (Fig. 4.5b), 'plaza' (Fig. 4.5c), and 'alley' (Fig. 4.5d).

Visual inspection of the feature value distribution maps confirms that positive coefficient features are more likely to have higher values in areas with higher median house prices, especially in the centre of London. Figure 4.5b reveals that the areas with visual characteristics that the CNN attaches the label 'embassy' to are likely to be in the centre of London. The same holds for 'plaza' (Fig. 4.5c) and 'alley' (Fig. 4.5d). As plazas, alleys surrounded by tall buildings and scenes like those captured in the example 'embassy' images (Fig. 4.4a) are more frequently found in city centres than residential areas, this accords with intuition.

For highly negative coefficients, we would expect the distribution of features across London to coincide with areas of lower median house prices, inversely to Fig. 4.5. Again, we investigate using the features from Fig. 4.4 as examples. Figure 4.6 shows a map of median house prices across London (Fig. 4.6a), in addition to maps of the prevalence of features across London areas for the three top negative coefficients from the Geograph model, 'industrial area' (Fig. 4.6b), 'motel' (Fig. 4.6c), and 'field (wild)' (Fig. 4.6d).

Examination of all three feature value distribution maps confirms that negative coefficient features tend to have lower values in areas with higher median house prices, and vice versa. Both industrial areas and fields are found in areas with a large amount of space, which are often outside of the city centre. Additionally, scenes similar to those in the example 'motel' tend to be found in residential areas, rather than city centres. Therefore, these maps also accord with intuition.

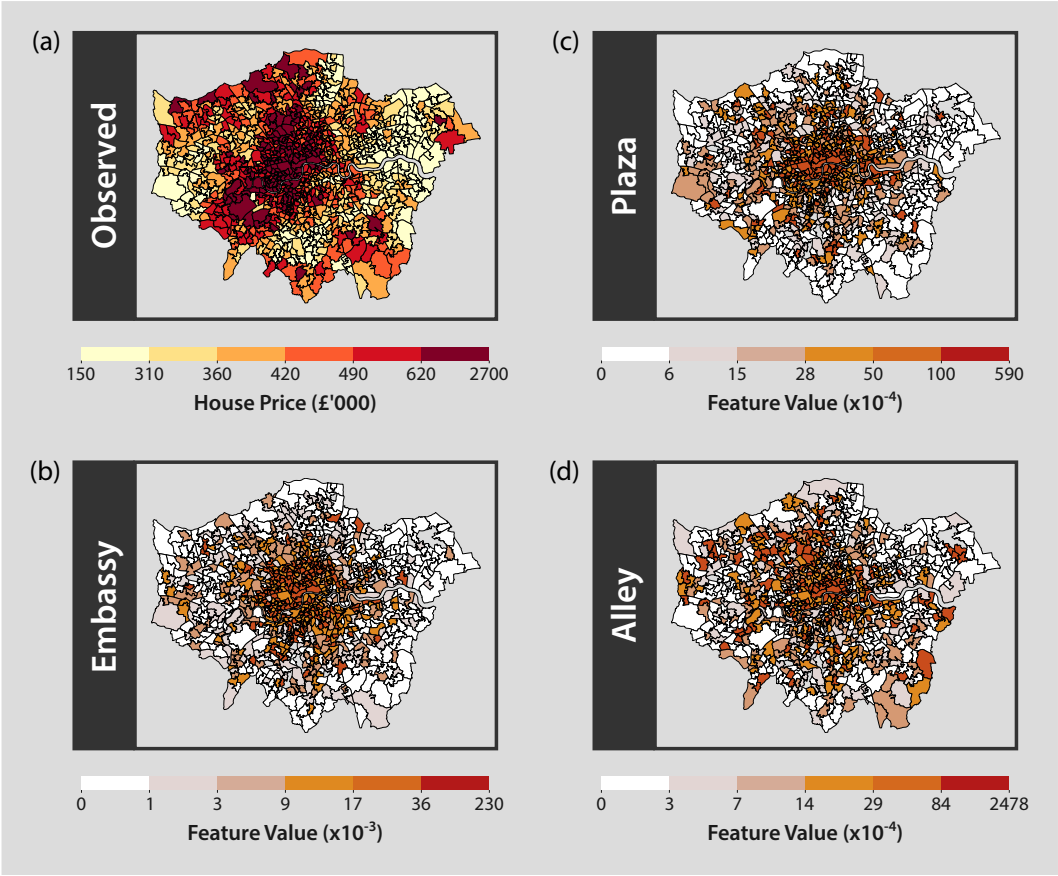

Figure 4.5: Prevalence of high positive coefficient features in London from Geograph images taken in 2015 and 2016. (a) Map of median house prices for each MSOA ('Observed'), as also shown in Fig. 4.2a. Darker areas relate to higher house prices. **(b)** Map of mean feature values for the feature 'embassy'. The feature values are extracted via the methodology outlined in Fig. 4.1. Since positive coefficients relate to visual features that are associated with higher median house prices, we would expect the features to have higher feature values in areas that correspond to higher median house prices. For all of the following maps, darker areas correspond to higher feature values, and lighter areas correspond to lower feature values. *(continues on the following page)*

Figure 4.5: *(continues from previous page)* Maps are again presented at MSOA granularity. **(c)** Map of mean feature values for the feature 'plaza'. **(d)** Map of mean feature values for the feature 'alley'. Scenes similar to those that the CNN labels as 'embassy' are often found in city centres, and less often in residential areas, as are plazas and alleys surrounded by tall buildings. As such, the feature value distribution maps follow intuition, since areas in the centre of London with higher median house prices tend to coincide with higher feature values for these features. Example images associated with high feature values for these features by the CNN can be found in Figs. 4.4a, 4.4b, and 4.4c respectively. For all maps, quantile breaks are implemented with six bins.

4.3.2 Estimating local house prices in 2015 and 2016 from the complete dataset of Geograph images

A key issue with restricting our analysis to Geograph images from 2015 and 2016 only is that we do not have a very high median number of images for each MSOA in London. Thus, to understand whether differences in median house prices can be inferred more accurately with more Geograph image data, we also estimate median house prices across London from the complete set of Geograph images. As described in Section 4.2.2, the full dataset contains 328,014 images uploaded to Geograph between 2005 and 2019, with a median of 151 images per MSOA, where all 983 MSOAs are represented. Recall that for our initial Geograph analysis (Section 4.3.1), we analyse 51,918 images with a median of 21 images per MSOA.

An important point to note is that the images in the complete dataset include images beyond 2016 up to 2019, whereas the house price data we consider are from 2015 and 2016 to maintain parity with the corresponding Google Street View analysis. This means that, in essence, 46,962 images from the future (14% of the dataset) are being used to produce the estimates of house prices. Nevertheless, this analysis should allow us to get some initial insight into whether having more Geograph images available would improve the performance of this method.

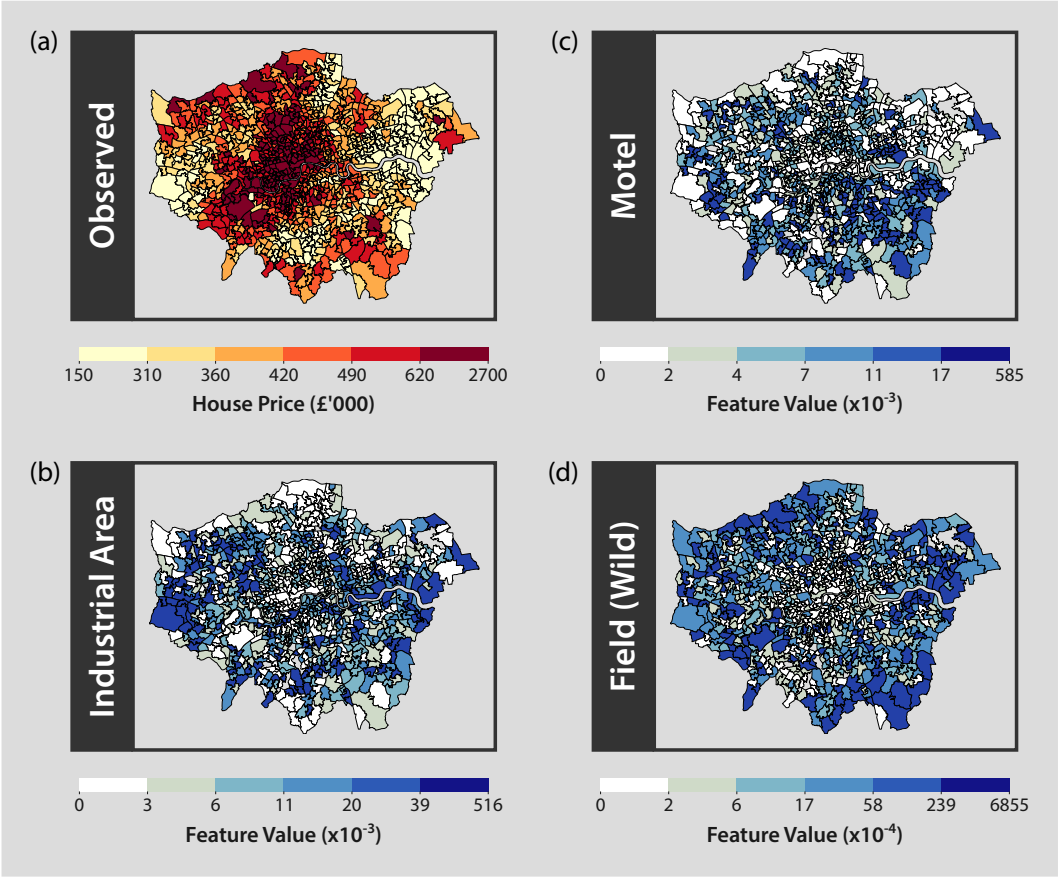

Figure 4.6: Prevalence of features with high negative coefficients in London from Geograph images taken in 2015 and 2016. **(a)** Map of median house prices for each MSOA ('Observed'), as also shown in Fig. 4.2a. Darker areas relate to higher house prices. **(b)** Map of mean feature values for the feature 'industrial area'. The feature values are extracted via the methodology outlined in Fig. 4.1. Since negative coefficients relate to visual features that are associated with lower median house prices, the features should have lower feature values in areas that correspond to higher median house prices. For all of the following maps, darker areas correspond to higher feature values, and lighter areas correspond to lower feature values. Maps are again presented at MSOA granularity. **(c)** Map of mean feature values for the feature 'motel'. **(d)** Map of mean feature values for the feature 'field (wild)'. Scenes resembling elements of the example 'motel' are often found in residential areas rather than city centres, and both industrial areas and fields require a large expanse of space that tends to be outside of city centres. *(continues on the following page)*

Figure 4.6: *(continues from previous page)* Here we see particularly that the features have lower values in the centre of London, where house prices tend to be higher. We also observe higher values in some specific areas with particularly low house prices: for example, 'industrial area' has particularly high values towards the far west and far east of London. Example images associated with high feature values for these features by the CNN can be found in Figs. 4.4d, 4.4e, and 4.4f respectively. For all maps, quantile breaks are implemented with six bins. Contains National Statistics data © Crown copyright and database right 2011. Contains Ordnance Survey data © Crown copyright and database right 2011. Geograph data © Geograph Project Limited and licensed for reuse under creativecommons.org/licenses/by-sa/2.0.

We present the observed and predicted median house prices from 2015 to 2016 in Fig. 4.7a and Fig. 4.7b. Note that the scales are slightly different to Fig. 4.2a and Fig. 4.2b, due to the fact that the estimates are different and we use quantile breaks for the maps. During the process of fitting the elastic net, we find the optimal value of *l1_ratio* to be 0.99999 with an α of 0.57 (to 2 decimal places). Again, we conduct a regression analysis comparing the observed median house price for each MSOA with the estimates produced by the elastic net using leave-one-out analysis. We find that this model is a considerable improvement on the model using Geograph photos restricted to 2015 and 2016 only ($R^2 = 0.18$, $N = 953$, $p < 0.001$) in Section 4.3.1, as this models results in a considerably larger coefficient of determination ($R^2 = 0.44$, $N = 983$, $p < 0.001$). Figure 4.7b also highlights the substantial improvement in estimating median house prices visually, as more details of the outskirts of London are identified in comparison to the analysis restricted to photos from 2015 and 2016 only (Fig. 4.2). For instance, the areas with lower median house prices in the east and west of London can be spotted easily.

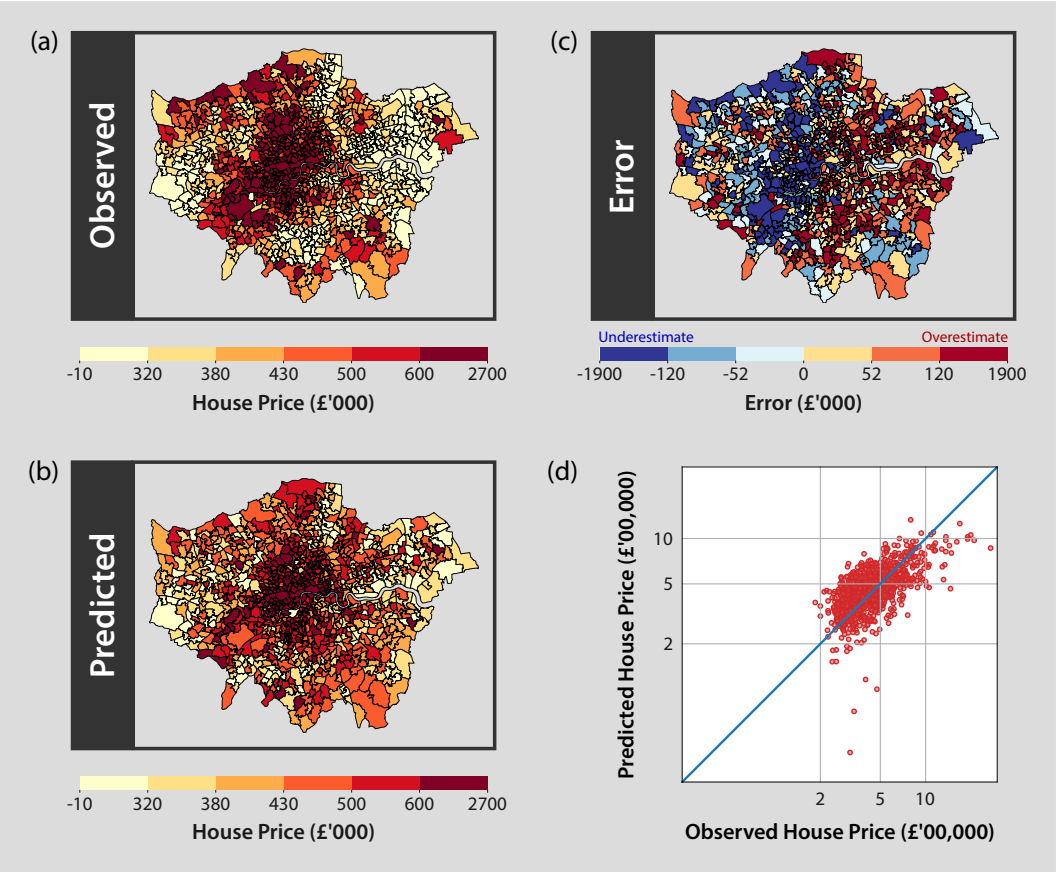

Figure 4.7: Evaluation of estimates of London house prices during 2015 and 2016 from all available Geograph images. **(a)** Map of median house prices for each MSOA ('Observed'). **(b)** Map of median house prices for each MSOA estimated from Geograph images ('Predicted'). Darker areas correspond to higher house prices in both maps. Again, we generate estimates using the methodology outlined in Fig. 4.1. We find that the performance of the model improves considerably in comparison to the analysis that considered Geograph images from 2015 and 2016 alone ($R^2 = 0.44$, $N = 953$, $p < 0.001$). In particular, the model appears to perform better at identifying areas of higher median house prices in south-west, north-west, and south-east of London. **(c)** Map of the errors in median house price estimates for each MSOA ('Error'). Blue denotes an underestimate and red denotes an overestimate. Median house prices in the east of London tend to be overestimated, whereas house prices in the south-west of London towards the centre tend to be underestimated, but the size of the errors is smaller than those from the Geograph analysis restricted to images from 2015 and 2016 only (Fig. 4.2). *(continues on the following page)*

Figure 4.7: *(continues from previous page)* **(d)** Log-log plot of observed and predicted median house prices. The red dots represent MSOAs of London and the blue line represents where perfect predictions should fall. Compared to the Geograph analysis using only 2015 and 2016 photos, the improvement in performance is clear and the difference in R^2 is reflected in the plot. We find that median house price estimates over £1,000,000 still tend to be low, but not as low as those from the aforementioned restricted Geograph analysis. For all maps, quantile breaks are implemented with six bins. Contains National Statistics data © Crown copyright and database right 2011. Contains Ordnance Survey data © Crown copyright and database right 2011. Contains Geograph data © Geograph Project Limited and licensed for reuse under creativecommons.org/licenses/by-sa/2.0.

The prediction errors of the model are shown in Fig. 4.7c. Median house prices in areas in the east of London and some areas in the west of London are still overestimated. Furthermore, median house prices in the north-west of London and south-west of London are generally underestimated. Despite this, the errors are notably smaller than those from the restricted Geograph analysis overall. The mean absolute error is £115,800 and the mean absolute percentage error is 24.75% (both to 4 significant figures). In contrast, the mean absolute error was £136,500 and the mean absolute percentage error was 29.18% for the Geograph analysis considering photos from 2015 and 2016 alone (both to 4 significant figures). However, for the analogous Google Street View analysis, the mean absolute error was £97,710 and the mean absolute percentage error was 20.24% (both to 4 significant figures). As such, the model using Google Street View photos outperforms both of the Geograph models.

Figure 4.7d displays the scatter plot of observed median house prices against predicted median house prices, and the improvement in median house price estimation in comparison to the analysis considering Geograph photos from 2015 and 2016 alone is highlighted once more. In summary, our results provide some initial evidence that increasing the number of images in a house price estimation model leads to improved performance. However, neither model using Geograph photos performs as well as the corresponding Google Street View model. This may be due to a greater number of Google Street View photos being available for analysis, as we used 518,808 Google Street View images and the median number of photos in each MSOA was 496. In comparison, we use 328,014 Geograph images for the analysis using all Geograph images, and the median number of photos in each MSOA is 151.

Figure 4.8 portrays the top 50 elastic net coefficients from the model. The coefficients are very similar to those extracted from the Geograph analysis restricted to images from 2015 and 2016 only, but occur in a different order. In other words, the relative size of the impact of features is dissimilar to the restricted Geograph analysis but similar features are positively and negatively associated with median house prices. As an example, 'courtyard' is the fourth largest positive coefficient in the restricted Geograph analysis (Fig. 4.3), but the top coefficient when considering all Geograph images in the analysis. Another example is the feature 'lawn', which is the fourth largest negative coefficient when only considering Geograph photos from 2015 and 2016 in the analysis, but the 44th largest negative coefficient when considering all Geograph images in the analysis.

Certain features are no longer present in the top 50 coefficients at all, such as 'mansion' and 'kennel'. However, a large proportion of available features are present overall. Of the 205 outdoor features available to the model, 187 are included (91%). It seems that using more images may result in fewer coefficients in the model, albeit only four fewer coefficients in comparison to the restricted Geograph model. The corresponding Google Street View model includes far fewer coefficients than both Geograph models, but also performs more accurately. This could be a result of having a greater median number of photos per MSOA. If there are more images to gain information from, the model can draw upon a more comprehensive visual landscape of London, so the most relevant visual features can be identified more precisely.

4.3.3 Estimating changes in local house price rank from 2005 and 2006 to 2015 and 2016 using Geograph images taken in 2015 and 2016

Do areas that have seen the greatest relative increases in median house prices have distinctive visual characteristics? As in Section 3.3.2, we calculate a change in median house price rank for each MSOA in London by using the methodology described in Section 4.2.5. We begin by only considering images taken and uploaded to Geograph in 2015 and 2016, and again analyse changes in house prices between 2005–2006 and 2015–2016. The changes in median house price rank range from -424.5 and 522.

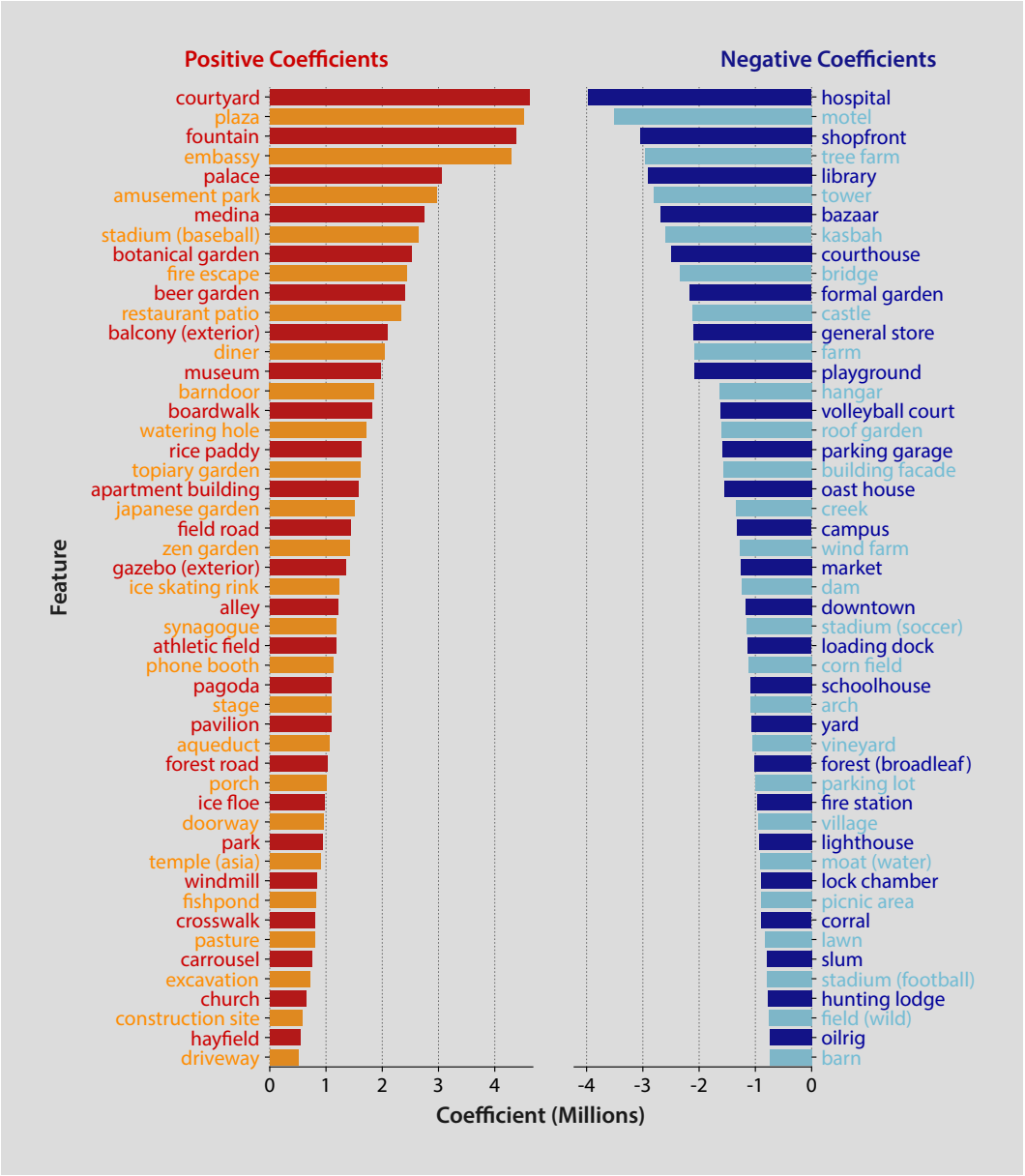

Figure 4.8: Top 50 elastic net coefficients for estimates of London house prices during 2015 and 2016 from all Geograph images. Positive coefficients reflect features that are associated with higher median house prices, and negative coefficients reflect features that are associated with lower median house prices. We train an elastic net model using the methodology outlined in Fig. 4.1 and depict the resulting top 50 positive and negative coefficients from the elastic net model. *(continues on the following page)*

Figure 4.8: *(continues from previous page)* The coefficients are similar to those extracted from the Geograph model considering photos from 2015 and 2016 only (Fig. 4.3), but occur in a different order. For instance, 'courtyard' is the fourth largest positive coefficient in the aforementioned Geograph analysis but the top coefficient when considering all Geograph images in the analysis.

The observed changes in median house price rank are shown in Fig. 4.9a. The estimated changes in median house price rank are generated via the methodology shown in Fig. 4.1 and depicted in Fig. 4.9b. The observed change in rank is the actual change in rank for each MSOA, and the predicted change in rank is the estimate of change in rank from Geograph photos taken in 2015 and 2016. For both maps, dark blue indicates a decrease in ranked price and dark red indicates an increase in ranked price between 2005–2006 and 2015–2016. When fitting the elastic net model, we find the optimal value of *l1_ratio* in this elastic net model to be 0.999999 with an α of 0.14 (to 2 significant figures).

Highlighted in the observed change in rank map (Fig. 4.9a) is the fact that areas surrounding the centre of London tend to decrease in ranked price, and areas in the centre of London, particularly towards the east, tend to increase in ranked price, as in the corresponding Google Street View analysis (Section 3.3.2). Comparing the maps of observed and predicted changes demonstrates that the elastic net model does capture some of this pattern. However, conducting a regression analysis after applying leave-one-out cross-validation underlines that the success of the model in estimating the true changes in rank is limited ($R^2 = 0.11$, $N = 953$, $p < 0.001$). This may again be due to the relatively low median number of Geograph photos per MSOA when considering imagery from 2015 and 2016 only.

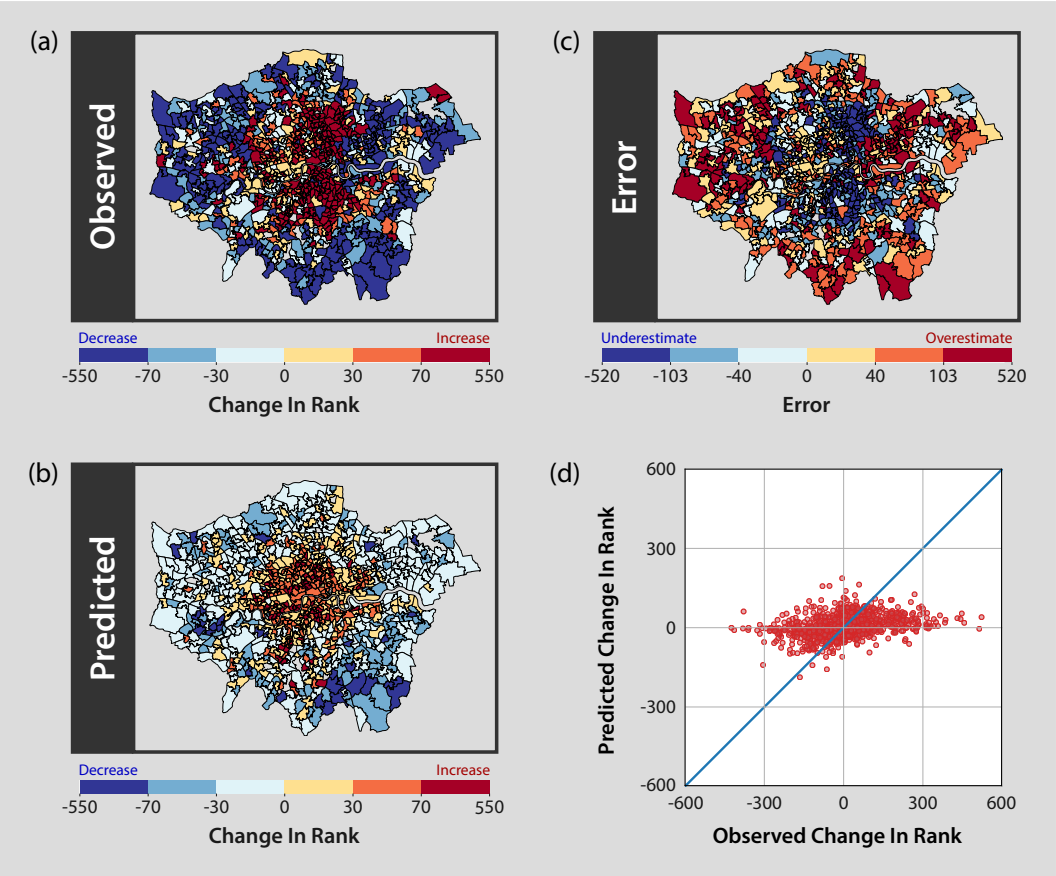

Figure 4.9: Inferring changes in house prices in London using Geograph images from 2015 and 2016. (a) Map of change in median house prices between 2005–2006 and 2015–2016 for each MSOA ('Observed'). We use the methodology described in Section 4.2.5 to obtain a change in rank for each MSOA in London. As with the corresponding Google Street View analysis (Fig. 3.9), blue denotes a decrease in ranked price and red denotes an increase in ranked price over the decade. **(b)** Map of change in median house prices for each MSOA estimated from Geograph images ('Predicted'). We generate estimates using the methodology outlined in Fig. 4.1. Again, blue denotes a decrease in ranked price and red denotes an increase in ranked price over the decade. A regression analysis shows that the model has very limited success in capturing the observed patterns in house price changes ($R^2 = 0.11$, $N = 953$, $p < 0.001$). *(continues on the following page)*

Figure 4.9: *(continues from previous page)* Visual inspection of the maps of observed and predicted changes in the median house price rank of local areas reveals that the elastic net model successfully infers greater relative increases in house price in the centre of London but generally underestimates the size of this effect. It does not capture many nuances in the spatial pattern beyond this. **(c)** Map of the errors in estimates of change in median house price rank for each MSOA ('Error'). Blue denotes an underestimate and red denotes an overestimate. The map again highlights that the model has underestimated the rise in ranks of areas in several parts of London, in particular towards the east, and has similarly not captured the full size of the fall in ranks in parts of outer London. **(d)** Scatter plot of observed and predicted change in rank. The red dots represent MSOAs of London and the blue line represents where perfect predictions should fall. Again, we see that the model has not captured a lot of the variation in changes in ranked house price. For all maps, quantile breaks are implemented with six bins.

Figure 4.9b demonstrates that the model correctly infers that areas in the outskirts of London decrease in ranked price and areas in the centre of London increase in ranked price, but the size of this effect is often underestimated. The model is unable to capture any further nuances in the spatial pattern. Visually, the prediction errors in Fig. 4.9c are almost an inverse of Fig. 4.9a; i.e. they show that the notable climb in rank in parts of the centre of London towards the east is underestimated, and highlight the inability of the model to capture much of the fall in rank in many parts of the outskirts of London. The model is also unable to capture a lot of the variation in the ranks, as is depicted in the scatter plot of observed versus predicted change in rank (Fig. 4.9d), and the change in rank is always estimated to fall between -280 and 280. This indicates that the information available to the model from Geograph images from 2015 and 2016 is not enough to make accurate estimates of local changes in house prices.

Figure 4.10 exhibits the elastic net coefficients from the model. Positive coefficients are connected to increases in median house price rank, and negative coefficients are connected to decreases in median house price rank. We note that the poor performance of the model appears to be accompanied by a very low number of non-zero coefficients being identified.

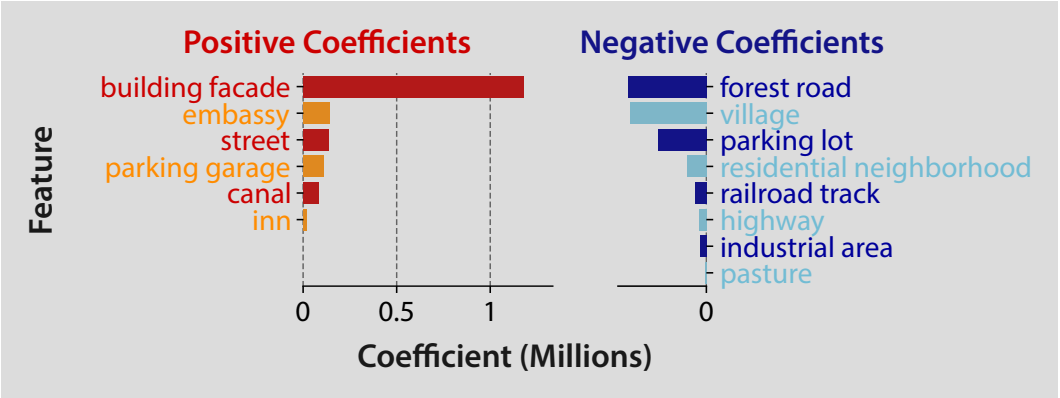

Figure 4.10: Elastic net coefficients for estimated changes in median house prices in London from Geograph images taken in 2015 and 2016. We rank London's MSOAs using the approach detailed in Section 4.2.5 and train an elastic net model to estimate the change in median house price rank between 2005–2006 and 2015–2016 using the methodology outlined in Fig. 4.1. The resulting coefficients from the elastic net model are depicted for all features where the coefficient is not zero. Positive coefficients reflect features that are associated with an increase in ranked price and negative coefficients reflect features that are associated with a decrease in ranked price. We note that far fewer non-zero coefficients are identified by this model. We also find that features associated with higher median house prices in the Geograph model considering images from 2015 and 2016 only (Fig. 4.3) are often associated with increases in median house price rank, for example 'embassy'. Similarly, features associated with lower median house prices in the restricted Geograph analysis are often associated with decreases in median house price rank, for example 'industrial area'.

We find that a number of features associated with higher and lower median house prices in the Geograph analysis restricted to photos from 2015 and 2016 alone are also associated with increases and decreases in median house price rank respectively. However, in such cases, the sizes of the effects are different. Example features include the positive coefficient 'embassy' and the negative coefficient 'industrial area'.

This result differs from what we see with the coefficients from the analogous Google Street View analysis. There, we found that coefficients that are positive when inferring median house prices are often negative when inferring the change in rank, and vice versa. It is not immediately clear what the reason for this difference is. We therefore investigate whether this pattern still holds when making use of the entire Geograph dataset.

4.3.4 Estimating changes in local house price rank from 2005 and 2006 to 2015 and 2016 using the complete dataset of Geograph images

From our previous Geograph analyses, we find that using a larger number of Geograph images to estimate median house prices across local areas of London leads to a more accurate house price estimation model. Is this also the case when estimating changes in median house prices across local areas of London? We begin our investigation by using the complete set of Geograph images once more to estimate the changes in median house price rank using the methodology described in Section 4.2.5.

Figure 4.11a depicts the observed change in rank, and Fig. 4.11b depicts the predicted change in rank from all of our Geograph images. Note that the scales are slightly different to those in the restricted Geograph analysis (Fig. 4.9) because we use quantile breaks in all maps and we need to account for different house price estimates. As we are using all Geograph images, we analyse 983 MSOAs, and again, the changes in ranked price range between -424.5 and 522. After deploying the elastic net, we find the optimal value of *l1_ratio* to be 0.99999 with an α of 0.024 (to 2 significant figures).

Based on the predicted change in rank map, an improvement in model performance is clear over the Geograph model restricted to photos from 2015 and 2016 only (Section 4.3.3). A regression analysis applying leave-one-out cross-validation supports this, although the model still has limited success in estimating true changes in rank ($R^2 = 0.18$, $N = 983$, $p < 0.001$). It still overgeneralises, estimating most increases in rank around the centre of London, but to a slightly lesser degree than the restricted Geograph model. Furthermore, the prediction errors (Fig. 4.11c) are smaller than those from the restricted Geograph analysis, and the elastic net model can capture slightly more variations in the changes. As seen in the scatter plot of observed versus predicted change in rank (Fig. 4.11d), the change in rank is now inferred to range between -300 and 300, rather than between -280 to 280 as for the restricted Geograph analysis. Overall, model performance is still poor, but including more image data in the model does improve the performance relative to the restricted Geograph analysis. Furthermore, the model performs worse than the corresponding Google Street View analysis ($R^2 = 0.27$, $N = 983$, $p < 0.001$; Section 3.3.2), but the fall in performance in comparison to the corresponding Google Street View analysis is similar to the fall in performance between the Google Street View and Geograph analyses for median house price estimation. Again, it is possible that the lower median number of images per MSOA is the cause of the discrepancy between the model performances.

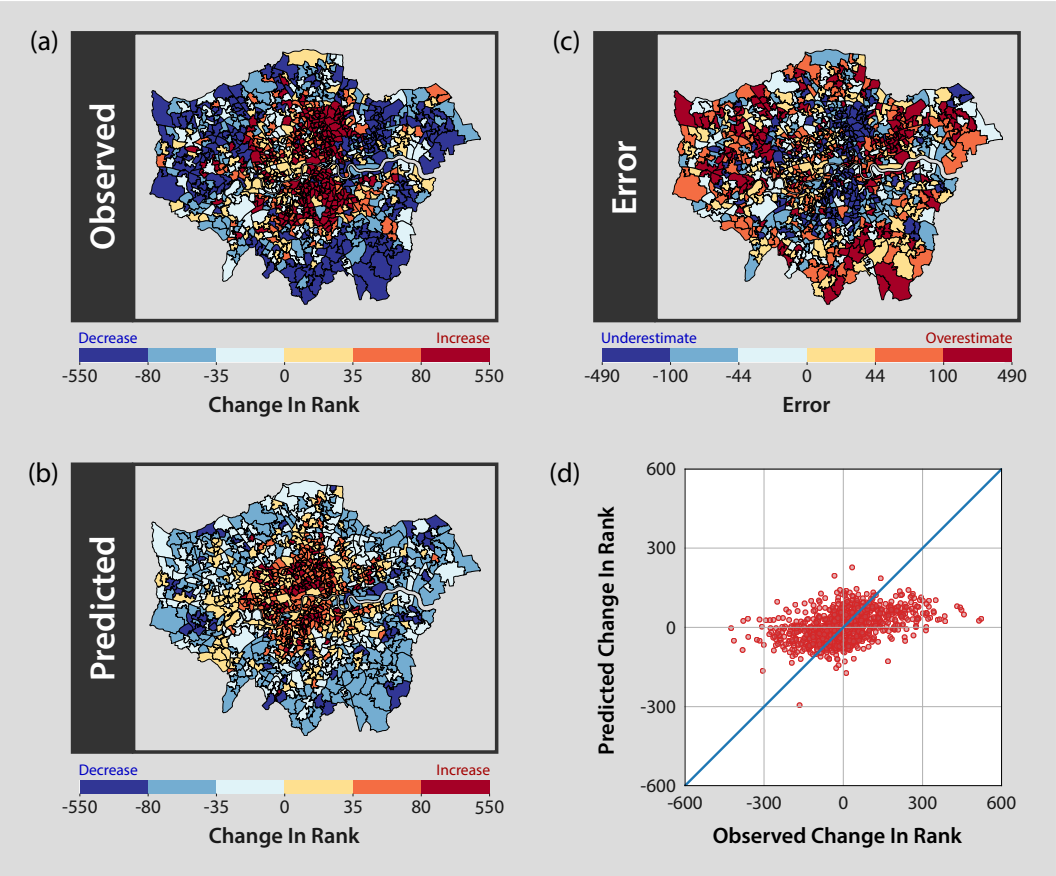

Figure 4.11: Inferring changes in house prices in London from all Geograph images. **(a)** Map of change in median house prices between 2005–2006 and 2015–2016 for each MSOA ('Observed'). Following the methodology outlined in Section 4.2.5, we obtain a change in rank for each MSOA in London. **(b)** Map of change in median house prices for each MSOA estimated from Geograph images ('Predicted'). We generate estimates using the methodology outlined in Fig. 4.1. For both maps, dark blue reflects a decrease in ranked price and dark red reflects an increase in ranked price. The model is an improvement on the corresponding Geograph analysis considering images from 2015 and 2016 only (Section 4.3.3). However, it still overgeneralises the change in rank across London, mostly inferring relative increases in price in the centre of London. *(continues on the following page)*

Figure 4.11: *(continues from previous page)* **(c)** Map of the errors in estimates of change in median house price rank for each MSOA ('Error'). Blue denotes an underestimate and red denotes an overestimate. The errors are less than those from the restricted Geograph analysis, but the areas that are underestimated and overestimated are similar. **(d)** Scatter plot of observed and predicted change in rank. The red dots represent MSOAs of London and the blue line represents where perfect predictions should fall. The change in rank estimates are better than those from the restricted Geograph analysis, now ranging between -300 and 300 and rank moving away from the mean, but model performance is still poor. A regression analysis confirms this ($R^2 = 0.18$, $N = 983$, $p < 0.001$). For all maps, quantile breaks are implemented with six bins. Contains National Statistics data © Crown copyright and database right 2011. Contains Ordnance Survey data © Crown copyright and database right 2011. Contains Geograph data © Geograph Project Limited and licensed for reuse under creativecommons.org/licenses/by-sa/2.0.

Figure 4.12 depicts the elastic net coefficients from the model. Again, fewer coefficients are non-zero when estimating the change in rank rather than median house prices. However, there are more non-zero coefficients for this model than for the restricted Geograph analysis (Fig. 4.10). It could be that increasing the number of images attached to each MSOA in London allows the elastic net to find more important connections between features and changes in rank.

As for the analogous Google Street View analysis (Section 3.3.2), we find that positive coefficients associated with higher median house prices in the Geograph model considering all photos (Section 4.3.2) are generally associated with decreases in ranked median house price here, and sometimes vice versa. We compare the coefficients between the models and find that the example features 'slum' and 'building facade' are negative coefficients when estimating median house prices but positive coefficients here. Similarly, the example features 'forest road' and 'formal garden' are positive coefficients when estimating median house prices but negative coefficients here.

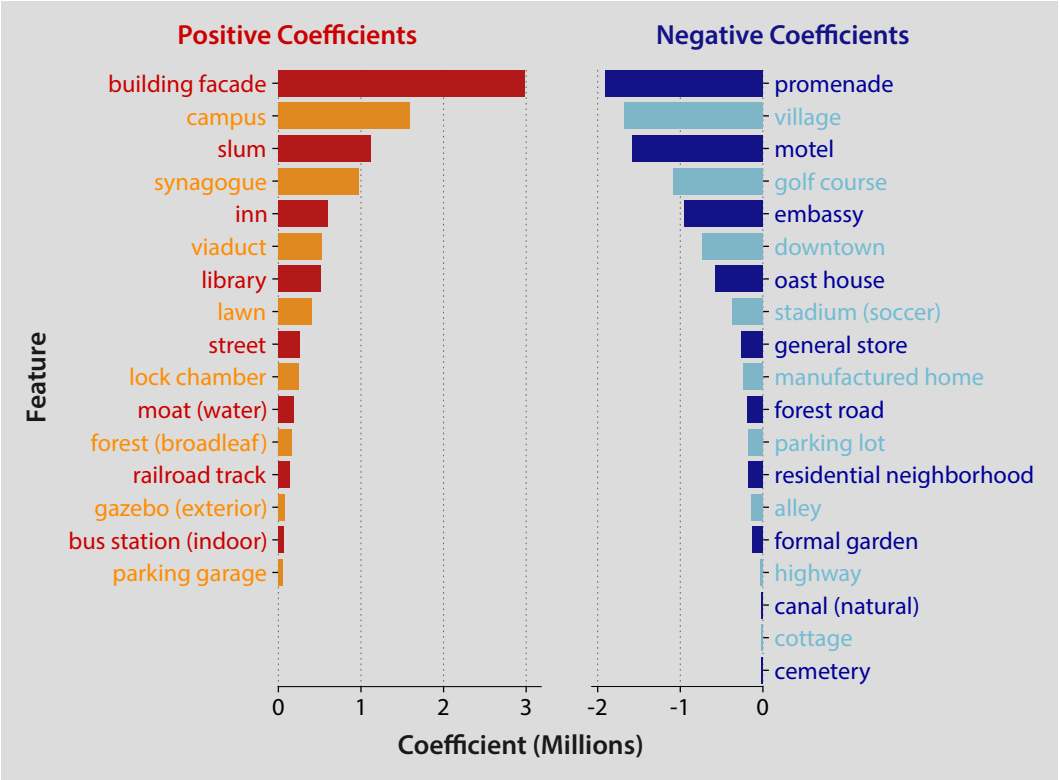

Figure 4.12: Elastic net coefficients for estimated changes in median house prices in London from all Geograph images. Using the approach detailed in Section 4.2.5, we rank the MSOAs in London and calculate the change in median house price rank between 2005–2006 and 2015–2016 for each MSOA. We then estimate these changes using the method outlined in Fig. 4.1. Positive coefficients reflect features that are associated with an increase in ranked house price (such that the area became relatively more expensive) and negative coefficients reflect features that are associated with a decrease in ranked house price. As was the case with the corresponding Google Street View analysis (Section 3.3.2), we discover that positive coefficients associated with higher median house prices are associated with decreases in ranked median house price. For example, if we compare the Geograph analyses using all photos, we find that the coefficient 'slum' is positive when estimating median house prices but negative when estimating changes in ranked median house price.

This phenomenon is similar to what we find in the analogous Google Street View Model. Again, some areas may have high median house prices but may not have climbed as highly in the ranks, whereas other areas may have low median house prices but have climbed high in the ranks. The swap in polarity of some of the coefficients would then reflect these areas. Moreover, it seems likely that the Geograph analysis using images from 2015 and 2016 only does not show this phenomenon simply due to an insufficient amount of Geograph images, since including more image data in the model, and consequently increasing the number of MSOAs that are well represented, does lead to this pattern.

4.4 Discussion

In this chapter, we try to determine whether local house prices across London can be inferred from an alternative street-level image source to Google Street View: namely, Geograph image data. Geograph images contain additional information on local neighbourhoods' off-road scenes such as parks, canals, and mountains, rather than focusing solely on streets. Our main analyses look at inferring median house prices in London from 2015 and 2016, but we also investigate the relative changes in median house prices in London over a decade. We conduct these analyses for Geograph photos taken in the same period as our house price data, specifically 2015 and 2016, as they are representative of the environment during that time. Furthermore, we look at our full dataset of Geograph images to see if we can improve the accuracy of the models.

We find that our model achieves limited success in estimating median house prices across areas in London from Geograph images taken between 2015 and 2016. The prediction errors from the model are large, and the model does not appear to gain enough information from the images to make accurate estimates. The lower volume of images per MSOA when considering Geograph photos taken between 2015 and 2016 in comparison to Google Street View photos could explain why we see such poor model performance. We also attempt to estimate local house prices from all Geograph images, without restricting the date on which the photo was taken to 2015 and 2016. This leads to an improvement in performance, but the estimates remain less accurate than those from the corresponding Google Street View model. It would therefore seem that including more image data improves the performance of the model.

Similar observations can be made for the model estimating relative changes in local house prices across London over a decade. When using Geograph photos

from 2015 and 2016, we find that the model has only limited success in inferring changes in median house price rank. We investigate whether performance improves if we consider all Geograph images, without restricting the date on which the photo was taken to 2015 and 2016. We find that this does indeed improve performance, but the model still does not perform as well as the corresponding Google Street View model. The difference in performance between the Google Street View model and the Geograph models may be due to the discrepancy in the number of photos analysed, as a greater number of Google Street View photos were analysed in comparison to the number of Geograph photos analysed.

In fact, the drop in accuracy when estimating changes in median house price rank instead of estimating median house prices is the same, regardless of whether Google Street View or Geograph images are used. Greater success could perhaps be achieved by comparing changes in images from 2005 and 2006 to 2015 and 2016. Although the Geograph dataset does have images from 2005 and 2006, it does not have a sufficient quantity of images to generate accurate estimates from. Fortunately, the amount of image data generated grows year on year, so it is likely that more Geograph data will be available in the future. Hence, a potential further step would be to estimate relative changes in local house prices at a future date with images from two well-represented time periods.

Another point to note is that the median house price changes are not normally distributed, so the models struggle with extreme values. Specifically, the distribution of changes in local house price rank is heavy-tailed, which reduces the accuracy of the model. As mentioned when discussing the Google Street View analyses (Section 3.4), a possible future step would be to change the type of model we use for estimating the changes in median house price rank to take the heavy-tails into account. As mentioned in the Google Street View chapter (Section 3.4), a neural network could be designed and used to estimate changes in ranked price.

In conclusion, we explore the use of a different large-scale street-level image dataset in conjunction with deep learning tools to infer local house prices and changes in local house prices. Would such inferences also be possible from internal images of houses? The visual information available from indoor imagery is quite distinct from Google Street View and Geograph. In Chapter 5, we investigate this question further.

Chapter 5

Inferring house prices from Zoopla images

5.1 Introduction

Both Chapter 3 and Chapter 4 demonstrate that we can successfully use information from outdoor imagery of London to infer local house prices. However, millions of images of the interior of houses are now also available. This is a result of the rise in property marketing websites, which has lead to huge swathes of data concerning the sales of properties, as estate agents and homeowners advertise their houses via these websites. When determining the value of a house, the interior characteristics undeniably contribute towards our judgement. In this chapter, we therefore investigate whether indoor imagery would be useful in a house price estimation model too.

For these analyses, we require a rich source of photographs of housing interiors, such as those available from Zoopla. The popular real estate website provides a platform for estate agents to advertise their properties, and those looking to find a place to live can then filter through the properties to find their homes. From the advertisements, Zoopla have been able to collect data regarding housing rentals and sales since their launch in 2008. Their data include substantial volumes of internal and external photos of houses, as well as house prices and rental values for properties.

Almost every house advertised on Zoopla contains multiple photos of the interiors of the house, since most houses have multiple rooms. Furthermore, Zoopla is frequently updated with properties from all across the UK, including London. Some properties are even advertised on the website multiple times over the years. Zoopla photos of the interiors of houses for sale could therefore potentially provide an alternative perspective on local house prices and their evolution.

Consequently, this chapter aims to combine the sheer volume of Zoopla images of houses in London with deep learning techniques to ask the following questions:

1. Can we automatically infer house prices from indoor Zoopla images?
2. Can we automatically identify areas which have recently seen relatively high increases in house prices from indoor Zoopla images?
3. How does the performance of the house price estimation methodology using indoor Zoopla images differ to the performance when using Geograph and Google Street View images?

5.2 Data and methods

5.2.1 House price data

In this chapter, we evaluate whether we can develop house price estimation models by analysing images of the interior of houses using the methodology outlined in Fig. 5.1. We then compare the results to those obtained from our outdoor image sources, Google Street View (Chapter 3) and Geograph (Chapter 4). For maximum comparability, we focus on property transactions from 2015 and 2016, and use the same house price data as we did for the Google Street View and Geograph analyses. Again, the house price data are skewed with a long positive tail, so we calculate the median house price for each MSOA in London.

Once more, we analyse both images and house prices at the Middle Layer Super Output Area (MSOA) level, where an MSOA contains a population of between 5,000 and 15,000 inhabitants, or between 2,000 and 6,000 households, with 983 MSOAs in London (Stokes, 2012).

5.2.2 Places CNN

As in the Geograph analysis, we extract features from our images using Places365. Each Zoopla image is processed with Places365 and for each image, we extract the values of 365 scene features into a feature vector. Since we are manipulating a large dataset, we use high-performance computing clusters managed by the Scientific Computing Research Technology Platform at the University of Warwick to process the images rapidly and efficiently by processing them in parallel. As in earlier analyses, the feature values extracted from Places365 for each image indicate the likelihood of each feature appearing in a given image. The feature values range from 0 to 1 and the feature vector sums to 1.

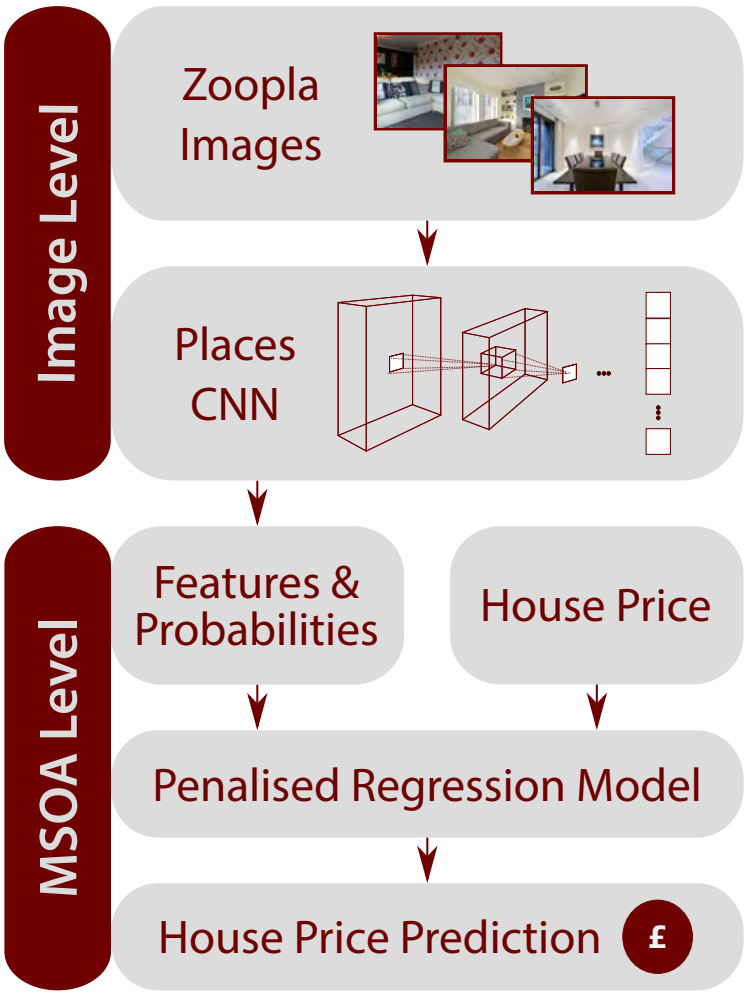

Figure 5.1: Deep learning framework. We adapt the deep learning framework from the Geograph analysis described in Chapter 4 to infer house prices from Zoopla imagery. We access all unique Zoopla images for house sales across the Greater London area from our dataset. We then allocate each Zoopla image to an MSOA and process each image using the Places365 CNN (Zhou et al., 2018). As in the Geograph analysis, given an input of an image, Places365 produces a vector of 365 feature values that indicate how likely it is that each of the Places features is present in the image. Example features include 'bedroom' and 'yard'. We then calculate the mean feature vector for each MSOA and train an elastic net to learn the relationship between the mean feature vector and the median house price of that MSOA. To test the estimates from the model, we use leave-one-out cross-validation and then compare the resulting predictions to the actual median house price for each MSOA. Contains images from Zoopla Limited, © 2018. Zoopla Limited. Economic and Social Research Council. Zoopla Property Data 2018. University of Glasgow - Urban Big Data Centre.

Unlike Google Street View and Geograph, where the images only contain outdoor visual information, both indoor and outdoor imagery is available from the Zoopla dataset. The images depict both the externals of a property, including building facades and gardens, and the internals of a property, such as a bedroom or a kitchen. Therefore, we create two subsets of the larger dataset, an indoor dataset and an outdoor dataset, so that we can initially focus specifically on the value of visual information. We describe these datasets further in Section 5.2.3.

Places scene features are all marked as 'indoor' or 'outdoor'. The images in the outdoor Zoopla dataset only contain outdoor visuals, whereas the images in the indoor dataset only contain indoor visuals. The complete Zoopla dataset contains all visual information available from our Zoopla data. To ensure that we are using appropriate features when analysing the images in the indoor dataset, we remove 205 outdoor features and are left with a feature vector of 160 indoor features. Conversely, for each image in the outdoor dataset, we remove 160 indoor features and are left with a feature vector of 205 outdoor features. We consider both indoor and outdoor features for the dataset containing both indoor and outdoor images.

In each of the three datasets, there are multiple images associated with each MSOA in London. Therefore, to obtain visual information at MSOA level for each dataset, we calculate the mean value of the feature vectors for all images in the dataset in each MSOA. More precisely, for each feature in each MSOA, we calculate the mean feature value from all images in the dataset that fall within that MSOA and build a mean feature vector for the MSOA comprising all mean feature values.

5.2.3 Zoopla images

We obtained Zoopla data from the Urban Big Data Centre. The properties included in the dataset are specifically those that go on the market. Details for each property include the first and last marketed date, whether the property is a rental or for sale, and a list of links to images of the property. The data cover advertisements of houses marketed up until June 2018, which was when we obtained the data. The data also include the dates on which the properties were first marketed, which start from January 2003.

Since we are estimating local house prices, we focus on properties advertised for sale, rather than those available for rental. We initially wanted to estimate local house prices in 2015 and 2016 using images from the same time frame, as with the Geograph analysis (Section 4.3.1). However, we find that a large volume of first and last marketed dates are missing in the Zoopla dataset. Some entries in the dataset had both dates, some had no dates, but many had either the first marketed date or

the last marketed date. We did not match a property to a date using one of these subsets of dates, as we would lose too much of the dataset. For example, 1,163,020 entries in the dataset were sales and 703,798 of those entries (61%) had image data available. Of those 703,798 entries, 154,225 (22%) of the entries had first marketed dates in 2015 and 2016, 212,335 (30%) of the entries had last marketed dates in 2015 and 2016, and only 137,412 (20%) of the entries had both first and last marketed dates in 2015 and 2016. In Chapter 4, we found that using more image data in a house price estimation model improves model performance. Therefore, we want to ensure that we have a large enough sample of properties and images to draw accurate conclusions from.

Furthermore, we find that numerous entries in the dataset are exactly the same except for the first or last marketed date and marketed price. The marketed dates usually follow on from each for a given property, or the dates are at least within a month of each other. For example, one property (ID 2466305) has an entry with a first marketed date of 5th January 2016, 16:07, and a last marketed date of 17th February 2016, 19:03. The same property then has another entry with a first marketed date of 17th February 2016, 21:08, and a last marketed date of 12th May 2016, 18:20. The price stated also differs for these two entries, as the price is initially £660,000, but increases to £685,000. Therefore, we establish that each time an advertisement is updated, it creates a new entry in the raw Zoopla dataset. Additionally, some of the properties are not in the HM Land Registry Price Paid Data (HM Land Registry, 2016), suggesting that some properties were advertised on Zoopla but did not sell. As such, many properties are represented multiple times in the dataset with different marketed prices.

There are alternative ways of tackling the issues of missing and duplicate data in the Zoopla dataset. For example, where there appear to be multiple entries in the Zoopla dataset, we could use the latest last marketed date, and where there is no last marketed date, we could use the first marketed date instead. However, this would add noise to the data and it would still be unclear which time frame the entry belongs to. For instance, if a property is on the market for more than a year and we are analysing properties last marketed in a specific year, the property may not be included in the analysis if the first marketed date is before the specified year, but the property may in fact have been last marketed in the specified year. As such, we would lose some information that could be valuable in our analysis.

In practice, if the data associated with a house advertised on Zoopla were made available as soon as the advert was released, the dates could be inferred from these streaming data. Here, where we have access to a corpus of historical Zoopla

data but less information on dates, we focus on determining whether any relationships can be found between Zoopla data and house prices, whilst largely ignoring the dates associated with the photos as we did for the full Geograph analysis in Chapter 4. We discuss this further in Section 5.4.

We start by identifying all unique properties from all time periods and access their associated photos to build our Zoopla dataset. Due to the large quantity of images available to access from Zoopla, we use the high-performance computing clusters managed by the Scientific Computing Research Technology Platform at the University of Warwick to retrieve them in a fast and efficient manner by retrieving the images in parallel. In total, we retrieve 3,727,890 unique Zoopla images relating to 478,948 advertisements. As described in the previous section, both indoor and outdoor imagery is available from the Zoopla dataset. Therefore, we can create two subsets of the larger dataset, an indoor dataset and an outdoor dataset, which will allow us to focus specifically on the value of indoor visual information.

The images are not labelled as indoor or outdoor, so we need to automatically determine their types. To automatically recognise whether an image is of the interior or exterior of a house, we take inspiration from the methodology implemented by Zhou et al. (2018) to determine the accuracy of Places365. When the Places365 CNN classifies an image, it actually determines a probability for each feature that refers to how confident it is that the feature label is the true one. If the true classification is among the top five features with the highest probabilities, then Zhou et al. (2018) deem the classification to be accurate. This method is known as the 'Top-5 accuracy' method. They also look at the 'Top-1 accuracy' to see if the feature with the highest probability is correct. For the variation of Places365 we use for our analyses, namely Places365-AlexNet, their testing reveals a Top-1 accuracy of 53.31% and a Top-5 accuracy of 82.75%. Since the Top-5 accuracy is substantially larger than the Top-1 accuracy, Zhou et al. (2018) consider Top-5 accuracy to be the more suitable measure of scene classification performance.

As such, we start by selecting the top five features extracted by Places365 that exhibit the highest feature values for each of the images in our dataset. Next, we calculate how many of the top five features are indoor features, since each feature label itself is labelled as indoor or outdoor. If three or more of the top five features are indoor features, the image is deemed an indoor image. Otherwise, it is deemed an outdoor image.

The approach results in an indoor images dataset containing 2,702,506 images and an outdoor images dataset containing 1,025,384 images. Each of these Zoopla datasets is considerably larger than the Google Street View and Geograph datasets.

The smaller outdoor Zoopla dataset already contains approximately double the number of images available in the Google Street View dataset and over three times as many images as the Geograph dataset (518,808 and 328,014 images respectively). We note however that for Google Street View, this difference in sample size does not reflect data availability but the way in which the photos were accessed.

The number of Zoopla images for each MSOA ranges from 455 to 7,139 with a median of 2,591 for the indoor dataset, and from 149 to 2,821 with a median of 973 for the outdoor dataset. For the entire dataset, the number of Zoopla images for each MSOA ranges from 619 to 9,907 with a median of 3,568. This represents a considerably larger volume of photos per MSOA in comparison to the Google Street View and Geograph datasets. For Google Street View, the number of images for each MSOA ranges from 88 to 2,164 with a median of 496, and for Geograph, the number of images ranges from 8 to 14,768 with a median of 151.

Both Google Street View and Geograph imagery contains outdoor visuals, and since we have access to outdoor Zoopla imagery, we additionally explore whether local house prices across London can be estimated well using outdoor Zoopla imagery. This allows us to compare model performance from different data sources when estimating house prices from outdoor imagery. Furthermore, in Chapter 4, we found that using a larger number of photos in a house prices estimation model improves performance. Given that we are in possession of such a large volume of Zoopla images, we may achieve better performance if we attempt to build such a model from all available images, rather than indoor or outdoor imagery alone. Thus, we also investigate how well local house prices across London can be estimated using all Zoopla images. Specifically, we estimate local house prices across London from indoor, outdoor, and both indoor and outdoor Zoopla imagery.

5.2.4 Elastic net model

As discussed earlier in Section 5.2.2, the visual characteristics of each MSOA in London are summarised by a feature vector. For visual information obtained from indoor Zoopla images, the feature vector is of length 160; for visual information obtained from outdoor Zoopla images, the feature vector is of length 205; and for visual information obtained from all Zoopla images, the feature vector is of length 365. When building each of the house price estimation models, the relevant feature vector acts as input to our regression. We use an elastic net model, as in the Google Street View and Geograph analyses, to help us manage correlations between features. To test and train our elastic net model, we use the leave-one-out approach detailed in the Google Street View analysis (Section 3.2.4).

5.2.5 Measuring change in house prices

In addition to estimating local median house prices in London, we also explore whether Zoopla imagery can shed light on how median house prices have changed in recent years. When estimating relative changes in house prices using Geograph imagery (Section 4.3.4), we achieved greater performance when using the full set of images, rather than the images restricted to the time period in question. Therefore despite our Zoopla dataset lacking dated images, we investigate how good the estimates of relative changes are when using Zoopla data. It would be better to investigate how areas had changed visually over time and compare this to relative changes in price, but this initial exploratory analysis may provide some sense of likely minimum performance when inferring changes in house price from visual information.

As with the Google Street View and Geograph analyses, we consider house price changes over a decade, from 2005 and 2006 to 2015 and 2016. Again, we rank MSOAs by median house price by setting the highest rank of 983 to the MSOA with the highest median house price, and the lowest rank of 1 to the MSOA with the lowest median house price. If two areas have the same median house price, we give them both the mean rank. The change in median house price rank is then calculated as the 2015–2016 rank minus the 2005–2006 rank. To interpret the change in ranked price, consider a change of -100. This would indicate a decrease in ranked median house price by 100 places over the decade, such that 100 more MSOAs now had higher median house prices.

To model the change in ranked price, we again employ an elastic net. Specifically, we take the relevant feature vector as input, generated from indoor, outdoor, or all Zoopla images, and predict the change in median house price rank.

5.3 Results

5.3.1 Estimating local house prices in London during 2015 and 2016 from indoor Zoopla imagery

Can we infer differences in median house prices across local areas of London from indoor Zoopla images? The observed and predicted median house prices are mapped and visualised in Fig. 5.2a and Fig. 5.2b, where a darker area represents a higher median house price. The observed median house price corresponds to the true median house price for each MSOA, and the predicted median house price corresponds to the estimate of median house price from indoor Zoopla images. We produce all estimates using the methodology outlined in Fig. 5.1. As with the Google Street View and

Geograph analyses, median house prices across the 983 MSOAs in London range from £69,460 to £1,270,000, and the observed median house price map in Fig. 5.2a illustrates that areas in the centre of London tend to have higher median house prices, whereas areas in the east of London tend to have relatively lower house prices.

To establish whether we can estimate median house prices from indoor Zoopla imagery, we fit an elastic net model. The optimal value of *l1_ratio* is 0.999999 with an α of 0.41 (to 2 decimal places). We also perform a regression analysis after using the leave-one-out approach to compare the true median house prices with the estimates generated by the elastic net for each MSOA. From the analysis, we find that the model can generate very good estimates of median house prices ($R^2 = 0.81$, $N = 983$, $p < 0.001$). In fact, the model performs notably better than the Google Street View ($R^2 = 0.56$, $N = 983$, $p < 0.001$) and Geograph ($R^2 = 0.44$, $N = 983$, $p < 0.001$) models estimating median house prices in London. This may be because the median number of indoor Zoopla images available per MSOA is high, so a large amount of visual information is available for each area. Recall that the number of indoor Zoopla images available per MSOA ranges from 455 to 7,139 with a median of 2,591 images. In contrast, the median number of photos available for each MSOA is 496 for our Google Street View dataset, and 151 for our complete Geograph dataset.

Mapping the estimates produced by the model in Fig. 5.2b visually confirms that the predictions of median house prices are close to the true values, and the model is able to capture variations in median house prices of MSOAs in London well. This is further supported by the map of prediction errors from the model, depicted in Fig. 5.2c, where blue denotes an underestimate and red denotes an overestimate of median house price. Visual inspection reveals that areas in the east of London are sometimes overestimated and areas in the west of London are sometimes underestimated. However, the errors are small. The mean absolute error is £67,800 and the mean absolute percentage error is 14.30%. In comparison, the mean absolute error and mean absolute percentage error for the Google Street View model was £97,710 and 20.24%, and for the Geograph model was £136,500 and 29.18% (all errors to 4 significant figures).

Figure 5.2d shows the scatter plot of observed against predicted median house prices, where the red dots represent MSOAs and the blue line represents where perfect predictions should fall. The plot highlights that the model does not struggle with extreme values like the Google Street View and Geograph models did. We do find that some of the highest median house prices are slightly underestimated, but the effect is subtle.

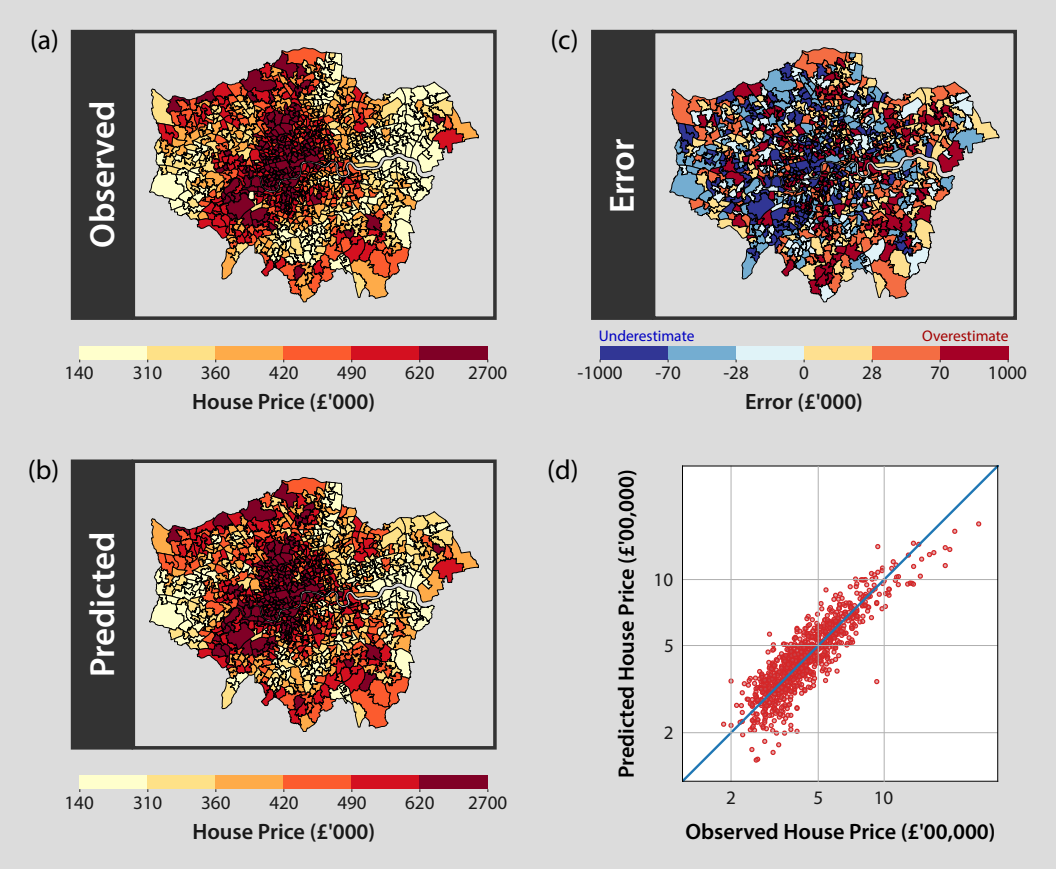

Figure 5.2: Evaluation of estimates of London house prices during 2015 and 2016 from indoor Zoopla images. (a) Map of median house prices for each MSOA ('Observed'). Darker areas correspond to higher house prices. **(b)** Map of median house prices for each MSOA estimated from Zoopla images ('Predicted'). Again, darker areas correspond to higher house prices. We generate estimates using the methodology outlined in Fig. 5.1. The model correctly identifies higher median house prices in the centre of London and lower median house prices in parts of the east of London. A regression analysis shows that the elastic net model performs very well at estimating local house prices ($R^2 = 0.81$, $N = 983$, $p < 0.001$). **(c)** Map of the errors in median house price estimates for each MSOA ('Error'). Blue denotes an underestimate and red denotes an overestimate in median house price. Median house prices around the east of London tend to be overestimated, whereas median house prices around the west of London tend to be underestimated. *(continues on the following page)*

Figure 5.2: *(continues from previous page)* **(d)** Log-log plot of observed and predicted median house prices. The red dots represent MSOAs in London and the blue line represents where perfect estimates would fall. We find that median house prices are inferred well, including higher median house prices. Comparatively, the Google Street View (Fig. 3.4d) and Geograph (Fig. 4.7d) models struggled with extreme values. For all maps, quantile breaks are implemented with six bins. Contains National Statistics data © Crown copyright and database right 2011. Contains Ordnance Survey data © Crown copyright and database right 2011. Zoopla Limited, © 2018. Contains data from Zoopla Limited. Economic and Social Research Council. Zoopla Property Data 2018. University of Glasgow - Urban Big Data Centre.

We depict all non-zero coefficients from the elastic net model in Fig. 5.3. For each coefficient, the absolute value indicates how big an effect the feature has on median house price estimates, and the sign indicates whether the effect is positive or negative. If the effect is positive, the feature is associated with higher median house prices, and if the effect is negative, the feature is associated with lower median house prices. For example, the model associates the features 'bedchamber', 'elevator lobby' and 'living room' with higher median house prices, and associates 'recreation room', 'waiting room', and 'bedroom' with lower median house prices. Although the sign of the coefficients generally follows intuition, some of the results appear contradictory. For example, a 'bedchamber' is a positive coefficient whereas a 'bedroom' is a negative coefficient, but they are both rooms that one would sleep in and are almost always present in a house. What is the distinction between these features?

5.3.2 Exploring coefficients from the house price estimation model using indoor Zoopla imagery

To explore the coefficients and understand how Places365 operates with indoor Zoopla imagery, we first look at a selection of top-scoring indoor images for top positive and negative coefficients. We pick the top three positive and negative coefficients and for each feature, examine three of the ten highest-scoring images with high feature values. Our analysis is presented in Fig. 5.4, where the feature values underneath each image are those extracted for the given feature using Places365 (Section 5.2.2). For each of the images, the corresponding feature is its top classification.

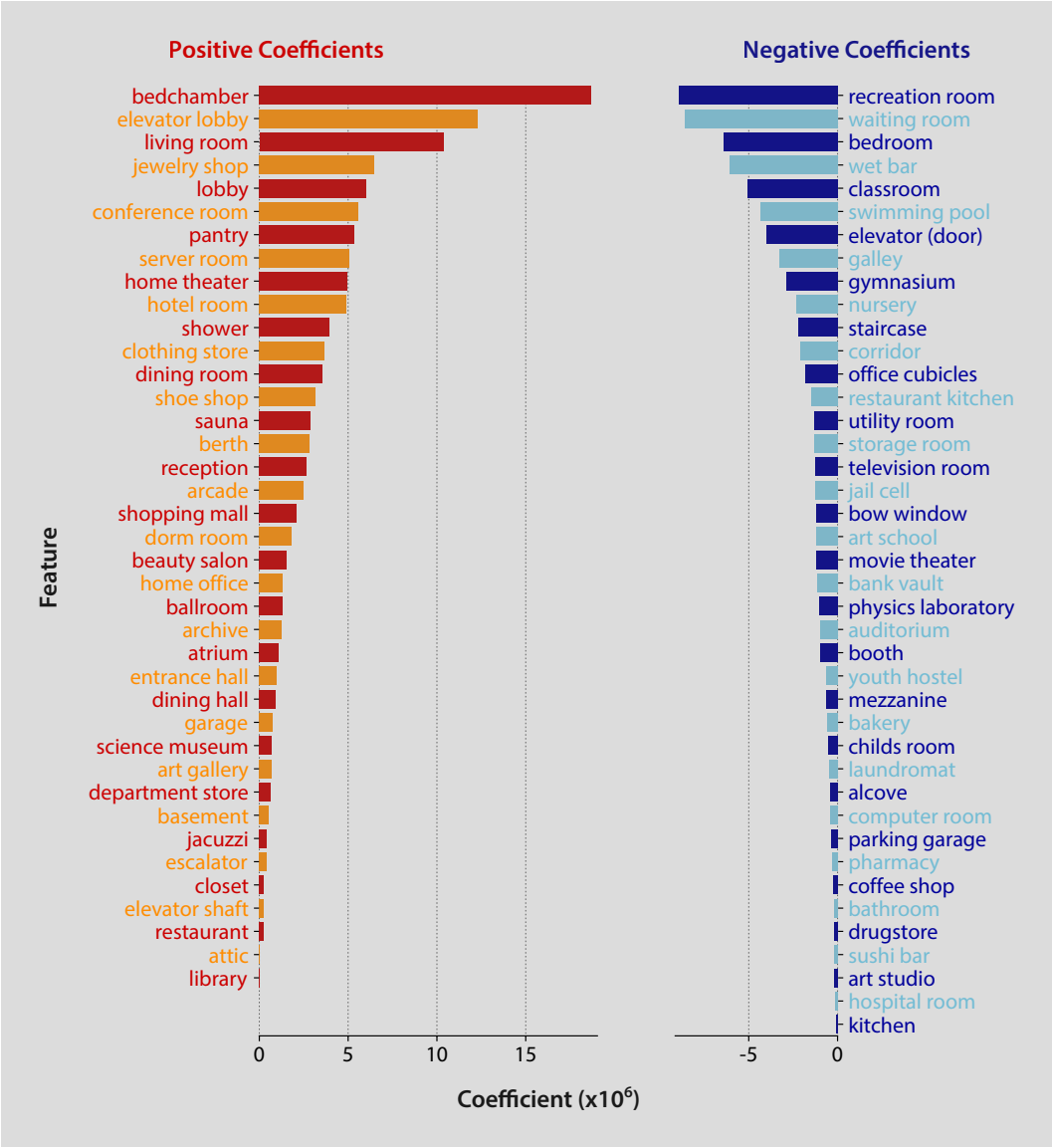

Figure 5.3: Elastic net coefficients for estimates of median house prices in London during 2015 and 2016 from indoor Zoopla images. We train an elastic net model using the methodology outlined in Fig. 5.1 and depict the resulting coefficients from the model for all features where the coefficient is not zero. *(continues on the following page)*

Figure 5.3: *(continues from previous page)* Positive coefficients relate to visual features that are associated with higher median house prices, whereas negative coefficients relate to visual features that are associated with lower median house prices. The features that the model has learnt to link with areas of higher median house prices include 'bedchamber', 'elevator lobby', and 'living room', and the features that the model has learnt to link with areas of lower median house prices include 'recreation room', 'waiting room', and 'bedroom'. Generally, we find that the sign of the coefficients accord with intuition but a few results appear contradictory. For example, it is unclear why a 'bedchamber' is linked with higher house prices, whereas a 'bedroom' is linked with lower house prices. We investigate this further in Section 5.3.2.

In general, Places365 assigns labels to images accurately and the labels themselves broadly reflect the contents of the images. Of the features shown in Fig. 5.4, only the images labelled 'waiting room' (Fig. 5.4e) are likely not to be waiting rooms, given that we know that we have retrieved images of housing. However, the rooms depicted do very much look like waiting rooms. In this housing context, it however seems more likely that these rooms are living rooms with minimal decor. Figure 5.4c shows images associated with the label 'living room', but these living rooms are more elaborately decorated. This example demonstrates that the CNN is able to identify nuances in indoor images of rooms. A further example includes the distinction between a 'bedchamber' (Fig. 5.4a) and a 'bedroom' (Fig. 5.4f). The images associated with both labels contain bedrooms, but those associated with a 'bedchamber' include a four-poster bed rather than the standard bed found in images labelled 'bedroom'.

Moreover, notice how the features with images of more elaborately decorated rooms ('living room', 'bedchamber') are positive coefficients in our house price estimation model, and the features with images of less elaborately decorated rooms ('waiting room', 'bedroom') are negative coefficients. It seems that the elastic net connects features describing well-furnished rooms with higher median house prices, and features describing less decorated rooms with lower median house prices. This is reasonable, given the likely differences in economic circumstances of people who can afford large four-poster beds and people who can afford very little furniture. The level of furnishing can give an indication of the financial status of those living in a particular house, which in turn provides some hints as to the price of the property.

Figure 5.4: Top indoor Zoopla images of London for top positive and negative coefficients. We identify the three features that are most positively and negatively related to median house prices in London according to the elastic net model. *(continues on the following page)*

Figure 5.4: *(continues from previous page)* By exploiting the feature vectors we extract from the indoor Zoopla images using the methodology outlined in Fig. 5.1, we identify three images with high feature values for each of the features. We can then get some insights into which shapes and patterns the CNN sees for each feature. **(a)** The images with a high score for the feature 'bedchamber' contain bedrooms with large, four-poster beds. The label accurately reflects the contents of the images. **(b)** The images with a high score for the feature 'elevator lobby' contain elevator lobbies, so are correctly labelled. **(c)** The images with a high score for the feature 'living room' contain well-furnished living rooms. Again, the label accurately reflects the contents of the images. **(d)** The images with a high score for the feature 'recreation room' contain a room with a pool table. As pool is a recreational activity, this label appears accurate. **(e)** The images with a high score for the feature 'waiting room' contain rooms with seating and minimal decorations. Without further context, it would appear entirely reasonable to suppose these were indeed waiting rooms. However, as we know that we have retrieved images of housing, it seems more likely that these are in fact sparsely decorated living rooms. **(f)** The images with a high score for the feature 'bedroom' contain bedrooms with neat, standard beds. Once again, the label accurately reflects the contents of the images. Generally, the images that exhibit high scores for a given feature have similar characteristics to each other, and are broadly labelled correctly. Furthermore, Places365 is able to identify nuances in indoor images of rooms, distinguishing 'bedchamber' from 'bedroom' for example. Contains images from Zoopla Limited, © 2018. Zoopla Limited. Economic and Social Research Council. Zoopla Property Data 2018. University of Glasgow - Urban Big Data Centre.

Since Places365 is able to distinguish differences between rooms from indoor Zoopla images, we want to better understand how the model operates. Which areas of London are associated with the features that represent the top positive and negative coefficients? In particular, we are interested in understanding which positive coefficients are associated with which areas of higher median house prices, such as some areas in the south-west of London, and which negative coefficients are associated with which areas of lower median house prices, such as some areas in East London.

We first examine feature value distribution maps for positive coefficients in Fig. 5.5, using the positive coefficients features showcased in Fig. 5.4 as examples. Visual inspection reveals that the features 'bedchamber' (Fig. 5.5b), 'elevator lobby' (Fig. 5.5c), and 'living room' (Fig. 5.5d) are all more prevalent in the centre of London, but the prevalence of the features varies across other areas in London that exhibit higher median house prices. In particular, 'bedchamber' exhibits higher

relative feature values in areas in the south-west of London with higher median house prices, similarly to 'living room', but is also more prevalent in areas in the south-east of London with higher median house prices, unlike 'elevator lobby' and 'living room'. Furthermore, 'elevator lobby' has higher relative feature values in areas in the north-east of London with higher median house prices.

Next, we examine the feature value distribution maps for negative coefficients in Fig. 5.6 using the negative coefficient features showcased in Fig. 5.4 as examples. We find that the features are more prevalent in areas with lower median house prices, but are also somewhat prevalent in areas with higher median house prices. The map for the feature 'recreation room' (Fig. 5.6b) reveals higher feature values in areas in the west of London where median house prices are lower, as does the map for the feature 'waiting room' (Fig. 5.6c). However, a 'waiting room' is also more prevalent in areas in the east of London where median house prices are lower, and less prevalent below the River Thames. The feature 'bedroom' (Fig. 5.6d) is more prevalent in the south-west of London, but is less prevalent in the north-west of London, where the local house prices are higher.

Some of the areas in which the feature 'recreation room' is highly prevalent are areas with higher median house prices, such as the south and north-west of London. This phenomenon could be explained by the fact that the features values themselves are an order of magnitude smaller than the other coefficients we map, so it is possible that recreation rooms are relatively uncommon in London. We investigate this further, and consider whether the importance of 'recreation room' is inflated because of its minimal presence in London. To do so, we find features whose mean feature value is within 0.005 of the mean feature value of 'recreation room', which is 0.009427 (to 4 significant figures). We then compare the coefficient values of the 10 features identified to determine whether they all have large magnitudes. We find that the features 'office' and 'playroom' both have low mean feature values in London but are ignored by our model. These counterexamples suggest that the weighting of 'recreation room' is not arbitrary.

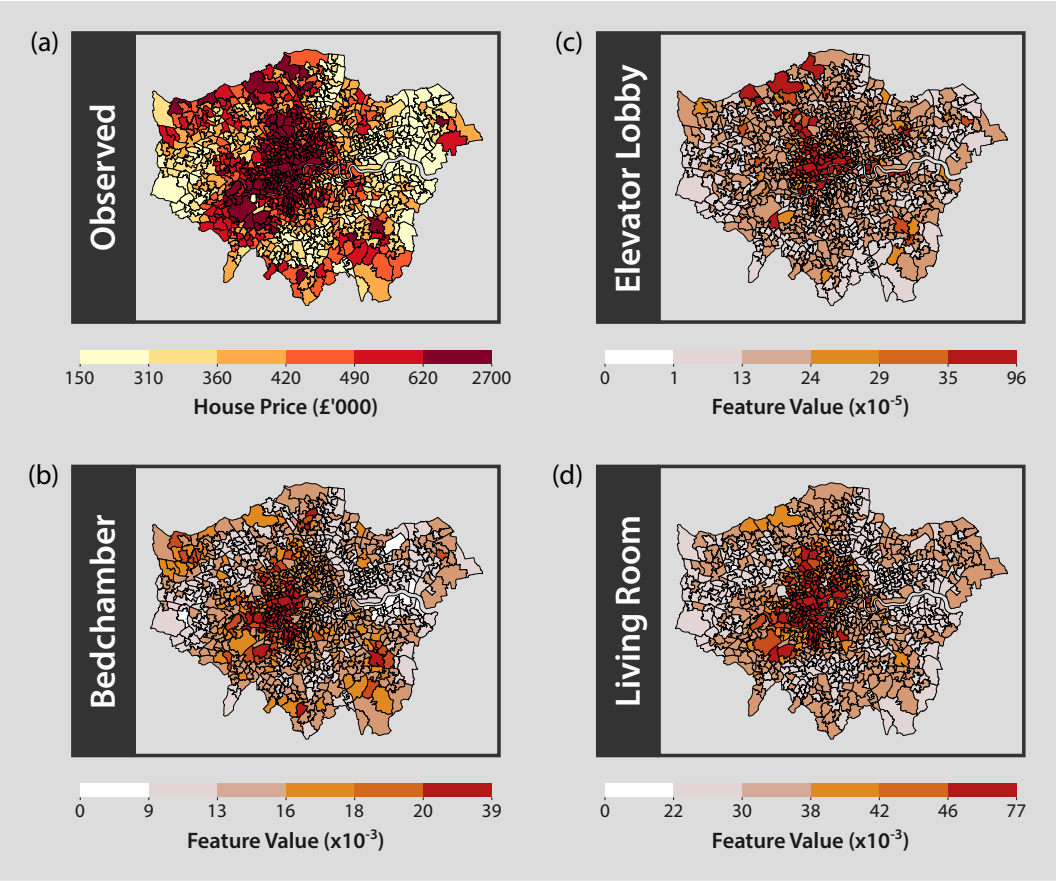

Figure 5.5: Prevalence of high positive coefficient features in London from indoor Zoopla images. **(a)** Map of median house prices for each MSOA ('Observed'), as also shown in Fig. 5.2a. Darker areas correspond to higher median house prices. **(b)** Map of mean feature values for the feature 'bedchamber'. The feature values are extracted via the methodology outlined in Fig. 5.1. We want to determine which positive coefficients are associated with which areas of higher median house prices. For all of the following maps, darker areas correspond to higher feature values and all maps are presented at MSOA granularity. Many of the areas that observe high median house prices seem to have house interiors that Places365 associates with the feature 'bedchamber', especially towards the south-west and south-east of London. **(c)** Map of mean feature values for the feature 'elevator lobby'. Again, areas corresponding to higher median house prices, the centre and north-east of London in particular, exhibit high mean feature values for the feature 'elevator lobby'. However, it is less prevalent in the south-east of London. *(continues on the following page)*

Figure 5.5: *(continues from previous page)* **(d)** Map of mean feature values for the feature 'living room'. Living rooms appear to be more prevalent in MSOAs that have higher median house prices, especially in the centre and south-west of London. We note that for all of our examples, the features are more prevalent in the centre of London. For all maps except (a), standard deviation breaks are implemented, whereas (a) uses quantile breaks, with six bins. Note that the maps therefore use different scales. Contains National Statistics data © Crown copyright and database right 2011. Contains Ordnance Survey data © Crown copyright and database right 2011. Contains data from Zoopla Limited, © 2018. Zoopla Limited. Economic and Social Research Council. Zoopla Property Data 2018. University of Glasgow - Urban Big Data Centre.

We further inspect the coefficients and feature values to determine whether we see high coefficients simply because the feature values are generally low. To begin, we calculate the mean feature value for each indoor feature across all MSOAs. We then identify the features whose mean feature value is in the top 10% of mean feature values calculated. We compare the coefficient values to the mean feature values of these 16 features. We find that three of the 16 features are in either the top 10 positive coefficients or in the top 10 negative coefficients (18.75%). Only one of the 16 features is set to zero. In comparison, 100% of the features in the bottom 10% of mean feature values are set to zero. Generally, it seems that features with high mean feature values also contribute the most to the estimates of median house prices in the model.

In addition to the feature value distribution maps, we also explore how well the model can characterise areas of London. We choose to compare the visual information available from two areas in London with different socioeconomic statuses. Specifically, we investigate newly gentrified Dalston (E02000365) (Davison, Dovey, & Woodcock, 2012), and family-oriented Acton (E02000258), as showcased in Fig. 5.7. For background, the population of Dalston consists of slightly more working professionals than Acton (Office for National Statistics, 2014b), whereas the population of Acton consists of more families than Dalston (Office for National Statistics, 2014a).

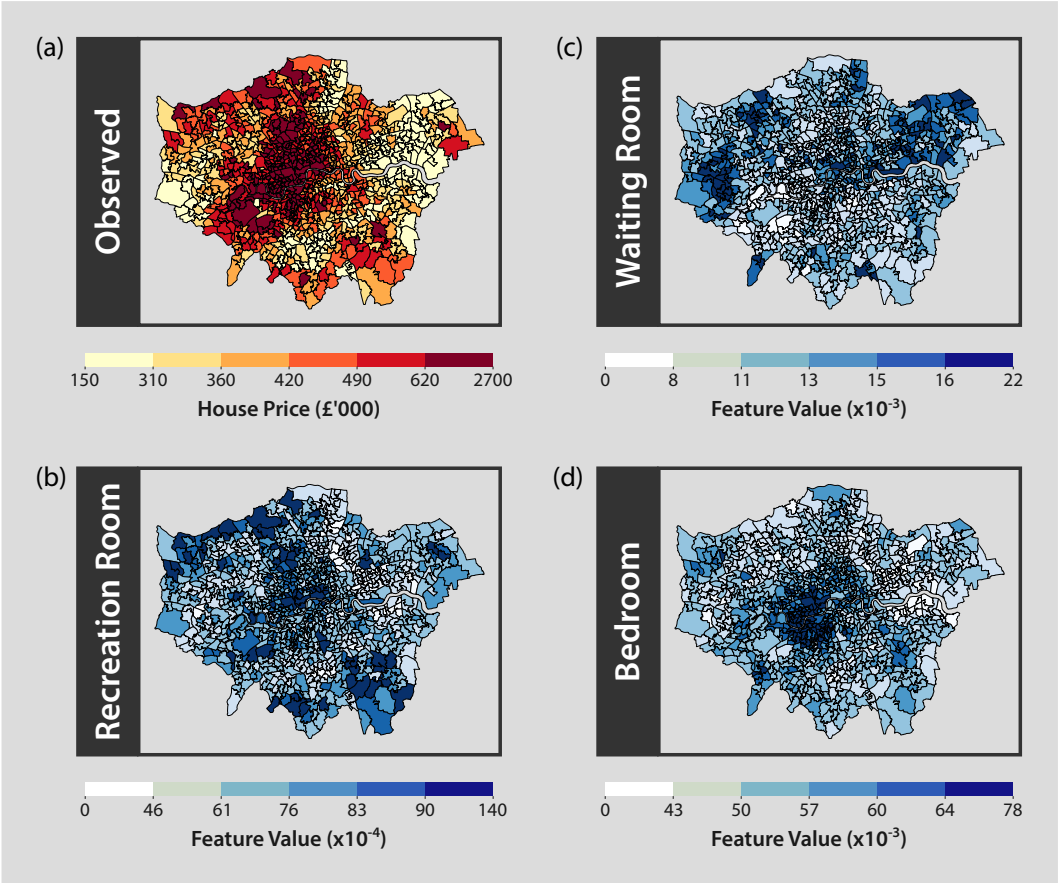

Figure 5.6: Prevalence of high negative coefficient features in London from indoor Zoopla images. (a) Map of median house prices for each MSOA ('Observed'), as also shown in Fig. 5.2a. Darker areas correspond to higher median house prices. **(b)** Map of mean feature values for the feature 'recreation room'. The feature values are extracted via the methodology outlined in Fig. 5.1. We want to determine which negative coefficients are associated with which areas of lower median house prices. For all of the following maps, darker areas correspond to higher feature values and all maps are presented at MSOA granularity. Visual inspection reveals that the feature 'recreation room' is more prevalent in areas with lower median house prices, such as the north-west of London. However, the feature values are smaller than the feature values of other negative coefficients by an order of magnitude, so it may be that there are generally not many recreation rooms in London. *(continues on the following page)*

Figure 5.6: *(continues from previous page)* **(c)** Map of mean feature values for the feature 'waiting room'. Areas corresponding to lower median house prices, especially towards the east and west of London, also correspond to the prevalence of 'waiting room' in London. It is also less prevalent below the River Thames. **(d)** Map of mean feature values for the feature 'bedroom'. The feature appears to be more prevalent in the south-west of London, but less prevalent in the north-west of London, where the local house prices are higher. For all of our examples, the features are more prevalent in areas that exhibit lower median house prices, but are also somewhat prevalent in areas which higher median house prices. For all maps except (a), standard deviation breaks are implemented, whereas (a) uses quantile breaks, with six bins. Note that the maps therefore use different scales. Contains National Statistics data © Crown copyright and database right 2011. Contains Ordnance Survey data © Crown copyright and database right 2011. Contains data from Zoopla Limited, © 2018. Zoopla Limited. Economic and Social Research Council. Zoopla Property Data 2018. University of Glasgow - Urban Big Data Centre.

To compare, we first obtain a mean feature vector for both MSOAs that includes both indoor and outdoor features. The process for extracting feature vectors is described in Section 5.2.2. We also calculate a total mean feature vector for London from all available Zoopla images, where the mean feature value for each feature is given by the sum of feature values from all images, divided by the number of images. Then for the two MSOAs in Dalston and Acton, we calculate the relative mean feature vector by subtracting the total mean feature vector for London from the mean feature vector of that MSOA. All feature vectors are of length 365, as they include both indoor and outdoor features extracted from Places365. Finally, we identify the ten features with the highest relative mean feature values and the ten features with the lowest relative mean feature values. This allows us to identify which visual characteristics are prevalent in each area and which are less prevalent, relative to the general visual characteristics of London.

Figure 5.7a displays the relative mean features for Acton and Fig. 5.7b displays the relative mean features for Dalston. The red features represent the top, most prevalent relative mean features and the blue features represent the lowest, least prevalent relative mean features. Next to the features of each area, we also visualise how prevalent the same features are in the other area. We can then compare how the relative mean feature values differ between the areas and determine whether the model can differentiate between areas with different socioeconomic statuses. The

relative mean feature values themselves can be interpreted as relating to how likely a given feature is to appear in a given area, compared to London generally. We also include four interior Zoopla images for each area in the figure, to provide an idea of common furnishings in each area.

By inspecting the features, we see that several features that are more prevalent in Dalston are less so in Acton, and vice versa. For example, the features 'yard' and 'house' are more likely to appear in family-oriented Acton but less likely to appear in newly gentrified Dalston, whereas the features 'balcony' and 'apartment building' are more likely in Dalston but less likely in Acton. This analysis is an illustration of how the model can identify different visual features in areas with different socioeconomic statuses.

5.3.3 Estimating changes in local house price rank in London from 2005 and 2006 to 2015 and 2016 using indoor Zoopla imagery

Can we identify which areas of London have become relatively more expensive or relatively cheaper over time? To answer this question, we model the change in median house price rank, as in Sections 3.2.5 and 4.3.3. The change in median house price rank for each MSOA is calculated using the methodology described in Section 5.2.5. We use the same house price data as in the Google Street View and Geograph analyses, and so the change in median house price rank again ranges from -424.5 to 522.

Figure 5.8a shows a map of the actual changes in median house price rank for each MSOA. Blue areas represent a decrease in ranked price and red areas represent an increase in ranked price over the decade. As noted with the Google Street View and Geograph analyses, we observe that areas towards the north-east and south-east of the centre of London exhibit increases in ranked price, whereas areas in the outskirts of London exhibit decreases in ranked price over the decade.

The predicted change in median house price rank is shown in Fig. 5.8b, where the estimated changes are generated using the methodology outlined in Fig. 5.1 and indoor Zoopla images. Again, blue areas represent a decrease in ranked price and red areas represent an increase in ranked price from 2005 and 2006 to 2015 and 2016. During the process of fitting our elastic net model, we find the optimal value of *l1_ratio* to be 0.99 with a corresponding α of 0.00060 (to 2 significant figures).

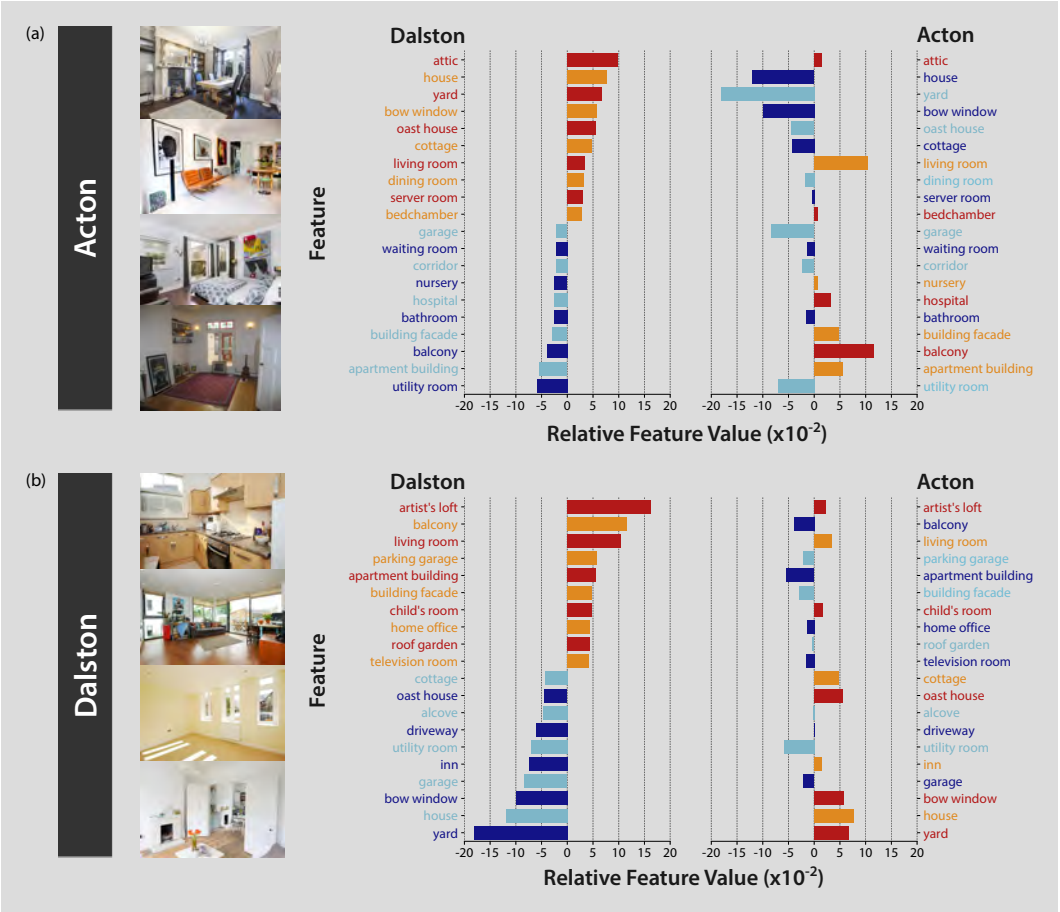

Figure 5.7: Comparing visual characteristics of two MSOAs with different socioeconomic statuses. Can the house price estimation model characterise different areas of London appropriately? We choose a newly gentrified MSOA in Dalston and a family-oriented MSOA in Acton, and compare their visual information. We extract feature vectors for each Zoopla image as described in Section 5.2.2. Using the full Zoopla dataset, including both indoor and outdoor Zoopla images, we calculate the total mean feature vector for London, or in other words, the mean feature value for all 365 features identified by Places365 for all Zoopla images in London. Then for both of our chosen MSOAs, we calculate a relative mean feature vector by subtracting the total mean feature vector for London from the mean feature vector of each MSOA. Finally, we identify the ten features with the highest relative mean feature values and the ten features with the lowest relative mean feature values. This allows us to identify the common and less common visual characteristics of each area, relative to the characteristics of the entirety of London. *(continues on the following page)*

Figure 5.7: *(continues from previous page)* **(a)** Relative mean features in Acton. Red denotes a higher relative mean feature value and blue denotes a lower relative mean feature value. The feature value itself relates to how prevalent a feature is in an area relative to the whole of London. The same relative mean features are also shown for Dalston for comparison. **(b)** Relative mean features in Dalston. Again, red denotes a higher relative mean feature value and blue denotes a lower relative mean feature value, and the same relative mean features are shown for Acton. It seems that the features 'yard' and 'house' are likely to appear in family-oriented Acton but unlikely to appear in newly gentrified Dalston, whereas the features 'apartment building' and 'balcony' are likely to be found in Dalston, but unlikely to be found in Acton. Overall, this analysis illustrates that our approach allows us to accurately characterise areas of London with different socioeconomic statuses. Contains images from Zoopla Limited, © 2018. Zoopla Limited. Economic and Social Research Council. Zoopla Property Data 2018. University of Glasgow - Urban Big Data Centre.

Comparing the map of predicted changes to the map of observed changes highlights that the model is able to estimate the changes in median house price rank reasonably well. A regression analysis comparing the observed and estimated changes in ranked price confirms the level of performance achieved by the model ($R^2 = 0.42$, $N = 983$, $p < 0.001$). In particular, the model captures the broad pattern that areas in the centre of London have seen increases in ranked price and that areas outside of the centre of London have seen decreases in ranked price. However, the size of the increases and decreases are often underestimated, as was the case with the Google Street View and Geograph models. This analysis confirms that we can achieve some success when estimating changes in local house prices without analysing changes in the visual appearance of different areas over time. It may be possible to achieve even greater success when comparing images from different time periods. We discuss this further in Section 5.4.

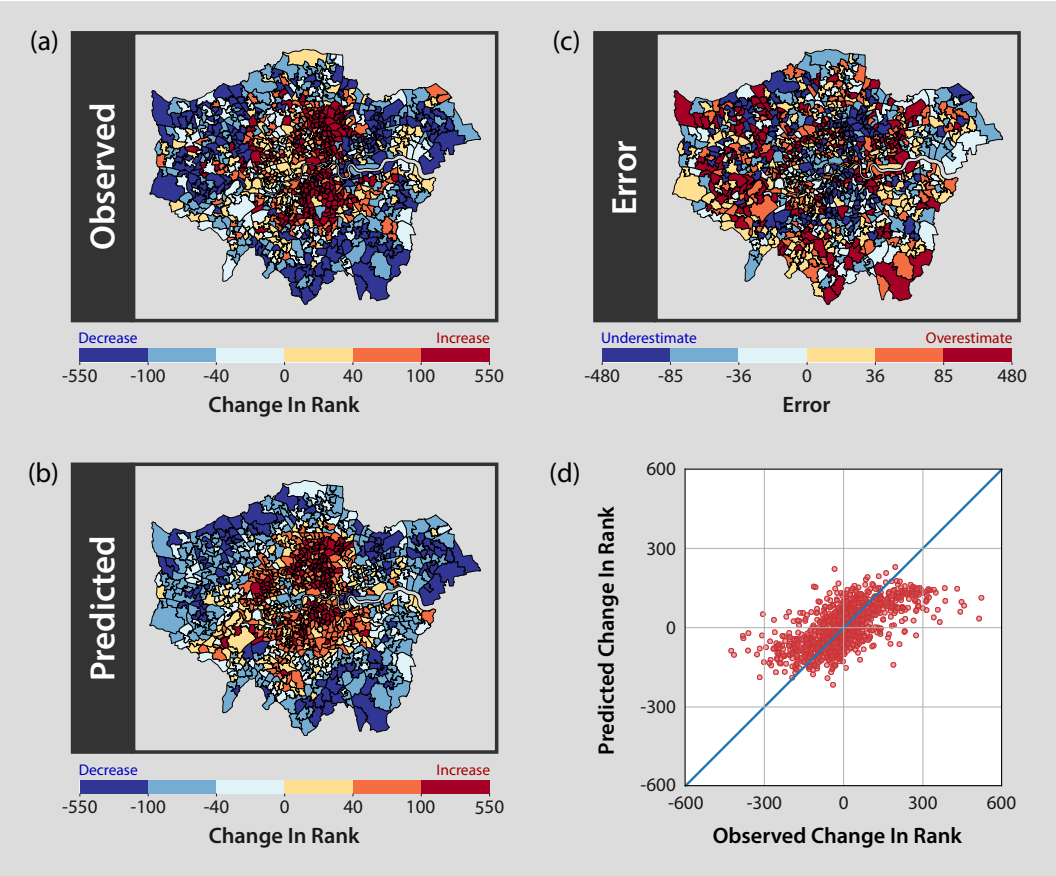

Figure 5.8: Inferring changes in house prices in London using indoor Zoopla images. (a) Map of change in median house prices between 2005–2006 and 2015–2016 for each MSOA ('Observed'). We obtain a change in median house price rank for each MSOA in London using the methodology described in Section 5.2.5. As with the Google Street View and Geograph analyses, blue denotes a decrease in ranked price and red denotes an increase in ranked price from 2005 and 2006 to 2015 and 2016. **(b)** Map of change in median house prices for each MSOA estimated from indoor Zoopla images ('Predicted'). We generate estimates using the methodology outlined in Fig. 5.1. Again, blue denotes a decrease in ranked price and red denotes an increase in ranked price over the decade. A regression analysis indicates that the model can capture the observed patterns across London reasonably well ($R^2 = 0.42$, $N = 983$, $p < 0.001$). This is evident from the maps of observed and predicted changes in the median house price rank, which show that the model is particularly good at identifying the increases in ranked price in the centre of London and decreases in ranked price in the outskirts of London. *(continues on the following page)*

Figure 5.8: *(continues from previous page)* However, the size of the increases and decreases are often underestimated. This phenomenon was also present with the analogous Google Street View and Geograph analyses. **(c)** Map of the errors in estimates of change in median house price rank for each MSOA ('Error'). Blue denotes an underestimate in the change in rank and red denotes an overestimate. Changes in median house price rank of areas in the outskirts of London, excluding the east, tend to be overestimated, whereas changes in median house price rank of the centre of London tend to be underestimated by the model. **(d)** Scatter plot of observed and predicted change in rank. The red dots represent MSOAs in London and the blue line represents where perfect predictions should fall. The map reveals that extreme values are not handled well by the model, especially when the change in median house price rank is greater than 300. For all maps, quantile breaks are implemented with six bins. Contains National Statistics data © Crown copyright and database right 2011. Contains Ordnance Survey data © Crown copyright and database right 2011. Contains data from Zoopla Limited, © 2018. Zoopla Limited. Economic and Social Research Council. Zoopla Property Data 2018. University of Glasgow - Urban Big Data Centre.

The prediction errors in Fig. 5.8c draw attention to the fact that the change in median house price rank of areas in the outskirts of London towards the west and south are often overestimated, whereas the change in median house price rank of areas in the centre of London towards the north and south are generally underestimated. Moreover, the scatter plot of observed versus predicted change in rank (Fig. 5.8d) highlights that estimates of the change in ranked price always fall between -300 and 300 in contrast to the true range of -425.5 to 522. More extreme changes of greater than 300 ranks are not estimated well. Again, this may be due to the distribution of changes in ranked price, which is heavy-tailed. Despite the model struggling with extreme values, its performance is notably superior to that seen when using Google Street View images ($R^2 = 0.27$, $N = 983$, $p < 0.001$) or Geograph images ($R^2 = 0.18$, $N = 983$, $p < 0.001$). This is presumably due either to the larger volume of images exploited to estimate the changes, or to the information available from indoor images.

Once again, we examine the non-zero coefficients in our elastic net model and depict them in Fig. 5.9. Positive coefficients represent features that the model associates with increases in ranked price over the decade, for example 'utility room' and 'staircase'. Conversely, negative coefficients represent features that the model associates with decreases in ranked price over the decade, for example 'bathroom' and

'dining room'. We discover that some of the features that are associated with higher median house prices are associated with larger relative decreases in ranked price, and vice versa, as we found in the corresponding Google Street View and Geograph analyses. For instance, 'bedchamber' and 'elevator lobby' are positive coefficients in the model estimating median house prices (Fig. 5.3) but negative coefficients in this analysis, while 'bedroom' and 'swimming pool' are negative coefficients in the model estimating median house prices but positive coefficients in this analysis. This could reflect the type of changes experienced by some local areas of London. For instance, some areas that see higher median house prices in 2015 and 2016 prices may not have climbed as highly in the ranks over the decade, whereas other areas that see lower median house prices in 2015 and 2016 may have climbed more highly in the ranks.

5.3.4 Estimating local house prices and changes in local house price rank in London from outdoor Zoopla imagery

In addition to inferring median house prices of local areas in London from indoor Zoopla imagery, we also attempt inference from outdoor Zoopla imagery. Performance of our house price estimation models when using indoor Zoopla images is far superior than when using Google Street View or Geograph images. Therefore, we want to establish whether using the information available from indoor image data improves the performance of our house price estimation model, or whether simply using Zoopla image data leads to an improvement.

One reason we might see better performance when using outdoor images from Zoopla rather than from Google Street View or Geograph is the number of images we have access to. Recall that we have 1,025,384 outdoor Zoopla images in our dataset with a median of 973 images per MSOA, while our indoor Zoopla images dataset contains 2,702,506 images with a median of 2,591 photos per MSOA. In contrast, we access 518,808 Google Street View images with a median of 496 photos per MSOA, and 328,014 Geograph images with a median of 151 photos per MSOA. Even when considering outdoor or indoor images alone, we have a far greater volume and median number of Zoopla images in each MSOA than for either of the other two data sources. Furthermore, the outdoor Zoopla images in our dataset include visual information resembling a mixture of Google Street View and Geograph images, as they consist of housing facades and gardens.

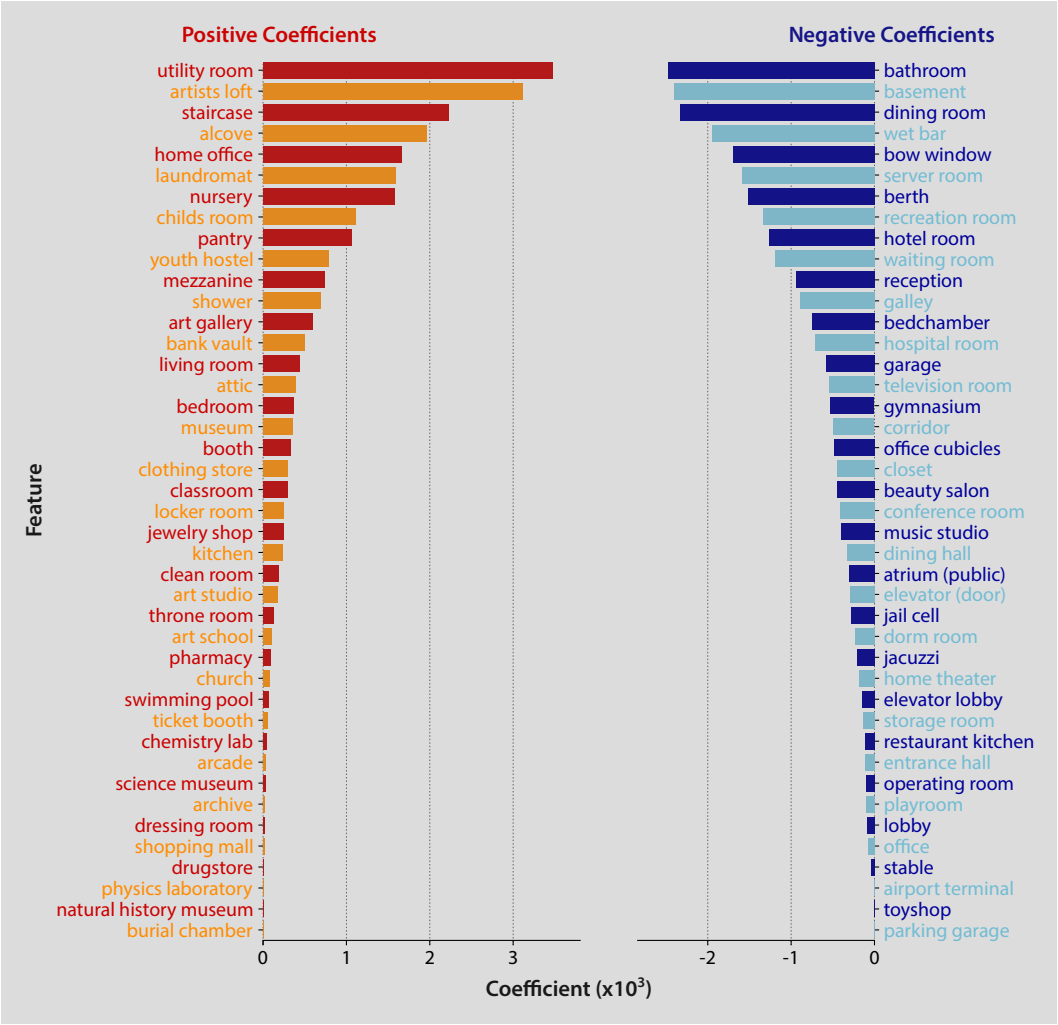

Figure 5.9: Elastic net coefficients for estimated changes in median house prices in London from indoor Zoopla images. We rank the MSOAs in London using the approach detailed in Section. 5.2.5. We then train an elastic net model to estimate the change in median house price rank between 2005–2006 and 2015–2016 using the methodology outlined in Fig. 5.1. The resulting coefficients from the elastic net model are depicted for all features where the coefficient is not zero. Positive coefficients relate to features associated with an increase in ranked house price over the decade, whereas negative coefficients relate to features associated with a decrease in ranked house price over the decade. The coefficients exhibit a similar pattern to that seen in the Google Street View analysis (Section 3.3.2), and to some extent in the Geograph analysis (Section 4.3.4). *(continues on the following page)*

Figure 5.9: *(continues from previous page)* Namely, several coefficients that are positively associated with median house prices are negatively associated with changes in median house price rank, for example 'bedchamber' and 'elevator lobby'. Likewise, several coefficients that are negatively associated with median house prices are positively associated with changes in median house price rank, for example 'bedroom' and 'swimming pool'.

We generate estimates of median house prices across areas of London from outdoor Zoopla images using the methodology outlined in Fig. 5.1. We conduct a regression analysis to compare the estimates to the true median house prices for each MSOA and find that the model performs well ($R^2 = 0.66$, $N = 983$, $p < 0.001$), although not as well as when using indoor Zoopla images ($R^2 = 0.81$, $N = 983$, $p < 0.001$; Section 5.3.1). Indeed, we find that the mean absolute error of the model is £81,090 and the mean absolute percentage error is 16.36%, whereas for the indoor Zoopla images model, the mean absolute error was £67,800 and the mean absolute percentage error was 14.30% (all errors to 4 significant figures). The model does however perform well in comparison to the corresponding Google Street View model (mean absolute error: £97,710; mean absolute percentage error: 20.24%; $R^2 = 0.56$, $N = 983$, $p < 0.001$) and Geograph model (mean absolute error: £136,500; mean absolute percentage error: 29.18%; $R^2 = 0.44$, $N = 983$, $p < 0.001$).

We also investigate whether we can generate better estimates of changes in median house price rank from 2005 and 2006 to 2015 and 2016 using outdoor Zoopla images rather than indoor Zoopla images. Again, the true change in median house price rank across this period ranges from -425.5 to 522. When employing the elastic net, we find the optimal value of *l1_ratio* to be 0.99 with an α of 0.0024 (to 2 significant figures). Performing a regression on the observed and predicted changes in ranked price demonstrates that the model performs reasonably well ($R^2 = 0.39$, $N = 983$, $p < 0.001$). Once again, the model performance is nearly as good as the indoor Zoopla images model ($R^2 = 0.42$, $N = 983$, $p < 0.001$; Section 5.3.3), and better than the Google Street View ($R^2 = 0.27$, $N = 983$, $p < 0.001$) and Geograph ($R^2 = 0.18$, $N = 983$, $p < 0.001$) models.

Altogether, it seems that using indoor Zoopla imagery leads to better performance than using outdoor Zoopla imagery to estimate median house prices and changes in median house price rank. However, we use fewer Zoopla images in the

outdoor image analysis compared to the analysis considering indoor images only. Moreover, we find that using either indoor or outdoor Zoopla imagery leads to improved model performance in comparison to using Google Street View or Geograph imagery. Given this observation, we ask whether using both indoor and outdoor Zoopla images in combination to estimate median house prices in London would result in further performance improvements.

5.3.5 Estimating local house prices in London during 2015 and 2016 from indoor and outdoor Zoopla imagery

As described in Section 5.2.3, our full Zoopla dataset of indoor and outdoor images contains 3,727,890 images with a median of 3,568 photos per MSOA, whereas our subset of indoor Zoopla images contains 2,702,506 images with a median of 2,591 photos per MSOA, and our subset of outdoor Zoopla images contains 1,025,384 images with a median of 973 images per MSOA. Can we improve our house price estimation model by using both interior and exterior Zoopla images?

We start by estimating median house prices for MSOAs in London during 2015 and 2016. Once again, we obtain estimates using the methodology outlined in Fig. 5.1. Applying the elastic net results in an optimal value of *l1_ratio* of 0.99999 with an α of 0.16 (to 2 significant figures). We also run a regression analysis comparing the observed median house price for each MSOA with the estimates produced by the elastic net using leave-one-out analysis and find that the model performs very well (mean absolute error: £65,300; mean absolute percentage error: 13.67%; $R^2 = 0.82$, $N = 983$, $p < 0.001$). This represents a marginal improvement over the analysis using indoor Zoopla images (mean absolute error: £67,800; mean absolute percentage error: 14.30%; $R^2 = 0.81$, $N = 983$, $p < 0.001$; Section 5.3.1).

We further discover that the model places more emphasis on indoor features than outdoor features. Of the top ten positive coefficients, 90% are indoor features, and of the top ten negative coefficients, 60% are indoor features. Moreover, the top three positive coefficients are 'bedchamber', 'living room', and 'hotel room', and the top three negative coefficients are 'waiting room', 'bedroom', and 'alcove', all of which are indoor features. The features 'bedchamber', 'living room', 'waiting room', and 'bedroom' were also top coefficients in the analysis using indoor Zoopla images only (Fig. 5.3). We note that we have almost three times as many indoor images in our dataset than outdoor images, and the median number of photos per MSOA is higher in the indoor images dataset than the outdoor images dataset. Hence, it may be that the model has found that there is more representative information regarding housing interiors for it to learn from.

We also investigate whether combining indoor and outdoor images impacts performance when inferring changes in median house price rank from 2005 and 2006 to 2015 and 2016 in London. Again, the true change in median house price rank across this period ranges from -425.5 to 522. When fitting an elastic net model to generate estimates of the changes, we find the optimal value of *l1_ratio* to be 0.99 with a corresponding α of 0.00077 (to 2 significant figures). Conducting a regression analysis comparing the observed relative changes for each MSOA with the estimates produced by the elastic net using leave-one-out analysis confirms that once again, using both indoor and outdoor Zoopla images results in a very slight improvement of model performance ($R^2 = 0.43$, $N = 983$, $p < 0.001$) over using only indoor Zoopla images ($R^2 = 0.42$, $N = 983$, $p < 0.001$). Again, of the top ten positive coefficients, 70% are indoor features, and of the top ten negative coefficients, 60% are indoor features.

5.4 Discussion

Due to the rise in property marketing websites, the availability of images reflecting the interior of houses has rapidly increased. Hence, the goal of this chapter is to explore a rich source of indoor imagery of London and learn whether local house prices can be inferred from them. We specifically explore photographic data from Zoopla, which contain images of housing interiors. Using all of the indoor Zoopla imagery at our disposal, we conduct two main analyses. Firstly, we attempt to infer median house prices in London during 2015 and 2016, and secondly, we investigate whether we can also infer relative changes in median house prices over a decade.

In addition to indoor Zoopla images, we have access to outdoor Zoopla imagery, which contain photos of building facades and property gardens. Therefore, we also look at conducting the same analyses with outdoor Zoopla images alone, as well as the full dataset of indoor and outdoor Zoopla images, to see how model performance compares to models using only indoor Zoopla images.

We successfully infer median house prices across local areas of London during 2015 and 2016 from indoor Zoopla images, and achieve superior model performance in comparison to inferences made from the outdoor image sources Google Street View and Geograph. We suggest that this success can be attributed to a large, representative volume of photos across areas of London, and potentially the information available from indoor imagery. Further analysis in which the number of photographs was held constant for the different data sources would be required to help distinguish between these two potential influences.

Furthermore, the CNN we use can distinguish between furnished and less furnished rooms. We find that well-decorated rooms are associated with higher median house prices, whereas more simple rooms are associated with lower median house prices. There is however still room for model improvement. Incorporating external factors influencing house prices that are not captured visually, such as the number of schools in an area or the mean distance of a property to a tube station, would be likely to improve model accuracy further (Efthymiou & Antoniou, 2013; Gibbons & Machin, 2003, 2005; Seo & Simons, 2009).

In addition to estimating local house prices from indoor Zoopla images, we also estimate relative changes in median house price of local areas in London from 2005 and 2006 to 2015 and 2016. Once again, we achieve superior model performance in comparison to inferences made from the outdoor image sources Google Street View and Geograph. However, similar model limitations exist when using indoor Zoopla images. For example, we are not making use of information regarding visual changes over time. The reason we do not explore how images evolve over time is the difficulty we experienced in pinpointing the exact dates that the images are taken from our dataset whilst retaining a sufficiently large amount of visual information. If future analyses could address this problem, they could separate the visual information available from different time periods and measure visual changes over time to investigate whether changes in visual appearance could be related to changes in median house price.

Since we also have access to outdoor Zoopla imagery, we investigate how well we can infer median house prices and relative changes in median house prices across areas in London using outdoor images from Zoopla instead. The outdoor Zoopla images contain visual information similar to both Google Street View and Geograph images, as they consist of housing facades and gardens. We again find that performance is better when using outdoor Zoopla images than when using photos from Google Street View and Geograph, although not as good as when using indoor Zoopla images. Again, this may be due to the discrepancy in the number of images used for each analysis.

We also investigate how performance is impacted if we use both indoor and outdoor Zoopla imagery together. We find that estimating median house prices and relative changes in median house prices across areas in London from all Zoopla images leads to only marginally improved accuracy relative to making inferences from indoor Zoopla images alone. It is not clear whether modelling with a larger volume of images improves accuracy of house price estimates up to a certain point, after which we see diminishing returns, or whether adding photos of a different kind brings limited benefits. To investigate this further, we could look at varying the sample size within

a set of homogeneous photographs (e.g., indoor photos alone).

To summarise, we find that indoor housing imagery can be used with deep learning tools to infer house prices across local areas of London, as well as relative changes in local house prices. Our findings further suggest that making use of both interior and exterior image data of properties only marginally improves estimation performance in comparison to using interior imagery alone. How would performance differ if we utilised visual information from multiple data sources at once? We investigate this question further in Chapter 6.

Chapter 6

Combining Google Street View, Geograph, and Zoopla images

6.1 Introduction

Throughout this thesis, we provide evidence suggesting that it is possible to infer local house prices in London from relevant visual imagery of the city. Both Chapter 3 and Chapter 4 investigate the use of street-level imagery of the outdoors, from visuals of building facades to off-road areas. Chapter 5 primarily focuses on the indoor imagery of housing interiors from Zoopla, with some additional analyses focused on the value of outdoor Zoopla imagery.

Overall, a diverse range of visuals is available to exploit from each data source, whether it be outdoor or indoor information. We also note that house price estimation models using each data source independently achieve varying levels of success. Finally, we discover that utilising visual information from both indoor and outdoor Zoopla images only marginally improves the performance of our house price estimation models in comparison to using only indoor images.

In this chapter, we consider all of our image sources as a collective. How does model performance compare if we harness the visual information contained in Google Street View, Geograph, and Zoopla images simultaneously?

6.2 Data and methods

6.2.1 House price data

In this chapter, we once again focus on property transactions from 2015 and 2016. Recall that the house price data are skewed with a long positive tail. Hence, we

calculate the median house price for each MSOA. As a reminder, a Middle Layer Super Output Area (MSOA) contains a population of between 5,000 and 15,000 inhabitants, or between 2,000 and 6,000 households, with 983 MSOAs in London (Stokes, 2012). Figure 6.1 illustrates our methodology.

6.2.2 Images of outdoor and indoor environments from different data sources

As described in previous chapters, we have access to images of outdoor and indoor environments from multiple data sources: namely, Google Street View (Chapter 3), Geograph (Chapter 4), and Zoopla (Chapter 5). We have more coverage of London for a longer period of time from both Geograph and Zoopla images, but only images from 2014 and 2015 from Google Street View images, due to our methodology. Therefore, we add extra Google Street View images to this analysis, in addition to the data that we analysed in Chapter 3, for comparability. Google Street View launched in 2007 and started capturing images of London from 2008, so we add Google Street View images from 2008 and 2009 to the original selection of images from 2014 and 2015 only. This allows us to use visual information that spans as far back as possible.

To access images from each data source, we use the methods described in their respective analysis chapters. More specifically, we use the methods described in Section 3.2.1 to access Google Street View images, in Section 4.2.2 to access all Geograph images, and in Section 5.2.3 to access both indoor and outdoor Zoopla images. When we refer to the Zoopla images, we are referring to the dataset and analyses using both indoor and outdoor images unless specified otherwise.

Since we are inferring median house prices at MSOA level, we collate the images at MSOA level. We then explore how the images are distributed across London at MSOA granularity in Fig. 6.2. For each map, a darker MSOA indicates a larger quantity of images from the named data source. The number of Google Street View images in each MSOA ranges from 176 to 4,328 with a median of 986 images per MSOA (Fig 6.2a); the number of Geograph images in each MSOA ranges from 8 to 14,780 with a median of 151 images per MSOA (Fig 6.2c); and the number of Zoopla images in each MSOA ranges from 617 to 9,907 with a median of 3,568 images per MSOA (Fig 6.2e). Our Zoopla dataset contains the most images with a total of 3,727,890 images of London, followed by our Google Street View dataset with 1,032,439 images of London, and finally, our Geograph dataset contains 273,915 images of London.

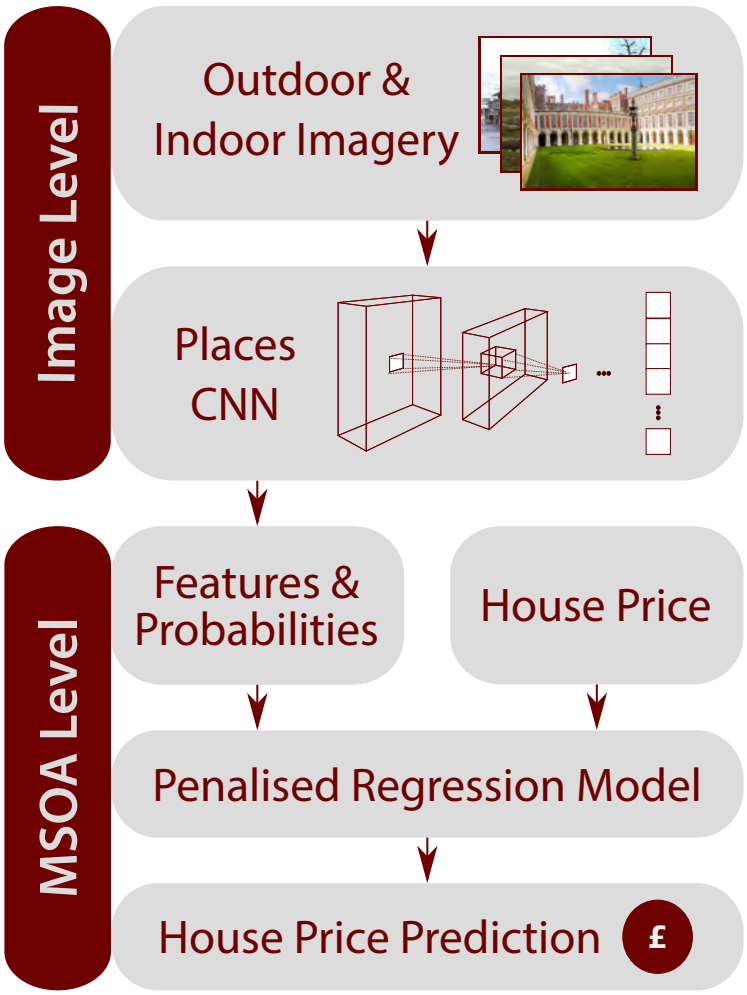

Figure 6.1: Deep learning framework. We adapt the deep learning framework from the Google Street View (Chapter 3), Geograph (Chapter 4), and Zoopla (Chapter 5) analyses to infer house prices from images. We access images from each data source as described in their respective chapters and assign each image to an MSOA. To process the images, we use the Places365 CNN (Zhou et al., 2018) once more, which produces a vector of 365 feature values for each image we give as input. Each of the 365 feature values relates to the likelihood of the corresponding Places feature occurring in the image. Example features that Places365 can identify include 'driveway' and 'hotel room'. We then calculate mean feature vectors for each MSOA and train an elastic net to learn the relationship between the mean feature vector and the median house price for a given MSOA. Finally, we test the estimates from the model using a leave-one-out cross-validation and compare the resulting predictions to the actual median house price for each MSOA. All Geograph photos are licensed for reuse under creativecommons.org/licenses/by-sa/2.0. Front photo © David Dixon; middle photo © Robert Lamb; back photo © Jim Osley.

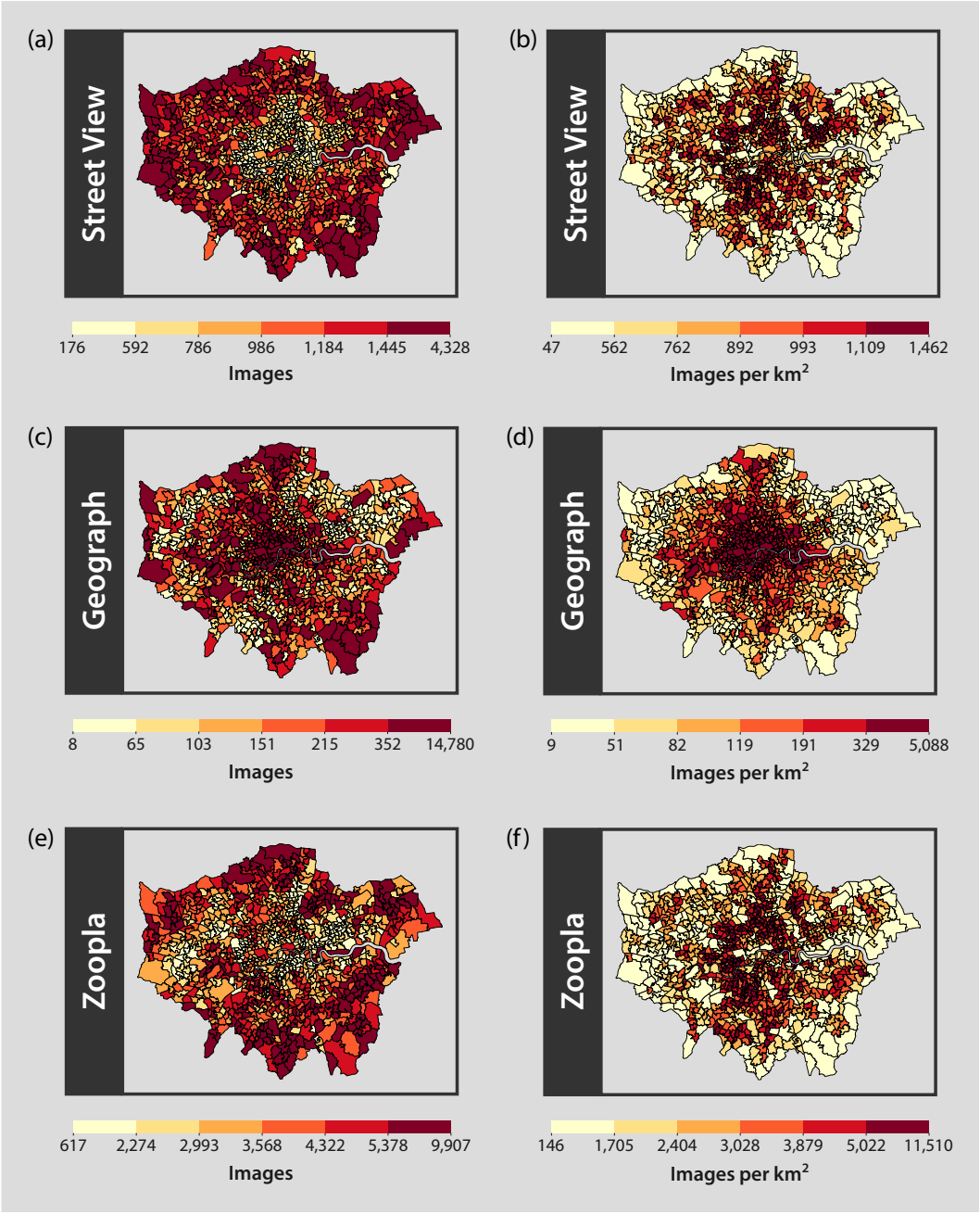

Figure 6.2: The distribution of images across London. How are images distributed across local areas of London for each of the data sources we are exploiting? **(a)** Map of the number of Google Street View images in each MSOA. A darker area represents a larger quantity of images in that area. *(continues on the following page)*

Figure 6.2: *(continues from previous page)* The number of images in each MSOA ranges from 176 to 4,328 with a median of 986 images per MSOA. **(b)** Map of the number of Google Street View images per km^2 for each MSOA. A darker area represents a larger number of images per km^2. Notice that the number of images per km^2 is not uniform. This is due to our methodology for retrieving Google Street View images (Section 3.2.1). **(c)** Map of the number of Geograph images in each MSOA. Again, a darker area represents a larger quantity of images in that area. The number of images in each MSOA ranges from 8 to 14,780 with a median of 151 images per MSOA. **(d)** Map of the number of Geograph images per km^2 for each MSOA. Again, a darker area represents a larger number of images per km^2. It is clear that there is a greater density of Geograph images in MSOAs in the centre of London. **(e)** Map of the number of Zoopla images in each MSOA. Once more, a darker area represents a larger quantity of images in that area. The number of images in each MSOA ranges from 617 to 9,907 with a median of 3,568 images per MSOA. **(f)** Map of the number of Zoopla images per km^2 for each MSOA. Again, a darker area represents a larger number of images per km^2. Visual inspection reveals that the photographs do not have uniform density across London. This is likely to reflect a greater density of houses, or at least houses for sale, in certain MSOAs. For all maps, quantile breaks are implemented with six bins. Contains National Statistics data © Crown copyright and database right 2011. Contains Ordnance Survey data © Crown copyright and database right 2011. Contains Street View data © 2017 Google. Contains Geograph data © Geograph Project Limited and licensed for reuse under creativecommons.org /licenses/by-sa/2.0. Contains data from Zoopla Limited, © 2018. Zoopla Limited. Economic and Social Research Council. Zoopla Property Data 2018. University of Glasgow - Urban Big Data Centre.

There are vast differences in the areas of different MSOAs in London. The largest MSOA in London has an area of 22 km^2, and the smallest an area of 0.29 km^2. London itself has an area of 1,600 km^2 with a median area of 1.1 km^2 (all areas to 2 significant figures). For this reason, in addition to examining the count of images for each data source per MSOA, we also calculate the number of images per km^2 for each MSOA. On the image density maps we produce for each of the data sources, a darker MSOA indicates a larger number of images per km^2 for that MSOA. The number of Google Street View images per km^2 in each MSOA ranges from 47 to 1,462 with a median of 892 (Fig 6.2b); the number of Geograph images per km^2 in each MSOA ranges from 9 to 5,088 with a median of 119 (Fig 6.2d); and the number of Zoopla images per km^2 in each MSOA ranges from 146 to 11,510 with a median of 3,028 (Fig 6.2f).

Given our method for collecting Google Street View images, we would expect the outer, larger MSOAs to have more coverage than the inner, smaller MSOAs. This does appear to be the case, as visible in Fig. 6.2a, but the corresponding density map reveals that the number of images per km^2 is not uniform (Fig. 6.2b). Recall that we defined a grid of squares over London of 100 metres by 100 metres and at the corner of each square, we took the latitude and longitude co-ordinates. We then identified the Google Street View panorama IDs nearest to the co-ordinate pairs to retrieve images. As such, the nearest panorama image may not necessarily be from the MSOA in which the co-ordinate pair is located.

The reason for this is that the distribution of roads across London is not uniform. For example, some areas in the centre of London near the River Thames have a higher density of roads than the remainder of London, whereas some areas in the east of London have a smaller network of roads, and more parks, rivers, and woods. As such, the distribution of Google Street View photos is also not uniform, as the photos are usually collected when driving a car around areas. If a Google Street View image is unavailable at an exact latitude and longitude pair, perhaps because there is no road there, the nearest location where an image is available could be in another MSOA. Therefore, the number of Google Street View images of London per km^2 is not uniform. Indeed, the MSOAs that have lower photo densities tend to correspond to areas with a lower density of roads.

Recall that Google Street View images contain visuals of streets and building facades, as well as snippets of the environment around buildings, and we know that the images cover the majority of the street network in London. As such, properties in woodland areas without streets are less likely to be captured by Google Street View, but a large variety of properties are represented by the images, from council houses to mansions. Therefore, we expect that house prices can be accurately reflected by this data source. Some information associated with house prices will inevitably be missing, however, since Google Street View images do not include visuals of the interiors of houses and are restricted to streets. Additionally, we would expect that the features that the elastic net associates with higher and lower house prices are representative of London.

From Fig. 6.2c, we observe that there is good coverage of Geograph images across MSOAs in London, but that there are more images per km^2 in the centre of London than in the outskirts of London (Fig. 6.2d), suggesting that more photos are taken in the centre. Geograph images are not restricted to visuals of streets, but also contain characteristics of the environment, such as parks and rivers, and the areas covered by Geograph images are those frequented by Geograph users, which is

a specific group of people. As such, the images may not be representative of London as a whole, but are representative of places that are popular with Geograph users. Therefore, we expect that houses prices may not be reflected as accurately with Geograph imagery in terms of London as a city, but the images will accurately reflect the house prices in areas popular with Geograph users, perhaps tourists or those living in the areas photographed. We would also expect that the features that the elastic net associates with higher and lower house prices are more environment-focused from Geograph imagery than Google Street View imagery.

The distribution of Zoopla images (Fig. 6.2e) shows that all MSOAs have a large volume of images, but comparing this distribution map to the distribution of images per km^2 of MSOA (Fig. 6.2f), we find that the photographs do not have uniform density across London. In fact, the density map for Zoopla images is similar to the density map for Google Street View images. This suggests that with Zoopla images, the photo density reflects the density of roads and houses for sale in London, and the Zoopla images are representative of areas that people tend to buy houses in.

We know that Zoopla images contain information about both the interiors and exteriors of houses, representing interior visuals that are not contained in Google Street View or Geograph images and exterior visuals that are partially contained in Google Street View and Geograph images. Due to the nature of the source, there is great coverage in areas of London where houses are, as Zoopla is a platform geared towards buying and selling housing. Therefore, we expect accurate conclusions to be drawn about house prices from this dataset. We further expect the addition of indoor data to provide good estimates of house prices, as the interiors may reflect the characteristics and feel of an area in London through decor and furnishing.

Each of the three data sources captures slightly different information about housing in London, and we need to be aware of the issues that may be associated with each data source to accurately interpret results based on them, as discussed in Section 2.1.1.2. Images accessed from Google Street View may contain personal and identifying information about people, such as the people themselves, if they are walking down a street at the same time that the street is being photographed by Google, or their license plates. Google have made the effort to automatically blur personal information to maintain the privacy of those that would be affected by the images and if some information is not obscured, it can be reported and rectified (Google, n.d.-a). The images may also be biased towards areas that contain large proportions of the street network, and would not capture properties for sale in other locations. However, since most houses for sale are on the street network, there is enough information to draw accurate conclusions from.

As mentioned previously, Geograph images are taken by members of the public that use Geograph, so the data are likely to have some population bias because Geograph users are not representative of the people living in London. However, the images do contain some information that is useful and we take the bias into account when interpreting our results. Since the photos are taken by members of the public, they do not always obscure the people that may be in the photographs, but this may not be an issue if appropriate permissions are gained. The quality of images also varies across the Geograph dataset, as the images can have any resolution, but they are generally high resolution images.

The quality of data is an issue in terms of Zoopla images too for the same reason. Additionally, images of properties on Zoopla tend to display the best features of a property in the best light, and may not always reflect the property accurately. The data themselves are also slightly biased towards houses that are bought relatively frequently to houses that do not change owners very often, which may be unrepresented by the data. This is especially the case if a house has not been resold for over 16 years, as the dataset spans 16 years.

Generally, areas across London are well represented by each of the data sources individually. They all operationalise in different ways but capture an essence of housing from a human perspective that is at least somewhat indicative of their prices. Although each of the sources are biased in some way, we do not currently have a standard data source that is representative of all areas of London, so we take the biases into consideration when drawing conclusions from the data sources. The biases will also need to be taken into consideration when using the data sources together. To understand how model performance would compare when using information from Google Street View, Geograph, and Zoopla together, we first create a dataset that contains details of all the images available to us. When combining all three data sources, we find that the number of images in each MSOA ranges from 1,117 to 21,351 with a median of 4,843 images per MSOA and a total of 5,034,244 images.

As well as investigating model performance when using combinations of visual information from all three data sources, we also investigate the use of combinations of parts of the data sources. Specifically, the combinations of images we study are:

1. Google Street View, Geograph, and Zoopla images
2. Google Street View and Geograph images
3. Google Street View and Zoopla images
4. Geograph and Zoopla images

6.2.3 Places CNN

In Chapter 4 and Chapter 5, we automatically extract visual features from all of the Geograph and Zoopla images we have access to using the updated Places CNN, Places365 (Zhou et al., 2018). However, in Chapter 3, we extract visual features from Google Street View images taken during 2014 and 2015 using the original Places CNN, Places205 (Zhou et al., 2014). To maintain consistency with the features extracted from Geograph and Zoopla images, we use Places365 to automatically extract visual features from our complete Google Street View dataset too.

To process the vast amount of Google Street View images in our dataset in parallel, we use high-performance computing clusters managed by the Scientific Computing Research Technology Platform at the University of Warwick, due to their ability to process data rapidly and efficiently. We use Places365 to extract the feature values of the 365 scene features that Places365 can recognise from images. Again, the feature values can be interpreted as the likelihood of each feature appearing in the given image. For each image, we store the feature values in a vector of length 365. Recall that the feature values range from 0 to 1 and the feature vector sums to 1.

For all of the images, we keep both indoor and outdoor features in our feature vectors. This is because all combinations of data sources except one include Zoopla images, which capture both indoor and outdoor scenes. For evaluation of the performance of the house price estimation model using only outdoor images from Google Street View and Geograph, we still keep all 365 features to maintain consistency with the other combinations of data sources we analyse.

Given that we are analysing house price data at MSOA level, we also aggregate the visual features data to MSOA level. There are multiple images available for each MSOA in London. We focus on two methods for aggregating the visual features to MSOA level:

1. **Mean features method:** We extract feature vectors from all of the images from the sources we are exploring. For example, if we were exploring the Google Street View and Geograph combination, we would have feature vectors for each image from those sources. Then for each MSOA, we collate the feature vectors extracted from the images allocated to the MSOA and calculate the mean feature vector. Each feature value in the mean feature vector is the mean feature value for the corresponding feature from all of the feature vectors associated with the MSOA. This leaves us with one feature vector of length 365. This is essentially the same method as we used in previous chapters where we analyse images from each of our data sources individually.

2. **Concatenated features method:** For each data source, we calculate the mean feature vector for each MSOA. For example, if we were investigating the combination of Geograph and Zoopla, we would have two mean feature vectors of length 365: one containing the visual features from Zoopla images alone, and one containing visual features from Geograph images alone. We then concatenate these feature vectors for each MSOA, resulting in a longer mean feature vector associated with each MSOA. For example, with the Geograph and Zoopla combination, we take the mean feature vectors from both sources and concatenate them for each MSOA. This leaves us with one feature vector per MSOA of length $365 \times 2 = 730$.

We explore the impact of the choice of method for aggregating visual features throughout the chapter.

6.2.4 Elastic net model

As in the Google Street View, Geograph, and Zoopla analyses, the visual characteristics of London for each MSOA are represented by a mean feature vector, where these vectors are calculated as described above. The feature vectors are then input into our penalised regression model to estimate local house prices in London. Once again, we use an elastic net model to help us manage correlations between the features. We train and test our elastic net model using the leave-one-out approach described in the Google Street View analysis (Section 3.2.4).

6.2.5 Measuring change in house prices

As well as comparing the performance of models that seek to estimate median house prices in London, we compare the performance of models that seek to estimate the relative change in median house prices in London. In the Google Street View, Geograph, and Zoopla analyses, we explore relative changes in median house prices from 2005 and 2006 to 2015 and 2016 at MSOA granularity. In this chapter, we follow the same approach and explore changes over the same decade. Recall that median house prices during 2015 and 2016 at MSOA level range from £69,460 to £1,270,000 with a median of £416,875, and median house prices during 2005 and 2006 at MSOA level range from £135,000 to £1,475,000 with a median of £234,000.

To rank MSOAs by median house price, we again set the highest rank of 983 to the MSOA with the highest median house price and the lowest rank of 1 to the MSOA with the lowest median house price. If two areas have the same median house price, we give them both the mean rank. The change in median house price rank is

then calculated by subtracting the 2005–2006 rank from the 2015–2016 rank. The changes in ranked price again range from -424.5 to 522 with a median of -4. For example, a change in ranked price of -100 indicates a decrease in rank of 100 places, hence a relative rank reduction in median house price over the decade.

Again, we model the change in median house price rank using an elastic net model. The relevant feature vector is given as input to the elastic net, which outputs a prediction for the change in ranked price.

6.3 Results

6.3.1 Estimating local house prices and changes in local house price rank in London using an extended set of Google Street View images

Before investigating the performance of our house price estimation model when using visual features from multiple data sources, we revisit a comparison of performance when using visual features from each of our data sources individually. As mentioned in Section 6.2.2, we previously used Google Street View images to estimate median house prices in MSOAs of London by extracting the visual features from a sample of images taken during 2014 and 2015 using the Places205 CNN (Chapter 3). Here, we refer to this as the original Google Street View model. In the original Google Street View model, we make use of 518,808 photos, where the number of photos in each MSOA ranges from 88 to 2,164 with a median of 496 photos per MSOA.

However in this chapter, we wish to also use visual features extracted from 513,631 Google Street View images taken during 2008 and 2009 in order to increase comparability between the periods of time covered by each of the three data sources, as discussed in Section 6.2.2. This leaves us with almost double the number of images, at 1,032,439 in total, where the number of images in each MSOA ranges from 176 to 4,328 with a median of 986 photos per MSOA. Furthermore, we extract visual features from all of the Google Street View images using the updated Places365 CNN, to align our feature extraction method with that used for the Geograph (Chapter 4) and Zoopla (Chapter 5) images. Does the performance of our model for median house price estimation differ when using this updated Google Street View model instead of the original Google Street View model?

To build the updated Google Street View model, we generate estimates of median house prices across MSOAs of London using the methodology outlined in Fig. 6.1. We find the optimal value of *l1_ratio* to be 0.9999999 with an α of 1.45 (to 3 significant figures) during the process of fitting the elastic net model. A regression

analysis comparing the estimates of median house prices with the true median house prices for MSOAs in London reveals that the updated model performs reasonably well ($R^2 = 0.46$, $N = 983$, $p < 0.001$) but not as well as the original model ($R^2 = 0.56$, $N = 983$, $p < 0.001$). The lower accuracy is clear when comparing the mean absolute error and mean absolute percentage error from both models. For the updated model, the mean absolute error and mean absolute percentage error are £114,000 and 24.9%, whereas for the original model, the mean absolute error and mean absolute percentage error are £97,700 and 20.2% (all errors to 3 significant figures).

In addition to estimating median house prices in London, we draw on the updated dataset and feature extraction approach to estimate changes in ranked median house prices across areas of London from 2005 and 2006 to 2015 and 2016 using the methodology outlined in Fig. 6.1. When fitting the elastic net, we find the optimal value of *l1_ratio* to be 0.9999 with a corresponding α of 0.00493 (to 3 significant figures). We compare the estimates of changes in ranked price with the true changes in ranked price for MSOAs in London via a regression analysis. Once again, we discover that the updated model can capture some changes in ranked median house price in areas of London ($R^2 = 0.22$, $N = 983$, $p < 0.001$) but does not perform as well as the original model ($R^2 = 0.27$, $N = 983$, $p < 0.001$).

From the Geograph and Zoopla analyses, we found that using more images to estimate median house prices and changes in median house prices improves model performance, even if the images are from an older time period. Therefore, it is surprising that this does not appear to be the case with Google Street View images. The cause for worsening performance could be that the Google Street View images taken during 2008 and 2009 in our dataset are not useful, or that the additional scene understanding (SUN) features extracted for the original Google Street View model in Chapter 3 played a key role in the performance of the model, or that using Places365 to extract visual features from Google Street View images is less useful than using Places205. We can determine whether either of the latter two explanations may be true by estimating median house prices and changes in median house prices across areas of London from Google Street View images taken during 2014 and 2015 only but extracting features using Places365.

In fact, using Places365 on Google Street View images taken during 2014 and 2015 alone leads to improved model performance when estimating median house prices ($R^2 = 0.59$, $N = 983$, $p < 0.001$) and changes in median house price rank ($R^2 = 0.30$, $N = 983$, $p < 0.001$). This performance is in fact marginally better than the performance of the original Google Street View models (median house price estimation: $R^2 = 0.56$, $N = 983$, $p < 0.001$; changes in median house price rank:

$R^2 = 0.27$, $N = 983$, $p < 0.001$).

Based on these results, we determine that model performance is worse when using all Google Street View images due to the information available from older images. We do not explore this further in this chapter. However, given that the Geograph model performed better when older photos were included, we suggest that the information from Google Street View images taken during 2008 and 2009 may be useful in conjunction with images from other data sources. For this reason, we continue to use the updated Google Street View dataset in models that combine multiple data sources. For the rest of the chapter, we therefore refer to the updated Google Street View analyses as simply the Google Street View analyses.

6.3.2 Estimating local house prices and changes in local house price rank in London using all available images and the mean features method

How useful is the visual information available from Google Street View, Geograph, and Zoopla images for estimating median house prices and changes in median house prices when combined together? We initially build models using the mean features method to extract visual characteristics from all of the images. As described in Section 6.2.3, we extract feature vectors from all of the images from all three data sources. For each MSOA, we then collate the feature vectors extracted from the images associated with the MSOA and calculate the mean feature vector. Recall that the number of images in each MSOA from the dataset consisting of all three data sources ranges from 1,117 to 21,351 with a median of 4,843 images per MSOA. We analyse a total of 5,034,244 images.

We first estimate median house prices across areas of London during 2015 and 2016 using all three data sources. We will refer to models using all three image sources as combination models. Employing the elastic net results in an optimal value of 0.999999 for *l1_ratio* with a corresponding α of 0.436 (to 3 significant figures). We discover that the combination model performs very well ($R^2 = 0.82$, $N = 983$, $p < 0.001$; mean absolute error: £65,700; mean absolute percentage error: 13.8%), almost matching the performance of the Zoopla model ($R^2 = 0.82$, $N = 983$, $p < 0.001$; mean absolute error: £65,300; mean absolute percentage error: 13.7%). It seems that including more images in this analysis does not improve model performance and may even slightly distract from the information gained from Zoopla imagery alone.

Next, we estimate changes in median house price rank from 2005 and 2006 to 2015 and 2016 across MSOAs in London. During the fitting of the elastic net, we

find the optimal value of *l1_ratio* to be 0.97 with an α of 0.000118 (to 3 significant figures). A regression analysis reveals that the combination model performs better than the Zoopla model by the smallest margin (combination model: $R^2 = 0.44$, $N = 983$, $p < 0.001$; Zoopla model: $R^2 = 0.43$, $N = 983$, $p < 0.001$).

6.3.3 Estimating local house prices and changes in local house price rank in London using all available images and the concatenated features method

How does house price estimation differ if we extract visual characteristics from images via the concatenated features method? As with the mean features method, we build models to estimate median house prices and changes in median house prices from Google Street View, Geograph, and Zoopla images together. As described in Section 6.2.3, we first calculate the mean feature vector for each MSOA for each data source. We then concatenate these feature vectors for each MSOA, resulting in a larger mean feature vector associated with each MSOA. Since we are concatenating features extracted from Google Street View, Geograph, and Zoopla, the feature vector we input into the elastic net is of length 1,095. Again, the number of images in each MSOA ranges from 1,117 to 21,351 with a median of 4,843 images per MSOA, and we analyse a total of 5,034,244 images, as with the analyses using the mean features method.

To estimate median house prices during 2015 and 2016 in MSOAs of London using all three data sources, we implement an elastic net model once more. We find the optimal value of *l1_ratio* to be 0.999999 with a corresponding α of 0.598 (to 3 significant figures). We also find that the model performs as well as the combination model using the mean features method (both models: $R^2 = 0.82$, $N = 983$, $p < 0.001$; both models, mean absolute error: £65,700; concatenated features method, mean absolute percentage error: 13.7%; mean features method, mean absolute percentage error: 13.8%). Overall, using concatenated features does not significantly affect performance in comparison to using mean features to estimate local house prices.

We also estimate changes in ranked median house price across local areas of London from 2005 and 2006 to 2015 and 2016. When employing the elastic net, we find the optimal value of *l1_ratio* to be 0.98 with an α of 0.000135 (to 3 significant figures). Once again, the results mirror those from the equivalent mean features model (both models: $R^2 = 0.44$, $N = 983$, $p < 0.001$).

6.3.4 Comparing models estimating median house prices in London during 2015 and 2016

We then compare the performance of models using one, two, or three data sources, and either the mean features or concatenated features method when building the models. We start by comparing the models that estimate median house prices during 2015 and 2016 across MSOAs in London. Table 6.1 shows the performance statistics for all of the models we build and discuss in this chapter. For each model, we report the combination of data sources used and whether the model uses the mean features method or concatenated features if more than one data source is used; the coefficient of determination (R^2); the mean absolute error (MAE, £) and the mean absolute percentage error (MAPE, %); and the total number of images we extract visual features from. The table is ordered in terms of model performance, where the best-performing model is first.

If we compare the models that only use one image source, we find that the use of Zoopla imagery leads to the model with the best performance ($R^2 = 0.82$, $N = 983$, $p < 0.001$) for MSOAs in London. The next best model uses Google Street View imagery ($R^2 = 0.46$, $N = 983$, $p < 0.001$) but is closely followed by the model using Geograph images ($R^2 = 0.44$, $N = 983$, $p < 0.001$). The differing performance may be due to the number of images we use to make inferences. Specifically, we make use of 3,727,890 Zoopla photos, 1,032,439 Google Street View photos, and only 273,915 Geograph photos of London, such that there is more visual information available from the Zoopla dataset for the model to learn from. Alternatively, the information available from Zoopla images may simply be more relevant and useful in visually capturing factors that relate to house prices across MSOAs in London.

Table 6.1: Model performance statistics for all models inferring median house prices in London during 2015 and 2016. We combine images from Google Street View, Geograph, and Zoopla in all possible variations, and generate estimates using the methodology outlined in Fig. 6.1. This table displays the performance statistics for each of the models. For each model, the table shows: the data source combination and if applicable, whether the model uses the mean features method or concatenated features method; the coefficient of determination (R^2), by which we order the table in order to highlight the models that perform most accurately; the mean absolute error (MAE, £); the mean absolute percentage error (MAPE, %); and the total number of images used to extract visual features. R^2, MAE and MAPE are presented to 3 significant figures. The model with the best fit uses visual information from Zoopla and Geograph images ($R^2 = 0.834$ $N = 983$, $p < 0.001$), and the difference between using mean features and concatenated features is negligible. However, this combination of imagery only leads to a slight improvement over Zoopla imagery alone ($R^2 = 0.821$, $N = 983$, $p < 0.001$), despite drawing on 1,032,439 more images. In terms of efficiency, it appears more sensible to estimate median house prices from Zoopla imagery alone, especially since all of the best-performing models include Zoopla images.

Model Input			Combination		Model Performance			Images
Google	**Geograph**	**Zoopla**	**Mean**	**Concatenated**	$\boldsymbol{R^2}$	**MAE (£)**	**MAPE (%)**	**Total**
X		X		X	0.834	64 100	13.5	4 760 329
X		X	X		0.830	64 900	13.7	4 760 329
		X			0.821	65 300	13.7	3 727 890
X	X	X	X		0.819	65 700	13.8	5 034 328
X	X	X		X	0.816	65 700	13.7	5 034 328
	X	X	X		0.807	66 600	14.0	4 001 889
	X	X		X	0.805	66 600	13.9	4 001 889
X	X		X		0.512	107 000	23.0	1 306 438
X	X			X	0.512	108 000	23.4	1 306 438
X					0.465	114 000	24.9	1 032 439
	X				0.444	115 000	24.6	273 999

Examination of the table reveals that the best-performing model uses Google Street View and Zoopla imagery ($R^2 = 0.834$, MAE = £64,100, MAPE = 13.5%). In essence, adding Google Street View images to Zoopla images boosts the number of outdoor images in the dataset, helping ensure that both the interiors and exteriors of properties are well represented. Whether the model uses the mean features method or the concatenated features method has a negligible impact on performance. Importantly however, we note that using both Google Street View and Zoopla images only slightly improves model performance in comparison to Zoopla imagery alone ($R^2 = 0.821$, MAE = £65,300, MAPE = 13.7%). Furthermore, only using Zoopla imagery reduces the number of images that need analysing by 1,032,439, which would make the process of building the model faster and more efficient.

In fact, all of the best-performing models use Zoopla imagery. The next best-performing model uses both Google Street View and Geograph photos, but there is a clear reduction in accuracy without the Zoopla photos ($R^2 = 0.512$, MAE = £107,000, MAPE = 23%). However, the Google Street View and Geograph model uses 2,421,452 less photos than the Zoopla model. It is therefore unclear whether insights available from indoor imagery are more valuable when estimating median house prices than those available from outdoor imagery, or whether we simply need a large number of images from one source.

In Table 6.2, we report the elastic net parameters and coefficient statistics for all of the models we build. For each model, we show the combination of data sources used and whether the model uses the mean features method or concatenated features if more than one data source is used, exactly like Table 6.1; the optimal value of *l1_ratio* and the corresponding α from fitting the elastic net; the total number of non-zero coefficients we extract from the elastic net; and of the top 50 positive and negative elastic net coefficients, the number of coefficients extracted from each data source if the model uses the concatenated features method. 'S' represents Google Street View features, 'G' represents Geograph features, and 'Z' represents Zoopla features. Once again, the table is ordered in terms of model performance, where the most accurate model is first.

Inspecting the elastic net parameters reveals that *l1_ratio* > 0.99999 for all models, so all of the models are essentially lasso models. As described in Section 2.2.2, these models focus on tackling overfitting more than tackling multicollinearity. Furthermore, the accompanying αs are generally very small.

Table 6.2: Model configuration for all models inferring median house prices in London during 2015 and 2016. We combine images from Google Street View, Geograph, and Zoopla in all possible variations, and generate estimates using the methodology outlined in Fig. 6.1. This table displays the model parameters and a breakdown of the coefficients for each of the models. For each model, the table shows: the data source combination and if applicable, whether the model uses the mean features method or concatenated features method; the optimal value of *l1_ ratio* from fitting the elastic net; the corresponding α; the total number of non-zero coefficients extracted from the model; and the number of coefficients extracted from each data source in the top 50 positive and negative coefficients, if the model uses the concatenated features method. When using concatenated features, as described in Section 6.2.3, we can distinguish between the features we extract from each data source. 'S' represents features extracted from Google Street View images, 'G' represents those extracted from Geograph images, and 'Z' represents those extracted from Zoopla images. Note that the α is given to 3 significant figures. All models show *l1_ ratio* > 0.99999, indicating that all of the elastic net models are close to lasso models. The αs are also generally small. If the model uses concatenated features when it uses Zoopla imagery, at least half of the features in the top 50 positive and negative coefficients are also Zoopla features.

Model Input			**Combination**		**Elastic Net Parameters**		**Coefficients**			
Google	**Geograph**	**Zoopla**	**Mean**	**Concatenated**	***l1_ ratio***	**α**	**Total**	**S**	**G**	**Z**
X		X		X	0.999 999	0.251	242	21		79
X		X	X		0.999 999	0.210	203			
		X			0.999 99	0.157	227			
X	X	X	X		0.999 999	0.436	210			
X	X	X		X	0.999 999	0.599	248	9	39	52
	X	X	X		0.999 999	0.551	200			
	X	X		X	0.999 999	0.675	222		41	59
X	X		X		0.999 999 9	10.5	52			
X	X			X	0.999 99	0.475	471	30	70	
X					0.999 999 9	1.45	90			
	X				0.999 99	0.758	302			

For models using the concatenated features method, we also report the number of coefficients extracted from each data source out of the top 50 positive and negative elastic net coefficients. These features have the most impact on the estimates we generate. We observe that each of the four models using the concatenated features method and Zoopla images are heavily influenced by Zoopla features extracted from Zoopla images. In the Google Street View and Zoopla model, 79 out of 100 features are Zoopla features; in the combination model, 52 out of 100 features are Zoopla features; and in the Geograph and Zoopla model, 59 out of 100 features are Zoopla features.

The impact of Zoopla features can be seen in Fig. 6.3, which depicts the elastic net coefficients we extract from the model using the concatenated features method with images from all three data sources. Since we use concatenated features, we need to distinguish between the features associated with each data source. In the figure, 'S' represents Google Street View, 'G' represents Geograph, and 'Z' represents Zoopla. The features extracted from Zoopla images dominate the top coefficients and are often indoor features. Of the top ten positive coefficients, six are extracted from Zoopla images, and of the top ten negative coefficients, seven are extracted from Zoopla images.

Moreover, we note that only a few of the top coefficients are extracted from Google Street View. In fact, only nine come from Google Street View, whereas 39 come from Geograph. This could be because the Geograph dataset includes images of both building facades and gardens, whereas the Google Street View dataset mainly includes visuals of building facades. As such, perhaps more useful information relating to local house prices is gained from more off-road visuals, rather than the exteriors of buildings. The Zoopla dataset also includes indoor imagery, which may complement the Geograph imagery well, providing a clearer landscape of houses and their surroundings for the elastic net to learn from.

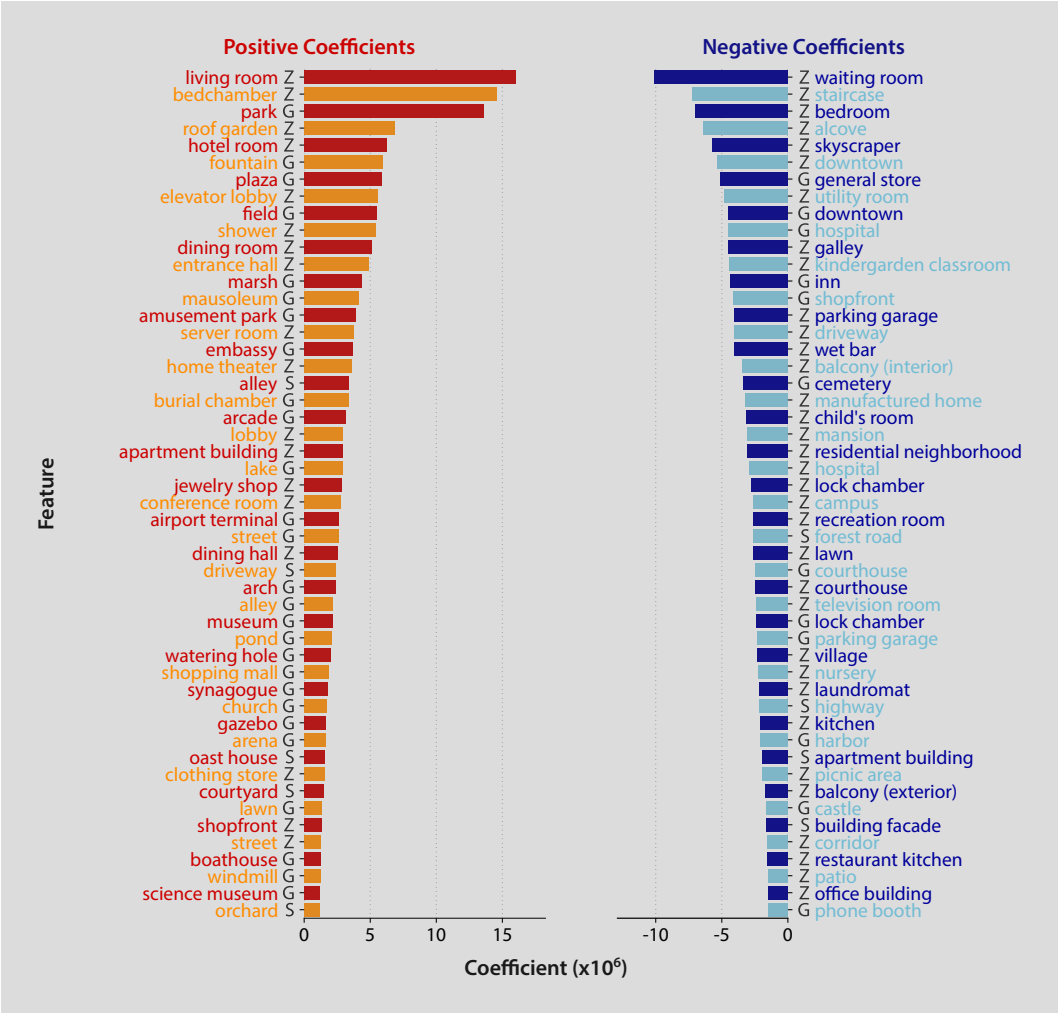

Figure 6.3: Top 50 elastic net coefficients for estimates of London house prices during 2015 and 2016 from all images using the concatenated features method. We train an elastic net model to estimate median house prices using the methodology outlined in Fig. 6.1 and display the resulting coefficients. We use the concatenated features method, as described in Section 6.2.3, where we independently calculate the mean feature vector for each MSOA for each data source, then concatenate these feature vectors for each MSOA. Using the concatenated features method, we can distinguish between the features we extract from each data source. 'S' represents Google Street View, 'G' represents Geograph, and 'Z' represents Zoopla. Positive coefficients reflect features that correspond to higher median house prices and negative coefficients reflect features that correspond to lower median house prices. *(continues on the following page)*

Figure 6.3: *(continues from previous page)* Most of the features deemed important by the elastic net are extracted from Zoopla images and tend to be indoor features. Of the top ten positive coefficients, six are extracted from Zoopla images, and of the top ten negative coefficients, seven are extracted from Zoopla images. None of the top ten coefficients and very few of the top 50 coefficients are extracted from Google Street View images.

6.3.5 Comparing models estimating changes in local house price rank in London from 2005 and 2006 to 2015 and 2016

We also compare the performance of models that estimate changes in median house price rank from 2005 and 2006 to 2015 and 2016 in MSOAs of London, again using one, two, or three data sources, and the mean features or concatenated features method in the models. The performance statistics for the models of change in ranked price are provided in Table 6.3. For each model, we report: the combination of data sources used and whether the model uses the mean features method or concatenated features method if multiple data sources are exploited; the coefficient of determination (R^2); and the total number of images we extract visual features from. The table is ordered in terms of model performance, where the most accurate model is first.

We first compare the models that only use one image source. In terms of performance, we observe the same pattern as we saw when estimating median house prices from 2015 and 2016. Namely, the model using Zoopla imagery performs best ($R^2 = 0.43$, $N = 983$, $p < 0.001$), followed by the model using Google Street View imagery ($R^2 = 0.22$, $N = 983$, $p < 0.001$), with the model using Geograph imagery achieving limited success ($R^2 = 0.18$, $N = 983$, $p < 0.001$).

Unlike median house price estimation models, the model that estimates the changes in median house price rank best uses imagery from Google Street View, Geograph, and Zoopla ($R^2 = 0.4438$). The second best model uses Geograph and Zoopla images ($R^2 = 0.4436$) and performs essentially identically to the best performing model. However, both models are only marginally better than the Zoopla model ($R^2 = 0.433$). Again, since the Zoopla model uses fewer images than the top two models, it may be more efficient to estimate changes from Zoopla images alone, as accuracy is not noticeably compromised.

Table 6.3: Model performance statistics for all models inferring changes in median house prices in London from 2005 and 2006 to 2015 and 2016. We combine images from Google Street View, Geograph, and Zoopla in all possible variations, and generate estimates using the methodology outlined in Fig. 6.1. This table displays the performance statistics for each of the models. For each model, the table shows: the data source combination and if applicable, whether the model uses the mean features method or concatenated features method; the coefficient of determination (R^2), by which we order the table in order to highlight the models that perform most accurately; and the total number of images used to extract visual features. Note that R^2 is presented to 3 significant figures. The model with the best fit uses visual information from all three data sources, but is almost identical to the model using Zoopla and Geograph imagery ($R^2 = 0.444$, $N = 983$, $p < 0.001$). Both models use the concatenated features method. However, this combination of imagery once again only leads to a small improvement over Zoopla imagery alone ($R^2 = 0.433$, $N = 983$, $p < 0.001$). Again, for efficiency, it would be more sensible to estimate changes in median house prices from Zoopla imagery alone, since all of the best-performing models include Zoopla images and the Zoopla dataset is smaller than any combination dataset, whether it be the Zoopla and Geograph dataset or Zoopla and Google Street View dataset.

Model Input			Combination		Model Performance	Images
Google	**Geograph**	**Zoopla**	**Mean**	**Concatenated**	$\boldsymbol{R^2}$	**Total**
X	X	X		X	0.444	5 034 328
	X	X		X	0.444	4 001 889
X		X		X	0.441	4 760 329
	X	X	X		0.439	4 001 889
X	X	X	X		0.438	5 034 328
		X			0.433	3 727 890
X		X	X		0.409	4 760 329
X	X		X		0.244	1 306 438
X	X			X	0.240	1 306 438
X					0.224	1 032 439
	X				0.187	273 999

Once again, all of the best-performing models use Zoopla photos. In most cases, whether concatenated features or mean features are used has little impact on model performance, other than for the model using Google Street View and Zoopla images, where the concatenated features approach leads to slightly better performance (concatenated features method: $R^2 = 0.44$, $N = 983$, $p < 0.001$; mean features method: $R^2 = 0.41$, $N = 983$, $p < 0.001$). If Zoopla imagery is not used in a model, there is a drop in performance once more, as the next best model uses Google Street View and Geograph photos ($R^2 = 0.244$, $N = 983$, $p < 0.001$).

Table 6.4 exhibits the elastic net parameters and coefficient statistics for all of the models estimating changes in median house price rank. For each model, we report: the combination of data sources used and whether the model uses the mean features method or concatenated features method if multiple data sources are used, exactly like Table 6.3; the optimal value of *l1_ratio* and the corresponding α from fitting the elastic net; the total number of non-zero coefficients we extract from the elastic net; and of the top 50 positive and negative elastic net coefficients, the number of coefficients extracted from each data source if the model uses the concatenated features method. Again, the table is ordered in terms of model performance, where the best-performing model is first.

These elastic net models are practically lasso models like the median house price estimation models, since the optimal *l1_ratio* values are always greater than 0.9. The *l1_ratio* values are more varied, though, and the corresponding αs are generally much smaller. For models using the concatenated features method, we also report the number of coefficients extracted from each data source out of the top 50 positive and negative elastic net coefficients, since the change in median house price rank estimates are most impacted by these features. Again, if models use Zoopla images and concatenated features, the elastic net coefficients are heavily influenced by Zoopla features. In the combination model, 63 out of 100 features are Zoopla features; in the Geograph and Zoopla model, 79 out of 100 features are Zoopla features; and in the Google Street View and Zoopla model, 73 out of 100 features are Zoopla features.

Table 6.4: Model configuration for all models inferring changes in median house prices in London from 2005 and 2006 to 2015 and 2016. We combine images from Google Street View, Geograph, and Zoopla in all possible variations, and generate estimates using the methodology outlined in Fig. 6.1. This table displays the model parameters and a breakdown of the coefficients for each of the models. For each model, the table shows: the data source combination and if applicable, whether the model uses the mean features method or concatenated features method; the optimal value of *l1_ratio* from fitting the elastic net; the corresponding α; the number of non-zero coefficients extracted from the model; and the number of coefficients extracted from each data source in the top 50 positive and negative coefficients, if the model uses the concatenated features method. When using concatenated features, as described in Section 6.2.3, we can distinguish between the features we extract from each data source. 'S' represents features extracted from Google Street View images, 'G' represents those extracted from Geograph images, and 'Z' represents those extracted from Zoopla images. Note that the α is given to 3 significant figures. Based on the values of *l1_ratio* > 0.9, the elastic net models are close to lasso models once again and the αs are also generally very small. As with the models estimating median house prices using concatenated features, if the models estimating changes in median house price rank use Zoopla imagery, most of the features in the top 50 positive and negative coefficients are also Zoopla features.

Model Input			**Combination**		**Elastic Net Parameters**		**Coefficients**			
Google	**Geograph**	**Zoopla**	**Mean**	**Concatenated**	***l1_ratio***	**α**	**Total**	**S**	**G**	**Z**
X	X	X		X	0.98	0.000 135	542	18	19	63
	X	X		X	0.95	0.000 077 7	548		21	79
X		X		X	0.98	0.000 127	355	27		73
	X	X	X		0.95	0.000 103	304			
X	X	X	X		0.97	0.000 118	298			
		X			0.99	9.000 767	146			
X		X	X		0.999 99	0.0163	15			
X	X		X		0.999	0.001 15	137			
X	X			X	0.999 999	0.0317	33	8	25	
X					0.9999	0.004 93	30			
	X				0.999 999	0.0223	37			

Again, the impact of Zoopla features can be seen in Fig. 6.4, which depicts the elastic net coefficients we extract from the model using the concatenated features method with images from all three data sources. Since we use concatenated features, we need to distinguish between the features associated with each data source. Of the top coefficients, 63 are features we extract from Zoopla images, whereas 18 come from Google Street View images and 19 come from Geograph images. The elastic net deems features extracted from Google Street View images and Geograph images less important than those extracted from Zoopla images. This could suggest that indoor visuals are more useful for estimating changes in local house prices than outdoor visuals, although it may also be an effect of the greater number of Zoopla images leading to a higher quality of information for these features.

Note that as in the earlier analyses in Chapters 3, 4, and 5, some features that are positive coefficients in models estimating median house prices are negative coefficients in models estimating changes in median house prices, and vice versa. Again, some areas with high median house prices may not have climbed as highly in the ranks, whereas other areas with low median house prices have climbed high in the ranks.

6.4 Discussion

Throughout this chapter, we study the impact of combining imagery from Google Street View, Geograph, and Zoopla on the performance of house price estimation models. The Google Street View imagery in our dataset consists of streets and building facades, whereas the Geograph imagery also depicts the environment surrounding buildings. The Zoopla imagery consists of pictures of house interiors, as well as some photos of housing facades. We examine the performance of models using one, two, or three of these data sources, in all possible combinations. For each combination of data source, we use the imagery to estimate median house prices in London during 2015 and 2016 at MSOA granularity, and to estimate relative changes in median house prices from 2005 and 2006 to 2015 and 2016 in MSOAs of London.

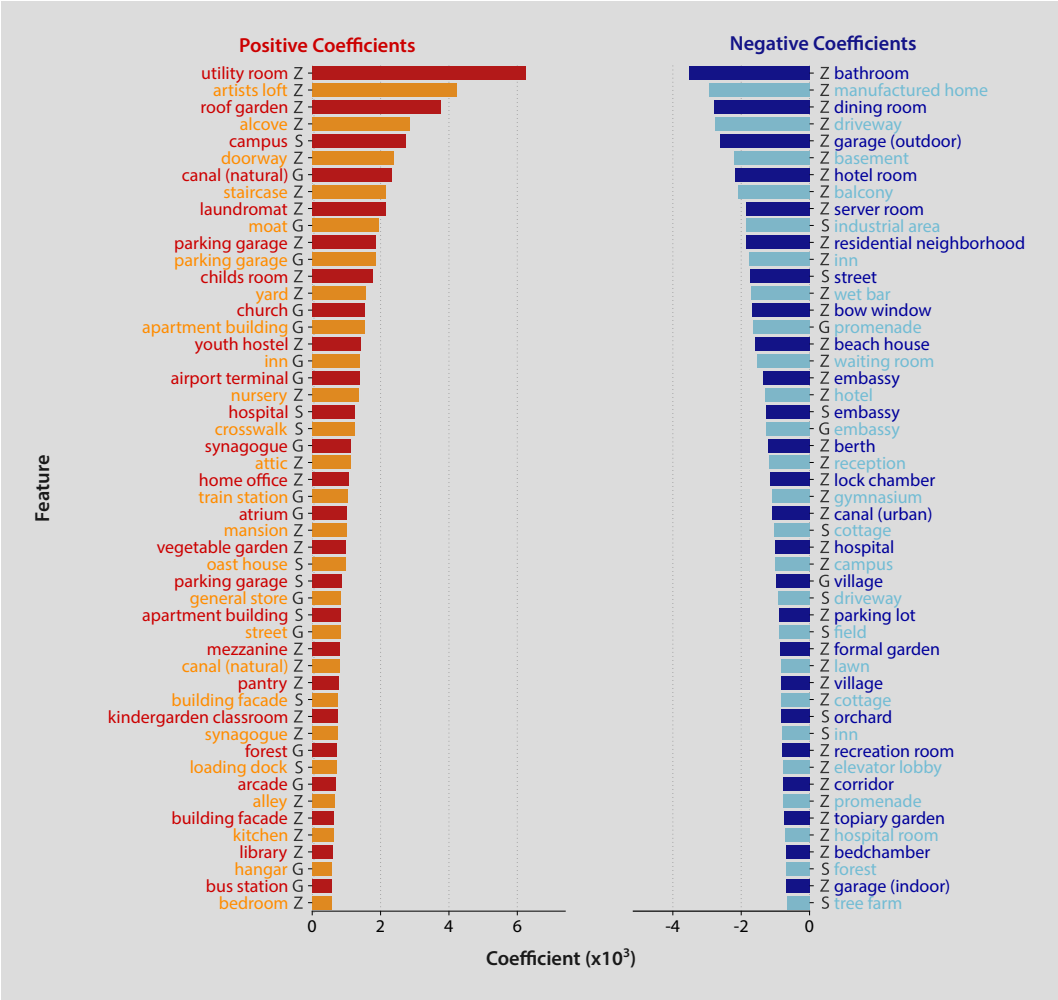

Figure 6.4: Top 50 elastic net coefficients for estimates of change in ranked median house price in London from all images using the concatenated features method. We train an elastic net model to estimate median house prices using the methodology outlined in Fig. 6.1 and show the resulting coefficients. We use the concatenated features method as described in Section 6.2.3, where we calculate the mean feature vector for each MSOA for each of the three data sources, then concatenate these feature vectors for each MSOA. Since we use concatenated features, we can distinguish between the features we extract from each data source. 'S' represents Google Street View, 'G' represents Geograph, and 'Z' represents Zoopla. Positive coefficients represent features that correspond to increases in median house price rank and negative coefficients represent features that correspond to decreases in median house price rank over the decade. *(continues on the following page)*

Figure 6.4: *(continues from previous page)* Once again, the features we extract from Zoopla images dominate the top coefficients. However, more features extracted from Google Street View images appear to play a somewhat important role than they did when estimating median house prices, leading to some reduction in importance of Geograph images.

When examining model performance from each data source independently, we discover that utilising older Google Street View images in conjunction with more recent Google Street View images actually hinders model performance in comparison to only using recent Google Street View images. We also find that the use of Zoopla imagery leads to the most accurate models, whether we are estimating changes in median house prices or median house prices themselves. Using Google Street View images alone results in the next best-performing models, followed by the models using Geograph imagery alone. This may be due to the difference in volume of images we possess for each data source, as we have almost 4 times as many Zoopla photos as Google Street View photos, and almost 14 times as many Zoopla photos as Geograph photos. To better understand the impact of sample size on model performance in future, we could use a small sample of images from each data source comprising the same number of images to build house price estimation models, and then compare performance.

To further compare model performance between Google Street View, Geograph, and Zoopla images, it would useful if the images we extract information from were from the same time period. Currently, our dataset consists of Google Street View images taken during 2008–2009 and 2014–2015, Geograph images taken up to November 2019 with 98% of photos taken after January 2000, and Zoopla images taken from January 2003 to June 2018. In future, we could look to retrieve image data from exactly the same time period for all three data sources. This could supplement the investigation into the relationship between sample size and model performance to precisely determine which factors affect model performance.

We next analyse how model performance differs if we make use of images from all three data sources at the same time. We observe that using all three data sources leads to model performance that is almost identical to that of the model using Zoopla images alone, despite the larger volume of images used. It is unclear whether the lack of further improvement reflects the fact that adding images to a model from a

different source has limited benefit, or whether there is little benefit in increasing the number of images above a certain volume.

Finally, we compare and contrast the house price estimation models we build from one, two, or three imagery data sources, in all possible combinations. When estimating median house prices, we find that the best-performing model uses Google Street View and Zoopla images. On the other hand, the model that uses images from all three data sources leads to the best performance when estimating changes in median house prices. However, for both analyses, the models improve upon the models using Zoopla images only by the narrowest of margins. In fact, all of the top-performing models use Zoopla imagery.

As such, from the models analysed here, it seems that using Zoopla imagery alone may be the most efficient approach for building house price estimation models. The best-performing models all use Zoopla imagery. By constructing a model that uses Zoopla imagery alone, with no additional data sources, we can build house price estimation models rapidly by using fewer photos but still achieving a high level of accuracy. Again, one possible explanation for the superior performance of models using Zoopla imagery is the larger sample of available images. As previously discussed, this could be controlled for in future analyses by taking samples of identical sizes from the three imagery sources. This would make it clearer whether the superior value of Zoopla imagery comes from the information the photos contain of housing interiors, or sample size alone.

Chapter 7

Discussion

7.1 Motivation

Whether it be individuals aiming for the property ladder or governments setting housing policies, the relevance of house prices to society is apparent at both a microeconomic and macroeconomic level. Unfortunately, obtaining accurate and timely information about house prices in England is challenging, as there exists a time delay of approximately two months between the occurrence of a property transaction and the availability of data on the transaction from the Land Registry. This time delay negatively impacts the ability of interested parties to make informed decisions regarding housing.

However, technology has advanced swiftly over the past few decades, now allowing for large and complex calculations that were not feasible in the decade prior. In turn, the development of deep learning tools, such as neural networks, has paved the way for high-calibre artificial intelligence. Specifically, CNNs have proven to be successful in image recognition tasks and can accurately extract visual information from images (LeCun et al., 2015). Additionally, the field of machine learning has produced various methodologies for working with large amounts of data, including the elastic net model, which is a linear model with a robust variable selection method that discards unimportant input variables (Zou & Hastie, 2005). Since CNNs can extract an immense number of visual characteristics from an image, using them together could be beneficial (Seresinhe et al., 2017).

The availability of image data is also growing rapidly, with well-known services such as Google Street View providing imagery of streets and buildings, and property websites such as Zoopla containing large volumes of photos of both housing interiors and exteriors. In this thesis, we present an exploratory analysis to determine the value

of the information that can be extracted from rapidly increasing large collections of imagery with deep learning tools with respect to house price estimation. If the source of imagery we utilised were updated more frequently than the Land Registry, it would be possible to generate faster indicators of house prices.

Traditional house price estimation models in computational social science literature use variables that relate to the value of a house to make house price predictions, including the number of bedrooms or the quality of nearby schools (Seo & Simons, 2009). However, the appearance of a house and its neighbourhood can also be indicative of its value. When humans judge the value of a house, we instinctively take into account what both the house and its surroundings look like. Yet, the researchers and practitioners studying house prices rarely include visual information in their models. This thesis outlines a methodology that could make it easier to gain insights from relevant imagery of housing in relation to house prices. Including visual details in a house price estimation model may be valuable to practitioners and researchers, and allow them to potentially improve their models.

The methodology we describe can also be applied to topics other than house prices. Social scientists may find value in learning and applying our methods to other sub-disciplines of social science. They could use the information available in imagery that is relevant to their topic to further understanding in their field. The focus of this thesis, however, is to understand whether different sources of imagery of houses can be exploited to infer house prices. By drawing on large volumes of photos taken at street or eye level, our research helps capture the human perspective in a computational model. Note that the foundations of the thesis are similar to those used in the field of social data science, as we start with data and draw conclusions from them, with the opportunity to inform theory afterwards.

7.2 Key results

Our first set of analyses in Chapter 3 explores whether we can infer house prices from Google Street View images. Google Street View images reflect a human perspective of street environments and consist of visuals of roads and building facades.

Throughout this thesis, we focus on estimating median house prices in London during 2015 and 2016 at MSOA granularity. To estimate median house prices, we first extract visual features from a set of 518,808 Google Street View images taken between 2014 and 2015 using the Places205 CNN (Zhou et al., 2014). We then use these visual features as input for an elastic net to make inferences about local house prices. We provide evidence suggesting that local house prices can be successfully

estimated from Google Street View images, and the model performs reasonably well. In addition to estimating median house prices, we also investigate whether we can estimate relative changes in median house prices over a decade. Namely, we utilise a ranking method to calculate the change in median house price rank from 2005 and 2006 to 2015 and 2016, and estimate this change from the same visual features as the median house price estimation model. Using the elastic net, we find that, at least to some extent, patterns in the changes can also be inferred from Google Street View images.

We then investigate whether this success can be replicated with an alternative data source of outdoor imagery. In Chapter 4, we therefore explore the usage of Geograph data to estimate house prices. Geograph is another frequently updated image source consisting of outdoor, street-level imagery, but images are not restricted to streets and include visuals of the surrounding environment, such as parks and rivers. Again, we estimate median house prices in London during 2015 and 2016 from Geograph imagery at MSOA granularity using an elastic net.

Unlike the Google Street View analyses, we use the updated and improved Places365 CNN (Zhou et al., 2018) to extract an even larger number of scene features from 51,918 Geograph images taken during 2015 and 2016. We discover that the model achieves only limited success when inferring median house prices in London. Upon inspection, we find that many areas across London are unrepresented by the images taken during 2015 and 2016. We therefore attempt to estimate median house prices from the entire dataset of Geograph imagery, covering photos taken up to November 2019 with 98% of photos taken after January 2000, and totalling 328,014 images. We find that the model shows a considerable improvement when using this larger dataset, but does not perform as well as the Google Street View model. This may be due to the greater number of Google Street View photos used for analysis (518,808) in comparison to the number of Geograph photos used for analysis (328,014).

Once more, we also attempt to estimate relative changes in median house price in London from 2005 and 2006 to 2015 and 2016 using visual features from Geograph and an elastic net. We do this for both the complete Geograph dataset and the subset of images taken between 2015 and 2016. Again, we discover that using the complete dataset leads to an improvement in model performance in comparison to using the subset of images, but both models achieve limited success and perform worse than the Google Street View model. However, the fall in performance in the Geograph analysis in comparison to the Google Street View analysis is similar to the fall in performance between the Google Street View and Geograph analyses for median house price estimation. Again, it is possible that the difference in the

number of images used for analysis is the cause of the discrepancy between the model performances.

Both of the data sources analysed so far comprise outdoor visual information. Would it be possible to infer local house prices from internal images of houses? To explore this question, in Chapter 5, we access large volumes of photographs of houses from the frequently updated property website Zoopla. Zoopla images depict both house interiors and house exteriors. Once again, we estimate median house prices in London during 2015 and 2016 at MSOA granularity. We first extract visual features from 2,702,506 indoor images of houses for sale using the Places365 CNN, and then use these features as predictors in an elastic net model. We discover that the model can generate very good estimates and is high-performing with indoor imagery alone. This model performs notably better than those built using Google Street View and Geograph images. Using the same indoor imagery to estimate relative changes in median house prices over a decade in London also leads to notably superior performance compared to the models using Google Street View and Geograph imagery. As with the Google Street View and Geograph analyses however, performance is not as good on this task as for estimation of median house prices.

As well as indoor Zoopla images, we also have access to 1,025,384 outdoor Zoopla images of building facades and back gardens. We find that these images can be used to estimate median house prices and changes in median house prices, with performance that improves upon that exhibited by the corresponding Google Street View and Geograph models. However, the models using indoor Zoopla imagery deliver better performance. We further study the effects of combining both interior and exterior visual information from Zoopla to estimate median house prices and changes in median house prices. The volume of images used in the models increases to 3,727,890. These models demonstrate a marginal improvement in comparison to models using indoor imagery alone. Based on this result, we consider combining the visual information from all three of our data sources, and ask how model performance would differ.

Hence, in Chapter 6, we examine the performance of models using one, two, or three of the data sources Google Street View, Geograph, and Zoopla in all possible combinations. Our methodology involves combining the visual features that we extract from each of our data sources using Places365, and using them as input to an elastic net to estimate median house prices during 2015 and 2016 and changes in median house prices from 2005 and 2006 to 2015 and 2016 in London at MSOA granularity.

Out of the combinations of data sources that we explore, we find that the

best-performing model for estimating median house prices makes use of both Google Street View and Zoopla imagery. However, the best-performing model for estimating changes in median house prices uses all three data sources. In fact, all of the best-performing models include visual information from Zoopla images and only marginally improve upon the models that use indoor Zoopla imagery alone. It seems that the most efficient approach for building house price estimation models would be to use indoor Zoopla imagery alone, with no additional data sources. We can then build house price estimation models rapidly by using fewer photos but still achieving a high level of accuracy.

7.3 Limitations and future work

Although we are successfully able to develop house price estimation models that perform well from images of houses and their surrounding neighbourhoods, there are some limitations to our analyses. A key limitation that impacts the accuracy of all of the models to some degree is their ability to handle more extreme values of local house prices and changes in local house prices. Since the house price data have a positive skew, estimating the logarithm of median house prices instead of median house prices directly could eradicate this issue with extreme values and improve model accuracy. In terms of estimating the changes in median house prices, the ranking method we use leads to changes with a heavy-tailed distribution. Using a different approach for inferring relative changes may be a future step to take.

For example, (Osorio, 2020) developed a library for the programming language R that is capable of performing robust estimation considering heavy-tailed distributions called 'heavy'. This library could be used to build a linear regression model that takes the features we extract from images as input to estimate the changes in median house price rank. Although this is similar to the methodology we use when building an elastic net model, the difference is that we can specify which heavy-tailed distribution should be considered when fitting the model, such as the Cauchy distribution. However, the normal distribution can also be chosen. This approach could lead to better estimates of the changes in ranked price.

Throughout this thesis, we use spatial data to model house prices in London at MSOA granularity, but we do not consider the impact of spatial autocorrelation on our results. Spatial autocorrelation is a measure of the similarity between nearby observations, which would be MSOAs in our case. Positive spatial autocorrelation indicates that similar values or areas are close together on a map, whereas negative spatial autocorrelation indicates that different values or areas are close together on

a map. Our data are likely to have some spatial autocorrelation present because MSOAs that are near each other are likely to share common characteristics. This is supported by our comparisons between Dalston and Acton in Fig. 5.7. If this is the case, our results will be impacted and we may have some redundancy. Researchers can extend our analyses in future by conducting a thorough analysis into the spatial autocorrelation in our models, and possibly adjusting the model to handle spatial autocorrelation if it is definitely present.

In terms of the high performance of models using Zoopla data, we suggest that it can be attributed to a large, representative volume of photos across areas of London, and potentially the information available from indoor imagery. We use the Places365 CNN for scene detection to extract features from the images, and it is able to capture some differences between housing interiors, such as differences between ordinary beds and four-poster beds. However, the interiors of homes contain many minutiae that cannot be captured by Places365, like the appliances in a kitchen or the furniture in a bathroom. As such, using an object detection CNN with indoor Zoopla data in addition to a scene recognition CNN could be helpful and allow us to identify objects that are related to house prices. Furthermore, data available from Zoopla are not restricted to images and also include details of houses advertised on Zoopla, such as the type of property on sale. Future work could look at developing individual models based on the type of housing of interest, from flats to houses.

A limitation of our analyses using the Zoopla dataset is that we are not able to attach sale and marketed dates to each property and instead use all of the image data from multiple time periods. If the data associated with properties advertised on Zoopla are made available as soon as the advert is released, then it could be dated in almost real-time. Furthermore, since the indoor Zoopla images lead to a successful house price estimation model, future research could also apply our methodology to other indoor imagery of housing. Li et al. (2018) present a dataset of simulated videos of the interiors of houses at a high-resolution and high frame-rate. The scenes have been rendered to look similar to real-world decorations, and contain different layouts of rooms and different lighting. Using our methodology to estimate local house prices from image stills from the synthetic videos would allow us to understand whether useful information can be extracted from synthetic imagery of housing interiors for local house price estimation, as was the case with real photographs of housing interiors.

When building models that use a combination of Google Street View, Geograph, and Zoopla images, we make use of all of the image data we have access to, but the sample sizes differ between the data sources. All of the best-performing

models use Zoopla imagery, and Zoopla is the data source with the largest sample size, followed by Google Street View, then Geograph. Furthermore, since we make use of all of the image data we have access to, we use images from many time periods, rather than specific images from set time periods. To better understand the impact of sample size and differing time periods on model performance, a next step would be to use a small sample of a set number of images from each data source during a set time period to build house price estimation models and compare performance.

Future research could also extend our analyses by using our methodology for images taken during a different time period. For example, indoor Zoopla images from 2019 and 2020 could be accessed and used to develop a model that estimates median house prices in MSOAs across London during 2019 and 2020. This would provide the opportunity to determine whether the models we build are robust and whether the results generalise to other points in time. As well exploring a different time period, future research could explore the visual differences over time between images from the past and images from the present. For example, visual features could be extracted from Geograph images from 2010 and 2020, and the differences in the feature values could be calculated and used to automatically infer changes in median house price rank. This may lead to models that perform well, as the visual differences would be accounted for.

Additionally, our work shows that local house prices can be estimated well at MSOA level, and lays the foundations for several extensions at other levels of granularity. For example, it may also be possible to estimate house prices of individual homes, rather than an entire area. With the Zoopla data in particular, it is clear which images are associated with which property. The latest last marketed date could be used to date the properties and our methodology could be used to estimate individual house prices. Another future step would be to apply our methodology to housing in areas outside of London. London is known to be more expensive than the remainder of England, so house price dynamics are likely to differ elsewhere.

In this thesis, we present an exploratory analysis that determines whether imagery of housing can be used to help estimate house prices. This could help individuals and organisations concerned with house prices to make decisions at a faster rate than is currently possible, if the imagery used for house price estimation is updated more frequently than official information. Evidently from our results, imagery of housing, especially from Zoopla, can be valuable for estimating local house prices. Furthermore, Zoopla data are updated more rapidly than house price data, as numerous properties are advertised on the marketing platform on a daily basis, rather than after an approximate delay of two months. As such, new images of properties

are frequently available from areas all across the UK and can be used for local house price estimation. However, the question remains open as to whether images from Zoopla are more valuable than delayed official information over time. To investigate this, future research could examine a time-series model that estimates local house prices based on official data and compare this model to a time-series model that also incorporates visual information, perhaps starting with a simple ARIMA model.

In future, a more thorough nowcasting analysis could also be helpful for determining whether the mean absolute errors that our models achieve are good, and whether the models we built improve upon waiting 2 months for official house price data. For instance, the Zoopla model that we build using imagery of both the indoors and outdoors of houses in London achieves a mean absolute error of £65,700. We would expect that using this model would be better than waiting 2 months for data because house prices can change rapidly. During the 2008 financial crash, for example, the entire economy changed, including a mass reduction in house prices. Making decisions over 2 months with out-of-date information in this circumstance could have negative consequences, and impact the economy further.

In general, our research follows the ideas that are used in the field of social data science. This means that we do not directly compare our results to outstanding theories associated with house prices. A future step would be to compare our models to models that take a more traditional approach and are formed based on a theory. We may then be able to understand which theories match our findings. Furthermore, the natural next step would be to combine image data with more traditional house price estimation models. Other models of house price estimation, such as hedonic pricing models, often include quantitative statistics relevant to housing, such as the distance to the nearest form of public transport or the number of bedrooms in a property (Higgins et al., 2019; Sirmans et al., 2006).By including visual features in addition to these statistics, a house price estimation model of greater accuracy could be developed.

In parallel to our studies, some researchers have combined the visual information available from Google Street View with deep learning to make inferences about neighbourhoods. For example, Gebru et al. (2017) use Google Street View imagery with a CNN to estimate socioeconomic attributes, such as income and education, of people in neighbourhoods across the US from motor vehicles encountered in these neighbourhoods, whereas Suel, Polak, Bennett, and Ezzati (2019) use Google Street View imagery to measure inequalities and monitor the impacts of policies in neighbourhoods of London. Furthermore, some researchers have begun the process of using visual information and statistics related to housing in house price estimation models.

Law, Paige, and Russell (2019) conduct many experiments using Google Street View imagery, as well as satellite imagery and descriptive statistics, to estimate house prices across London. They compare a range of modelling techniques, including a linear hedonic pricing model and a generalised additive model, and highlight the benefits of using all accessible information in a house price estimation model.

Whilst Law et al. (2019) focuses on the use of outdoor imagery in conjunction with statistical data about housing, a recent paper takes the analyses one step further in Nouriani and Lemke (2022). They build a dataset containing indoor, outdoor, and satellite imagery of housing around Minnesota, USA, and compare numerous methods to extract visual information from them. As well as extracting visual information, they classify the level of luxury of a room and complement the visual information with descriptive statistics of a home to develop house price estimation models at an individual level. Of the six regression methods compared, the gradient boosting machine performed most accurately, over the elastic net method, and the model outperformed the pre-existing house valuation tool 'Zillow's Zestimate'. These results suggest that a similar methodology could be implemented in England using Zoopla imagery and further data.

Although research conducted in parallel to ours investigates the use of Google Street View images for house price estimation, our research offers insights into the performance of local house price estimation models using images from Geograph and Zoopla as well. The Geograph images in particular contain visuals of streets, like Google Street View, but also scenes of off-road areas such as parks, canals, and mountains. Furthermore, the research conducted in parallel to ours tends to focus on estimating house prices, whereas our research offers insights into relative changes in local house prices, in addition to estimating local house prices themselves.

7.4 Conclusion

In this thesis, we take initial steps towards understanding whether street-level imagery can be valuable when estimating local house prices and changes in local house prices. Our results demonstrate how utilising image data reflecting a human perspective of houses in conjunction with deep learning tools can facilitate the development of successful house price estimation models.

Our novel methodology could be beneficial for social scientists looking to answer other questions with relevant image data, and help further the understanding of human behaviour. Our methods also have the potential to improve the understanding of house prices for researchers interested in them, as a combination of visual and

non-visual characteristics could improve house price estimation models.

Furthermore, as the speed of collection of such imagery increases, our approach could equip national and local policymakers with more timely information on house prices, avoiding the traditional two month delay usually imposed upon this key economic indicator. In turn, this would enable policymakers to make more informed decisions about economic policy.

References

Alanyali, M., Moat, H. S., & Preis, T. (2013). Quantifying the relationship between financial news and the stock market. *Scientific Reports*, *3*, Article 3578.

Alanyali, M., Preis, T., & Moat, H. S. (2016). Tracking protests using geotagged Flickr photographs. *PLoS ONE*, *11*(3), Article e0150466.

Alis, C. M., & Lim, M. T. (2013). Spatio-temporal variation of conversational utterances on Twitter. *PLoS ONE*, *8*(10), Article e77793.

Alis, C. M., Lim, M. T., Moat, H. S., Barchiesi, D., Preis, T., & Bishop, S. R. (2015). Quantifying regional differences in the length of Twitter messages. *PLoS ONE*, *10*(4), Article e0122278.

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., ... Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, *8*, Article 53.

Ameen, S., & Vadera, S. (2017). A convolutional neural network to classify American Sign Language fingerspelling from depth and colour images. *Expert Systems*, *34*, Article e12197.

Anthony, M., & Bartlett, P. L. (1999). *Neural network learning: theoretical foundations*. Cambridge University Press.

Arietta, S. M., Efros, A. A., Ramamoorthi, R., & Agrawala, M. (2014). City forensics: Using visual elements to predict non-visual city attributes. *IEEE Transactions on Visualization and Computer Graphics*, *20*(12), 2624–2633.

Askitas, N. (2016). Trend-spotting in the housing market. *Cityscape: A Journal of Policy Development and Research*, *18*(2), 165–178.

Atkins, A., Niranjan, M., & Gerding, E. (2018). Financial news predicts stock market volatility better than close price. *Journal of Finance and Data Science*, *4*(2), 120–137.

Badarinza, C., & Ramadorai, T. (2018). Home away from home? Foreign demand and London house prices. *Journal of Financial Economics*, *130*(3), 532–555.

Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse

news and opinion on Facebook. *Science*, *348*(6239), 1130–1132.

Barchiesi, D., Moat, H. S., Alis, C., Bishop, S., & Preis, T. (2015). Quantifying international travel flows using Flickr. *PLoS ONE*, *10*(7), Article e0128470.

Barchiesi, D., Preis, T., Bishop, S., & Moat, H. S. (2015). Modelling human mobility patterns using photographic data shared online. *Royal Society Open Science*, *2*(8), Article 150046.

Barlacchi, G., De Nadai, M., Larcher, R., Casella, A., Chitic, C., Torrisi, G., ... Lepri, B. (2015). A multi-source dataset of urban life in the city of Milan and the Province of Trentino. *Scientific Data*, *2*, Article 150055.

Beeston, G. P., Leon Urrutia, M., Halcrow, C., Xiao, X., Liu, L., Wang, J., ... Park, K. (2014). Humour reactions in crisis: A proximal analysis of Chinese posts on Sina Weibo in reaction to the salt panic of March 2011. In C.-W. Chung, A. Broder, K. Shim, & T. Suel (Eds.), *Proceedings of the 23rd International Conference on World Wide Web (WWW '14 Companion)* (pp. 1043–1048). Seoul, Korea.

Begiazi, K., & Katsiampa, P. (2019). Modelling UK house prices with structural breaks and conditional variance analysis. *Journal of Real Estate Finance and Economics*, *58*, 290–309.

Bhatt, D., Patel, C., Talsania, H., Patel, J., Vaghela, R., Pandya, S., ... Ghayvat, H. (2021). CNN variants for computer vision: History, architecture, application, challenges and future scope. *Electronics*, *10*(20), Article 2470.

Bhavnani, R., & Choi, H. J. (2012). Modeling civil violence in Afghanistan: Ethnic geography, control, and collaboration. *Complexity*, *17*, 42–51.

Biljecki, F., & Ito, K. (2021). Street view imagery in urban analytics and GIS: A review. *Landscape and Urban Planning*, *215*, Article 104217.

Blondel, V. D., Decuyper, A., & Krings, G. (2015). A survey of results on mobile phone datasets analysis. *EPJ Data Science*, *4*, Article 10.

Bogin, A. N., & Doerner, W. M. (2019). Property renovations and their impact on house price index construction. *Journal of Real Estate Research*, *41*(2), 249–284.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*(1), 1–8.

Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D. I., Marlow, C., Settle, J. E., & Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. *Nature*, *489*, 295–298.

Bordino, I., Battiston, S., Caldarelli, G., Cristelli, M., Ukkonen, A., & Weber, I. (2012). Web search queries can predict stock market volumes. *PLoS ONE*,

7(7), Article e40014.

Borge-Holthoefer, J., Rivero, A., & Moreno, Y. (2012). Locating privileged spreaders on an online social network. *Physical Review E*, *85*(6), Article 066123.

Botta, F., Moat, H. S., & Preis, T. (2015). Quantifying crowd size with mobile phone and Twitter data. *Royal Society Open Science*, *2*(5), Article 150162.

Botta, F., Moat, H. S., & Preis, T. (2020). Measuring the size of a crowd using Instagram. *Environment and Planning B: Urban Analytics and City Science*, *47*(9), 1690–1703.

Botta, F., Moat, H. S., Stanley, H. E., & Preis, T. (2015). Quantifying stock return distributions in financial markets. *PLoS ONE*, *10*(9), Article e0135600.

Botta, F., Preis, T., & Moat, H. S. (2020). In search of art: rapid estimates of gallery and museum visits using Google Trends. *EPJ Data Science*, *9*, Article 14.

Bourassa, S. C., Hoesli, M., & Oikarinen, E. (2016). Measuring house price bubbles. *Real Estate Economics*, *47*(2), 534–563.

Bowers, K. J., Johnson, S. D., & Pease, K. (2004). Prospective hot-spotting: The future of crime mapping? *British Journal of Criminology*, *44*(5), 641–658.

Brady, R. R. (2014). The spatial diffusion of regional housing prices across U.S. states. *Regional Science and Urban Economics*, *46*, 150–166.

Breiman, L. (1995). Better subset regression using the nonnegative garrote. *Technometrics*, *37*(4), 373-384.

Brown, J. P., Song, H., & McGillivray, A. (1997). Forecasting UK house prices: A time varying coefficient approach. *Economic Modelling*, *14*(4), 529–548.

Butler, D. (2013). When Google got flu wrong. *Nature*, *494*, 155-156.

Capocci, A., Servedio, V. D. P., Colaiori, F., Buriol, L. S., Donato, D., Leonardi, S., & Caldarelli, G. (2006). Preferential attachment in the growth of social networks: the internet encyclopedia Wikipedia. *Physical Review E*, *74*(3), Article 036116.

Caputi, T. L., Leas, E., Dredze, M., Cohen, J. E., & Ayers, J. W. (2017). They're heating up: Internet search query trends reveal significant public interest in heat-not-burn tobacco products. *PLoS ONE*, *12*(10), Article e0185735.

Carneiro, G., Nascimento, J., & Bradley, A. P. (2015). Unregistered multiview mammogram analysis with pre-trained deep learning models. In N. Navab, J. Hornegger, W. Wells, & A. Frangi (Eds.), *Lecture notes in computer science: Vol. 9351. Medical image computing and computer-assisted intervention – MICCAI 2015.* (pp. 652–650). Springer.

Case, B., Clapp, J., Dubin, R., & Rodriguez, M. (2004). Modeling spatial and temporal house price patterns: A comparison of four models. *Journal of Real*

Estate Finance and Economics, *29*, 167–191.

Cheng, X., Fang, L., Hong, X., & Yang, L. (2017). Exploiting mobile big data: Sources, features, and applications. *IEEE Network*, *31*(1), 72–79.

Chetty, R., Jackson, M. O., Kuchler, T., Stroebel, J., Hendren, N., Fluegge, R. B., ... Barberá, P. (2022a). Social capital II: determinants of economic connectedness. *Nature*, *608*, 122–134.

Chetty, R., Jackson, M. O., Kuchler, T., Stroebel, J., Hendren, N., Fluegge, R. B., ... Barberá, P. (2022b). Social capital I: measurement and associations with economic mobility. *Nature*, *608*, 108–121.

Chin, W. W. (1998). The partial least squares approach to structural equation modeling. In G. A. Marcoulides (Ed.), *Modern methods for business research* (pp. 295–336). Psychology Press.

Chinazzi, M., Davis, J. T., Ajelli, M., Gioannini, C., Litvinova, M., Merler, S., ... Vespignani, A. (2020). The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science*, *368*(6489), 395–400.

Choi, H., & Varian, H. (2012). Predicting the present with Google Trends. *Economic Record*, *88*, 2–9.

Colizza, V., Barrat, A., Barthelemy, M., Valleron, A.-J., & Vespignani, A. (2007). Modeling the worldwide spread of pandemic influenza: baseline case and containment interventions. *PLoS Medicine*, *4*(1), Article e13.

Colizza, V., Barrat, A., Barthélemy, M., & Vespignani, A. (2006). The role of the airline transportation network in the prediction and predictability of global epidemics. *Proceedings of the National Academy of Sciences*, *103*(7), 2015–2020.

Columbro, C., Eudave, R. R., Ferreira, T. M., Lourenço, P. B., & Fabbrocino, G. (2022). On the use of web mapping platforms to support the seismic vulnerability assessment of old urban areas. *Remote Sensing*, *14*(6), Article 1424.

Conte, R., Gilbert, N., Bonelli, G., Cioffi-Revilla, C., Deffuant, G., Kertesz, J., ... Helbing, D. (2012). Manifesto of computational social science. *European Physical Journal Special Topics*, *214*, 325–346.

Cook, S. (2005). Detecting long-run relationships in regional house prices in the UK. *International Review of Applied Economics*, *19*(1), 107–118.

Curme, C., Preis, T., Stanley, H. E., & Moat, H. S. (2014). Quantifying the semantics of search behavior before stock market moves. *Proceedings of the National Academy of Sciences*, *111*(32), 11600–11605.

Curme, C., Stanley, H. E., & Vodenska, I. (2015). Coupled network approach to predictability of financial market returns and news sentiments. *International Journal of Theoretical and Applied Finance*, *18*(07), Article 1550043.

Curme, C., Zhuo, Y. D., Moat, H. S., & Preis, T. (2017). Quantifying the diversity of news around stock market moves. *Journal of Network Theory in Finance*, *3*(1), 1–20.

Data Science Lab. (n.d.). *ScenicOrNot.* Warwick Business School. Retrieved September 12, 2022, from https://scenicornot.datasciencelab.co.uk.

Davies, T., & Johnson, S. D. (2014). Examining the relationship between road structure and burglary risk via quantitative network analysis. *Journal of Quantitative Criminology*, *31*, 481–507.

Davison, G., Dovey, K., & Woodcock, I. (2012). "Keeping Dalston different": Defending place-identity in East London. *Planning Theory & Practice*, *13*(1), 47–69.

De Nadai, M., Vieriu, R. L., Zen, G., Dragicevic, S., Naik, N., Caraviello, M., ... Lepri, B. (2016). Are safer looking neighborhoods more lively? A multimodal investigation into urban life. In A. Hanjalic et al. (Eds.), *Proceedings of the 24th ACM International Conference on Multimedia (MM '16)* (pp. 1127–1135). Amsterdam, The Netherlands.

De Domenico, M., Lima, A., Mougel, P., & Musolesi, M. (2013). The anatomy of a scientific rumor. *Scientific Reports*, *3*, Article 2980.

Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., ... Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, *113*(3), 554–559.

Demšar, U., Reades, J., Manley, E., & Batty, M. (2018). Revisiting the past: Replicating fifty-year-old flow analysis using contemporary taxi flow data. *Annals of the American Association of Geographers*, *108*(3), 811–828.

Deville, P., Linard, C., Martin, S., Gilbert, M., Stevens, F. R., Gaughan, A. E., ... Tatem, A. J. (2014). Dynamic population mapping using mobile phone data. *Proceedings of the National Academy of Sciences*, *111*(45), 15888–15893.

Distinguin, I., Rous, P., & Tarazi, A. (2006). Market discipline and the use of stock market data to predict bank financial distress. *Journal of Financial Services Research*, *30*, 151–176.

Doerner, W. M., & Leventis, A. (2013). Distressed sales and the FHFA house price index. *Journal of Housing Research*, *24*(2), 127–146.

Doersch, C., Singh, S., Gupta, A., Sivic, J., & Efros, A. (2012). What makes Paris look like Paris? *ACM Transactions on Graphics*, *31*(4), Article 101.

Douglass, R. W., Meyer, D. A., Ram, M., Rideout, D., & Song, D. (2015). High resolution population estimates from telecommunications data. *EPJ Data Science*, *4*, Article 4.

Dufaux, F. (2021). Grand challenges in image processing. *Frontiers in Signal Processing*.

Duffy, B., Smith, K., Terhanian, G., & Bremer, J. (2005). Comparing data from online and face-to-face surveys. *International Journal of Market Research*, 615–639.

Eetemadi, A., Rai, N., Pereira, B. M. P., Kim, M., Schmitz, H., & Tagkopoulos, I. (2020). The computational diet: A review of computational methods across diet, microbiome, and health. *Frontiers in Microbiology*, *11*, Article 393.

Efthymiou, D., & Antoniou, C. (2013). How do transport infrastructure and policies affect house prices and rents? Evidence from Athens, Greece. *Transportation Research Part A: Policy and Practice*, *52*, 1–2.

Feichtenhofer, C., Pinz, A., & Zisserman, A. (2016). Convolutional two-stream network fusion for video action recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1933–1941). Las Vegas, NV, USA.

Feng, Y., & Jones, K. (2015). Comparing multilevel modelling and artificial neural networks in house price prediction. In Y. Leung, D. Guo, & C. Chen (Eds.), *2015 2nd IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services (ICSDM)* (pp. 108–114). Fuzhou, China.

Ferrara, E. (2012). A large-scale community structure analysis in Facebook. *EPJ Data Science*, *1*, Article 9.

Ferwerda, B., Schedl, M., & Tkalcic, M. (2016). Using Instagram picture features to predict users' personality. In *Proceedings, Part I, of the 22nd International Conference on Multimedia Modeling - Volume 9516 (MMM 2016)* (pp. 850–861). Miami, FL, USA.

FHFA. (n.d.). *FHFA House Price Index.* Retrieved September 12, 2022, from https://www.fhfa.gov/DataTools/Downloads/Pages/House-Price-Index.aspx.

Flick, C. (2015). Informed consent and the Facebook emotional manipulation study. *Research Ethics*, *12*(1), 14–28.

Flickr. (n.d.). *The App Garden.* Retrieved September 12, 2022, from https://www.flickr.com/services/api.

Fortunato, S., Bergstrom, C. T., Börner, K., Evans, J. A., Helbing, D., Milojević, S., ... Barabási, A.-L. (2018). Science of science. *Science*, *359*(6379), Article eaao0185.

Fox, J. (2015). *Applied regression analysis and generalized linear models.* Sage Publications.

Frenken, K., Hardeman, S., & Hoekman, J. (2009). Spatial scientometrics: Towards

a cumulative research program. *Journal of Informetrics*, *3*(3), 222–232.

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, *33*(1), 1–22.

Friedman, J., Hastie, T., Tibshirani, R., Narasimhan, B., Tay, K., Simon, N., & Qian, J. (2022, April 15). *Package 'glmnet'.* The Comprehensive R Archive Network. https://cran.r-project.org/web/packages/glmnet/glmnet.pdf.

Frost, J. (2018, Jun 01). *7 classical assumptions of ordinary least squares (OLS) linear regression.* Statistics By Jim. https://statisticsbyjim.com/regression/ols-linear-regression-assumptions/.

Gebru, T., Krause, J., Wang, Y., Chen, D., Deng, J., Aiden, E. L., & Fei-Fei, L. (2017). Using deep learning and Google Street View to estimate the demographic makeup of neighborhoods across the United States. *Proceedings of the National Academy of Sciences*, *114*(50), 13108–13113.

Geograph. (n.d.-a). *Geograph.* Retrieved September 12, 2022, from https://www.geograph.org.uk.

Geograph. (n.d.-b). *Geograph British Isles :: Academic & Research Image DataSets.* Retrieved September 12, 2022, from http://data.geograph.org.uk/datasets.html.

Gibbons, S. (2004). The costs of urban property crime. *The Economic Journal*, *114*(499), F441–F463.

Gibbons, S. (2015). Gone with the wind: Valuing the visual impacts of wind turbines through house prices. *Journal of Environmental Economics and Management*, *72*, 177–196.

Gibbons, S., Heblich, S., Lho, E., & Timmins, C. D. (2016). *Fear of fracking? the impact of the shale gas exploration on house prices in Britain* (Working Paper No. w22859). National Bureau of Economic Research (NBER).

Gibbons, S., & Machin, S. (2003). Valuing English primary schools. *Journal of Urban Economics*, *53*(2), 197–219.

Gibbons, S., & Machin, S. (2005). Valuing rail access using transport innovations. *Journal of Urban Economics*, *57*(1), 148–169.

Gibbons, S., Peng, C., & Tang, C. K. (2021). Valuing the environmental benefits of canals and canal restoration using house prices. *Land Economics*, *97*(4), 858–874.

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, *457*, 1012–1014.

Goddard, J. B. (1970). Functional regions within the city centre: a study by factor analysis of taxi flows in Central London. *Transactions of the Institute of British Geographers*, *49*, 161–182.

González, M. C., Hidalgo, C. A., & Barabási, A.-L. (2008). Understanding individual human mobility patterns. *Nature*, *453*, 779–782.

González-Bailón, S., Borge-Holthoefer, J., Rivero, A., & Moreno, Y. (2011). The dynamics of protest recruitment through an online network. *Scientific Reports*, *1*, Article 197.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning.* MIT Press.

Google. (n.d.-a). *Blur or remove 360 imagery and Photo Paths.* Retrieved December 12, 2022, from https://support.google.com/maps/answer/7011973?hl=en-GB&co=GENIE.Platform%3DDesktop.

Google. (n.d.-b). *Earth in 360 degrees.* Retrieved September 12, 2022, from https://www.google.com/earth/education/tools/street-view/#: :text=Google%20collects%20Street%20View%20imagery,a%20single%20360-degree%20image.

Google. (n.d.-c). *FAQ about Google Trends data.* Retrieved September 12, 2022, from https://support.google.com/trends/answer/4365533?hl=en.

Google. (n.d.-d). *What is Street View?* Retrieved September 12, 2022, from https://www.google.com/streetview.

Greater London Authority. (n.d.). *Postcode directory for london.* Retrieved April 15, 2017, from https://data.london.gov.uk/dataset/postcode-directory-for-london.

Griew, P., Hillsdon, M., Foster, C., Coombes, E., Jones, A., & Wilkinson, P. (2013). Developing and testing a street audit tool using Google Street View to measure environmental supportiveness for physical activity. *International Journal of Behavioral Nutrition and Physical Activity*, *10*, Article 103.

Grunerbl, A., Muaremi, A., Osmani, V., Bahle, G., Ohler, S., Troster, G., ... Lukowicz, P. (2015). Smartphone-based recognition of states and state changes in bipolar disorder patients. *IEEE Journal of Biomedical and Health Informatics*, *19*(1), 140–148.

Gutiérrez-Roig, M., Preis, T., Seresinhe, C. I., Letchford, A., & Moat, H. S. (2017). *Predicting urban inequality using landscape features.* Unpublished manuscript. Universty of Warwick.

Hagan, M. T., Demuth, H. B., Beale, M. H., & De Jesús, O. (2014). *Neural network design.* Martin Hagan.

Hagenau, M., Liebmann, M., & Neumann, D. (2013). Automated news reading: Stock price prediction based on financial news using context-capturing features. *Decision Support Systems*, *55*(3), 685–697.

Hara, K., Le, V., & Froehlich, J. (2013). Combining crowdsourcing and Google Street View to identify street-level accessibility problems. In W. E. Mackay, S. Brewster, & S. Bødker (Eds.), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)* (pp. 631–640). Paris, France.

Hassanpour, S., Tomita, N., DeLise, T., Crosier, B., & Marsch, L. A. (2018). Identifying substance use risk based on deep neural networks and Instagram social media data. *Neuropsychopharmacology*, *44*, 487–494.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction.* (2nd ed.). Springer.

Hausmann, A., Toivonen, T., Slotow, R., Tenkanen, H., Moilanen, A., Heikinheimo, V., & Di Minin, E. (2017). Social media data can be used to understand tourists' preferences for nature-based experiences in protected areas. *Conservation Letters*, *11*, Article e12343.

Hebb, D. O. (1949). *The organization of behavior: A neuropsychological theory.* Wiley.

Heikinheimo, V., Minin, E. D., Tenkanen, H., Hausmann, A., Erkkonen, J., & Toivonen, T. (2017). User-generated geographic information for visitor monitoring in a national park: A comparison of social media data and visitor survey. *ISPRS International Journal of Geo-Information*, *6*(3), Article 85.

Higgins, D. M., Rezaei, A., & Wood, P. (2019). The value of a tram station on local house prices: an hedonic modelling approach. *Pacific Rim Property Research Journal*, *25*(3), 217–227.

HM Land Registry. (2016, June 14). *How to access HM Land Registry Price Paid Data.* https://www.gov.uk/guidance/about-the-price-paid-data.

HM Land Registry. (2020, November 2). *HM Land Registry data.* https://www.gov.uk/government/statistics/quality-assurance-of-administrative-data-in-the-uk-house-price-index/hm-land-registry-data.

Hoerl, A., & Kennard, R. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55–67.

Hou, Y., Zhang, H., & Zhou, S. (2015). Convolutional neural network-based image representation for visual loop closure detection. In *2015 IEEE International Conference on Information and Automation (ICIA)* (pp. 2238–2245). Lijiang, China.

Israel, M. (2015). *Research ethics and integrity for social scientists.* Sage Publications.

Jaderberg, M., Simonyan, K., Vedaldi, A., & Zisserman, A. (2015). Reading text in the wild with convolutional neural networks. *International Journal of Computer Vision*, *116*(1), 1–20.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: with applications in R.* Springer.

Jendryke, M., Balz, T., & Liao, M. (2017). Big location-based social media messages from China's Sina Weibo network: Collection, storage, visualization, and potential ways of analysis. *Transactions in GIS*, *21*(4), 825–834.

Jendryke, M., Balz, T., Mcclure, S. C., & Liao, M. (2017). Putting people in the picture: Combining big location-based social media data and remote sensing imagery for enhanced contextual urban information in Shanghai. *Computers, Environment and Urban Systems*, *62*, 99–112.

Johnson, N., Carran, S., Botner, J., Fontaine, K., Laxague, N., Nuetzel, P., ... Tivnan, B. (2011). Pattern in escalations in insurgent and terrorist activity. *Science*, *333*(6038), 81–84.

Johnson, N. F., Zheng, M., Vorobyeva, Y., Gabriel, A., Qi, H., Velasquez, N., ... Wuchty, S. (2016). New online ecology of adversarial aggregates: ISIS and beyond. *Science*, *352*(6292), 1459–1463.

Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., & Fei-Fei, L. (2014). Large-scale video classification with convolutional neural networks. In *2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1725–1732). Columbus, OH, USA.

KartaView. (n.d.). *Easy mapping.* Retrieved September 12, 2022, from https://kartaview.org/landing.

Kim, Y. S., & Rous, J. J. (2012). House price convergence: Evidence from US state and metropolitan area panels. *Journal of Housing Economics*, *21*(2), 169–186.

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, *110*(15), 5802–5805.

Kraemer, M. U. G., Yang, C.-H., Gutierrez, B., Wu, C.-H., Klein, B., Pigott, D. M., ... Scarpino, S. V. (2020). The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science*, *368*(6490), 493–497.

Kramer, A. D. I., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, *111*(24), 8788–8790.

Krings, G., Calabrese, F., Ratti, C., & Blondel, V. D. (2009). Urban gravity: a model for inter-city telecommunication flows. *Journal of Statistical Mechanics: Theory and Experiment*, *2009*, Article L07003.

Kristoufek, L. (2013). BitCoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the internet era. *Scientific Reports*, *3*,

Article 3415.

Kristoufek, L., Moat, H. S., & Preis, T. (2016). Estimating suicide occurrence statistics using Google Trends. *EPJ Data Science*, *5*, Article 32.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In F. Pereira, C. J. Burges, L. Bottou, & K. Q. Weinberger (Eds.), *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1 (NIPS'12)* (pp. 1097–1105). Lake Tahoe, NV, USA.

Kröger, J. L., Miceli, M., & Müller, F. (2021). *How data can be used against people: A classification of personal data misuses.* SSRN.

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), Article 54.

Lazer, D., Pentland, A. S., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., ... Van Alstyne, M. (2009). Computational social science. *Science*, *323*(5915), 721–723.

LeCun, Y. (1989). Generalization and network design strategies. In R. Pfeifer, Z. Schreter, F. Fogelman, & L. Steels (Eds.), *Connectionism in perspective.* Elsevier.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*, 436–444.

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural Computation*, *1*(4), 541–551.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2324.

Lefever, S., Dal, M., & Matthíasdóttir, A. (2007). Online data collection in academic research: advantages and limitations. *British Journal of Educational Technology*, *38*, 574–582.

Le Moignan, E., Lawson, S., Rowland, D. A., Mahoney, J., & Briggs, P. (2017). Has Instagram fundamentally altered the 'family snapshot'? In G. Mark et al. (Eds.), *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)* (pp. 4935–4947). Denver, CO, USA.

Letchford, A. (2018, March 31). *Streetview.* GitHub. https://github.com/robolyst/streetview.

Letchford, A., Moat, H. S., & Preis, T. (2015). The advantage of short paper titles. *Royal Society Open Science*, *2*(8), Article 150266.

Letchford, A., Preis, T., & Moat, H. S. (2016a). The advantage of simple paper

abstracts. *Journal of Informetrics*, *10*(1), 1–8.

Letchford, A., Preis, T., & Moat, H. S. (2016b). Quantifying the search behaviour of different demographics using Google Correlate. *PLoS ONE*, *11*(2), Article e0149025.

Li, W., Saeedi, S., McCormac, J., Clark, R., Tzoumanikas, D., Ye, Q., ... Leutenegger, S. (2018). *InteriorNet: Mega-scale multi-sensor photo-realistic indoor scenes dataset.* arXiv.

Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., & Alsaadi, F. E. (2017). A survey of deep neural network architectures and their applications. *Neurocomputing*, *234*, 11–26.

Lu, B., Charlton, M., Harris, P., & Fotheringham, A. S. (2014). Geographically weighted regression with a non-Euclidean distance metric: a case study using hedonic house price data. *International Journal of Geographical Information Science*, *28*(4), 660–681.

Ma, J., Jiang, X., Fan, A., Jiang, J., & Yan, J. (2020). Image matching from handcrafted to deep features: A survey. *International Journal of Computer Vision*, *129*, 23–79.

MacKerron, G., & Mourato, S. (2013). Happiness is greater in natural environments. *Global Environmental Change*, *23*(5), 992–1000.

Manley, E., Zhong, C., & Batty, M. (2016). Spatiotemporal variation in travel regularity through transit user profiling. *Transportation*, *45*(3), 703–732.

Mapillary. (n.d.). *Make betters maps.* Retrieved September 12, 2022, from https://www.mapillary.com/.

Marques-Toledo, C. d. A., Degener, C. M., Vinhal, L., Coelho, G., Meira, W., Codeço, C. T., & Teixeira, M. M. (2017). Dengue prediction by the web: Tweets are a useful tool for estimating and forecasting dengue at country and city level. *PLoS Neglected Tropical Diseases*, *11*(7), Article e0005729.

Maruf, H. A., Meshkat, N., Ali, M. E., & Mahmud, J. (2015). Human behaviour in different social medias: A case study of Twitter and Disqus. In J. Pei, F. Silvestri, & J. Tang (Eds.), *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015 (ASONAM '15)* (pp. 270–273). Paris, France.

Mastrandrea, R., Fournet, J., & Barrat, A. (2015). Contact patterns in a high school: A comparison between data collected using wearable sensors, contact diaries and friendship surveys. *PLoS ONE*, *10*(9), Article e0136497.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, *5*, 115–133.

McKenna, R., Weinand, J. M., Mulalic, I., Petrović, S., Mainzer, K., Preis, T., & Moat, H. S. (2021). Scenicness assessment of onshore wind sites with geotagged photographs and impacts on approval and cost-efficiency. *Nature Energy*, *6*, 663–672.

McLaren, N., & Shanbhogue, R. (2011). Using internet search data as economic indicators. *Bank of England Quarterly Bulletin*, *51*(2), 134–140.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2022). A survey on bias and fairness in machine learning. *ACM Computing Surveys*, *54*(6), Article 115.

Mestyán, M., Yasseri, T., & Kertész, J. (2013). Early prediction of movie box office success based on Wikipedia activity big data. *PLoS ONE*, *8*(8), Article e71226.

Meta. (2022, July 27). *Meta reports second quarter 2022 results.* https://investor.fb.com/investor-events/event-details/2022/Q2-2022-Earnings/default.aspx.

Miller, S., Moat, H. S., & Preis, T. (2020). Using aircraft location data to estimate current economic activity. *Scientific Reports*, *10*, Article 7576.

Miller, S., Preis, T., Mizzi, G., Bastos, L. S., Gomes, M. F. d. C., Coelho, F. C., ... Moat, H. S. (2022). Faster indicators of chikungunya incidence using Google searches. *PLoS Neglected Tropical Diseases*, *16*(6), Article e0010441.

Mingers, J., & Leydesdorff, L. (2015). A review of theory and practice in scientometrics. *European Journal of Operational Research*, *246*(1), 1–19.

Mitra, P., Sanyal, A., & Choudhuri, S. (2017). Nowcasting real estate activity in India using Google Trend data. *Reserve Bank of India Occasional Papers*, *38*(1), 1–32.

Mizzi, G., Preis, T., Bastos, L. S., Gomes, M. F. d. C., Codeço, C. T., & Moat, H. S. (2021). *Faster indicators of dengue fever case counts using Google and Twitter.* arXiv.

Moat, H. S., Curme, C., Avakian, A., Kenett, D. Y., Stanley, H. E., & Preis, T. (2013). Quantifying Wikipedia usage patterns before stock market moves. *Scientific Reports*, *3*, Article 1801.

Moat, H. S., Olivola, C. Y., Chater, N., & Preis, T. (2016). Searching choices: Quantifying decision-making processes using search engine data. *Topics in Cognitive Science*, *8*(3), 685–696.

Moat, H. S., Preis, T., Olivola, C. Y., Liu, C., & Chater, N. (2014). Using big data to predict collective behavior in the real world. *Behavioral and Brain Sciences*, *37*(1), 92–93.

Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P., & Tita, G. E. (2011). Self-exciting point process modeling of crime. *Journal of the American*

Statistical Association, *106*(493), 100–108.

Morancho, A. B. (2003). A hedonic valuation of urban green areas. *Landscape and Urban Planning*, *66*(1), 35–41.

Mullett, T. L., & Stewart, N. (2016). Implications of visual attention phenomena for models of preferential choice. *Decision*, *3*(4), 231–253.

Münnix, M. C., Shimada, T., Schäfer, R., Leyvraz, F., Seligman, T. H., Guhr, T., & Stanley, H. E. (2012). Identifying states of a financial market. *Scientific Reports*, *2*, Article 644.

Narayan, P. K., & Bannigidadmath, D. (2017). Does financial news predict stock returns? New evidence from islamic and non-islamic stocks. *Pacific-Basin Finance Journal*, *42*, 24–45.

Nayak, P. (2019, October 25). Understanding searches better than ever before. *Google*. https://blog.google/products/search/search-language-understanding-bert/.

Neuhold, G., Ollmann, T., Rota Bulo, S., & Kontschieder, P. (2017). The Mapillary Vistas Dataset for semantic understanding of street scenes. In *2017 IEEE International Conference on Computer Vision (ICCV)* (pp. 4990–4999). Venice, Italy.

Nguyen, D. T., Kim, K. W., Hong, H. G., Koo, J. H., Kim, M. C., & Park, K. R. (2017). Gender recognition from human-body images using visible-light and thermal camera videos based on a convolutional neural network for image feature extraction. *Sensors*, *17*(3), Article 637.

Noguchi, T., Stewart, N., Olivola, C. Y., Moat, H. S., & Preis, T. (2014). Characterizing the time-perspective of nations with search engine query data. *PLoS ONE*, *9*(4), Article e95209.

Nogueira, S., Sechidis, K., & Brown, G. (2018). On the stability of feature selection algorithms. *Journal of Machine Learning Research*, *18*(174), 1–54.

Nouriani, A., & Lemke, L. (2022). Vision-based housing price estimation using interior, exterior & satellite images. *Intelligent Systems with Applications*, *14*, Article 200081.

Odgers, C., Caspi, A., Bates, C., Sampson, R., & Moffitt, T. (2012). Systematic social observation of children's neighborhoods using Google Street View: a reliable and cost-effective method. *Journal of Child Psychology and Psychiatry*, *53*(10), 1009–1017.

Oestmann, M., & Bennöhr, L. (2015). Determinants of house price dynamics. What can we learn from search engine data? *Review of Economics*, *66*(1), 99–127.

Office for National Statistics. (2014a, January 23). *Concealed family status by family*

type by dependent children by age of Family Reference Person (FRP). https://www.nomisweb.co.uk/census/2011/lc1110ew.

Office for National Statistics. (2014b, May 23). *Employment status (Workplace population).* https://www.nomisweb.co.uk/census/2011/wp601ew.

Olivola, C. Y., Moat, H. S., & Preis, T. (2019). Using big data to map the relationship between time perspectives and economic outputs. *Behavioral and Brain Sciences*, *42*, Article E206.

Osland, L. (2010). An application of spatial econometrics in relation to hedonic house price modeling. *Journal of Real Estate Research*, *32*(3), 289–320.

Osorio, F. (2020, July 19). *Robust estimation using heavy-tailed distributions.* GitHub. https://github.com/faosorios/heavy.

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.

Pain, N., & Westaway, P. (1997). Modelling structural change in the UK housing market: A comparison of alternative house price models. *Economic Modelling*, *14*(4), 587–610.

Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015). Deep face recognition. In X. Xie, M. W. Jones, & G. K. L. Tam (Eds.), *Proceedings of the British Machine Vision Conference (BMVC)* (p. Article 41). Swansea, UK.

Patterson, G., Xu, C., Su, H., & Hays, J. (2014). The SUN attribute database: Beyond categories for deeper scene understanding. *International Journal of Computer Vision*, *108*, 59–81.

Pavel, M., Jimison, H. B., Korhonen, I., Gordon, C. M., & Saranummi, N. (2015). Behavioral informatics and computational modeling in support of proactive health management and care. *IEEE Transactions on Biomedical Engineering*, *62*(12), 2763–2775.

Paül i Agustí, D. (2021). Mapping gender in tourist behaviour based on Instagram. *Journal of Outdoor Recreation and Tourism*, *35*, Article 100381.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Penner, O., Pan, R. K., Petersen, A. M., Kaski, K., & Fortunato, S. (2013). On the predictability of future impact in science. *Scientific Reports*, *3*, Article 3052.

Petersen, A. M., Riccaboni, M., Stanley, H. E., & Pammolli, F. (2012). Persistence and uncertainty in the academic career. *Proceedings of the National Academy of Sciences*, *109*(14), 5213–5218.

Piškorec, M., Antulov-Fantulin, N., Novak, P. K., Mozetič, I., Grčar, M., Vodenska, I., & Šmuc, T. (2014). Cohesiveness in financial news and its relation to market volatility. *Scientific Reports*, *4*, Article 5038.

Poole, M. A., & O'Farrell, P. N. (1971). The assumptions of the linear regression model. *Transactions of the Institute of British Geographers*, *52*, 145–158.

Preis, T., Botta, F., & Moat, H. S. (2020). Sensing global tourism numbers with millions of publicly shared online photographs. *Environment and Planning A: Economy and Space*, *52*(3), 471–477.

Preis, T., Kenett, D. Y., Stanley, H. E., Helbing, D., & Ben-Jacob, E. (2012). Quantifying the behavior of stock correlations under market stress. *Scientific Reports*, *2*, Article 752.

Preis, T., & Moat, H. S. (2014). Adaptive nowcasting of influenza outbreaks using Google searches. *Royal Society Open Science*, *1*(2), Article 140095.

Preis, T., & Moat, H. S. (2015). Early signs of financial market moves reflected by Google searches. In B. Gonçalves & N. Perra (Eds.), *Social phenomena: From data analysis to models* (pp. 85–97). Springer.

Preis, T., Moat, H. S., Bishop, S. R., Treleaven, P., & Stanley, H. E. (2013). Quantifying the digital traces of Hurricane Sandy on Flickr. *Scientific Reports*, *3*, Article 3141.

Preis, T., Moat, H. S., & Stanley, H. E. (2013). Quantifying trading behavior in financial markets using Google Trends. *Scientific Reports*, *3*, Article 1684.

Preis, T., Moat, H. S., Stanley, H. E., & Bishop, S. R. (2012). Quantifying the advantage of looking forward. *Scientific Reports*, *2*, Article 350.

Preis, T., Schneider, J. J., & Stanley, H. E. (2011). Switching processes in financial markets. *Proceedings of the National Academy of Sciences*, *108*(19), 7674–7678.

Quercia, D., Schifanella, R., & Aiello, L. M. (2014). The shortest path to happiness: recommending beautiful, quiet, and happy routes in the city. In L. Ferres, G. Rossi, V. Almeida, & E. Herder (Eds.), *Proceedings of the 25th ACM Conference on Hypertext and Social Media (HT '14)* (pp. 116–125). Santiago, Chile.

Radicchi, F., Fortunato, S., & Castellano, C. (2008). Universality of citation distributions: Toward an objective measure of scientific impact. *Proceedings of the National Academy of Sciences*, *105*(45), 17268–17272.

Radicchi, F., Fortunato, S., Markines, B., & Vespignani, A. (2009). Diffusion of scientific credits and the ranking of scientists. *Physical Review E*, *80*(5), Article 056103.

Redmon, J., & Farhadi, A. (2016). *YOLO9000: Better, faster, stronger.* arXiv.

Reid, E. (2020, February 06). *A look back at 15 years of mapping the world.* Google. https://blog.google/products/maps/look-back-15-years-mapping-world.

Rodriguez, S. (2021, December 14). *Instagram surpasses 2 billion monthly users while powering through a year of turmoil.* CNBC. https://www.cnbc.com/2021/12/14/instagram-surpasses-2-billion-monthly-users.html.

Rossi, L., Boscaro, E., & Torsello, A. (2018). Venice through the lens of Instagram: A visual narrative of tourism in Venice. In P.-A. Champin, F. Gandon, L. Médini, M. Lalmas, & P. G. Ipeirotis (Eds.), *Companion Proceedings of the The Web Conference 2018 (WWW '18)* (pp. 1190–1197). Lyon, France.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*, 533–536.

Rundle, A., Bader, M., Richards, C., Neckerman, K., & Teitler, J. (2011). Using Google Street View to audit neighborhood environments. *American Journal of Preventive Medicine*, *40*(1), 94–100.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... Fei-Fei, L. (2015). ImageNet large scale visual recognition challenge. *International Journal of Computer Vision*, *115*, 211–252.

Sachin, R., Sowmya, V., Govind, D., & Soman, K. P. (2017). Dependency of various color and intensity planes on CNN based image classification. In S. M. Thampi, S. Krishnan, J. M. C. Rodriguez, S. Das, M. Wozniak, & D. Al-Jumeily (Eds.), *Advances in intelligent systems and computing: Vol. 678. Advances in Signal Processing and Intelligent Recognition Systems.* (pp. 167–177). Springer.

Schootman, M., Toor, A., Cavazos-Rehg, P., Jeffe, D. B., McQueen, A., Eberth, J., & Davidson, N. O. (2015). The utility of Google Trends data to examine interest in cancer screening. *BMJ Open*, *5*, Article e006678.

scikit-learn. (n.d.). *sklearn.linear_model.ElasticNet.* Retrieved August 13, 2017, from https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.ElasticNet.html.

Seo, Y., & Simons, R. A. (2009). The effect of school quality on residential sales price. *Journal of Real Estate Research*, *31*(3), 307–328.

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

Seresinhe, C. I., Preis, T., MacKerron, G., & Moat, H. S. (2019). Happiness is greater in more scenic locations. *Scientific Reports*, *9*, Article 4498.

Seresinhe, C. I., Preis, T., & Moat, H. S. (2015). Quantifying the impact of scenic environments on health. *Scientific Reports*, *5*, Article 16899.

Seresinhe, C. I., Preis, T., & Moat, H. S. (2016). Quantifying the link between art and property prices in urban neighbourhoods. *Royal Society Open Science*, *3*(4), Article 160146.

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), Article 170170.

Sharkawy, A.-N. (2020). Principle of neural network and its main types: Review. *Journal of Advances in Applied & Computational Mathematics*, *7*, 8–19.

Shen, J., Brdiczka, O., & Liu, J. (2015). A study of Facebook behavior: What does it tell about your neuroticism and extraversion? *Computers in Human Behavior*, *45*, 32–38.

Simonyan, K., & Zisserman, A. (2014). Two-stream convolutional networks for action recognition in videos. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, & K. Q. Weinberger (Eds.), *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 1 (NIPS'14)* (pp. 568–576). Montreal, Canada.

Sims, S., Dent, P., & Oskrochi, G. R. (2008). Modelling the impact of wind farms on house prices in the UK. *International Journal of Strategic Property Management*, *12*(4), 251–269.

Sirmans, G. S., MacDonald, L., Macpherson, D. A., & Zietz, E. N. (2006). The value of housing characteristics: A meta analysis. *Journal of Real Estate Finance and Economics*, *33*, 215–240.

Skarlatidou, A., Ludwig, L., Solymosi, R., & Bradford, B. (2021). Understanding knife crime and trust in police with young people in East London. *Crime & Delinquency*, Article 001112872110298.

Sloan, L., & Morgan, J. (2015). Who tweets with their location? Understanding the relationship between demographic characteristics and the use of geoservices and geotagging on Twitter. *PLoS ONE*, *10*(11), Article e0142209.

Solymosi, R., Ashby, M. P. J., Cohen, T., & Sidebottom, A. (2017). *Alternative denominators in transport crime rates.* SocArXiv.

Solymosi, R., Bowers, K., & Fujiyama, T. (2015). Mapping fear of crime as a context-dependent everyday experience that varies in space and time. *Legal and Criminological Psychology*, *20*(2), 193–211.

Solymosi, R., Bowers, K. J., & Fujiyama, T. (2017). Crowdsourcing subjective perceptions of neighbourhood disorder: Interpreting bias in open data. *British Journal of Criminology*, *58*(4), 944–967.

Solymosi, R., Cella, K., & Newton, A. (2017). Did they report it to stop it? A realist evaluation of the effect of an advertising campaign on victims' willingness to

report unwanted sexual behaviour. *Security Journal*, *31*, 570–590.

Sornette, D., Woodard, R., & Zhou, W.-X. (2009). The 2006–2008 oil bubble: Evidence of speculation, and prediction. *Physica A: Statistical Mechanics and its Applications*, *388*(8), 1571–1576.

Souma, W., Vodenska, I., & Aoyama, H. (2019). Enhanced news sentiment analysis using deep learning methods. *Journal of Computational Social Science*, *2*, 33–46.

StatCounter. (2022, August). *Desktop search engine market share worldwide - August 2022.* https://gs.statcounter.com/search-engine-market-share/desktop/worldwide/#monthly-202207-202207-bar.

Steenbruggen, J., Tranos, E., & Nijkamp, P. (2015). Data from mobile phone operators: A tool for smarter cities? *Telecommunications Policy*, *39*(3), 335–346.

Stigler, S. M. (1981). Gauss and the invention of least squares. *Annals of Statistics*, *9*(3), 465–474.

Stokes, P. (2012, November 23). *2011 census: Population and household estimates for small areas in England and Wales, March 2011.* Office for National Statistics. https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/2011censuspopulationandhouseholdestimatesforsmallareasinenglandandwales/2012-11-23.

Suel, E., Polak, J. W., Bennett, J. E., & Ezzati, M. (2019). Measuring social, environmental and health inequalities using deep learning and street imagery. *Scientific Reports*, *9*, Article 6229.

Sulis, P., & Manley, E. (2018). Exploring similarities and variations of human mobility patterns in the city of London. In B. A. Mohamed, B. A. Abdelhakim, & Y. Ali (Eds.), *Proceedings of the 3rd International Conference on Smart Data and Smart Cities (ISPRS)* (pp. 51–58). Tetouan, Morocco.

Sun, T., Wang, Y., Yang, J., & Hu, X. (2017). Convolution neural networks with two pathways for image style recognition. *IEEE Transactions on Image Processing*, *26*(9), 4102–4113.

Takhteyev, Y., Gruzd, A., & Wellman, B. (2012). Geography of Twitter networks. *Social Networks*, *34*(1), 73–81.

Tenkanen, H., Di Minin, E., Heikinheimo, V., Hausmann, A., Herbst, M., Kajala, L., & Toivonen, T. (2017). Instagram, Flickr, or Twitter: Assessing the usability of social media data for visitor monitoring in protected areas. *Scientific Reports*, *7*, Article 17615.

Tesfatsion, L. (2002). Agent-based computational economics: growing economies from the bottom up. *Artificial Life*, *8*(1), 55–82.

Tetlock, P. C. (2010). Does public financial news resolve asymmetric information? *Review of Financial Studies*, *23*(9), 3520–3557.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288.

Trendl, A., Stewart, N., & Mullett, T. L. (2021). The role of alcohol in the link between national football (soccer) tournaments and domestic abuse - Evidence from England. *Social Science & Medicine*, *268*, Article 113457.

Trevisan, M., Vassio, L., Drago, I., Mellia, M., Murai, F., Figueiredo, F., ... Almeida, J. M. (2019). Towards understanding political interactions on Instagram. In C. Atzenbeck, J. Rubart, & D. E. Millard (Eds.), *Proceedings of the 30th ACM Conference on Hypertext and Social Media (HT '19)* (pp. 247–251). Hof, Germany.

Tumasjan, A., Sprenger, T., Sandner, P., & Welpe, I. (2010). Predicting elections with Twitter: What 140 characters reveal about political sentiment. *Proceedings of the International AAAI Conference on Weblogs and Social Media*, *4*(1), 178–185.

Twitter. (2022, July 22). *Twitter announces second quarter 2022 results.* [Press release]. https://s22.q4cdn.com/826641620/files/doc_financials/2022/q2/Final_Q2'22_Earnings_Release.pdf.

Vandeviver, C. (2014). Applying Google Maps and Google Street View in criminological research. *Crime Science*, *3*, Article 13.

Weibo. (2020, May 19). *Weibo reports first quarter 2020 unaudited financial results.* [Press release]. http://ir.weibo.com/news-releases/news-release-details/weibo-reports-first-quarter-2020-unaudited-financial-results.

Weinberg, D. (2014). Cityscape data sources for U.S. housing research, part 1: Public sector data sources. *Cityscape: A Journal of Policy Development and Research*, *16*(3), 131–147.

Widgrén, J. (2016). *Predicting housing prices with Google searches in Finland* (ETLA Reports No. 63). Helsinki, Finland: The Research Institute of the Finnish Economy.

Wilamowski, B. (2009). Neural network architectures and learning algorithms. *IEEE Industrial Electronics Magazine*, *3*(4), 56–63.

Wood, S. A., Guerry, A. D., Silver, J. M., & Lacayo, M. (2013). Using social media to quantify nature-based tourism and recreation. *Scientific Reports*, *3*, Article 2976.

Woolley, S. C. (2016). Automating power: Social bot interference in global politics. *First Monday*, *21*(4).

Yao, J., & Fotheringham, A. S. (2016). Local spatiotemporal modeling of house prices: A mixed model approach. *Professional Geographer*, *68*(2), 189–201.

Yasseri, T., Kornai, A., & Kertész, J. (2012). A practical approach to language complexity: A Wikipedia case study. *PLoS ONE*, *7*(11), Article e48386.

Yasseri, T., Sumi, R., & Kertész, J. (2012). Circadian patterns of Wikipedia editorial activity: A demographic analysis. *PLoS ONE*, *7*(1), Article e30091.

Yasseri, T., Sumi, R., Rung, A., Kornai, A., & Kertész, J. (2012). Dynamics of conflicts in Wikipedia. *PLoS ONE*, *7*(6), Article e38869.

Ye, Z., Newing, A., & Clarke, G. (2020). Understanding Chinese tourist mobility and consumption-related behaviours in London using Sina Weibo check-ins. *Environment and Planning B: Urban Analytics and City Science*, *48*(8), 2436–2452.

Zhang, Y., Li, B., & Hong, J. (2016). Understanding user economic behavior in the city using large-scale geotagged and crowdsourced data. In J. Bourdeau, J. A. Hendler, R. N. Nkambou, I. Horrocks, & B. Y. Zhao (Eds.), *Proceedings of the 25th International Conference on World Wide Web (WWW '16)* (pp. 205–214). Montréal, Québec, Canada.

Zheng, S., Wang, J., Sun, C., Zhang, X., & Kahn, M. E. (2019). Air pollution lowers Chinese urbanites' expressed happiness on social media. *Nature Human Behaviour*, *3*, 237–243.

Zhong, C., Arisona, S. M., Huang, X., Batty, M., & Schmitt, G. (2014). Detecting the dynamics of urban structure through spatial network analysis. *International Journal of Geographical Information Science*, *28*(11), 2178–2199.

Zhong, C., Manley, E., Müller Arisona, S., Batty, M., & Schmitt, G. (2015). Measuring variability of mobility patterns from multiday smart-card data. *Journal of Computational Science*, *9*, 125–130.

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.

Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., & Oliva, A. (2014). Learning deep features for scene recognition using Places database. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, & K. Q. Weinberger (Eds.), *Proceedings of the 27th International Conference on Neural Information Processing Systems (NIPS'14)* (pp. 487–495). Montréal, Québec, Canada.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic-net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *67*(2), 301–320.