Quantifying Human Behaviour With Online Images

by

Merve Alanyali

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Warwick Business School

June 2018

Contents

List of Tables

List of Figures

Acknowledgments

First of all, I would like to express my deepest and warmest gratitude to Professor Tobias Preis and Professor Suzy Moat who have been more than just supervisors to me since my MSc years. Without their endless support and guidance you would not be reading this thesis. It has been such a valuable experience for me to be working with you but most importantly getting to know such amazing academics and people. I am also very grateful to the University of Warwick (Chancellor's International Scholarship Scheme) and The Alan Turing Institute (the enrichment year programme) for financially supporting my studies.

I am also very thankful to all my friends in the Data Science Lab, Complexity Centre, Warwick Business School and The Alan Turing Institute for their valuable feedback and comments which immensely helped me shape and improve my work. I would also like to thank them for being great companions and for keeping me sane (maybe!) during stressful times.

I am thankful to every single person who has touched my life in many ways. I will probably need to write another thesis to fit everyone in but there are a few names I want to mention before proceeding to the serious content. My special thanks go to:

My very first teacher, Sehrayi Con, for her persistence to make me enjoy going to school and building strong ground for my (never-ending) journey of education,

All my teachers at Izmir MEV Koleji, Izmir Ozel Turk Fen Lisesi for all the science and life lessons they taught,

My professors at the Computer Science Department of Izmir University of Economics for helping me establish foundations of my academic career (even though this might have meant making my university years pretty much a living hell),

All my true friends who share my happiness and sadness and send their constant love and support even they are in a country far far away (maybe not that far, but still) (Merve, Ozge, Selin, Duygu, Begum, Gokhan, Alican, Ekincan and many more.),

Teoman aka bafir for being a true companion of all the fun(!) we have to endure in this life,

Andrea for bringing me "after lunch coffee" in his Italian way and most of all being

the best *ngg* friend (also for proofreading my thesis),

Diana for being my very first true friend in England (life at Warwick would have been dull without you!),

Chanuki and Ben for being the ultimate source of energy and fun and making sure I do not fall behind social life (extra thanks to Ben for his super human proofreading skills and Chanuki for, apart from being a true friend, also being a great colleague and collaborator),

Busem for being the bestest of best friends from miles away (best sister-from-a-different-mum I could have asked for),

#dogsofinstagram, especially the_daily_ollie for brightening my dull days with their fluffy posts,

"The gang" for keeping my inner(!) child alive,

My family on earth and in heaven (both mine and Tom's) for their unconditional love and infinite support (and for all my loved ones in heaven - my thoughts are always with you),

And let me stop here. I have been trying to write this part for so long that I can say it is harder than writing the entire thesis. Because there are no words to express what these two people mean to me. I will try anyway knowing that words will not do any justice to them.

Mum - without your limitless support and motivation I would not have become the person I am now. Can not thank you enough for everything you have done (and keep doing) for me. You are my true inspiration!

Tom (or shall I say Dr Rafferty?) - without you by my side (especially during the big resets) in academic life and beyond I cannot image how I would have coped. Thank you for "being the unique solution to my ordinary differential equation".

Merve

Declarations

This work has been composed by myself and has not been submitted for any other degree or professional qualification.

- Chapter 3 has been published in Alanyali et al. (2016). At the time of writing this thesis, it has received six citation and received "Honourable Mentions" at 2016 Association of British and Turkish Academics (ABTA) Doctoral Researcher Awards under Management & Social Sciences category.
- The work done in Chapters 3 and 4 has been presented together in more than ten conference talks, including five invited talks. It has also been presented as a poster at three events.
- The work done in Chapters 4 and 5 will be submitted for publication.
- At the time of writing this thesis, the work from Chapter 5 has been presented in three invited meet-up and conference talks.
- Part of the work presented in Chapter 6 has been submitted to *Royal Society Open Science* and is currently under review.

To know that you do not know is the best.

To think you know when you do not is a disease.

Recognising this disease as a disease is to be free of it.

Lao Tzu

Abstract

From online searches to social media posts, our everyday interactions with the Internet are creating vast amounts of data. Large volumes of this data can be accessed rapidly at low cost, opening up unprecedented possibilities to monitor and analyse social processes and measure human behaviour.

As Internet connectivity has continued to improve, photo-sharing platforms such as *Instagram* and *Flickr* have gained widespread popularity. At the same time, considerable advances have been achieved in the power of computers to analyse the contents of images. In particular, deep learning based methods such as convolutional neural networks have radically transformed the performance of systems seeking to identify objects in images, or classify the contents of a scene.

Here, we showcase a series of studies in which we seek to quantify various aspects of human behaviour by exploiting both the large quantities of photographic data shared online and recent developments in computer vision. Specifically, we investigate whether data extracted from photographs shared on *Flickr* and *Instagram* can help us track global protest outbreaks; estimate the income of inhabitants living in different areas of London and New York; and predict the occurrence of noise complaints in New York City.

Our findings are in line with the striking hypothesis that information extracted through automatic analysis of photographs shared online may help us measure human behaviour, whether in individual cities or across the globe.

CHAPTER 1

Introduction

Developments in technological devices are placing them at the very heart of our daily routines, and changing many aspects of our lives. From mobile phones to computers, the widespread usage of such devices and the online services they connect us to are generating vast amounts of data documenting everyday behaviour at a national or even global scale.

As a consequence of improved connectivity, sharing visual media has become ubiquitous in recent years. More and more online posts, particularly on social media channels, have shifted from being solely text based to include multimedia, such as videos and photographs. Here, by exploiting the vast amount of photographs shared online, we present a series of studies investigating how state-of-the-art image analysis methods can be applied on this new form of data in order to detect global events, estimate socioeconomic statistics and predict the location of non-emergency incidents.

In Chapter 2, we cover a wide range of example studies in the emerging field of Computational Social Science. We provide an extensive discussion of how previous research utilised data extracted in numerous forms from online platforms including search engines and social media channels in order to gain insights into human behaviour. We also present a detailed discussion of the advances in image analysis algorithms with a primary focus on deep learning methods.

Over the last few decades, we have witnessed an increased number of protests emerging across countries and continents, sometimes leading to political change or mass casualties. During times of protests, online users turn to social media channels to organise protests, mobilise people and spread information. This increased usage of social media is generating large amounts of data and creating almost real-time reports of protest outbreaks around the world. In Chapter 3, we analyse textual data attached to a large set of photographs shared on *Flickr* to investigate whether it is possible to use this data to track protest outbreaks across the globe. We quantify the relationship between the number of pictures on *Flickr* uploaded with a tag containing the word “protest” in 34 different languages and the number of protest related news articles published in the online version of the newspaper *The Guardian*. We find that greater numbers of protest tagged pictures correspond to higher proportions of protest related news articles.

In addition to text based data, *Flickr* offers a rich set of information: the photographs themselves. Ignited by advances in computational power and an increased number of data sets available online, deep learning architectures especially convolutional neural networks have proved their power in numerous image analysis problems including classification. In Chapter 4, we therefore extend our initial analysis from Chapter 3 by incorporating data extracted from *Flickr* photographs using a convolutional neural network based framework. Our findings provide evidence that a higher number of pictures automatically classified as being protest related by our custom-built classifier is linked to a higher proportion of protest related news articles in the newspaper *The Guardian*.

A portrayal of the socioeconomic status of a country is immensely crucial for policy makers. For decades, the main source of such information has been surveys conducted by national agencies. Despite offering rich and valuable information, running such labour intensive exercises can be immensely costly. Furthermore, results are usually released with a delay and therefore do not necessarily reflect the current status of a city or a country. In Chapter 5, we demonstrate how visual characteristics of images shared on *Instagram* can help us create a spatial income profile of two major cities, London and New York City. Our findings set an example that automatic analysis of online pictures may give us insight into key socioeconomic attributes of metropolitan areas around the world.

Modern cities are plagued by a myriad of problems. Some of these, such as criminal activity, are handled by emergency services. However, there are other types of problems that affect the smooth functioning of a city, such as noisy neighbourhoods, faulty traffic lights or illegal parking. In recent years, in order to rapidly resolve such problems, a number of cities have introduced systems to help citizens report issues they encounter. A key example is New York City's 311 services. In Chapter 6, we show that data on complaints reported to the 311 services can be used not only to monitor problems the city is currently facing but also to predict where related problems may be reported next. Finally, we investigate whether we can create a similar early warning mechanism for noise related complaints by analysing photographs shared on *Flickr*. Our results suggest that appropriate analysis of data generated in urban settings and the photographs shared online could create early signals of locations in which future incidents will be reported.

CHAPTER 2

Background

From communication to transportation and shopping to daily exercise, technological devices reside at the very centre of modern life. The widespread usage of these devices and the online services they offer are creating strikingly detailed data on everyday behaviour. These gigantic streams of online data tend to be available at high speed and low cost, offering new ways to measure aspects of individual and group behaviour at a national or even global scale. Researchers therefore have begun to investigate whether this fast growing online data can be used as a practical supplement to the information extracted from traditional methods used to study human behaviour. This has given rise to a new field of data-driven research, often referred to as Computational Social Science or Social Data Science (Conte et al., 2012; King, 2011; Lazer et al., 2009; Moat et al., 2014).

In the first part of this chapter we demonstrate a wide range of studies that fall within the field of Computational Social Science. We provide a detailed discussion of how these studies exploit online data extracted from numerous channels to provide insights into various aspects of human behaviour. In the second part of this chapter, we present a review of the methods used for analysing images, which is useful for the work we present in the following chapters.

2.1 Computational Social Science

The increasing quantities of available data documenting human behaviour is opening up new ways to address problems arising in social sciences, even offering possibilities to tackle problems that were previously intractable using traditional data sources. From search engines to social media platforms, a diverse set of channels are contributing to this expanding data generation process. In recent years with improved Internet connectivity, the form of online data has shifted from being solely text based to multimedia such as pictures and videos. Ignited by the vast quantities of online data in many forms, an increasing number of studies have been undertaken in the emerging field of Computational Social Science (Conte et al., 2012; King, 2011; Lazer et al., 2009; Moat et al., 2014). By drawing on the methods developed across a wide range of disciplines such as computer science and

statistics, these studies aim to create a new form of “mass ethnography” (Bentley et al., 2014).

In the following sections, we will focus on individual examples operating on a diverse set of online data. Despite the amount of data available online, it is naive to assume that an online user profile is an exact representation of the entire demographic. We therefore provide a discussion on potential biases that online data might incorporate as well as the pitfalls that such biases might cause in any analysis. Data ethics and ownership are other crucial issues that need to be considered when working with online data. Thus, we include a brief review touching these issues before finalising our discussion on Computational Social Science.

2.1.1 Internet as an information source

The proliferation of technology is changing multiple aspects of our daily lives. A significant change is happening in the way we collect information. Information gathering is a crucial step in the decision making process, as it enables us to refine the coarse prior knowledge upon which we base our initial opinion (Simon, 1955). Building upon the vast amount of easily accessible information, the Internet has become the ultimate information source for individuals making decisions in the modern world (Moat et al., 2016). As more and more people turn to the Internet in search of information, an increasing number of studies exploit this increased online activity as a proxy of collective consciousness.

Over the last few decades, search engines have made a big impact on how we search for information. Constantly improving their indexing and searching algorithms, they provide users with quick and effective ways to retrieve information. *Google*, without a doubt, is one of the most popular search engines. In addition to helping online users find information, it also makes search volume data publicly available via its *Google Trends* service. Search terms are given a relative popularity, which is calculated by normalising the search frequency of a specific term with the total search volume coming from that location over a certain period of time. Relative search volumes are available since 2004 with a weekly granularity, however, if the requested historical data is closer to the time of the request, then the time granularity can go down to days or even hours. Motivated by the large amount of publicly available and easily accessible data combined with a user friendly interface that also provides basic visualisation of the underlying data, *Google Trends* has been a popular source of data among researchers.

Drawing attention to the importance of real time monitoring, Choi and Varian (2012) focused on nowcasting - in other words estimating the real time values of car sales, holiday spendings and US employment claims using search volume data from *Google*. A similar study was conducted by Askitas and Zimmermann (2009) on forecasting unemployment rates in Germany by exploiting the frequency of job-related and unemployment-related search phrases submitted to *Google*. Using *Google Trends* data, Preis et al. (2012) performed a global study to create a “future orientation index” which is a measure indicating

whether online users tend to search for information about the future rather than the past. Their findings suggest that nations with a higher Gross Domestic Product (GDP) tend to focus more on the future than the past. Building on this study, Noguchi et al. (2014) investigated whether time-perspectives for nations change in relation with their GDP. Letchford et al. (2016) used *Google Correlate* to compare search behaviour across US states. They illustrated how search behaviour varied with demographic variables such as infant mortality rates, providing evidence that search data might offer insight into the concerns of different demographics.

Changes in stock markets have an impact on the lives of many individuals both from the financial sector and beyond. Understanding and predicting the behaviour of this complex system therefore has obvious benefits. Hence, a number of studies have focused on using *Google Trends* data to create early warning signals before stock market moves. Preis et al. (2013b) demonstrated that changes in the number of finance related terms submitted to *Google* can be used as indicators of stock market movements. They constructed a hypothetical trading strategy to buy or sell the Dow Jones Industrial Average (DJIA) by using the search volume of a wide range of terms related to the stock markets which which they refer to as the "*Google Trends* strategy". They showed that the *Google Trends* strategy implemented for the search volume of the term "debt" within a three week window would have increased the portfolio by 326% whereas the buy and hold investment strategy yielded an increase of only 16%. Similar results were found when using *Google* searches for the names of Dow Jones companies to anticipate movements in the value of the company's stock (Preis and Moat, 2015). A separate study showed how data from *Google* searches can be used in portfolio selection and risk diversification (Kristoufek, 2013b).

Another prominent scenario in which search engine data has proved to be useful is monitoring public health. Timely detection of disease activity is crucial for taking rapid actions to prevent the further spread of the disease. A whole body of research has focused on creating timely estimates for the spread of influenza like epidemics using search data. Ginsberg et al. (2009) provided evidence that spread of the influenza like diseases can be predicted by analysing the search volume of flu related words on *Google* It was also turned into a *Google* service; "*Google Flu Trends*" which was releasing an estimate for the number of cases of an epidemic disease such as flu and dengue. This service no longer releases new estimates however historical data is available to download.

Despite being a big success when published, *Google Flu* trends failed between 2012-2013 by overestimating the flu spread. Alternative studies were published discussing the potential reasons causing the failure as well as enhancing the initial method to show that search engine data may provide significant information on the key health indicators (Lazer et al., 2014; Preis and Moat, 2014). Kristoufek et al. (2016) detailed another application of search data in the area of public health, investigating whether *Google* data can be used to improve estimates of suicide occurrence statistics.

There are a number of other studies that exploit data from alternative search engines or tools. Goel et al. (2010) used data from *Yahoo* search results to predict consumer

activity including box office revenue, sales of video games and music charts. In Bordino et al. (2012), the authors extracted search volume for the queries made to *Yahoo* related to the companies listed in NASDAQ-100. They found evidence of a positive correlation between stock related search volume and the trading volume of the same stocks on following days.

Instead of focusing on search volume data from a single search engine, Ettredge et al. (2005) used WordTracker, which extracts data from the "Web's largest meta search engines". In Hulth et al. (2009), the authors took a different approach by analysing the volume of search queries sent to a Swedish medical website to show the potential of web queries in syndromic surveillance.

Search engines are not the only source that online users turn to when seeking information. *Wikipedia* is a web-based free encyclopedia created and updated by millions of volunteers around the world. With more than 16 billion page views in March 2018 (Wikimedia, 2018), *Wikipedia* is one of the most visited websites on the World Wide Web. The numbers of visits and edits for each individual page are recorded and made publicly available. Just like the data on search volumes, this large dataset has quickly become popular among researchers as a valuable source of information documenting online activity of *Wikipedia* users. Moat et al. (2013) provided evidence that the amount of traffic attracted by finance related *Wikipedia* pages can be used as early indicators of stock market moves. Another study utilised *Wikipedia* edit logs in order to investigate the behaviour of multilingual users, which are defined as users editing *Wikipedia* pages in multiple languages (Hale, 2014). The authors showed that on average, multilingual editors tend to be 2.3 times more active in editing compared to monolingual editors. Another study illustrated how the popularity of a film can be predicted before its release by analysing *Wikipedia* activity (Mestyán et al., 2013).

Several studies have exploited data generated by combined interactions with both *Google* and *Wikipedia*. Kristoufek (2013a) showed that fluctuations in BitCoin price are positively correlated with both BitCoin related search volume on *Google* and page views on *Wikipedia*, and demonstrated a strong bidirectional causal link between the BitCoin price dynamics and the change in search and page view frequencies. Another example of research utilising data from both *Google* and *Wikipedia* is Curme et al. (2014). In their paper, the authors analysed the *Google* search volume of a large set of keywords that were extracted by analysing the entire set of articles in the English version of *Wikipedia*. They showed that keywords related to politics or business are linked to movements in stock markets.

Online users do not only get information through explicit search attempts but also actively or passively via receiving news broadcasts. As a result of digitisation, most newspapers also publish online editions. These are updated regularly throughout the day presenting readers with news that is as up to date as possible. Several studies have been conducted by exploiting this new form of this traditional resource. Alanyali et al. (2013) sought to investigate whether there is a link between the interest of a company in finan-

cial news and that company's stock in the stock market. They focused on the companies that are listed in DJIA and for each company extracted the daily number of mentions of its name in the *Financial Times*. Figure 2.1 depicts the correlation between daily mentions and transaction volume of the company's stock. They showed the existence of a positive correlation between the number of mentions and the transaction volume. They unveiled a similar relationship between the number of mentions and the absolute return price of a company's stock, however they found no evidence of a relationship between the number of mentions and the return price once the direction of the change is taken into account.

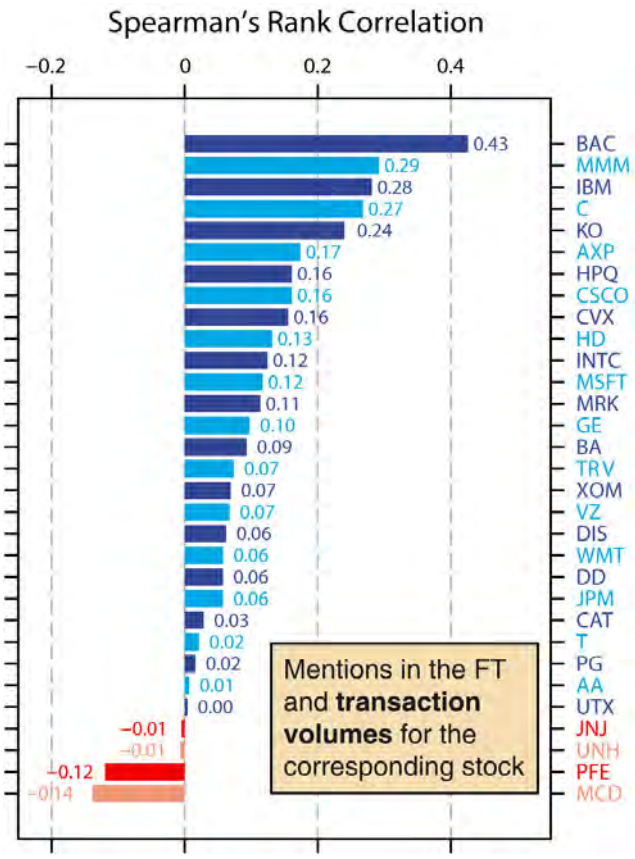

Figure 2.1: Correlation between the daily number of mentions of a company name and transaction volume of a company's stock.
Figure taken from Alanyali et al. (2013).

Data extracted from online news articles are also widely used as a proxy in the absence of reliable ground truth data. Braha (2012) analysed civil unrest across 170 countries over a 90 year period. As the main source of unrest data, they used articles from the *New York Times*. They showed that the distribution of unrest events that happened over any given year can be modelled using a dynamical model highlighting the similarity between the spread of the unrest events and dynamics of the spread of the other events

such as natural disasters and epidemics.

2.1.2 Internet as a communication channel

Apart from how we gather information, yet another facet of our daily lives that is experiencing a major makeover due to the changes introduced by technological devices is the way we interact with one another. Social interactions either professional or personal which used to be in person have been shifting online. With the emergence of social media channels, forms and tools of communication have been changing remarkably. Nowadays, it is common to hold international meetings online without the need to travel between countries; connect with geographically-distant friends for face-to-face conversations; or share daily snapshots of our lives with family, friends or even people that we haven't met and may never meet.

As a result of widespread usage of social media, the large amounts of data generated are offering insights into what people are thinking along with snapshots of what is happening around the world at a national or even global scale. Hence, an increasing number of studies have been conducted to understand the dynamics of these social media platforms (Liu et al., 2018; Traud et al., 2012; Vázquez et al., 2002). Furthermore, in the hope of measuring collective consciousness, social media has also become a main focus for researchers, as well as being a top agenda item for many businesses.

Soon after its creation in 2004 as a tool for an online multiplayer game, *Flickr* has been turned into an image and video hosting platform, qne is now one of the most popular image sharing platforms on the Web. *Flickr* is providing a platform where people can upload and manage their pictures. It is a social network where users can follow other users, create groups, and like and comment on each others photos and videos. When uploading visual media content, *Flickr* allows users to include a user-defined textual tag, title and description as well as further information about the media such as the date and location where it was taken, which are typically added automatically if the media is being uploaded from a mobile device.

In order to provide easy access to the large amounts of pictures shared online *Flickr* offers an open API that enables non-commercial users and developers to exploit the database of public *Flickr* data. Not surprisingly, the existence of this API has placed *Flickr* in the focus of many scientific studies.

Preis et al. (2013a) used data from *Flickr* to investigate the relationship between online activity and natural disasters, namely the Hurricane Sandy disaster in 2012. They extracted pictures uploaded with hurricane-related text attached to them, such as "hurricane", "sandy" and "hurricane sandy" and then normalised these occurrences with the total daily number of pictures uploaded to *Flickr*. Once compared to the atmospheric pressure data from New Jersey, US, they unveiled a striking relationship between hurricane-related pictures uploaded to *Flickr* and atmospheric pressure. Visual inspection also echoes the existence of a significant relationship between the *Flickr* activity and atmospheric pressure

(Figure 2.2).

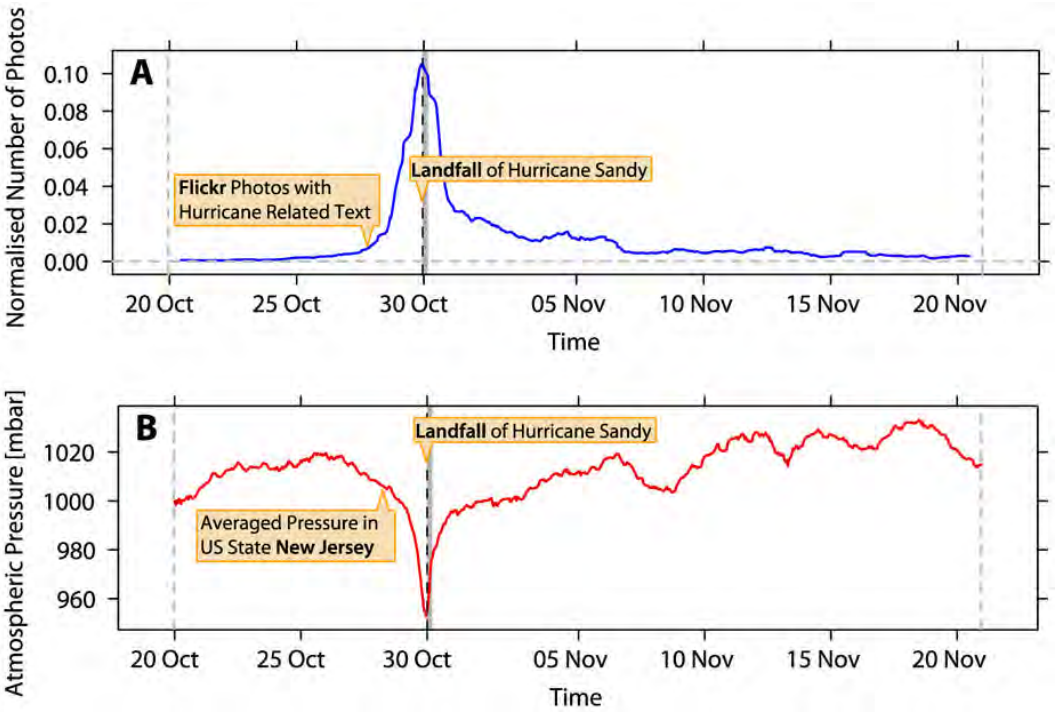

Figure 2.2: Comparison between the proportion of Hurricane Sandy related *Flickr* pictures and the atmospheric pressure.
(A) The number of Hurricane Sandy related pictures normalised with the daily number of *Flickr* pictures at an hourly granularity. (B) The change in the atmospheric pressure in New Jersey, US at hourly granularity. Figure taken from Preis et al. (2013a).

Geotagged data can also be extremely useful to identify characteristics of different areas across cities. In Aiello et al. (2016), the authors analysed tags of 17 million *Flickr* pictures taken between 2005 and 2015 in order to create maps of sounds around the streets of London and Barcelona.

Exploiting the broad spatial and temporal coverage offered by the pictures uploaded to *Flickr*, Barchiesi et al. (2015a) focused on quantifying international travel flows to the UK using *Flickr* pictures. The authors extracted user profiles with pictures taken in the UK between 2008 and 2013. For each user, by analysing the location of the other pictures they uploaded to *Flickr*, the authors identified their potential country of origin. Using this data, they generated estimates of traffic flows to the UK, and compared them with the data collected via International Passenger Survey to provide evidence of a significant link between the estimated numbers and the official statistics. Figure 2.3 depicts the relationship between the official statistics and the *Flickr* based estimates of UK visitors from 28 countries which are represented by the flag of the corresponding country. A different study on modelling mobility also exploited data from *Flickr* users to show that human mobility patterns obey Levy flights (Barchiesi et al., 2015b). Similarly, Wood et al. (2013) estimated

where the visitors to 836 recreational sites across 31 countries come from by analysing the profiles of the users who shared a picture from these recreational sites.

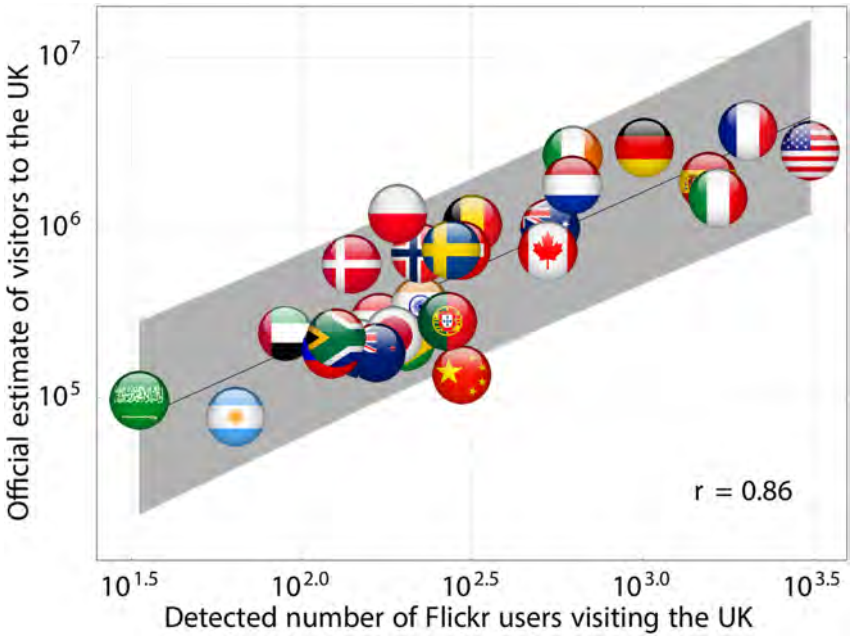

Figure 2.3: Link between the official statistics and *Flickr* based estimates of the number of UK visitors.
The authors demonstrated a significant correlation between the number of visitors from each one the 28 countries represented here by their flags and the number of users extracted from *Flickr* (r = 0.86, *N* = 28, $p < 0.001$, Pearsons correlation test). Figure taken from Barchiesi et al. (2015a).

Other studies focusing on *Flickr* data have investigated whether tags and visual characteristics of the images can help identify spatial and temporal patterns such as famous landmarks and significant events (Kennedy et al., 2007), analysed general patterns of tag usage focusing on pictures from the university groups (Angus et al., 2008), explored the relationship between house prices and art (Seresinhe et al., 2016), and used data on *Flickr* photographs to inform estimates of the beauty of the environment (Seresinhe et al., 2018).

Shortly after the launch of the multimedia sharing platform *Flickr*, in 2006 a new form of microblogging website was introduced into the social media space. *Twitter* started mainly as text-based social media platform originally restricting its users to 140 character posts called “tweets”. Now, as wel as enjoying a higher character limit, users can share text and multimedia, follow or send direct messages to other users, and like and respond to others’ tweets. Just like *Flickr*, *Twitter* also provides an API that makes a selection of public tweets available to download.

Fuelled by the vast amount of data provided by the *Twitter* API, a large number of studies have been conducted that aim to gain insights into collective consciousness and

human behaviour. Steinert-Threlkeld et al. (2015) used 14 million geolocalised tweets from 16 countries during the period of the Arab Spring in 2010-2011 to show a significant link between certain hashtags and the number of protests the following day. Exploiting data on *Twitter* activity recorded over two months in and around the Milan area, Botta et al. (2015) showed that it is possible to estimate the crowd size within a certain area. Alis et al. (2015) investigated whether *Twitter* data can provide quantitative evidence of regional differences in how talkative people are. Ciulla et al. (2012) used geoatagged *Twitter* data to map where fans of individual contestants of the TV show American Idol are populated both within US and abroad as well as predicting the outcome of the show. In Bollen et al. (2011a), the authors showed a significant correlation between socio-economic, cultural and political events and the public mood extracted from a corpus of nearly 10 million tweets. *Twitter* data was also used to gain insights into the movement of stock markets. Bollen et al. (2011b) provided evidence of a relationship between changes in public mood and fluctuations of Dow Jones Industrial Average (DJIA) values. They also showed that predictions of the DJIA are significantly more accurate once the public mood, which was extracted from a large set of tweets, was included.

In recent years, online political propaganda, as a supplement to standard offline campaigns has an extensive role in a leader's campaign programme. A selection of studies therefore have used *Twitter* to analyse public opinion and predict outcomes of political elections. Focusing on the German federal elections, Tumasjan et al. (2010) analysed a set of *Twitter* messages referring to a political party or a politician and extracted the sentiment of these messages. Their findings highlight the widespread usage of *Twitter* for political discussions as well as showing a relationship between sentiment and political programmes. In Conover et al. (2011), the authors analysed the underlying tweet-retweet network from the two opposing sides of a political discussion of the US Congressional Elections. They showed that users tend to retweet posts of other users that share similar political views while there is a very limited connection between left-leaning and right-leaning sympathisers. On the other hand, several studies have criticised applications that overemphasise predictions of electoral outcomes using *Twitter* data. For instance, Gayo Avello et al. (2011) performed analysis which have proven to perform well in predicting election results using *Twitter* data but failed to find a correlation between their findings and the actual electoral outcome. Hence they concluded that findings of analyses based on *Twitter* data should be interpreted with caution due to the potential bias introduced by the fact that the *Twitter* user base is not an exact representation of the entire population.

Another popular social media platform is *Facebook*, which was originally released around the same time as *Flickr*, although it wasn't until 2006 that it was opened to the general public. Nowadays, 68% of Americans use *Facebook* and 75% of them report that they use the platform on a daily basis Smith and Monica (2018). Although it has been widely used since its public launch, not many studies have been conducted using *Facebook* data, as unlike the previous platforms discussed above, there is no public API allowing data access. The only way to extract data is via applications, mainly the Graph Explorer API

provided by *Facebook*, however, it only lets users download the data that they have access to; their own profile, data from their friend network and public profiles and pages. Due to the restricted access to this gigantic *Facebook* dataset, among the studies exploiting data from social media, *Facebook* does not have the lion's share.

In order to analyse the influence and spread of information on social networks, Aral and Walker (2012) exploited *Facebook* data from 1.3 million users. They revealed striking findings suggesting that younger users are more susceptible to influence than older users and married users are the least susceptible group in using the product that was offered. In addition to the traditional interview-based methods, Boichak (2017) used *Facebook* to analyse civilian resistance networks during the Ukraine conflicts.

Facebook have themselves attracted negative media attention with the controversial study they conducted on the user profiles without the user's consent (Kramer et al., 2014) as well as the security breach regarding to Cambridge Analytica case.

Starting as an application only for iOS devices in 2010 and bought by *Facebook* in 2012, *Instagram* is a photo sharing platform with more than 800 million active monthly users and 40 billion total pictures (Instagram, 2017). It allows users to share visual content such as pictures and short videos either publicly or to a user-defined audience. The pictures can be uploaded with extra information embedded such as a timestamp indicating when the picture was taken or a georeference showing where the picture was taken. In addition, users can indicate the name of the place where the picture was taken by choosing from a list of location names or by creating their own location name.

Instagram provides an API that allows access to data from public profiles however, there has been a major change in the Terms of Use, effective since June 1, 2016, making data access far more restricted. Especially before the change took place, a number of studies were published using *Instagram* data to investigate human behaviour.

Hochman and Manovich (2013) compared visual aspects of *Instagram* pictures from 13 different cities. By analysing pictures from Tel Aviv taken over a three month period as a case study, they showed how online photographs can offer social, political and cultural insights. Similar studies have also been conducted using visual elements of *Instagram* data to trace visual cultural rhythms in New York City and Tokyo (Hochman and Schwartz, 2012) and to identify elements that differentiate one city from another by using Paris as an example (Doersch et al., 2012).

Motivated by the fact that human faces play a crucial role in communication, Bakhshi et al. (2014) provided evidence that *Instagram* pictures with faces receive more engagement regardless of the subject's age and gender. According to their study where they analysed one million pictures shared on *Instagram*, they showed that pictures with faces are 38% more likely to get a like and 32% more likely to receive comments compared to pictures that do not contain a face. Weilenmann et al. (2013) analysed how *Instagram* can be used to measure visitor experience at Gothenburg Natural History Museum. They showed how visitors use *Instagram* to create their own exhibits by regrouping and readjusting the museum environment.

A considerable number of studies combined data extracted from different social media platforms in order to investigate whether diversifying the data would bring additional input as well as to compare the dynamics and power of distinct platforms in answering their research problems. For instance, In Zhang et al. (2017), the authors used data extracted from *Twitter* and *Facebook* to analyse and compare how politicians use different social media channels to communicate their party programme during the 2016 US presidential campaign. Another example is by Quercia et al. (2015), a similar study to Aiello et al. (2016), where the authors extracted geotagged pictures from *Flickr* and *Instagram* together with georeferenced tweets from *Twitter* to generate a map of smells in London and Barcelona.

2.1.3 Internet as a crowdsourcing platform

In addition to information gathering and communication, another innovation the Internet has introduced is crowdsourcing platforms. Traditional techniques in analysing human behaviour involve rigorous surveys, interviews and laboratory experiments with controlled conditions. Although they provide rich and valuable information, conducting studies at scale can be very costly as well as having inherent difficulties such as finding participants. Crowdsourcing platforms offer scientists working with the most complex system – humans – an alternative medium to orchestrate surveys at large scale at ease.

In order to analyse the relationship between wellbeing and environmental factors, MacKerron and Mourato (2013) created a smartphone application that asks its users to report on their mood several times a day, and records their answers along with their location. They found that users feel happier when they are around natural environments, while also reporting limitations on drawing conclusions on causal relationships. In a different study, Seresinhe et al. (2015) utilised data from a crowdsourcing platform that gathers ratings of "scenicness" for photographs of areas across Great Britain. Their results showed that perceived beauty of an environment may have an effect on our wellbeing.

Amazon's *Mechanical Turk* is another famous example of a crowdsourcing platforms. Unlike the previous examples, Mechanical Turk is a micro-task website where stakeholders can upload tasks with a specific price for the human users to complete. The tasks can vary with the most common cases being creating annotated datasets for computer vision or natural language processing tasks. Kittur et al. (2008) discussed the usage of *Mechanical Turk* in user studies highlighting the importance of formulating the tasks. Widely used image benchmark datasets ImageNet, Places Database and SUN Database, which will be discussed in the following parts of this thesis also benefited from the large user base of the *Mechanical Turk* platform.

Another sector that benefits immensely from crowdsourcing platforms is the commercial sector. It is crucial for a business to be part of a crowdsourced review platform due to numerous reasons such as increasing their online presence, measuring customer satisfaction and interacting with their customers. This two way interaction is creating an

enormous amount of data. As in the examples discussed in the previous sections, some of these platforms make their data publicly available via APIs. One example of such a platform is *Yelp*.

Yelp is a crowdsourcing platform designed to collect users' reviews of businesses that include, but are not limited to, rating businesses by giving them a score out of 5, writing free-text reviews and uploading pictures of a given business. In addition to crowd sourcing reviews for businesses, Yelp provides its users a range of extra options such as posting on forum pages, finding nearby events or as on social media platforms, connecting with friends. *Yelp* published an extensive dataset on the data science challenge platform Kaggle, comprising more than 5 million reviews from over 150 000 businesses spanning 11 major cities as well as releasing a restaurant photo classification challenge on the same platform in 2016. Apart from the data made available on Kaggle, through the free API *Yelp Fusion*, *Yelp* provides access to its data on more than 50 million businesses in 32 countries Yelp (2018) which makes it an attractive resource for researchers.

Luca (2016) analysed a set of restaurant reviews on *Yelp* together with restaurant data from the Washington State Department of Revenue in order to investigate whether online reviews have an effect on restaurant demand. Another study focused on review fraud (Luca and Zervas, 2016). The authors analysed reviews that are identified as fake by *Yelp*'s filtering algorithm. They present four main findings including that a it is more likely for a restaurant to commit review fraud given that it does not have many reviews and when it has recently received a bad review. McAuley and Leskovec (2013) utilised reviews posted on *Yelp* to suggest including review text together with the ratings to improve the performance of a recommender system.

2.1.4 Issues with online data and privacy

Studies presented in the previous sections set an example of how online data can be used to shed light on various questions arising in social sciences, however, the limitations of these gigantic datasets should be carefully considered and reported. One important limitation is the bias in usage. Online user groups are not an exact representation of the offline population groups. The lack of demographic information therefore forms a critical impediment in using online data to analyse human behaviour and collective consciousness. For instance, Internet usage varies profoundly between different age groups and countries yielding an uneven representation of different groups online. Hence, several studies have been conducted to infer user demographics on social media through analysing users' posts. In Sloan et al. (2015), the authors compared the age, occupation and social class of the *Twitter* users in the UK, deduced from their tweets and profiles, with the official figures on population demographics extracted from the 2011 census data. Edwards et al. (2013) provided a detailed discussion on using social media data for social research by arguing its limitations. They highlight the potential of this new source of data together with the other online data sources and promote their usage as a supplement to traditional social research

methods.

Training data is crucial when building supervised machine learning systems. Even though the designer of the system knows the details of the underlying algorithm, training data can make the end product biased. For instance, Microsoft launched a chat bot called Tay on *Twitter* with the training set being tweets of the other users. It was shut down after only 16 hours after posting a number of offensive tweets. On a blog post shared on Microsoft's official website, Peter Lee, Corporate Vice President of Microsoft AI and Research apologised for the incident and noted "AI systems feed off of both positive and negative interactions with people" highlighting the importance of the underlying training data in creating a fair machine learning system (Lee, 2016).

Although providing rich information about human behaviour and collective consciousness, high granularity data retains several dangers. In particular, the metadata attached to these large datasets documenting human behaviour may contain sensitive information that raises concerns about privacy. Previous studies provided evidence that it is possible to uniquely identify the majority of the individuals in an anonymised dataset from mobility patterns (De Montjoye et al., 2013) and credit card metadata (De Montjoye et al., 2015). A whole body of research therefore has been formed to effectively anonymise data in order to protect the privacy of the data subjects (Cormode, 2011; Zhu et al., 2010) using various techniques such as differential privacy (Chen et al., 2011).

One final issue that needs attention is the data ownership. The majority of the digital traces we leave behind are being recorded by online platforms and technological devices we interact with. This raises a big ethical question about who the actual owner of the data is: the person who generates the data or the platform on which the data has been generated. Another issue that needs to be addressed carefully is automated systems. For instance, who is liable when an autonomous vehicle has an accident? In order to address these issues and more as well as to provide guidelines for the data controllers, General Data Protection Regulation (GDPR) in EU law has been introduced which also brought discussions. Numerous studies in the field of data ethics are therefore being conducted to discuss the advantages and limitations of the GDPR.

Wachter (2018) presented guidelines to protect data subjects' identities and privacy by providing two example cases on how the regulations can be applied. Much discussion has been shaped around the GDPR's focus on "right to explanation", which involves data-driven automated decision making process being explained to the individuals. In their paper, Wachter et al. (2017) proposed three goals for automated decision making including the "right to explain" and discuss to which extent they are supported by the GDPR.

2.2 Image analysis and deep learning

As discussed in the previous section, large quantities of online data are being generated through daily interaction with everyday technological devices. In order to gain insights into everyday human behaviour, we need an automatic way to analyse these gigantic datasets

to extract meaningful information.

Creating intelligent machines has been a dream for hundreds of years. A passion to build machines that can help with laborious work, understand commands or images, and assist scientists with their research has led to the creation of Artificial Intelligence (AI). AI, which is formed of a wide range of topics with rapidly increasing applications, is the general name for an intelligent software. Figure 2.4 depicts the relationship between different subfields of AI.

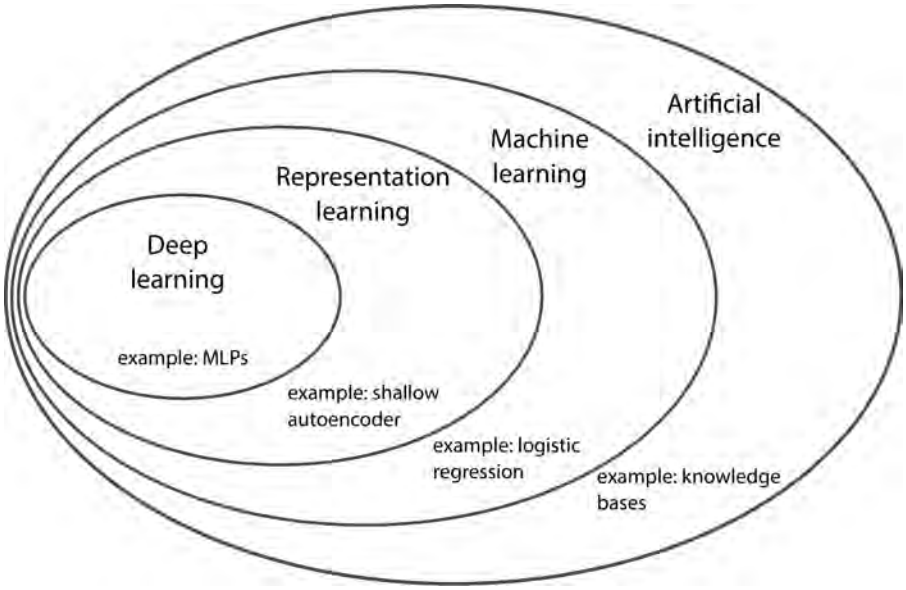

Figure 2.4: A Venn diagram illustrating the relationship between different subfields of AI. Each subfield contains an example application. The sketch is adapted from Goodfellow et al. (2016).

The early applications of AI involved hard-coding knowledge to the computers, which is not feasible for most cases. Researchers therefore came up with a set of methods that enable computers to extract their own knowledge from a given dataset. These group of methods are known as “machine learning algorithms”. Throughout this thesis, we will exploit various machine learning algorithms such as logistic regression and elastic net.

In machine learning algorithms, creating representations of the data is the key. These data representations, which are called features, enable the machine learning algorithm to identify similarities and differences across different categories. Traditionally, feature extraction involves careful hand-engineering of the features where domain expertise is a necessity. It is fair to say that the introduction of representation learning methods was a major breakthrough for the advancement of machine learning algorithms.

Representation learning approaches come with a promise to reduce manual intervention in feature extraction to a minimum level by enabling machines to learn the suitable representation of an underlying dataset. Their power comes from their adaptable

nature to new tasks with minimal human intervention. A number of representation learning approaches have been developed over the years including shallow autoencoders and restricted Boltzmann machines. Although these methods have proven to be useful in many cases, a main problem still remains: they cannot capture the unseen factors behind the features. In their book, Goodfellow et al. (2016) defined these factors as the "concepts or abstractions that help us make rich sense of the rich variability in the data". There is however, a special set of representation learning algorithms that addresses this problem.

Deep learning is a type of representation learning used in both supervised and unsupervised learning problems. Deep learning architectures are composed of several layers that are trained to extract features using the output from the previous layer. Each layer is composed of simple modules called neurons that create multiple levels of representations and abstractions of the input data. These modules increase selectivity and invariance of representation aiming to capture the underlying complex patterns ingrained in the input data.

Like the other representation learning approaches, deep learning architectures take raw data as an input. The main difference however comes from how they identify features layer by layer in a hierarchical manner. As we move higher up in the deep architecture, each feature is defined through its relation to simpler features, identified in the preceding layers. For instance, let us assume we have a deep network trained to detect a particular object. Initial layers would identify simpler and more generic features such as edges followed by motifs whereas the latter layers will detect parts of the object and finally detect the main object.

There are various types of deep architectures exploiting different characteristics of the input data of different forms. A typical and probably the simplest example of such architectures is the feedforward neural network or multilayer perceptron (MLP). Due to their simple yet efficient way of learning complex non-linear mappings from the raw input data, previous studies have exploited MLPs to address a number of problems including speech recognition (Bourlard and Wellekens, 1989; Waibel et al., 1990) and face detection (Sung and Poggio, 1998), facial expression recognition (Zhang et al., 1998) as well as creating recommender systems (Alashkar et al., 2017; Huang et al., 2015).

MLPs also play a very crucial role as they form the basis of the most widely used deep learning applications. One example of these applications is natural language processing. In the last decade, many natural language processing applications exploit a special type of a deep architecture that uses MLPs as a "conceptual stepping stone" (Goodfellow et al., 2016): recurrent neural networks (RNNs) (Rumelhart et al., 1986). RNNs prove their success with a diverse range of problems such as speech recognition (Graves et al., 2013; Mikolov et al., 2010), time series prediction (Connor et al., 1994) and object and gesture tracking in videos (Ng et al., 2015) where the underlying dataset is of a sequential form.

Another field that immensely benefits from the perks of MLP is computer vision. Ballard and Brown (1982) defined computer vision as "the enterprise of automating and integrating a wide range of processes and representations used for vision perception",

which is formed of a wide range of sub-fields such as video tracking, pose estimation, image processing and most importantly for our studies; image analysis.

Image analysis is the general name for a collection of methods aiming to extract meaningful information from 2D images generally by using machine learning and image processing techniques. Although comprehending photographic information might seem to be a trivial task for us humans, it has proved to be a challenging problem for computers. Benefiting from the hierarchical nature of the algorithms, deep learning methods achieve ground breaking performance in renowned image classification and object detection challenges.

In addition to the new state-of-the-art algorithms, deep learning methods proposed decades ago such as convolutional neural networks are experiencing a resurgence of interest owing to improvements in the processing power of computers as well as the availability of extensive datasets for training, these deep architectures are outperforming image analysis methods using traditional features such as Fisher Vectors (Chatfield et al., 2014) and SIFT features (Krizhevsky et al., 2012).

In the next section, we will have a closer look at the convolutional neural networks that we will be widely using throughout this thesis.

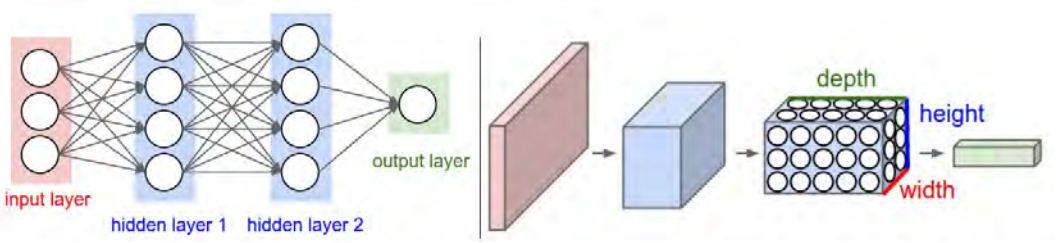

Figure 2.5: Structure of a regular neural network (left) versus convolutional neural network (right).
CNNs organise their neurons in a 3D structure where every layer transforms a 3D input to a 3D output. Here, on the right sketch, the red block represents the input image where width and heigh represents the size of the input image and depth is three due to three colour channels; red, green and blue. Figure taken from `https://cs231n.github.io/convolutional-networks/`.

2.2.1 Convolutional neural networks: an overview

CNNs are designed for input data that has a lattice-like structure. This means that CNNs are specialised for processing data that comes in the form of multiple arrays, such as colour images which have three channels, red, green and blue, with 2-D grid of pixel values. Figure 2.5 provides a toy example of a CNN architecture in comparison with a regular neural network architecture, such as an MLP. By arranging their neurons in the form of the 3D input images, at each layer CNNs transform a 3D input to a 3D output with the third dimension being the colour channel.

In order to better illustrate the advantages of using CNNs, we first need to introduce their building blocks.

2.2.2 Building blocks of a CNN

2.2.2.1 Convolutional layer

The very first layer and one of the main building blocks of a CNN is the convolutional layer. Before explaining the details of this layer, we first need to describe what convolution is and in order to do so, we will use the example from Goodfellow et al. (2016).

Convolution is simply a mathematical operation on two functions. For example, imagine a noisy laser sensor providing the position of a vehicle, $x(t)$, at time t. In order to reduce the effect of noise and get a better estimate of the current position of the vehicle, we can take the average of several measurements over different time t. However, we also need to give more emphasis on the recent measurements, which can be done by introducing a weight function $w(a)$ where a denotes the recency of the measurement. We then calculate the weighted average of measurements at every moment which will give us a smooth estimate of the vehicle's current position, $s(t)$:

$$s(t) = \int_{-\infty}^{\infty} x(a)w(t-a)da. \tag{2.1}$$

This operation is called convolution and is typically shown with an asterisk

$$(x * w)(t). \tag{2.2}$$

In CNN terminology, *x* is the input, *w* is referred to as the kernel or convolutional filter in image processing and the output is called the feature map.

Unlike the laser example, data on a computer is discrete. We therefore change the integral to summation, and the convolution operation becomes:

$$s(t) = (x * w)(t) = \sum_{a=-\infty}^{\infty} x(a)w(t-a). \tag{2.3}$$

Here, the input is a multidimensional array, for instance an image, and the kernel is a multidimensional array of parameters. In Figure 2.6, we depict a toy example of a 2D input array convolved with a filter, i.e. the filter is sliding over the input array to compute dot products. The size of the output is then determined by the size of the input, filter and stride with which we move the filter. For instance, considering this example, an input array of size 4×4 convolved with a filter of size 3×3 with stride 1, will produce a 2×2 output. We can generalise this relationship as follows:

$$Size_{output} = \frac{(Size_{input} - Size_{filter})}{Size_{stride}} + 1. \tag{2.4}$$

Alternatively, in order to have a better control of the size of the output, we can

introduce zero-padding, which is an approach to add zeros around the borders of a matrix, to increase the size. By incorporating the size of padding, the general formula of the output size then changes to:

$$Size_{output} = \frac{(Size_{input} - Size_{filter} + 2Size_{padding})}{Size_{stride}} + 1 \tag{2.5}$$

In cases where we want to preserve the input size in the output, it is a common approach to have stride 1 with zero padding of $(Size_{filter} - 1)/2$.

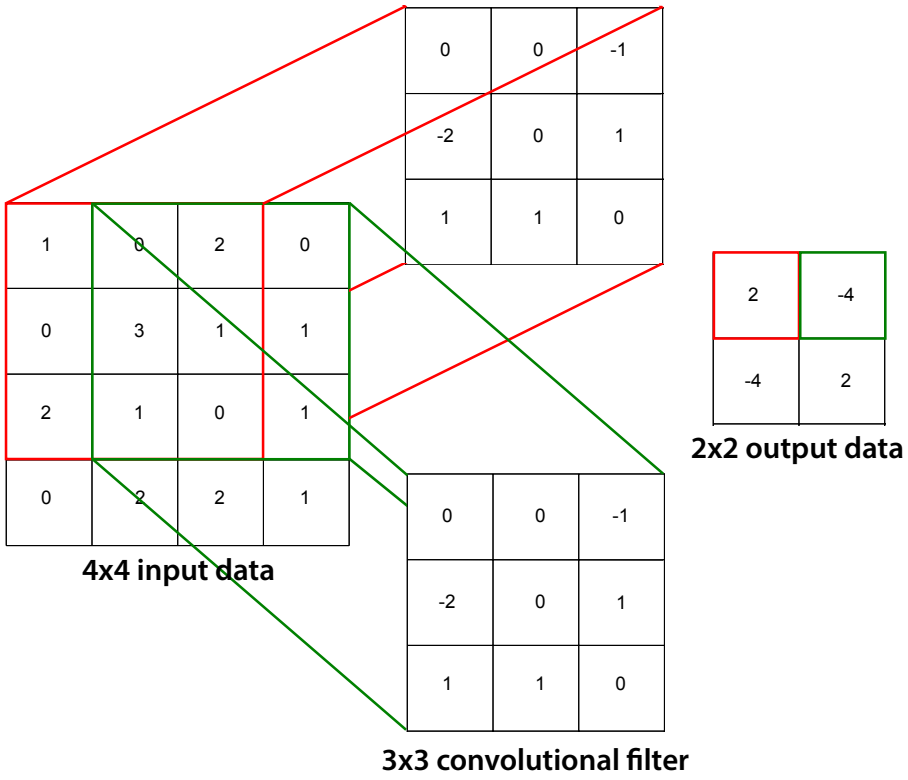

Figure 2.6: Toy example of a convolution operation.
A 4×4 input is convolved with a 3×3 filter where we compute the sum of dot products at each cell. We repeat the same operation at each position by moving the filter by one cell at a time (stride 1).

The convolutional layer as the name suggests consists of a set of three dimensional "convolution" filters that are learned from the input data. The forward-pass of this layer involves each filter to convolve across the input image. Each of these filters will produce a separate activation map, also called a feature map.

The output of the layer is constructed by stacking these maps which are as many the number of filters. Figure 2.7 shows an example of creating one activation map from an input of size $4 \times 4 \times 3$ using a $3 \times 3 \times 3$ filter. If we have *k* number of filters, then the output would be of size $2 \times 2 \times 2$. In real applications, filter size is usually limited to be orders

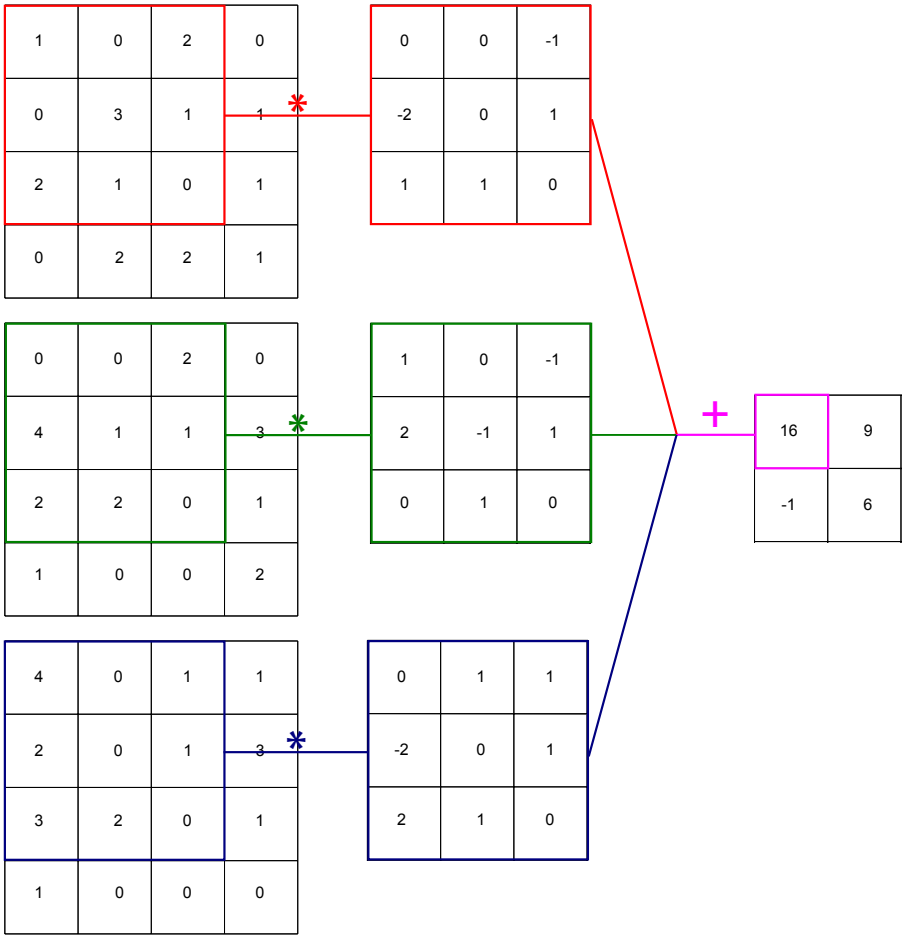

Figure 2.7: Creating the activation map of a 3D input.

We convolve three dimensional input data, which is also called an input volume with a filter of size $3 \times 3 \times 3$ to generate an output volume of size 2×2. The depth of the output volume is the same as the number of filters, for instance if we have *k* filters of size $3 \times 3 \times 3$, then the output volume would be of size $2 \times 2 \times k$.

of magnitude smaller than the input size to reduce the number of parameters required in the architecture. For instance, standard neural nets such as MLPs, are fully connected meaning that in each layer, every input unit is directly connected to every output unit. This type of architecture does not scale well with large images causing long running times as well as increased storage needs to store the expanding number of parameters. However, by utilising local connectivity with filters of a size much smaller than the input, CNNs reduce the time and space required to train the network without a critical performance loss.

The other critical feature of the convolutional layer is parameter sharing. In contrast to fully connected architectures, which require learning every parameter shared between an input unit, such as a pixel in an image, and an output unit, CNNs enable sharing of the weights. Instead of learning weights of each connection at each location, they learn one set that will sweep over the entire input matrix. Considering the previous example, for a framework with *k* number of filters, for each unit in the convolutional layer output, the CNN will learn one set of weights. Hence, with a filter size of $3 \times 3 \times 3$, the network will need to learn $3 \times 3 \times 3 \times k$ weights in total for the specified convolutional layer. This has no effect on the run time however it reduces the number of parameters greatly.

The final important advantage of incorporating convolution in a neural network structure is the equivariance in representations also referred to as translation invariance. If we move some pixels in an image to a different location, the convolution will still produce the same output but in a different location. This makes CNNs location invariant. However, this can't be generalised to invariance in scale or rotation. This feature is useful especially when we want to identify similar patterns repeated across the image such as edges or certain objects.

These are the three core features that the convolution operation brings to neural networks. However, there are certain scenarios where special adjustments need to be made, such as images centred to human faces. In such cases, we need to learn different features in different areas of the face like eyes or mouth. Hence, it is useful to implement a slightly different version of a convolutional layer for instance by relaxing the shared weights restriction. Nevertheless, in general, by exploiting the advantage of the sparse connections via local connectivity, shared weights and equivariant representations, convolutional layers help to identify intricate relationships embedded in the underlying data, and are one of the core building blocks of a CNN architecture.

2.2.2.2 Non-linearity

A convolutional layer is a layer where linear operations take place. In CNNs, convolutional layer is always followed by a non-linear layer, which sometimes is referred to as the detector stage. The non-linearity is introduced by a so-called activation function which is applied to each component within a feature map. Although initial papers included these non-linearities with a hyperbolic tangent function (LeCun et al., 1989), in more recent models, this has been replaced by a rectified linear unit (ReLU). ReLU is a simple non-linear function that

simply picks the maximum between zero and a given value.

Previous studies suggest that compared to other non-linear functions used in a CNN setting, ReLU is better at finding the minima during training and is also better at bringing networks without unsupervised pretraining to a similar level to networks with pretraining (Glorot et al., 2011). Models with ReLU are also shown to be faster to train especially with a very large training set (Krizhevsky et al., 2012). Using ReLU also simplifies backpropagation as well as avoiding saturation issues.

2.2.2.3 Pooling

Another important element that forms the backbone of a CNN is the pooling layer. Running on an individual feature map, pooling combines nearby features into one with a chosen operator such as max-pooling, i.e. picking the maximum feature value given a set of features, or sum-pooling which involves summing all feature values that fall within the pooling frame.

By creating a summary of nearby features, a pooling layer helps to create features that are robust to small changes in the input. A pooling layer also helps to avoid overfitting via gradually reducing the number of parameters as well as the computational time. It is also very useful for tackling input images of varying sizes. A pooling layer can create subsets of the same size that can then be transformed to classification layer; for instance regardless of the original dimensions, the pooling layer can create four sets of features by focusing on the quadrants of each image.

However, pooling layers are not universally popular. In Springenberg et al. (2014), the authors dropped their pooling layer and instead utilised larger strides in the convolutional layer in order to further reduce the input dimensions.

2.2.2.4 Normalisation

Normalisation is another building block of a CNN. Various types of normalisation over different levels including feature map level or image level have been proposed however, to the minimal improvement the layer brings to the CNN's performance, normalisation has mostly been abandoned.

2.2.2.5 Combining layers

Convolutional, non-linear and pooling layers, in some cases with normalisation layer are the main building blocks of a CNN architecture. In practice, different architectures are created by stacking several layers composed of a combination of these building blocks before adding the final fully-connected and classification layers.

2.2.3 Training and knowledge transfer

Once we create the network architecture, the next step is to train the network, i.e. to learn the parameters. Like most of the other deep learning architectures, CNNs are also trained

using supervised learning. Let us take a classification problem as an example. The training process includes passing an image to a network as the input, for which the architecture will then generate a set of scores representing the likelihood that the image belongs to each of a set of categories. The image will then be grouped under the category with the highest score. In order to determine whether the image is classified into the right category, we compute an objective function, also referred to as the loss function, which is the distance between the actual set of scores and scores calculated by the network. Taking the objective function into account, the network will then make necessary adjustments to its parameters aiming to minimise the error between the actual and generated output scores.

To correctly adjust the parameters, for each weight the learning framework will create a gradient vector that measures the amount of change in the error when there is a slight change in the weight. The weight vector will then be updated in the opposite direction of this gradient vector. In LeCun et al. (2015), the authors portrayed the objective function generated by the mean objective values calculated for each training instance as a "hilly landscape in the high-dimensional space of weight values" where the negative gradient vector will take the objective value close to minimum.

One of the most common methods utilised in calculating gradient vectors is Stochastic Gradient Descent (SGD). Using a small subset of the training data, the SGD algorithm computes the output and errors that will later be used to update the weights of the network. This process will then be repeated for a number of small sets created from the initial training set until the algorithm converges, i.e. the error becomes very small.

Despite its simplicity, the performance of the SGD is competitive with more intricate optimisation algorithms (Bottou and Bousquet, 2008). SGD's capability of calculating a good set of weights considerably quickly has made it one of the most preferred approaches among practitioners.

If the neurons use fairly smooth activation functions, which have continuous derivatives, then the gradients of an objective function can be computed with a method called backpropagation. Drawing on the chain rule for derivatives, the main idea behind the backpropagation method is as follows: for a given neuron the derivative of an objective with respect to the input can be computed from the gradient of the objective with respect to the output of the neuron. It is calculated for every neuron within a layer and applied over and over again to pass the gradients through all layers, as the name suggests working its way backwards from output to the input layer. With a series of forward and backwards passes, the weights of CNN including filters can then be trained using a gradient-based method utilising backpropagation.

The backpropagation algorithm had been mostly abandoned by the machine learning computer vision communities due to a disbelief that these algorithms suffer from getting stuck at poor local minima. However, later studies have shown that in practice, large networks do not suffer from the local minima problem (Choromanska et al., 2015). This is one of multiple reasons why CNNs hadn't been adopted by the computer vision community till recently.

In addition to optimisation and backpropagation, another crucial part of the training process is to decide on the training set. The training set is a key part of the learning process as it is the ultimate source of knowledge made available to the network. The quality and scale of such sets therefore can have a huge impact on the performance of a CNN.

Today's digital era with the ever-increasing number of online images, combined with crowdsourcing platforms has helped to addressed the lack of good quality big image data for developing image based algorithms by giving rise to the creation of large image sets. ImageNet is an example of such open source image sets (Deng et al., 2009). By 2010, ImageNet reached 14 197 122 annotated images from 21 841 categories based on the WordNet hierarchy (Kilgarriff and Fellbaum, 2000). An initial set of images constructed via sending image search queries to multiple search engines using the words and phrases from the WordNet dataset. Then, the set has gone through a cleaning process on Amazon Mechanical Turk (AMT), which is a crowdsourcing platform with a paid user-base dedicated to solve various tasks such as labelling images uploaded by the paying users. Finally, in order to minimise human error, a final control system employing multiple users to label the same image has been introduced.

However, ImageNet is not the only large-scale image database available for practitioners. Different from ImageNet where objects are the main focus, in order to address the scene classification problem, Zhou et al. (2014) have created the Places database using the semantic categories from the Scene Understanding (SUN) dataset (Xiao et al., 2010). This database has been formed following a similar approach to ImageNet by constructing an initial set of images extracted via sending queries to search engines followed by manual elimination of noise with AMT followed by a final control case. The Places database consists of more than 10 million images from 434 scene categories covering about 98% of the different places that a person can come across in the world (Zhou et al., 2014).

Different from the category-based image sets such as the ImageNet and Places datasets, where each picture is labelled with a single category to create deeper scene understanding a new image set that enables multi attribute representations has been created: the SUN attribute dataset (Patterson et al., 2014). Complementing scene categories including the labels under the SUN database (Xiao et al., 2010), the SUN attribute dataset has been formed of 102 discriminative attributes identifying varied facets of a scene such as surface properties, lighting and spatial layout.

However, in real life finding clean, annotated training sets at scale is non-trivial. Although CNNs perform well on a number of computer vision problems once trained on large datasets, they generally overfit hence perform with poor generalisation when they are trained on a limited set of training data. In cases where there is a lack of training data, it is therefore not feasible to build and train a network from scratch.

Owing to the adaptable nature of CNNs, knowledge that has already been gained by training a CNN on millions of annotated training images can be transferred to new generic tasks. There are two different ways to repurpose a network that has already been trained on a large dataset: fine tune the network using a smaller training data or utilising

the output from one of the fully connected layers as features to train a classifier that can be as simple as a linear classifier. Previous studies have shown evidence that using CNNs as feature extractors or fine tuning a pretrained network creates powerful image descriptors hence performs better in a wide range of recognition tasks such as object detection and scene recognition even though the new problem is different from the original task (Chatfield et al., 2014; Donahue et al., 2014; Girshick et al., 2014; Sharif Razavian et al., 2014).

Exploiting a pretrained network also has a runtime advantage. Due to the need of computing and storing large numbers of parameters, training a CNN from scratch can be a computationally intensive task requiring days of GPU time and/or the use of multiple GPU cores. However transfer learning methods especially when CNNs are used as feature extractors are similar to training a classifier using hand crafted features yet they benefit from the powerful image representations learned from a large set of training images.

2.2.4 Applications

In recent years, CNNs have become one of the most popular deep learning approaches especially among the computer vision community. Although the proliferation of these architectures in object and speech recognition wasn't until recently, CNNs were first introduced nearly three decades ago (LeCun et al., 1989). Inspired by the biological structures from visual neuroscience (Cadieu et al., 2014; Hubel and Wiesel, 1962), LeCun et al. (1989) proposed a neural network architecture containing convolutional filters connected to local patches. Trained via backpropagation, this new architecture was able to successfully detect handwritten zip codes from a dataset provided by the U.S. Postal Service.

Since then CNNs have excelled in diverse applications from phoneme recognition (Waibel et al., 1990) to digit classification (LeCun et al., 1998) and localisation of faces in images (Vaillant et al., 1994). However, until recently, they weren't as popular among the computer vision and machine learning community. With increased computational power thanks to GPUs that are easier to programme, larger training data made available with the help of crowdsourcing platforms as well as distortion of the already available training data together with the new methods such as ReLU and dropout, CNNs have their breakthough by nearly halving the error rate compared to the nearest best performing model in ILSVRC 2012 (Krizhevsky et al., 2012).

Since then a growing number of studies especially on recognition and detection tasks have been conducted exploiting the adaptable and easy-to-train nature of the CNNs addressing a wide range of problems arising in computer vision. The majority of the recognition and detection tasks nowadays are utilising convolutional neural networks (Garcia and Delakis, 2004; Karpathy and Fei-Fei, 2015; Sermanet et al., 2013; Simonyan and Zisserman, 2015).

Crowley and Zisserman (2014) exploited a pretrained network as a feature extractor. By creating a small set of natural images extracted from image searches, they trained an object-category classifier to identify different objects in paintings. Seresinhe et al. (2017)

used transfer learning to adapt a CNN initially trained on Places database in order to compute a “scenicness score” for images. Gebru et al. (2017) estimated a demographic map of the US by automatically analysing *Google Street View* images via CNN based framework. They built and trained a CNN to detect the year, make and model of motor vehicles around different neighbourhoods in order to estimate socioeconomic characteristics including income, race and education as well as voting patterns in the presidential elections.

In addition to the applications in image analysis, there are a wide range of other applications exploiting CNN architectures including sentence classification (Kim, 2014), creating automatic subtitles from reading lips in videos (Chung and Zisserman, 2018), speech recognition (Sainath et al., 2013) and building recommender systems (Gong and Zhang, 2016; Kim et al., 2016). In the rest of this thesis, we will utilise different CNN architectures to automatically analyse online images to quantify human behaviour.

CHAPTER 3

Tracking Protests Using Geotagged Flickr Photographs

3.1 Introduction

In recent years, news reports have described a number of prominent outbursts of protests in countries around the world, in some cases leading to political change. During the time of protests, much media attention has been focused on the increasing usage of social media to coordinate and provide instantly available reports on these protests (Arthur, 2011; Branigan, 2009; Christie-Miller, 2014). Digital traces that are generated as a consequence of the online activity around protests therefore serve as a fruitful information source for scientists to shed light onto the dynamics of these collective movements via both qualitative and quantitative analyses.

In Boichak (2017), the authors conducted interviews with battlefront volunteer groups as well as performing a network analysis to highlight the usage of the social media platform *Facebook* by battlefront volunteers in helping to stop the spread of the military conflict in Ukraine. A similar study again focusing on the crisis in Ukraine exploited data from *VKontakte* (VK), which is a social networking platform widely used in Ukraine (Gruzd and Tsyganova, 2015). The study provided evidence towards a link between online and offline social networks formed during protests. Another research based on exploiting VK data concentrated on the "Fair Movement" protests in St. Petersburg, Russia (Koltsova and Selivanova, 2015). The authors analysed data from 12 000 online users and 200 offline participants to find a link between online activity and offline participation.

A large number of studies utilised *Twitter* in order to analyse the dynamics of the "Occupy Wall Street" movement (Conover et al., 2013), "Gezi" protests in Turkey (Budak and Watts, 2015), and 15-M movement in Spain (González-Bailón et al., 2011). In Steinert-Threlkeld et al. (2015), the authors exploited 14 million geotagged tweets to analyse Arab Spring protests across 16 countries to uncover a significant correlation between the social media activity and protest movements.

As a result of improved connectivity, posts to social media sites are steadily begin-

ning to shift from solely text based reports to sharing of visual media such as photographs and videos. Here, we explore whether the data created through such widespread usage of online services may offer a valuable new source of measurements of behaviour during protests. Specifically, we investigate whether data on photographs uploaded to the photo sharing website *Flickr* can be used to identify protest outbreaks around the world.

3.2 Data

3.2.1 *Flickr* data

We analyse a large corpus of metadata on the 24 944 764 geotagged photographs taken and uploaded to *Flickr* between 1st January 2013 and 31st December 2013. We retrieved data on image uploads to *Flickr* by accessing the *Flickr* API in January 2014, and downloading data in JSON format using R 3.0.1. The metadata we analyse comprise a wide range of information on where and when a photograph was taken, information about the photographer, as well as user chosen title, description and tags for each photograph, and the URL from which the photograph can be downloaded.

For each geotagged photograph, we retrieve data on both the time and the place at which the photograph was taken. For each week, for each of the 242 countries and regions listed in Table A1 in Appendix A, as well as the United Kingdom and the United States, we determine how many photographs were taken and uploaded with the word "protest" in English in either the title, photograph description or photograph tag. We also translate the word "protest" into 33 further languages, by accessing the "Protest" article on the English language *Wikipedia*, and using the title of all articles on versions of *Wikipedia* which are not in English, but which are linked as translations of the article. The complete list of translations is provided in Table A2. The counts of photographs taken and shared on *Flickr* throughout 2013 in each of the 244 countries and regions are listed in Table A3.

The overall number of photos taken and uploaded to *Flickr* in different countries and regions may differ. To account for this, we extract the total number of photos taken and uploaded during each week in 2013 for each of the 244 countries and regions analysed. We consider a week as starting on a Monday and ending on a Sunday. Using these counts, we normalise the weekly counts of photographs taken in each country and region in each week, by dividing the number of photographs labelled with a word signifying "protest" by the weekly count of all photos taken in the same country and region.

3.2.2 *The Guardian* data

To determine whether we can find any evidence that changes in the number of protest-tagged photographs taken and uploaded to *Flickr* correspond to changes in the number of protest outbreaks, we require data on when and where protests have occurred. Such ground truth data can be difficult to obtain. Most studies of civil unrest therefore rely on

data from newspaper reports as a proxy for ground truth (Braha, 2012; Compton et al., 2014; Steinert-Threlkeld et al., 2015). Following this approach, here we determine how many protest related articles for each of the 244 countries and regions were published in the online edition of *The Guardian* in each week in 2013.

We retrieved data on articles in the online edition of *The Guardian* via *The Guardian* Developer Toolbox in January 2016. We deem an article as protest related if it is tagged with the word "protest", and we deem an article as covering news related to one of the 244 countries and regions analysed if it is tagged with the country and region's name. To account for differences in coverage of news in different countries and regions by *The Guardian*, we also determine the total number of articles published in each week and tagged with each country and region's name. In total, we analyse data on 178 730 articles from *The Guardian*. The counts of *The Guardian* articles published in 2013 and tagged with each of the of the 244 country and region names are listed in Table A4. We note that, *The Guardian* uses a different tagging system for articles relating to the United Kingdom and the United States. For this reason, we determine the number of articles relating to the United Kingdom by counting articles tagged with the names "England", "Scotland", "Wales" and "Northern Ireland", and we determine the number of articles relating to the United States by counting articles listed under the section "us-news".

3.3 Analysis and results

In order to model the relationship between *Flickr* user activity and protest outbreaks, we build a logistic regression panel model with the outcome variable to be whether a *The Guardian* article is protest related or not. To control for underlying differences in the number of protests in a given country and week, we include country and week as fixed effects in our model. We note that in this analysis, we focus on the relationship between *Flickr* activity and protest reports within the same week. Future analyses may wish to investigate whether photographic data can be used to predict protest activity before it occurs.

We use data on reports of protests in the online edition of *The Guardian* as an approximation of the ground truth of when and where protest outbreaks occurred. For each of the 244 countries listed in Table A1, for each month in 2013, we calculate the number of *The Guardian* articles tagged with the country's name. In Figure 3.1, we depict the percentage of articles for each country and each month which were also tagged with the word "protest". Patterns which can be visually identified in the data reflect known major protest events in 2013: for example, protest outbreaks in both Brazil and Turkey can be observed in June 2013.

We examine to which extent data on the number of photographs tagged with the word "protest" and uploaded to *Flickr* reflect the ground truth data extracted from *The Guardian*. Again, for each of the 244 countries listed in Table A2, for each month in 2013, we calculate the total number of geotagged photographs taken and uploaded to *Flickr*. In Figure 3.2, we visualise the percentage of photographs for each country and each month

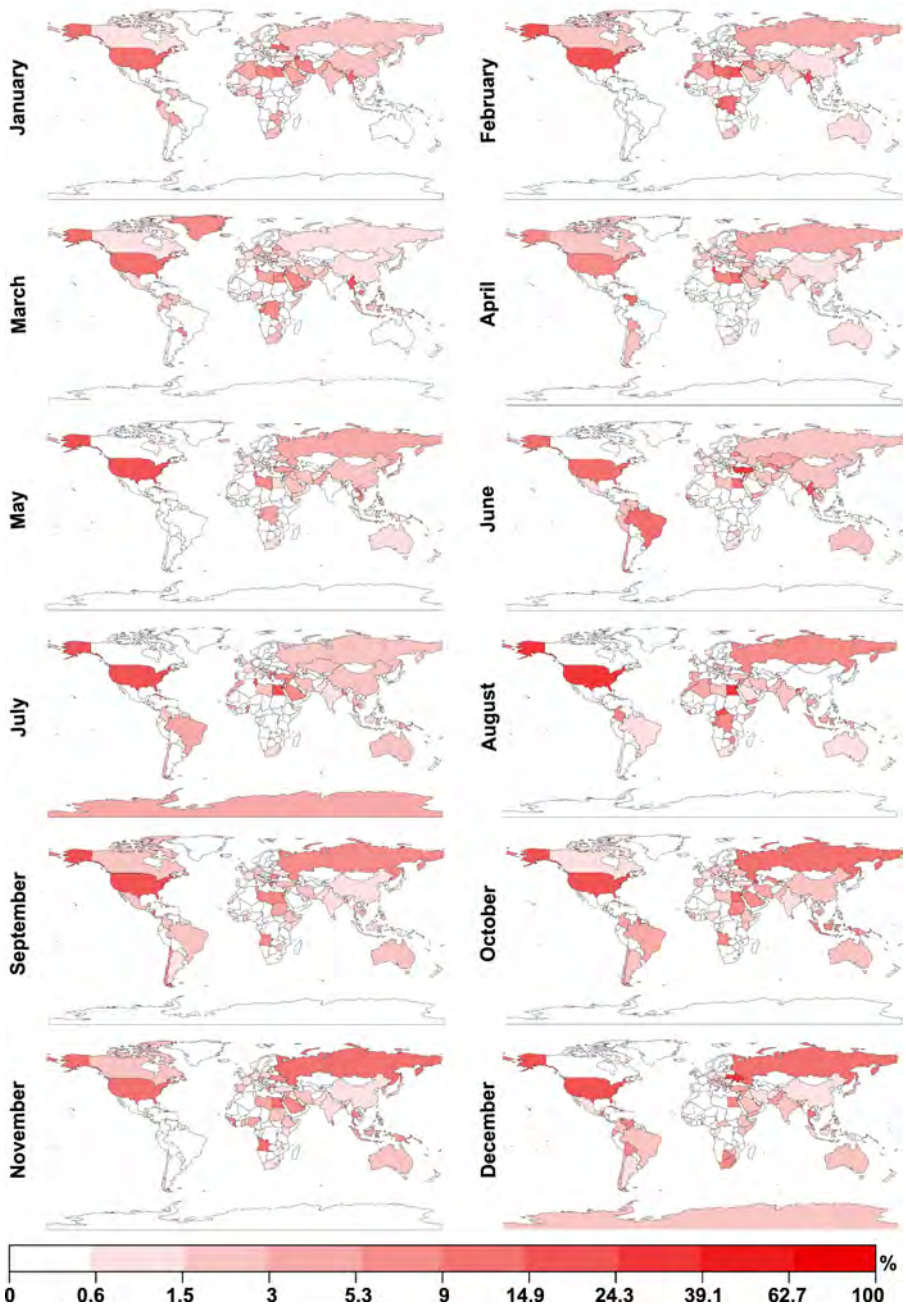

Figure 3.1: Reports of protests in 2013 in the online edition of *The Guardian*. We use data on reports of protests in the online edition of *The Guardian* as an approximation of the ground truth of when and where notable protest outbreaks occurred. For each of the 244 countries for each month in 2013, we calculate the number of *The Guardian* articles tagged with the country and region's name. Here, we depict the percentage of articles for each country and each month which were also tagged with the word "protest". Patterns which can be visually identified in the data reflect known major protest events in 2013: for example, protest outbreaks in both Brazil and Turkey can be observed in June 2013. Equal breaks are calculated for the logarithmically transformed percentages.

which were also labelled with a word signifying "protest" in one of the 34 languages identified above and listed in Table A2. Visual inspection suggests that while there are clear differences between the spatio-temporal distributions of "protest" labelled *Flickr* photographs and "protest" labelled articles in *The Guardian*, some key similarities can also be identified, such as an increase in "protest" labelled *Flickr* photographs in Brazil and Turkey in June 2013.

To determine whether we can find statistical evidence of a relationship between the number of "protest" labelled photographs taken and uploaded to *Flickr* and reports of protests in *The Guardian*, we consider both datasets at weekly granularity. For each week in 2013, for each country, we calculate the number of geotagged photographs taken and uploaded to *Flickr* which are labelled with the character sequence "protest" in 34 different languages, and normalise this count by the total number of geotagged photographs taken and uploaded to *Flickr* in that week and country. To analyse the relationship between the data mined from *Flickr* and reports of protests in *The Guardian*, we build a logistic regression panel model. To account for unobserved differences in coverage between countries and weeks, we include country and week as fixed effects.

Our results suggest that a greater normalised number of "protest" labelled *Flickr* photographs in a given week and country corresponds to a greater proportion of *The Guardian* articles about that country being tagged with the word "protest" (*Flickr* predictor: β = 2.95, *SE* = 0.31, *z* = 9.48, *N* = 12 932, $p < 0.001$). The odds ratio corresponding to the weekly fraction of "protest" tagged photos is 19.08, 95% *CI*: [10.37 – 35.09]. This implies that if we fix the country and week effects, increasing the normalised number of "protest" tagged *Flickr* pictures by 0.1 will increase the odds of a protest related *The Guardian* article by 34%.

For comparison, we construct a simple baseline model which captures differences in protest frequency between countries, and differences in protest frequencies across different weeks, by building a logistic regression panel model with country and week as fixed effects, leaving out the *Flickr* predictor. We find that the model including data on the normalised number of "protest" labelled *Flickr* photographs allows us to account for more variance in the proportion of *The Guardian* articles tagged with the word "protest" than this simple baseline model of differences between different countries and different weeks (*McFadden* R^2 for baseline model = 0.337, *McFadden* R^2 for *Flickr* model = 0.344, $\chi^2(1)$ = 84.55, $p < 0.001$, Likelihood Ratio Test).

3.4 Summary and discussion

We investigate whether data on photographs uploaded to the photo sharing website *Flickr* may be of use in identifying protest outbreaks. We analyse 25 million photos uploaded to *Flickr* in 2013 across 244 countries, and determine for each week in each country what proportion of the photographs are tagged with the word "protest" in 34 different languages. We find that higher proportions of "protest"-tagged photographs in a given country in a

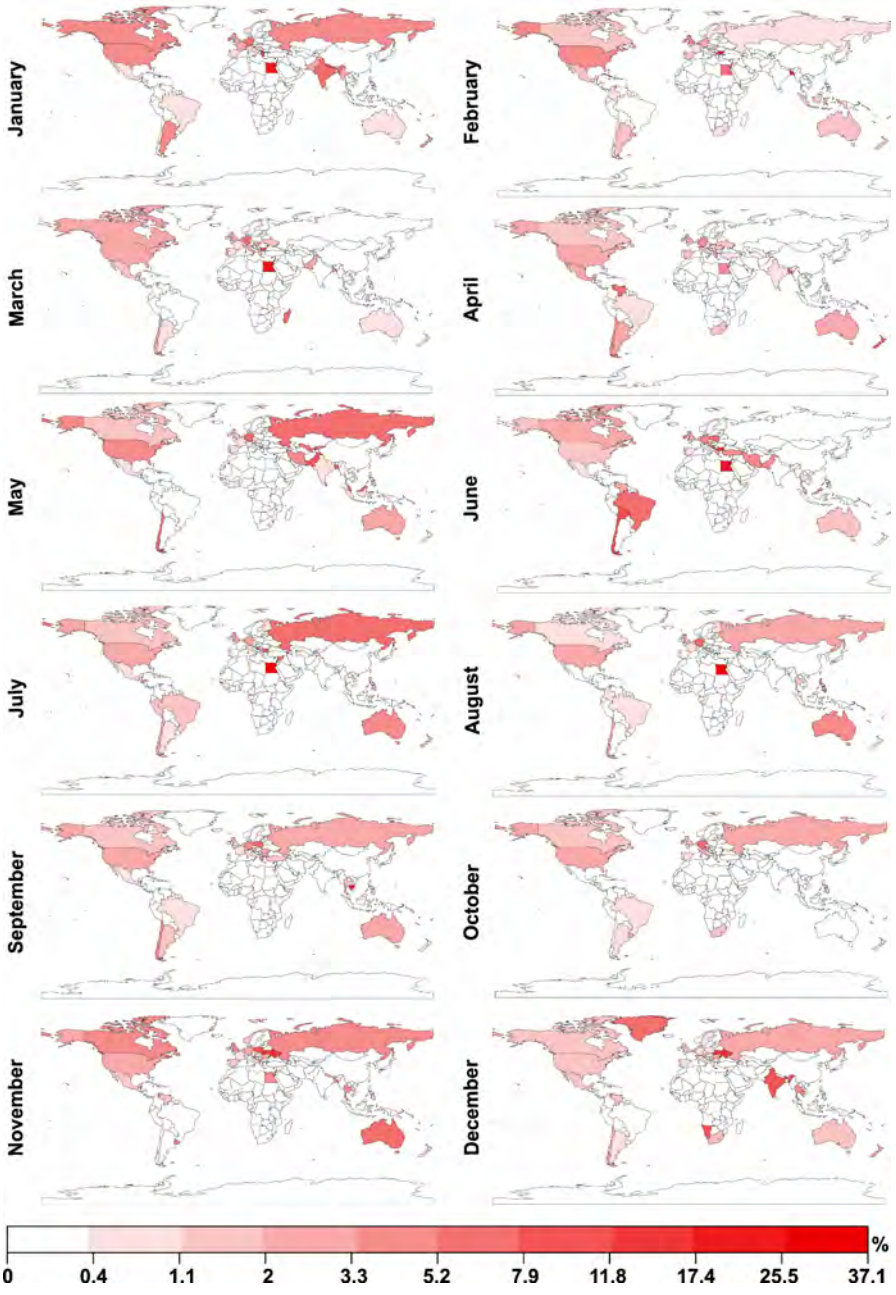

Figure 3.2: Locations of *Flickr* photographs labelled with "protest" in 2013. We investigate to what extent data on the number of photographs tagged with the word "protest" and uploaded to *Flickr* reflect the ground truth data extracted from *The Guardian*. For each of the 244 countries for each month in 2013, we calculate the total number of geotagged photographs taken and uploaded to *Flickr*. Here, we visualise the percentage of photographs for each country and each month which were also labelled with the character sequence "protest". Visual inspection suggests that while there are clear differences between the spatio-temporal distributions of "protest" labelled *Flickr* photographs and "protest" labelled articles in *The Guardian*, some key similarities can also be identified, such as an increase in "protest" labelled *Flickr* photographs in Brazil and Turkey in June 2013. Equal breaks are calculated for the logarithmically transformed percentages.

given week correspond to greater numbers of reports of protests in that country and week in the newspaper *The Guardian*. These results are in line with the striking hypothesis that data on photographs uploaded to *Flickr* may help us identify protest outbreaks.

In line with other studies of civil unrest, our analysis uses data from newspaper reports of protests as a proxy for ground truth data on protest occurrences (Braha, 2012; Compton et al., 2014; Dos Santos et al., 2014). As a result, we cannot rule out the possibility that *Flickr* users are posting photographs labelled with a word signifying "protest" as a result of reading an article about protests in their country in *The Guardian*, or another news source. We posit however that the geotagged nature of the *Flickr* photographs we analyse makes it less likely that such an explanation may hold, in contrast with simple time series analyses of online behaviour on services such as *Google* or *Twitter*, where searching behaviour or tweets may reflect reactions to news articles. With this caveat in mind, our results are consistent with the hypothesis that data on photographs posted to *Flickr* may help us to identify protest outbreaks around the world.

CHAPTER 4

Using Deep Learning to Detect Protest Outbreaks With Flickr Photographs

4.1 Introduction

Previous studies, including the work we presented in Chapter 3, have investigated whether it is possible to analyse human behaviour at a global scale by using data attached to the online photographs, such as text attached to the photographs including user defined picture title and tags (Alanyali et al., 2016; Barchiesi et al., 2015b; Preis et al., 2013a; Wood et al., 2013). However, relying on textual data presents a number of issues. First of all, language is one of the biggest constraints when performing a text based analysis on a global scale. Secondly, many photographs are uploaded with no text attached, whereas they often contain information on when and where they were taken particularly if they were captured by a smart phone camera. Here, we therefore extend the work we presented in the previous chapter. We investigate whether we can build a system that can automatically analyse the visual content of online pictures to identify protest outbreaks around the world.

For a long time, machine learning techniques applied to image analysis required considerable domain expertise to carefully design a feature extractor which would create a powerful representation of an image, so that a detector can later detect or classify certain patterns in the input image (LeCun et al., 2015). As discussed in Chapter 2, traditional image analysis methods have recently been outperformed by representation learning techniques which automatically determine the important features of an image from raw data for efficient detection or classification. Fuelled by the increasing power of the computational sources such as the fast graphics processing unit (GPU) as well as the large amount of annotated training sets available, these techniques are bringing the performance of machines to a level similar to visual perception of humans (Chatfield et al., 2014; Sharif Razavian et al., 2014).

In this chapter, by exploiting the state-of-the-art deep learning methods, we extend our work from Chapter 3 to investigate whether we can build a system to automatically analyse the visual content of *Flickr* pictures for identifying protest outbreaks around the world.

4.2 Data retrieval and preprocessing

We analyse a corpus of 24 944 764 million geotagged and publicly available photographs uploaded to the photo sharing platform *Flickr* in 2013. We extracted *Flickr* data in JSON format via the *Flickr* API in January 2014 using R 3.0.1. The dataset contains diverse information on pictures including where and when they were captured alongside text data such as a user defined title, picture tag and description.

In order to create the training dataset, we implement an automated search using the *Bing* Image Search API. The search phrase to form the positive training set was a given country name followed the word "protest", for instance "England protest". We downloaded the top 150 pictures returned by the search engine and repeated the same steps for each of the countries listed in Table B1 in Appendix B. For all the countries except Algeria, Micronesia, Kyrgyzstan and Tunisia, which returned less than 150 search results, we extracted 150 picture links.

To create the negative training set, we follow a similar approach as in Crowley and Zisserman (2014). For each of the countries, we extract the top 150 pictures returned by the search phrase formed with the country named followed by "things NOT protest", for instance "England things NOT protest". This fetches pictures indexed with the word "things" while omitting the ones which are indexed with "protest". For all the countries listed in Table B1, except Micronesia, North Korea and Western Sahara, we extract 150 image links for the negative training set. We should also note that 360 of these image links were broken. After discarding these pictures, we have 29 040 pictures remaining in the negative training set.

We then create the second set of training data using *Flickr* pictures taken and uploaded in 2012. We first extract all pictures containing the word "protest" either in the picture title, tag or description to serve as the positive training set. This process enabled us to extract 43 946 positive training samples. We then created a random subset from the remaining images matching the number of pictures in the positive training set.

For each *Flickr* photo taken and uploaded in 2013, we extract data on when and where it was taken along with its URL. We also retrieve the text data attached to each picture, such as the title, description and tags selected by the user. By the time we performed our analyses, some of the pictures that we have the metadata for had been removed by the users. This corresponds to 15% data loss compared to the total nunmber of pictures used in the final analysis in Chapter 3. For those pictures, while creating the models for comparison, we also discard the corresponding text data we already have to ensure consistency. For each week and each of the 244 countries and regions listed in Table A1 in Appendix

A, we determine how many pictures are automatically classified as protest-related. For comparison purposes, for each week and location, we also determine the number of pictures that were uploaded with the word “protest” in 34 different languages. The complete list of translations are provided in Table A2. The list of total number of pictures taken and uploaded to *Flickr* in 2013 from each country and region is in Table B2.

Due to numerous reasons including differences in Internet and *Flickr* usage across countries as well as usage variances across time, the number of pictures taken and uploaded to *Flickr* is not the same between different locations and different weeks. In order to minimise this effect, we normalise the counts of protest-related *Flickr* pictures in our models, by dividing counts of protest-related photographs from a given week and location by the total number of pictures from the same week and location. We should note that in all of our analyses, a week starts on Monday ending on Sunday.

As an indicator of the number of protest outbreaks around the world, we use counts of protest-related articles in the online version of *The Guardian* as a proxy for ground truth. We extracted data on articles in the online edition of *The Guardian* via *The Guardian Open Platform* in January 2016. We searched for articles which contained the word “protest” in the article content and were tagged with the name of the country or region as listed in Table A1, with the exception of the United Kingdom (where *The Guardian* newspaper is based) and the United States (for which area *The Guardian* has a digital edition, leading to some differences in article labelling practices). For the United Kingdom, we therefore searched for England, Scotland, Wales and Northern Ireland, whereas for the United States we looked for articles listed under the section “us-news”. To account for differences in news coverage between the countries and regions by *The Guardian*, for each week we extracted the total number of news articles containing the name of the country or region within the news content. For each country or region, the total number of *The Guardian* articles covering the news from the given location are listed in Table A4. When extracting the total number of news articles from the United Kingdom, we again searched for England, Scotland, Wales and Northern Ireland. For the United States, we investigate how many articles were published under the section “us-news”.

4.3 Methods

4.3.1 Training an initial classifier to detect protest scenes

For each picture taken and uploaded to *Flickr* in 2013, we retrieve data on both the time and place at which the picture was taken. For each week and each of the 244 countries and regions listed in Table A1, we determine how many pictures were potentially taken during a protest. In order to decide whether a picture was taken at a protest or not, we build a convolutional neural network based classifier.

In Chapter 2, we have provided example studies showing that in the absence of a large reliable training set, output from one of the fully-connected layers of CNNs can be

used as feature vectors (Chatfield et al., 2015; Crowley and Zisserman, 2014; Sharif Razavian et al., 2014), where they successfully perform better than using hand crafted image features in computer vision tasks such as image classification (Chatfield et al., 2014).

Figure 4.1a depicts the general architecture of a CNN. The final fully connected layer, commonly followed by a softmax function provides the classification results. Hence, the output of the network is a score denoting, given a set of classes, which class does the given image belong. Here, we use the output from the penultimate layer of the pretrained VGG-M-128 network (Chatfield et al., 2014) as our feature vector. We highlight the part of the network that we use as a feature extractor with a red box in Figure 4.1a.

In order to train a CNN based classifier, two sets of data are required, which are positive and negative training sets. These datasets are used to teach the classifier what kind of information is distinctive across different categories and what features are common within a certain category. This enables the CNN to capture the important aspects of visual data which can later be used to detect an object or a scene. The positive training set contains the samples of scenes or images with the object to be detected, whilst the negative training set consists of random images that do not contain the target object or scene information. It is then possible to run the CNN over both the positive and negative training sets to extract the feature vectors to be fed to a linear Support Vector Machine (SVM). The SVM will learn the common features within a category as well as the differences between categories and classify any given image into one of the two categories, which in our case are protest-related and non-protest-related (Figure 4.1b).

There is however no pre-annotated data available to train a classifier to automatically detect protest outbreaks. Following Chatfield et al. (2015), we therefore use readily available natural images which are annotated by the search engine *Bing*. We create positive and negative training sets by using images returned from an automated search using the *Bing* Image Search. Further details about extracting data from *Bing* is provided under Section 4.2. We then extract feature vectors of both positive and negative training sets and use them to train the CNN based classifier. We refer to this classifier as the *Bing* classifier.

In order to test the performance of the classifier, we form a test set using the *Flickr* images taken and uploaded in 2014. We download 1 500 images tagged with the word “protest” to serve as positive test samples. For the negative test set, we extract another 1 500 images that do not contain the word “protest” in their tag, title or description. We then run the classifier trained on *Bing* images over the test set which grouped them either as protest-related or non-protest-related. Figure 4.2 summarises the performance of the classifier by showing the number of correctly classified photographs in blue and the number of misclassified photographs in red. The classifier is able to categorise 64% of the protest pictures correctly. However, it misclassifies more than one third of the pictures in the negative test set, labelling them as protest-related.

4.3.2 Training a refined classifier to detect protest scenes

To check whether we can improve the performance of the classifier by using a different training set, we trained a second classifier by using *Flickr* pictures taken and uploaded in 2012. Similar to the test set, we use pictures tagged with the word "protest" as the positive training set whilst creating a random subset of pictures not containing the word "protest" in the text data attached. We call this the *Flickr* classifier. We then run the *Flickr* classifier over the test set. The *Flickr* classifier performs much better with fewer pictures in the negative training set labelled as protest-related compared to the *Bing* classifier (Figure 4.2).

As both of these training sets are created automatically, there might be some noise in the data which could affect the final performance of the classifier. We therefore investigate whether running one of the classifiers over the positive training set for the other classifier might help remove the potential errors in the training dataset. We first run the classifier trained on *Flickr* data over the positive training data created using *Bing*. We then discard the pictures labelled as non-protest-related by the classifier and create a subset by just keeping the pictures that are labelled as protest-related. Using the new positive training set, we then train a new refined classifier. We call this the refined *Bing* classifier. Similarly, we repeat the same procedure and run the *Bing* classifier over the *Flickr* positive training set to create a second refined classifier. We call this the refined *Flickr* classifier. Figure 4.1c visualises the work flow of how we train a classifier in two steps. For both scenarios, we then check the performance of the refined classifiers by running them over the test set. The two rightmost subplots in Figure 4.2a summarise the performance of the refined classifiers trained in two steps. Although these refined classifiers display similar performance, both outperforming the initial *Bing* and *Flickr* classifiers, the refined *Flickr* classifier labelled fewer pictures in the negative training set as protest-related, therefore demonstrating higher accuracy. Considering the overall performance of all four classifiers, we decide to select the refined *Flickr* classifier to be used in the rest of our analyses.

4.3.3 Creating Receiver Operating Characteristic (ROC) curves

Next, we evaluate the performance of the classifier by calculating a ROC curve. For each image in the test set, we know the actual label and we also know the score returned by the classifier, which is a linear SVM trained using features of the images in the training set extracted with a CNN. We determine a picture to be protest-related by gradually changing the threshold values. Lower threshold means labelling more pictures as protest-related whereas higher threshold means labelling fewer pictures as protest-related. For n unique SVM scores, we would have $n+1$ threshold values. For each threshold, we will calculate sensitivity and specificity scores. Sensitivity, also known as the true positive rate, measures how many protest-related pictures are labelled as protest-related by the classifier whereas specificity, true negative rate measures how many non-protest pictures are labelled as non protest-related by the classifier. Using these values, we can then draw a ROC curve by

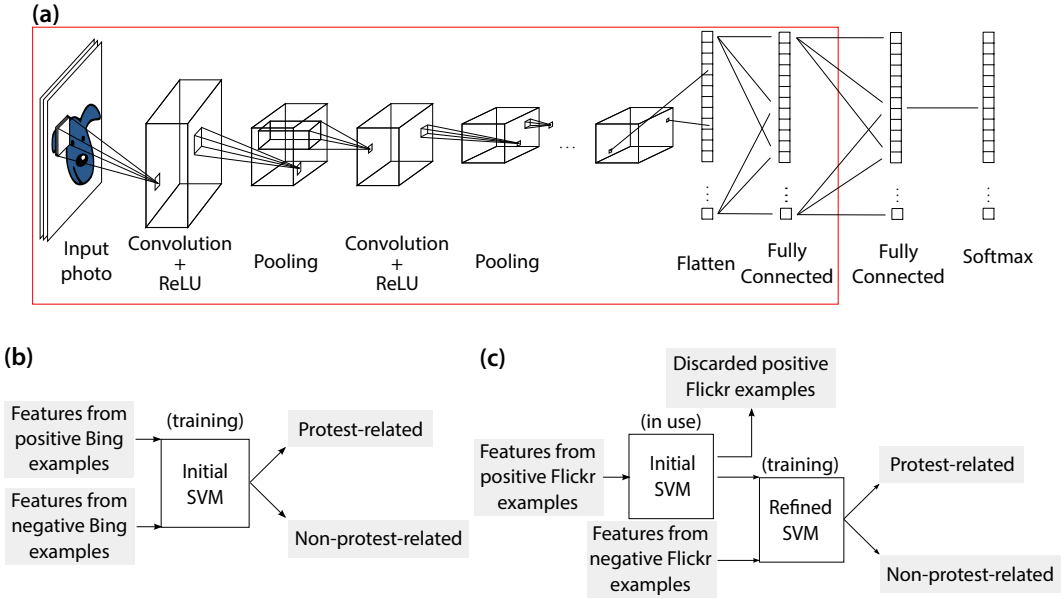

Figure 4.1: Work flow for training the CNN based classifier.
(a) A toy model of a convolutional neural network. CNNs are formed by layers of neuron-like structures that are similar to neurons in the human body. Here, we remove the final softmax layer keeping the part of the network marked with the red box and use the output of the network as feature vectors. The sketch is adapted from the figure at `https://uk.mathworks.com/discovery/convolutional-neural-network.html`. **(b)** General work flow of training a classifier where CNNs are used as feature extractors. Once features from both positive and negative training sets are retrieved, we pass them to a support vector machine with a linear kernel, which will learn the features that are common across the samples in the positive training set and the differences between positive and negative training sets. Our training procedure then has two steps. In the first step, we train the initial SVM by using the features extracted from the *Bing* training set. We call this the *Bing* classifier. **(c)** The full classifier training workflow. In the second step of the training process, we use the *Bing* classifier, which was trained as shown in Figure 4.1b, to eliminate the noise in the positive training set which is formed by using the images extracted from *Flickr*. We discard the pictures in the positive training set if they are not labelled as protest-related by the *Bing* classifier. Finally, we pass the refined positive training set with the negative training set created by the data from *Flickr* to the refined SVM in order to train the main classifier, which we call the refined *Flickr* classifier.

plotting sensitivity values against specificity values. To determine whether this classification provides us with useful results, we use the Area Under Curve (AUC) metric, where AUC values closer to 1 suggest a more accurate classification. Figure 4.2b visualises the performance of the two step classifier using a ROC curve. together with the performance expected from randomly labelling images as protest-related or not which is shown with the black line and has an expected AUC of 0.5. The classifier performs with 83% accuracy and 0.877 ± 0.012 AUC with a 95% confidence interval, suggesting that the classifier correctly labels images at a level above random labelling.

We then run the refined *Flickr* classifier over the entire set of photos taken and uploaded to *Flickr* in 2013. Figure 4.3 depicts sample images that are automatically grouped by the classifier. Photos with a blue frame are classified as protest-related while photos with a red frame are classified as non protest-related. Visual inspection suggests that the classifier is capable of detecting protest scenes even with protest signs in different languages highlighting the language independent nature of analysing picture content. These photographs also demonstrate that the classifier can differentiate between different crowds not just classifying every picture with a group of people as a protest scene.

4.3.4 Computing Akaike Information Criterion (AIC) weights

Akaike Information Criterion weights, AICw, are transformed AIC values which makes the observed difference between the raw AIC values easier to compare. They can be interpreted as the probability of a model given the data, when choosing a model from the set of models for which the AICws have been calculated. In order to obtain the AICw values for the models we created, we use the following formula introduced in Wagenmakers and Farrell (2004):

$$AICw_i = \frac{exp\{-\frac{1}{2}\Delta_i(AIC)\}}{\sum_{j=1}^{N} exp\{-\frac{1}{2}\Delta_j(AIC)\}} \tag{4.1}$$

where $\Delta_i(AIC)$ is the difference between the AIC value of the i^{th} model and lowest AIC value among the *N* models under consideration. For every model we create, we then calculate the AICw value.

4.4 Analysis and results

For each week, we extract the number of pictures classified as protest-related taken in each one of the 244 countries and regions. However, the number of users and therefore the number of photographs taken at each location as well as each week might differ. In order to reduce the potential bias which might be caused by the variations in the *Flickr* usage, we divide the number of weekly pictures classified as protest-related by the total number of pictures taken and uploaded in the same week and location. To investigate whether we can identify protest outbreaks using data mined from *Flickr*, we need data on

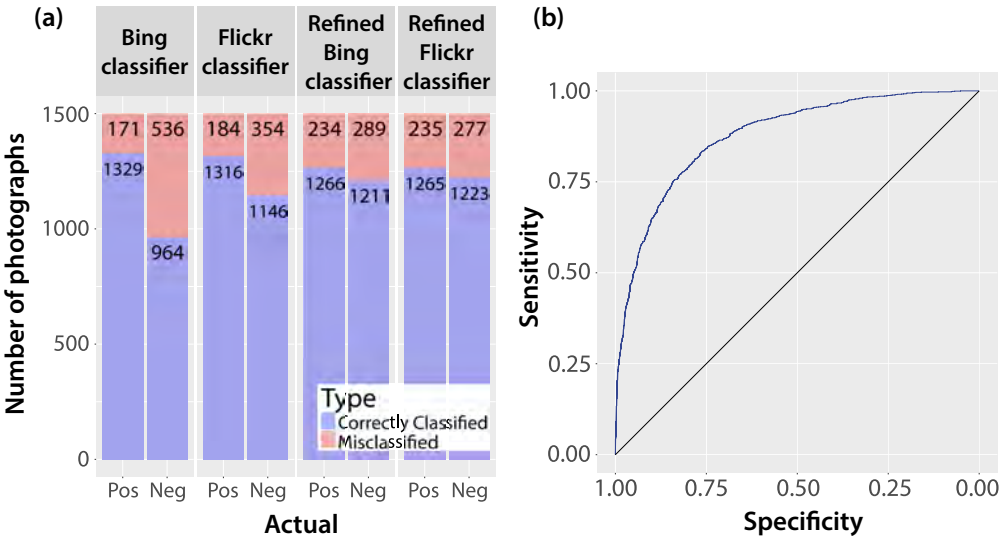

Figure 4.2: Evaluating the performance of the classifier.
(a) Performance summary of the four classifiers over the test set of 1 500 protest-related positive and 1 500 non-protest-related negative images. The test results suggest that the *Flickr* classifier outperforms the *Bing* classifier, whilst the refined classifiers outperform them both. Although the refined classifiers display similar performance, the refined *Flickr* classifier produces slightly fewer misclassified protest-related pictures compared to the refined *Bing* classifier. We therefore pick the refined *Flickr* classifier as our main classifier for the rest of the analysis. **(b)** We measure the performance of the classifier by systematically adjusting the threshold values to determine whether a picture is protest-related or non-protest-related. For each picture, we calculate two metrics of quality to determine the accuracy of the classifier: sensitivity and specificity. Sensitivity is the proportion of images classified as protest-related which appear to be protest-related, whilst specificity refers to the proportion of images classified as non-protest-related which are indeed not. Using these metrics, we can then plot a Receiver Operating Characteristic (ROC) curve, where the closer the Area Under Curve (AUC) to one, the better is the performance of the classifier. Our analyses suggest that the classifier performs well with an AUC value of 0.877 ± 0.012 and 83% accuracy at a level above randomly grouping the images, represented by the black line which has an AUC of 0.5.

Figure 4.3: Sample set of pictures automatically grouped by the classifier.
Pictures with a blue frame are automatically classified as protest-related whilst the pictures with a red frame are automatically classified as non-protest-related. Visual inspection suggests that the classifier is capable of detecting general protest scenes across different countries. For example, the classifier does not depend on protest signs being presented in a particular language. We also note that the classifier can differentiate between a protesting crowd and an ordinary crowd of people, where examples of the latter are labelled as non-protest-related. Picture credits are listed in Appendix B.1.

the number of protest outbreaks that took place in 2013. Previous studies analysing social unrest exploit data collected from news outlets as a proxy to extract the number of protest outbreaks around the world (Braha, 2012; Compton et al., 2014; Dos Santos et al., 2014; Steinert-Threlkeld et al., 2015). We therefore extract the proportion of the articles covering protests in the online edition of *The Guardian*.

To quantify the relationship between *Flickr* pictures classified as protest-related and reports of protests in *The Guardian*, we build a logistic regression panel model. In order to account for unobserved differences in coverage from different weeks and locations, in addition to the *Flickr* predictor, we include two fixed effects which are location (country and region) and time (weeks) in order to account for unobserved differences in coverage from different weeks and locations. We will refer to this model as the "image model". Our results suggest that an increase in the normalised number of *Flickr* pictures classified as protest-related is linked to an increase in the proportion of the online articles in *The Guardian* covering protest news (*Flickr* predictor: β = 1.02, *SE* = 0.18, *z* = 5.65, *N* = 12 932, $p <$ 0.001). We find that, when keeping the time and location fixed effects constant, a 0.1 increase in the normalised number of the pictures classified as protest-related is linked to an 11 % increase in the odds of an article in *The Guardian* covering protest-related news.

To analyse the performance of the image model, like in Chapter 3, we build a simple baseline model using a logistic regression panel model for comparison. The baseline model only includes the two fixed effects: location (country and region) and time (weeks). The baseline model therefore only accounts for certain countries or regions generally receiving larger amounts of news coverage about protests, or certain weeks generally seeing larger numbers of articles about protests in *The Guardian*. A comparison of this baseline model to the image model which includes information we extract by analysing *Flickr* pictures reveals that the image model is better at capturing the changes in the proportion of *The Guardian* articles covering protest-related news (*McFadden's* R^2 for baseline model = 0.353, *McFadden* R^2 for *Flickr* model = 0.356, $\chi^2(1)$ = 29.54, $p <$ 0.001, Likelihood Ratio Test).

Aiming to create a language independent method to identify protest outbreaks, in this chapter we exploit visual information embedded in the online pictures to build the image model. In our earlier work presented in Chapter 3, we focused on analysing textual data attached to the pictures uploaded online which we will refer to as the "tag model". Once compared to the baseline model, our results suggest that in capturing proportions of protest-related news articles, the tag model has a similar yet better performance than the image model (*McFadden* R^2 for baseline model = 0.353, *McFadden* R^2 for tag model = 0.357, $\chi^2(1)$ = 45.84, $p <$ 0.001, Likelihood Ratio Test).

We create a third and final model, which we will refer to as the "combined model" to investigate whether we can achieve better performance by combining the information extracted from both text data and image content. In comparison to the baseline model, the combined model is better at modelling the change in the proportion of the protest-related articles published in the newspaper *The Guardian* (*McFadden* R^2 for baseline model =

Table 4.1: Model comparison results.

To make comparisons between the models, we calculate McFadden's adjusted R^2 for each model. Using this measure, models are penalised with respect to the number of predictors, to help guard against overfitting. We also calculate the AIC weights, which can be interpreted as the probability of a model given the data, when calculated and normalised across all models under consideration (Wagenmakers and Farrell, 2004). Our results suggest that the best model of global protest activity is the model that seeks to identify photographs of protests on *Flickr* using both text data attached to images and by processing image content itself.

Model	**Adjusted $\mathbf{R^2}$**	**AIC**	**AIC Weights**
Baseline	0.353	8598.8	< 0.001
Image model	0.356	8571.3	< 0.001
Tag model	0.357	8555.0	< 0.001
Combined model	0.358	8543.2	0.997

0.353, *McFadden* R^2 for combined model = 0.358, $\chi^2(1)$ = 59.61, $p <$ 0.001, Likelihood Ratio Test).

To provide a further comparison of the quality of these models, we calculate their Akaike Information Criterion (AIC) weights (Wagenmakers and Farrell, 2004). Given a set of models for comparison, the AIC weights can be interpreted as the probability of each model given the data. The AICws for all models considered sum to 1.

Table 4.1 depicts the AICw for each model. Our analysis shows that the AICw of the combined model using both image and text data is nearly 1, whereas the AICws of all other models, including the tag and image models are nearly 0. These findings align with the results of the adjusted R^2 analysis, suggesting that the best model of global protest activity is the model that seeks to identify photographs of protests on *Flickr* using both text data attached to images and by processing image content itself.

4.5 Summary and discussion

In summary, automatic analysis of the pictures uploaded online might help us to identify protest outbreaks around the world. Previous studies using online images have mainly focused on analysing textual data attached to the pictures (Alanyali et al., 2016; Barchiesi et al., 2015b; Preis et al., 2013a; Wood et al., 2013). Recent improvements in computer vision provide new methods to analyse images taking a step forward to bring machines to a level of perception similar to humans' (Chatfield et al., 2014; LeCun et al., 2015). This opens up new avenues to exploit the potential of online images to provide fast, cheap and language independent ways of measuring human behaviour around the world. Here, we seek to investigate whether images shared online provide valuable information in detecting

protest outbreaks worldwide.

We train a two-step classifier using CNNs to extract feature vectors of natural images from *Bing* search engine and photographs uploaded to *Flickr*. According to the ROC curves and AUC, the classifier works better compared to randomly labelling photographs as protest-related or not. We then run the classifier over a large corpus of online pictures taken and uploaded to the photo sharing platform *Flickr* in 2013 across 244 countries and regions. For each week and location, we determine the ratio of the pictures that are automatically classified as protest-related. In order to compare whether the number of photographs classified as protest-related are inline with the number of protest outbreaks around the world, we use articles in the online edition of *The Guardian*. We model the relationship between protest labelled *Flickr* pictures and *The Guardian* articles covering protests by building a logistic regression panel model. We find that an increase in the normalised number of *Flickr* pictures automatically classified as protest-related corresponds to a rise in the proportion of the protest-related articles published in *The Guardian*.

We then compare the model using information from analysing image content with the model using information extracted from the text data attached to the *Flickr* images. These two models have a very similar performance, but we find that the model containing results from text analysis performs slightly better. Nevertheless, the model which combines information from both text and image analysis proves to be the best model capturing changes in the number of protest-related online articles in *The Guardian*. Our results suggest that combining information extracted from the image content alongside the metadata attached to the pictures shared online may improve the estimates of the number of protest outbreaks around the world.

We note that the change in the variation of the protest-related articles in *The Guardian* captured by using *Flickr* photographs are still not extremely different from the baseline model. One reason might be that the majority of the change is captured by the fixed effects. We can also argue that the improvement might be limited due to the pictures falsely classified as protest-related. In order to reduce the number of misclassified pictures, we have tried different training sets as well as introducing a filtering step. However, like other classifiers, the final classifier is prone to misclassification. This means that we get more pictures labelled as protest-related than there are at the same time missing out some of the actual protest pictures. This may reduce the strength of the meaningful signal coming from the *Flickr* photographs to analyse whether we can capture the change in the number of articles in the newspaper *The Guardian* covering protests.

We also posit that the *Flickr* dataset itself inherently has a bias coming from the usage of the platform. Usage of this social media platform varies across different regions as well as different time periods. In our models, we have tried to incorporate this difference. However, it is likely that the performance of the models are affected especially in cases which there is no data shared from a certain location or over a certain time period. Besides, the lack of ground truth data might also constitute a problem.

Unfortunately, there is no global database that tracks the number of protest out-

breaks happening hence most of the studies focusing on social unrest used news outlets as an alternative data source (Braha, 2012; Compton et al., 2014; Steinert-Threlkeld et al., 2015). We therefore cannot rule out the possibility that there might be cases where a protest outbreak might not be covered by *The Guardian.*

With these drawbacks in mind, our results are in line with the striking hypothesis that photographs shared on social media platforms can provide insights into protest activity around the world. These findings illustrate the new opportunities that exist to use online images as a source of cheap, rapid and language independent measurements of global collective human behaviour.

CHAPTER 5

Estimating Socioeconomic Attributes Using Instagram

5.1 Introduction

As discussed in Chapter 2, the popularisation of cameras embedded in mobile devices gave rise to social photography. With their easily accessible nature as well as high quality cameras, mobile devices are now often preferred over traditional cameras, especially in "everyday photojournalism" (Okabe, 2004).

In recent years, one of the trends everyday photojournalism introduced in our lives is to share food pictures on social media channels. According to a food survey conducted by one of the leading supermarket chains in the UK, in 2016 one in five Britons has uploaded a picture of food either on a social media platform or shared it via messaging channels (Waitrose, 2017). This new trend has gaven rise to a large number of food-related pictures shared on social media channels. Motivated by the increasing amount of data created as a consequence of this trend, in this chapter, we exploit a set of food pictures from the picture sharing platform *Instagram* taken at restaurants in London. We provide results from an initial attempt to estimate restaurant ratings posted on *Yelp* by automatic analysis of the picture content. We then extend the scope of our analysis to investigate whether pictures shared on *Instagram* can unveil information about the socioeconomic status of a city.

Policy makers traditionally collect information on social statistics by surveying populations; often via the use of a census. Despite serving as a rich and valuable source of information on the socioeconomic status of populations, orchestrating surveys at a national scale is expensive and labour intensive. For instance, the census in the UK is decennial, meaning that data from a national census will be used to make decisions over that ten year period. Thus, as time passes the information becomes a more outdated description of the status of a country rather than describing its current state. Researchers have therefore been investigating whether data generated as a result of a global surge in using technological devices can be used as a complementary source to the traditional information outlets

in real time monitoring of a city or a country.

In previous studies, analysing data from search engines (Choi and Varian, 2012; Ettredge et al., 2005) and social media platforms (Antenucci et al., 2014; Bollen et al., 2011a) has emerged as a promising way of creating indicators of key socioeconomic measures. Additionally, a number of studies have exploited different forms of data including data generated by mobile phone usage (Blumenstock and Eagle, 2010), night lights observed from satellite images (Pinkovskiy and Sala-i Martin, 2014) and images extracted from *Google Street View* (Gebru et al., 2017) to shed light on population statistics.

In this chapter, after investigating whether we can estimate a restaurant's rating by using pictures of food taken at the restaurant, we discuss how food-related pictures shared over a six month period on *Instagram* can be used to infer income patterns in London. We later extend this analysis by incorporating the entire set of *Instagram* pictures to investigate whether we can create better estimates of income in London. Finally, in order to verify whether these results are consistent across other cities, we investigate whether we can estimate income in New York City by analysing visual characteristics of six months worth of pictures taken around New York City.

5.2 Data retrieval and preprocessing

5.2.1 London data

5.2.1.1 *Instagram* data

We collected metadata on 6 117 318 photos uploaded to *Instagram* which are publicly shared by users between September 2015 and February 2016 within the Greater London area, via the *Instagram* API.

After the 2001 nationwide census in the UK, in order to enhance reporting of local statistics in England and Wales, new population areas called Super Output Areas (SOA) were defined (ONS, 2012). SOAs are further grouped into two subcategories taking into account population and the number of households: Lower layer Super Output Area (LSOAs) and Medium layer Super Output Area (MSOAs). MSOAs are composed of groups of LSOAs and contain between 2 000 and 6 000 households. Their boundaries stayed the same until 2011 when some modifications were made due to the changing size of the areas.

In the rest of this chapter, for the London part of our analysis, we use MSOAs as spatial units. Figure 5.1 depicts an overview of the *Instagram* dataset visualised at MSOA level.

Visual inspection of Figure 5.1 suggests that the majority of the *Instagram* pictures are grouped around areas such as Westminster, Camden, Hackney and Greenwich where the most iconic London attractions reside. Almost 2 million pictures in our dataset were taken in Westminster alone. However, there is another popular *Instagram* spot that might initially come as a surprise considering it is substantially further away from central London.

Figure 5.1: Total number of *Instagram* pictures per MSOA shared during a six-month period between September 2015 and February 2016.
The majority of the pictures are clustered around the centre of London, with a few exceptions including Hillingdon, where Heathrow Airport is located. Numbers are shown in logarithmic scale. Colour breaks are created using the k-means clustering algorithm.

This is Hillingdon, where Heathrow Airport is located. Although not evenly distributed, the *Instagram* dataset has a good coverage of London with a minimum of 29 pictures per MSOA.

5.2.1.2 *Yelp* data

We retrieved restaurant data from Greater London area using the *Yelp* API in June 2017. We should underline that *Yelp*'s definition of a restaurant contains several subcategories including bistros, cafes as well as a long list of world cuisines. These are all listed in Table C1 in Appendix C. For consistency, we will use the term "restaurant" in the rest of this chapter to refer to restaurants together with all these subcategories.

We extract restaurant data via the *Yelp* API which provides various endpoints including '/business/search' that returns businesses matching the information specified by the input parameters and '/business/id' that returns extended information including name, geo location as well as the reviews. In order to get detailed information on restaurants, we need the unique business IDs. These are unique identifiers defined by *Yelp* to distinguish between different restaurants and different branches of the same restaurant chain. We therefore create an automated request by using the 'business/search' endpoint. We divide the Greater London area into 250×250 meter square grid cells, and for each cell extract the centre coordinates. Then, for each coordinate pair, we make a search request by setting the radius to 250 meters. This covers a larger area than the grid cell. However, before proceeding to the next step, we eliminate duplicate entries. We note that the current API only returns data from businesses with at least one review on the system.

Once we collect all the Yelp specific IDs, for each restaurant, we make another GET request using the '/business/id' endpoint to extract detailed information. After removing restaurants without geolocation information, we finally have 16 372 restaurants from the Greater London area in our dataset. Figure 5.2a shows the total number of restaurants per MSOA. There are 44 MSOAs with no restaurant information in the *Yelp* dataset which are indicated by the grey shaded areas. Visual inspection hints at the existence of a similar pattern to *Instagram* pictures where the majority of restaurants are clustered around the central London area, though they are concentrated around a much smaller radius.

In order to develop a better understanding of the spread of the restaurants around central London, we pick the 20 MSOAs with the highest number of restaurants. Figure 5.2b visualises data at restaurant level where each circle shows a restaurant. The icon size is proportional to the number of reviews posted on *Yelp* about a given restaurant at the time we extracted the dataset. Circles with darker colour represent restaurants with higher ratings while the lighter shades represents lower ratings. Among the selected 20 MSOAs, we also highlight the three restaurants with the highest numbers of reviews, which are *Dishoom* with 801 reviews, *Duck & Waffle* with 407, and *Sketch* with 401 reviews posted on *Yelp*.

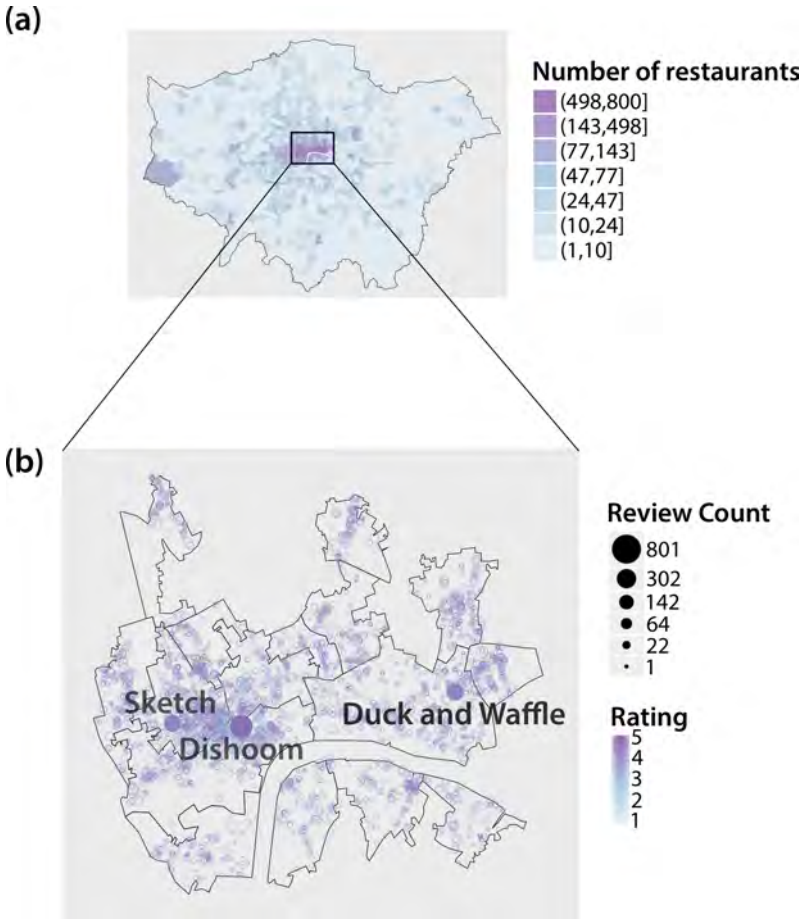

Figure 5.2: Number of restaurants in the *Yelp* dataset.
(a) Figure depicting the total number of restaurants per MSOA. As in the *Instagram* picture, the majority of the restaurants in the dataset are clustered around the central London area. (b) The magnified area of central London covering 20 MSOAs with the highest number of restaurants. Within this area, the top 3 restaurants with the largest number of reviews are marked on the map. Colour breaks for both figures and review count breaks are calculated using *k*-means clustering.

5.2.1.3 Combining *Instagram* and *Yelp* data

In order to identify which *Instagram* pictures are taken at restaurants we need to overlay *Instagram* and *Yelp* datasets with respect to location.

Location names on *Instagram* are created using public location pages on *Facebook*. In addition, if users cannot find the place they are searching for on the list, they can create a new location name again via *Facebook*. The list of location names is very diverse, ranging from very specific names such as the name of a local business or a socio-cultural attraction spot, to being high level such as the name of a town or city. On the other hand, the *Yelp* dataset contains geolocation information and the name of the restaurant. Although they both provide geographical coordinates of the data points, coordinates from the same location might differ slightly between the two datasets. This difference might lead to matching images to the wrong restaurants. It is therefore a non-trivial task to merge *Instagram* and *Yelp* datasets by matching the coordinates. Hence, we merge these two datasets by matching location names in *Instagram* with the restaurant names in *Yelp*.

Due to variations in the naming convention of different platforms, *Instagram* and *Yelp* names have certain differences. In order to handle these inconsistencies, before merging we therefore convert all location names in both datasets to lowercase, replace "&" with "and", strip multiple white spaces and then remove all non alphanumeric characters.

As different branches of the same restaurant have the same name, once we merge the datasets, some entries have multiple matches. To address this problem, we calculate the Haversine distance between the matched data points which assumes spherical earth by ignoring ellipsoidal effects when calculating the distance between two coordinate pairs. We compute the Haversine distance using the *distHaversine* function from *geosphere* package in R (Hijmans et al., 2015).

Once we calculated the distance between each matched coordinate pair, we pick the pair with the smallest distance and call it a match. However, if the pair still has a distance more than 50 meters, we then assume that the match is not valid and remove the data point from our analysis. We have a final set of 134 898 *Instagram* images taken at 4 035 restaurants.

5.2.1.4 The household income data

In this study, we use MSOA level household income estimates released by the Greater London Authority (GLA) in 2015. GLA estimates have been calculated using results from multiple surveys such as Understanding Society, Annual Survey of Hours and Earnings (ASHE) and the survey of personal income (GLA, 2015). Part of the model used to create estimates relies on house prices and data from the UK Census, which is conducted once per decade.

The GLA release estimates at various administrative levels in the Greater London area including Super Output Areas (SOAs). Here, we exploit MSOA level median income estimates for 2011/12 modelled by the GLA. For simplicity, for the rest of this chapter we

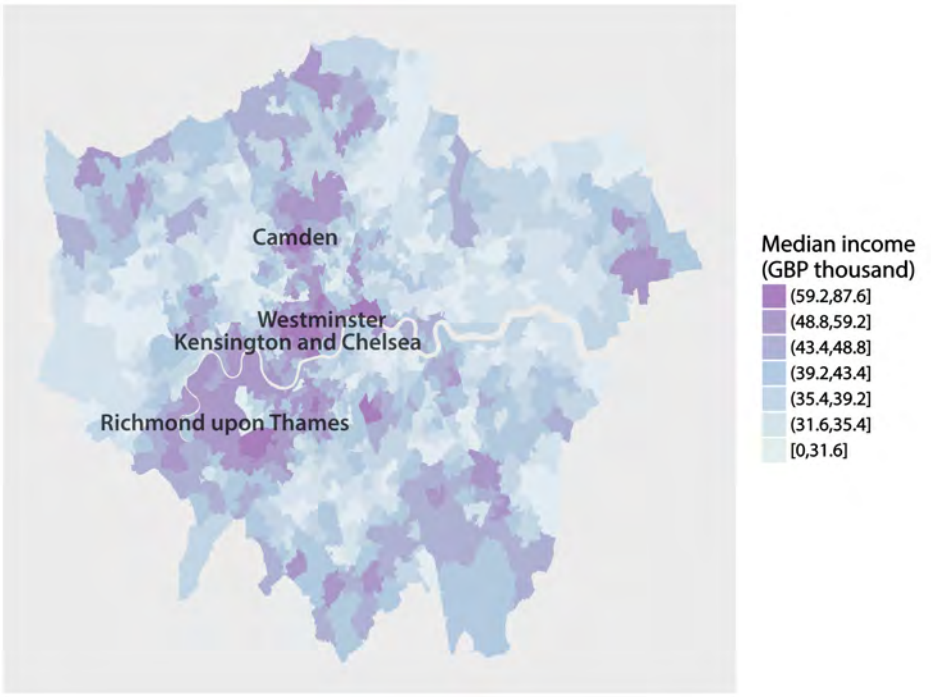

Figure 5.3: MSOA level median household income estimates.
Visual inspection suggests that high income areas, which are represented by darker colours, such as Westminster, Kensington and Chelsea, and Richmond upon Thames are clustered around a path following the Thames in the west. On the other hand, lower income areas are scattered around London. Colour breaks are calculated using the *k*-means clustering algorithm.

refer to these values as "actual" income values saving the word "estimate" for the output of our analyses and use "London" when referring to the Greater London area including the City of London.

Figure 5.3 depicts income distributions across MSOAs in London. The highest income areas are represented with darker colours and are mainly clustered around a flipped L-shape including Westminster, Kensington and Chelsea towards Richmond upon Thames in the south west, and stretching up to Camden in the north, as marked on the map. While most of the high income areas are clustered around the city centre, MSOAs with lower income seem relatively more spread out around London.

5.2.2 New York City data

5.2.2.1 *Instagram* data

In order to download *Instagram* pictures from New York City, we follow a similar procedure as extracting pictures from London. We download *Instagram* pictures taken around New

York City over a six month period between January and March 2015 and from July to September 2015. The dataset was downloaded in the first quarter of 2016. After clipping the New York City area, the final dataset in total contains 5 528 910 pictures from 2 101 census tracts.

Census tracts are geographic areas generally hosting 4 000 inhabitants on average (US Census Bureau, 2010). The census tract boundaries we use for this chapter were developed for the 2010 census. New York City has a total of 2 168 census tracts each with an average land area of 90 acres (NYC City Planning, 2015). Figure 5.4 depicts the distribution of the *Instagram* pictures at census tract level taken in New York City over a six month period. Not surprisingly, touristic places such as Central Park and the Statue of Liberty as well as other outdoor green spaces such as Prospect Park stand out with a relatively high number of pictures shared on *Instagram*. As in London, apart from the tourist hot spots, airports are also areas where many *Instagram* users tend to take pictures.

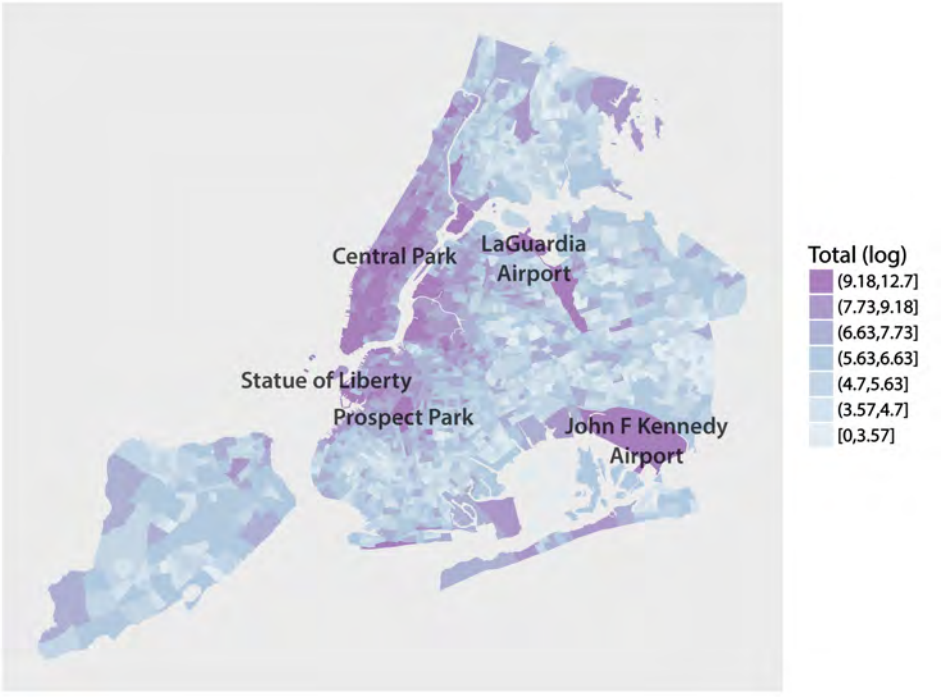

Figure 5.4: Total number of *Instagram* pictures per census tract taken in New York City over a six-month period.
The majority of the pictures are clustered around tourist hot spots as well as large parks. As in London, airports also appear to be popular spots among *Instagram* users. The number of pictures is shown in logarithmic scale and colour breaks are created using the *k*-means clustering algorithm.

5.2.2.2 Household income data

In the final part of this chapter, we focus on estimating income values using *Instagram* pictures in New York City. In terms of spatial units, for New York City we utilise census tracts. However, unlike the MSOAs in London, not all census tracts have income data associated with them as some of them only contain non residential areas such as airports, parks and cemeteries. Figure 5.5 depicts the distribution of the median annual income across New York City. Darker colours represent higher income areas such as Carnegie Hill, Little Italy, Midtown South and Union Square.

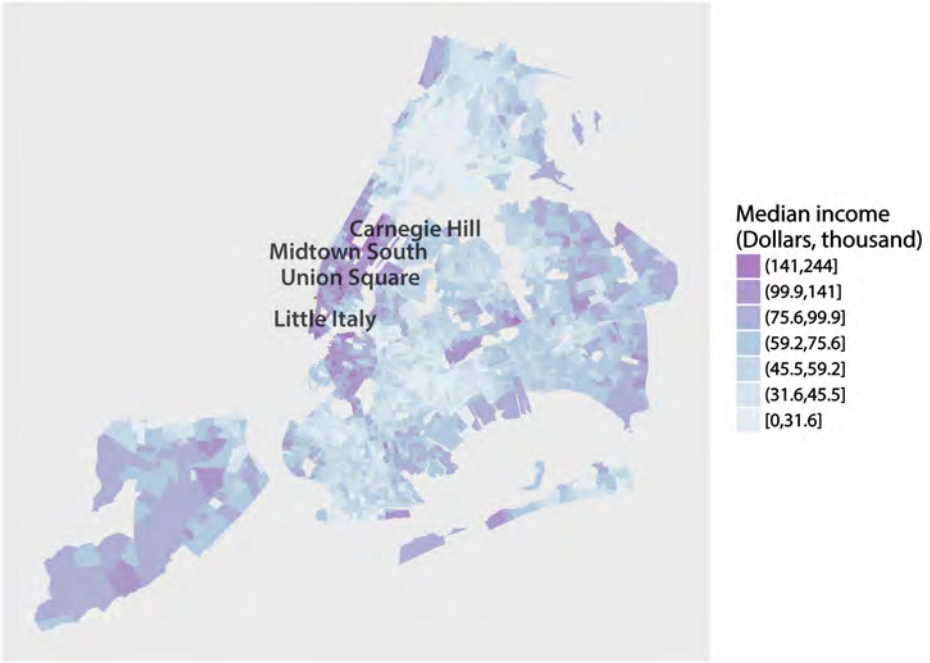

Figure 5.5: Median income of New York City at census tract level.
The map illustrates the distribution of income across New York City. Census tracts with high income including the census tracts enclosing the areas Carnegie Hill, Little Italy, Midtown South and Unique Square, which are marked on the map, are represented by darker shades. Census tracts without income data are shaded in grey and colour breaks are created using k-means clustering algorithm.

5.3 Methods

5.3.1 Creating feature vectors

In Chapter 2 we discussed how CNNs can identify important features directly from the raw image data for various computer vision tasks including object detection and scene classi-

fication. Apart from being trained or tuned for specific needs, we have provided example studies showing how CNNs trained on a specific dataset can efficiently be used as feature extractors on a different dataset.

In the first part of this chapter, in order to train a detector to identify pictures of food, as in Chapter 4, we use the output from the penultimate layer of a network that was initially trained on the ImageNet dataset. Here, we utilise the entire network including the fully connected layer which was previously discarded. The output from this final layer then goes through a softmax function that computes a score between 0 and 1 denoting how likely it is that a given image belongs to one of the 1000 ImageNet categories the CNN has initially been trained on. If a category's score is closer to 0, then the category is less likely to be linked to a given image, whereas the closer the score is to 1, the more likely it is that the category is related to the input image. The final output of a CNN is therefore a 1000 dimensional vector formed of scores per category computed by the softmax function. By using scores generated in the softmax layer, for a given input image, we create a 1000-dimensional feature vector where each dimension represents an ImageNet category. For each image we create a feature vector where individual features are ImageNet categories. Figure 5.6 provides a visual example of how we create feature vectors of the pictures. Due to licensing restrictions, we use photographs uploaded to the photo sharing platform *Flickr* to illustrate our methodology. In Figure 5.6, each bar chart shows ImageNet categories with the three highest and three lowest scores returned by the CNN. Higher scores indicate that the corresponding category is better at describing the given image compared to categories with lower scores.

In order to get a better holistic understanding of the scene captured in an image, we need to extract further information than the existence of individual objects that is what we extract by using a CNN trained on the ImageNet dataset. As a second set of features, we therefore exploit a CNN which has been trained on the Places-365 Standard dataset. More information on these benchmark datasets are provided in Chapter 2. Using this new CNN, for each image we create a 365 dimensional feature vector where each feature is a category from the Places-365 Standard dataset.

Finally, to further broaden the scene representation, we use a CNN trained on the SUN attribute dataset to create 102-dimensional feature vectors. To summarise, in this chapter, we exploit three different sets of feature vectors formed by the categories from the ImageNet, Places-365 and SUN attribute datasets each capturing different aspects of the images provided as an input.

5.3.2 Training a classifier to recognise pictures of food

In order to analyse whether there is a relationship between food pictures posted on the photo sharing website *Instagram* and the restaurant reviews shared on the crowdsourcing platform *Yelp*, we first need to identify food-related pictures. To extract food pictures on *Instagram* automatically, as in Chapter 4 we build a CNN based classifier. When training

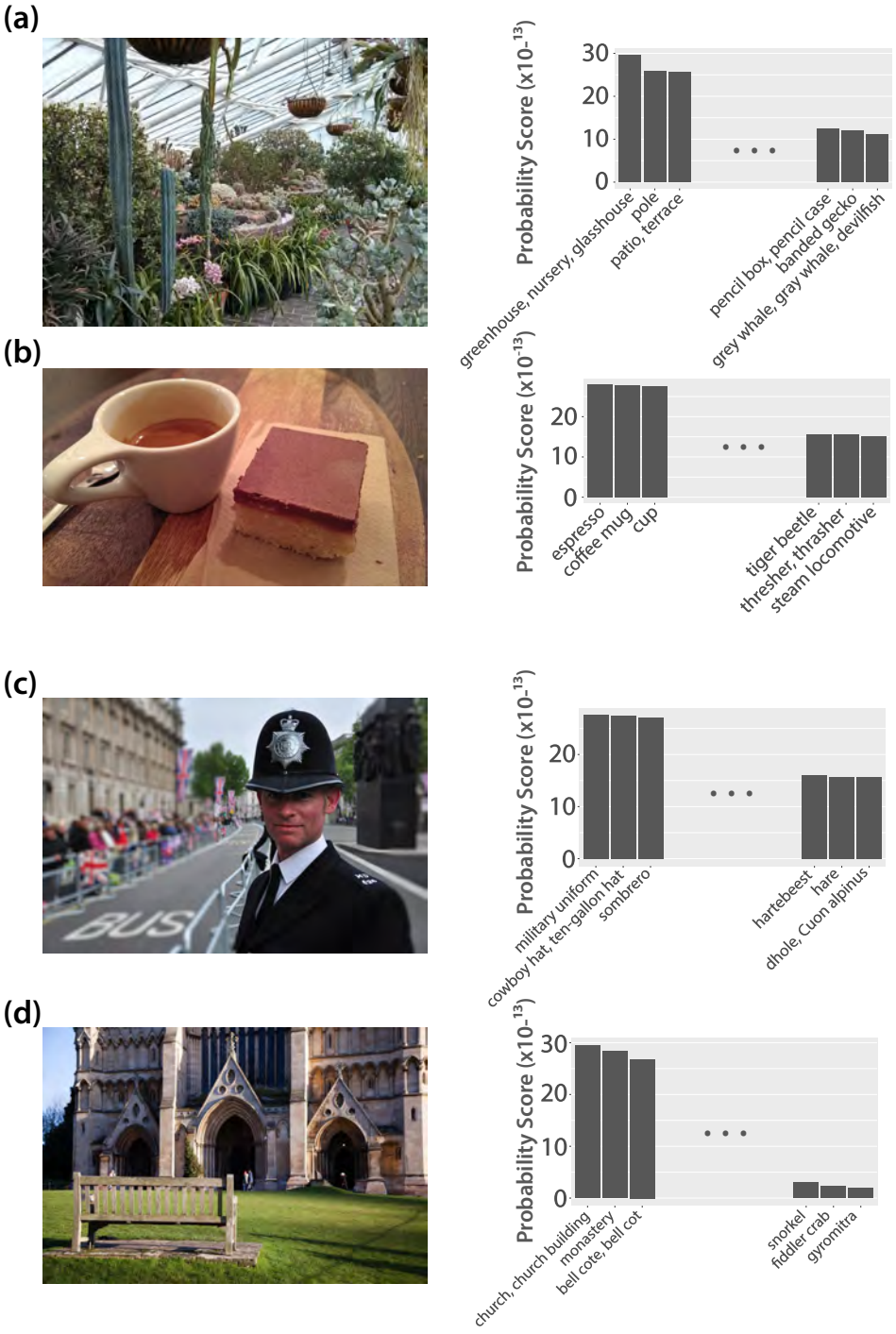

Figure 5.6: Sample images with their feature vectors.
We create feature vectors for the sample images downloaded from *Flickr* by using probability scores extracted from the VGG-M-128 model which has initially been trained on the ImageNet dataset. Bar charts depict three most and least likely categories that the CNN has grouped the pictures into. To enhance the visibility of the inter category probability variations, we rescaled probability values to $[0, 10^{-13}]$. Picture credits are provided in Appendix C.1.

a classifier, we need to provide a set of positive training images, which in our case are pictures of food and a separate set of negative training images, which is a random set of pictures that are not food-related. Utilising these positive and negative training sets, the classifier can then learn to identify similarities and differences between objects within the same class while distinguishing variations between different classes. We should note that we define food-related pictures to be pictures of either food or drink.

It is however non-trivial to find annotated open datasets that suit individual image analysis problems. Following a similar approach as in Chapter 4, we therefore create an automated search to create both positive and negative training sets. The search query we use to form the positive training set consists of a country name followed by the word "food"; for instance "England food". On the contrary, in order to retrieve results that are not indexed as being food-related by the search engine *Bing*, we form a search query by using the country name followed by the phrase "NOT food". We iterate over the entire set of countries downloading the top 150 search results for each query. The full list of countries used in this analysis is provided in Table B1 in Appendix B. This process lets us create a positive training set with 20 658 pictures and a negative training set with 19 065 pictures. For both the positive and negative training sets, we then create feature vectors by extracting the output from the pretrained network's penultimate layer.

The number of training images we have here is not sufficient to train a CNN from scratch which would require millions of training images to achieve high performance as well as to avoid overfitting. However, as discussed in Chapter 2, owing to their adaptive nature, pretrained CNNs can be used to extract features that can then be passed to train a classifier such as an SVM (Chatfield et al., 2014; Girshick et al., 2014; Sharif Razavian et al., 2014). This type of hybrid learning does not require training sets as big as those required to train a network from scratch while still exploiting the advantages of using a deep net to create powerful image descriptors (Donahue et al., 2014). Here, we use the output from the penultimate layer of VGG-M-128 as feature vectors which will be fed to an SVM with a linear kernel.

In order to test the performance of the food classifier, we create a separate test set by downloading pictures taken and uploaded to *Flickr* in 2014 with the word "food" included in either the picture title, tag or description. We also extract a set of images again taken and uploaded to *Flickr* in 2014 but without the word "food" attached. We then manually go through both sets to eliminate pictures that appear to be under the wrong category. After the cleaning, in the end, we get 1 000 positive and 1 000 negative test images.

We test the performance of the classifier over this set of test images. The classifier performed with 95% accuracy and 98% precision on the test set detecting 910 of the 1 000 food-related pictures while correctly labelling 985 out of 1 000 non-food-related pictures. Figure 5.7a summarises these performance results where correct classifications are represented by blue and misclassifications by red. We evaluate the performance of the classifier using a ROC curve by tuning its sensitivity over the results of a 5-fold cross validation on the training set.

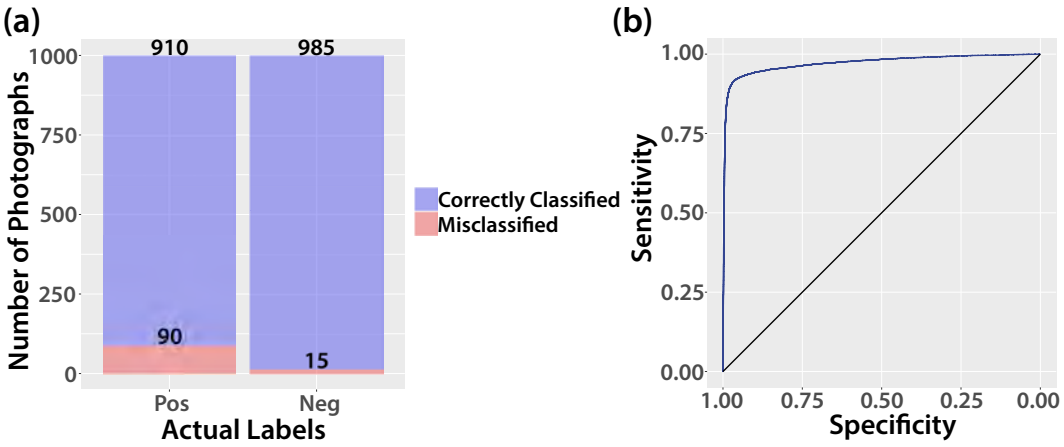

Figure 5.7: Evaluating the performance of the food classifier.
(a) Performance summary of the food classifier over the test set with 1 000 positive and 1 000 negative images. The classifier correctly identified 910 pictures as food-related. Less than 10% of the food pictures were not detected. With only 15 false positives, the classifier correctly identified almost 99% of the negative test images. **(b)** ROC curve of the food classifier created with 5-fold cross validation of the training set. The classifier performs with an AUC value of 0.972 ± 0.002 which is considerably better than the performance that would be expected when randomly categorising an image as food-related or not which would have an AUC of 0.5 and is represented by the black line in the figure.

As illustrated in Figure 5.7b, the classifier performs much better than randomly labelling images as food-related or not, yielding an AUC value of 0.972 ± 0.002. Randomly-labelled images would be expected to have an AUC of 0.5 which is represented by the black line in the figure. Finally, we run the classifier over the entire set of *Instagram* pictures which automatically categorises 781 664 pictures as being food-related.

5.4 Analysis and results

5.4.1 Quantifying the relationship between food pictures and restaurant ratings

To investigate the relationship between food pictures shared on *Instagram* and restaurant reviews posted on *Yelp*, we need to match the pictures of food with the restaurants they were taken at. Thus, as described in Section 5.2.1.3, we overlay these two sets of data by comparing *Instagram* location names with *Yelp* restaurant names.

We first analyse whether there is a link between the number of food pictures taken at a certain restaurant and the number of reviews *Yelp* users leave for that specific restaurant. We find that a greater number of food-related photos taken in a restaurant corresponds to a higher number of user reviews for that restaurant posted to *Yelp* (τ = 0.301, $p < 0.001$, N = 4 035, Kendall's rank correlation).

We also examine whether there is a relationship between the number of pictures and restaurant ratings. We find that there is a significant yet weak correlation between the number of pictures of food taken at a restaurant and the restaurant's rating, such that restaurants where more photographs of food have been posted tend to receive higher ratings (τ = 0.067, $p < 0.001$, *N* = 4 035, Kendall's rank correlation). These initial analyses suggest that higher volumes of food-related pictures posted from a specific restaurant might be used as an indicator for the number of reviews about that restaurant posted on *Yelp*, although we note again that the relationship between food-related pictures and restaurant ratings is particularly weak.

However, this preliminary investigation is based solely on analysing the number of food-related pictures posted to *Instagram*. No consideration is made of the kind of food that *Instagram* users have chosen to photograph. We therefore extend our analysis to examine whether we can uncover a potential relationship between the content of the pictures and the restaurant ratings.

The first step in any image analysis task is to find a suitable image representation. In this section, as image representations, we use the output from the pretrained CNN based on the VGG-M-128 architecture. For each image, by using the scores returned by the CNN, we therefore create a 1000-dimensional feature vector where each feature dimension represents a category.

We then group pictures with respect to the restaurants they were taken in and calculate the mean score per category. Hence, we create one feature vector per restaurant that contains a mean score for each one of the 1 000 features that are indeed the ImageNet categories.

To examine whether we can estimate a restaurant's rating by using feature vectors created from food-related pictures, we build an elastic net model by using restaurant ratings as the output variable and individual features as predictors. We fit a model by using data from all but one restaurant. With the fitted model, we estimate the rating of the excluded restaurant. We repeat the same process to create estimated ratings for each one of the 4 035 restaurants. We should note that, since *Yelp* collects ratings and reviews for individual branches separately, we too consider each branch of the same restaurant individually.

In order to analyse how well our estimated restaurant ratings coincide with the actual ratings on *Yelp*, we calculate the correlation between the estimated ratings and actual ratings. In addition, to measure how much of the variance in the restaurant ratings is captured by the changes in feature vectors, we calculate R^2 statistics as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N} (y_i - \overline{y})^2} \tag{5.1}$$

where y_i is the actual restaurant rating and $\hat{y}_i$ is the estimated rating of the i^{th} restaurant as determined by the elastic net, *N* is the number of restaurants, and $\overline{y}$ is the mean restaurant

rating.

The initial investigation of performance suggests the existence of a weak yet significant signal of restaurant ratings from the content of the food-related pictures (τ = 0.247, $p < 0.001$, N = 4 035, Kendall's rank correlation). However, we find that only 9% of the variance in the actual restaurant ratings can be explained by the changes in the feature vector (R^2 = 0.094).

Not all restaurants in our processed dataset have large numbers of photographs uploaded to *Instagram*. We therefore introduce a threshold to eliminate restaurants with fewer pictures, in order to check whether exclusion of these restaurants helps improve the estimates. We test the performance of our approach for threshold values between 10 and 90 photographs. In addition, we also try setting the threshold value to 7 pictures, which is the median number of pictures per restaurant, and 33 photographs, the mean number of pictures per restaurant. Figure 5.8 depicts the change in the correlation between the actual and estimated restaurant ratings as we tune the threshold value.

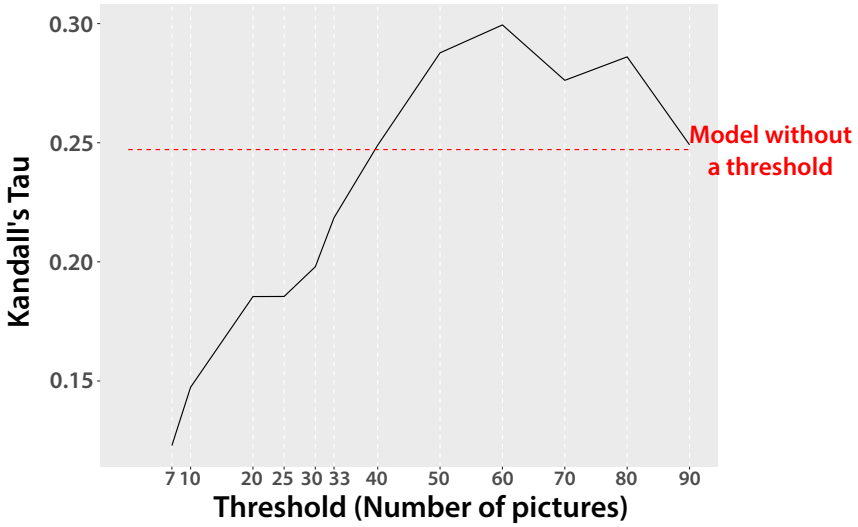

Figure 5.8: Change in the correlation between actual and estimated values with respect to the threshold.
Not all restaurants in our processed dataset have large numbers of photographs uploaded on *Instagram*. We therefore introduce a threshold, which is the minimum number of food-related pictures shared from a restaurant, to eliminate restaurants with fewer pictures to analyse whether exclusion of these restaurants helps improve the estimates. The model including all restaurants before introducing a threshold is represented by the red dashed line. The model achieves its best performance when the threshold value is 60 images.

When we first introduce the threshold, the model performs worse than the model including all restaurants which in Figure 5.8 is represented by the red dashed line. The performance stays below the initial level as we increase the threshold value. However, after the threshold exceeds 40 photographs, new models start to outperform the initial model that includes all restaurants. The best model is observed when the threshold is

Table 5.1: Change in the performance of the elastic net model as we adjust the lower limit of the number of food-related photographs shared on *Instagram* per restaurant.

Increasing the threshold means that fewer restaurants are taken into consideration, as shown under the column *N*. We first compute estimated ratings without imposing a lower limit on the number of pictures per restaurant. The ratings estimated by the model including all restaurants have a relatively strong link to the actual ratings, as listed under τ. As we increase the threshold, the model performance keeps decreasing until the threshold is 40 pictures. The model reaches the best performance when the threshold is equal to 60 photographs, which is highlighted in bold. At this point data loss becomes beneficial when the underlying signal coming from the remaining restaurants in the analysis is stronger than the noise. As we keep increasing the threshold beyond 60 pictures, the number of restaurants used in the analysis decreases as does the accuracy of estimations. Finally, the highest threshold performs very similar to the initial model where we considered the entire restaurant data. *p* values have been FDR corrected to account for the fact that multiple correlation analyses have been carried out.

Threshold	*N*	R^2	τ	*p*-value
0	4 035	0.094	0.247	< 0.001
7	2 100	0.047	0.024	< 0.001
10	1 789	0.036	0.147	< 0.001
20	1 180	0.057	0.185	< 0.001
30	897	0.060	0.198	< 0.001
33	841	0.077	0.219	< 0.001
40	724	0.101	0.249	< 0.001
50	614	0.143	0.288	< 0.001
60	**527**	**0.162**	**0.299**	$\mathbf{< 0.001}$
70	449	0.122	0.276	< 0.001
80	403	0.130	0.286	< 0.001
90	353	0.097	0.249	< 0.001

equal to 60 images (R^2 = 0.16) where the link between estimates and actual values is the strongest (τ = 0.30, $p < 0.001$, *N* = 527, Kendall's rank correlation). As we increase the threshold, the model performance declines. However, models built with data from fewer restaurants above the threshold of 60 pictures still provide better estimates compared to the initial model using all restaurants. Table 5.1 summarises the change in the performance of the model as we adjust the minimum required number of pictures taken per restaurant. The table also includes the number of remaining restaurants used in the analysis after introducing a threshold on the number of pictures shared from a restaurant.

To summarise, we exploit data from 781 664 food pictures and 16 372 restaurants to uncover the relationship between photographs of food and restaurant reviews. We start our analysis at restaurant level by merging image and restaurant datasets to examine the link between the activity on *Instagram* and customer ratings.

Our initial results suggest that the number of food-related *Instagram* photographs taken at a particular restaurant can be an indicator of the number of reviews posted on *Yelp* about that restaurant. A similar relationship exists between the number of food-related pho-

tographs and the restaurant's rating. However, the correlation is much weaker compared to the link between the amount of food-related pictures and the number of *Yelp* reviews. This suggests that restaurants where a higher number of food-related pictures are shared do not necessarily have a higher rating on *Yelp*.

We then include information on image content to investigate whether we can estimate a restaurant's rating given the photographs of food taken at that restaurant. Our findings suggest that information extracted from automatic analysis of the image content can create estimates that are line with the actual restaurant ratings. We show that the estimates of ratings can be improved by limiting the model to just those restaurants that have a certain number of food-related photographs shared on *Instagram*. However, we also find that this minimum number should be carefully chosen as it might lead to a loss of important information while enhancing the noise, which may explain the poor performance of the first set of threshold values tested.

These results are in line with the initial hypothesis that photographs of food uploaded to *Instagram* can give us insights into the rating of the restaurant where the pictures were taken. We provide a first example of how online photographic data can be used to gain insights into the behaviour of restaurant customers in a metropolitan city. Our findings highlight the potential of the content shared on *Instagram* to be used as an additional feedback mechanism for businesses especially in cases where people prefer to reflect their opinion on social media channels rather than crowdsourcing platforms. Building on our results, future work might examine whether it is possible to predict the variation in customer opinion as well as ratings by exploiting a longitudinal dataset of restaurant ratings and reviews where there is more comprehensive information on the change of feedback over time.

5.4.2 Estimating household income for London using photographs shared on *Instagram*

In the first part of our study, we analyse photographs of food from *Instagram* taken at restaurants in London to gain insights into the restaurant ratings on *Yelp*. However, these online pictures might reveal further information about the area in which they were taken.

Previous studies, including the work cited under Section 5.1, have demonstrated that online pictures can be utilised to infer socioeconomic statistics (Arietta et al., 2014; Naik et al., 2017). For instance, Gutiérrez-Roig et al. (preprint) demonstrate that the spatial income distribution of several cities around the world can be inferred by analysing *Google Street View* images using convolutional neural networks. Similarly, by automatically identifying characteristics of vehicles present in *Google Street View* images, Gebru et al. (2017) generated estimates of demographics and socioeconomic statistics including race, education and income from 200 cities in the US. In this section, motivated by these examples, we extend our previous work to investigate whether photographs of food uploaded to *Instagram* within London can be used to infer key socioeconomic measurements, namely

income. For this analysis, we exploit the MSOA-level median household income per London area released by the Greater London Authority.

For each food-related picture shared on *Instagram*, following the approach explained in Section 5.3.1, we first create a feature vector by using the VGG-M-128 architecture trained over ImageNet. We group these pictures and the corresponding feature vectors with respect to which MSOA they were taken. For each MSOA, we then create one feature vector formed of the mean scores per ImageNet category. In order to compute estimates of household income, we build an elastic net model to compute income estimates.

MSOA level income data shows a log-normal behaviour with a positive skew, like income values in general. However, the underlying model in the elastic net we utilise is linear regression which requires output variables to follow a normal distribution. Before fitting the model, we therefore log transform income values so that they are heuristically close to normal distribution and will not violate the normality of residuals assumption of linear regression in addition that we do not predict negative income values.

Setting log transformed income as the output variable and individual features - i.e. ImageNet categories as predictors - we fit an elastic net model by leaving one MSOA out. In order to compute the estimated income of the omitted MSOA, we pass its feature vector to the fitted elastic net model and take the inverse logarithm of the output to revert the values back from the log-scale. We then repeat the same procedure for all of the 983 MSOAs located within the London area in order to create an estimated income value for each MSOA.

To evaluate the performance of our approach, we compare the estimated income values with the actual income values by calculating the correlation and R^2 statistics. Even though the changes in the features extracted from pictures of food can only capture 8% of the variance in the income values ($R^2 = 0.08$), our results hint at the existence of a significant relationship between our estimations and the actual income values (τ=0.224, $p <$0.001, *N*=983, Kendall's rank correlation).

In order to gain a deeper understanding of how individual features affect the computation of income estimates, we fit an elastic net model using data from the entire set of MSOAs and analyse the coefficients of each predictor. Our initial analysis suggests that this model retains 270 out of the 1 000 features when estimating income values. Figure 5.9 depicts categories with the ten largest positive and ten largest negative coefficients. These initial findings show that if a feature vector has higher scores for the categories represented by blue, then the corresponding picture is likely to be taken in a higher income area. Conversely, if a feature vector contains higher scores for the categories shown in red, then the corresponding picture was most likely taken in a lower income area.

Initial comparison of these coefficients also suggest that pictures taken in higher income areas tend to be related with more decoration oriented categories such as *tray*, *vase* or *flowerpot* whereas pictures from lower income areas likely to be linked to food-related categories such as *carbonara*, *hot dog* or *mashed potatoes*. Categories with higher

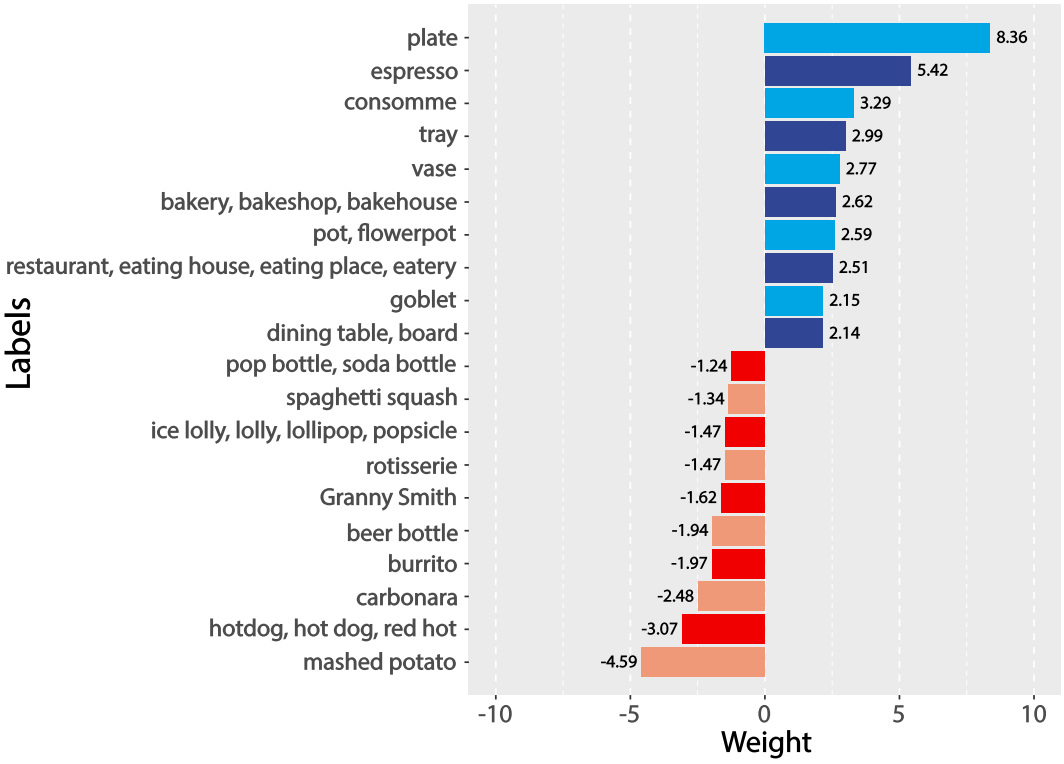

Figure 5.9: Features with ten largest and ten smallest coefficients.
In estimating the income, scores of the features depicted in blue have a positive contribution whereas scores of the features shown in red have a negative contribution. This means that if a picture has higher scores for the blue categories then it is more likely that it was taken in a higher income area whereas if a picture has higher scores for the red categories it is more probable that it was taken in a lower income area. In addition to this, the top ten features appear to be related more to the surrounding objects such as *vase* and *flower pot*, while the bottom ten features appear to be more food-related such as *hot dog* and *mashed potatoes*.

coefficients include eateries such as *bakery* and *restaurant* whereas the *grocery store* category has a very high negative coefficient indicating that pictures capturing grocery stores are more likely to be taken in lower income areas. The patterns visible from this initial analysis might indicate that food-related pictures taken in higher income areas do not only focus on food but also focus on what is surrounding the food while pictures taken in lower income areas are mainly picturing the food. All in all, these results highlight the different characteristics of food-related photographs taken in areas with different household income.

As discussed earlier, analysing pictures of food for estimating economic indicators such as income is a promising idea since food-related photographs may contain visual cues about the area in which they were taken. However, food-related pictures constitute less than 12% of our entire dataset. Thus, it is quite likely that we lose a considerable

amount of information hidden in the rest of the pictures that might be valuable to better represent characteristics of the neighbourhoods where the pictures were captured. We therefore include the entire set of *Instagram* pictures to investigate whether we can create more powerful models identifying the income variations between MSOAs in London.

For each MSOA, we first update feature vectors by calculating the mean probability per category using the entire *Instagram* dataset. We then calculate estimated income values again by building an elastic net model. We call this model the "ImageNet model" since to form the feature vectors, we use a CNN which has been trained on the ImageNet dataset. The initial results indicate that in comparison to the model using solely food-related pictures, the ImageNet model built on the entire set of *Instagram* pictures is better at explaining the variations in the actual income values (R^2 = 0.28) and the relationship between the estimated and actual income values is much stronger (τ = 0.392, $p < 0.001$, *N* = 983, Kendall's rank correlation).

The ImageNet model shows that extracting information about the objects present in a picture can provide key observations into the income of the area where the picture was taken. However, to gain further understanding about the area, we need additional scene features that can give us more holistic insights into the image. We therefore investigate whether changing the nature of the underlying dataset, which the CNN has been trained on, can provide us with more information to create better neighbourhood descriptors, and hence better income estimates.

Using the same approach as explained above, we extract a new set of features by using a CNN trained on the Places-365 Standard dataset. For each image, we extract a 365 dimensional feature vector with the corresponding scores for each category under the Places-365 dataset. We then group these feature vectors to create one feature vector per MSOA with the mean scores per category. We call this model the "Places model". Initial results suggest that in comparison with the ImageNet model, feature vectors used to build the Places model are better at modelling the change in income values between MSOAs (R^2 = 0.31) with estimations more closely linked to the actual values (τ = 0.410, $p < 0.001$, *N* = 983, Kendall's rank correlation). Less than 9% of the Places-365 categories are regularised by elastic net when estimating the median income values per MSOA.

We then investigate whether we can further identify the distinct properties of the images by combining the ImageNet and Places models to exploit their different characteristics. For each image, we therefore combined feature vectors from the ImageNet and Places models to create a higher dimensional vector with 1 365 features and again group them with respect to MSOAs. Using this new set of extended features, we create a new model which we refer to as the "ImageNet+Places" model.

We find that the combined ImageNet and Places features serve as better predictors for estimating income compared to using these features separately (τ = 0.429, $p < 0.001$, *N* = 983, Kendall's rank correlation) By exploiting the discriminative characteristics of both the ImageNet and Places models, the ImageNet+Places model captures 34% of the variance of the median income at MSOA level (R^2 = 0.34).

Table 5.2: Performance scores for the elastic new models created using different feature sets.

Models combining two or more different sets of features perform better than the models using only one set of features. The ImageNet+Places model, highlighted in bold, is the best performing model with the strongest link between the estimated and actual income values.

Model	R^2	τ	*p*-value	Nonzero-Coefficients
ImageNet	0.28	0.392	< 0.001	719
Places	0.31	0.410	< 0.001	348
ImageNet + Places	**0.34**	**0.429**	$\mathbf{< 0.001}$	**887**
SUN	0.14	0.277	< 0.001	76
Places + SUN	0.32	0.416	< 0.001	259
ImageNet + SUN	0.27	0.388	< 0.001	552
Combined	0.33	0.428	< 0.001	347

We also investigate whether we can capture different aspects of scenes by introducing a different set of scene features. As discussed in Chapter 2, the SUN Attribute Dataset has initially been created to complement the SUN Places database which the categories under the Places Database are based on. We therefore create a new set of vectors with 102 features using a CNN trained on the SUN attributes dataset. Before combining with the Places model to check whether these two sets of features will complement each other and give better income estimates, we build a model using solely the SUN features which we call the "SUN model". Initial performance analysis indicates that the SUN model cannot capture the variance in household income across London ($R^2 = 0.14$) as much as the ImageNet+Places (R^2 = 0.34), Places (R^2 = 0.31) or the ImageNet (R^2 = 0.28) models though there is a significant correlation between the estimated and actual income values (τ = 0.277, $p < 0.001$, *N* = 983, Kendall's rank correlation).

We then examine whether we can better capture scene characteristics by constructing an extended set of features exploiting features from both the SUN and Places models. We build a new model using these combined feature vectors which we call the "Places+SUN model". As expected, this combined model performs better (R^2 = 0.32), compared to the Places and SUN models creating better income estimates (τ = 0.416, $p <$ 0.001, *N* = 983, Kendall's rank correlation).

We also create a separate set of features by combining ImageNet categories and SUN attributes calling the new combined model the "ImageNet+SUN model". There is a significant link between the estimated income values calculated by the ImageNet+SUN model and the actual household income (τ = 0.388, $p < 0.001$, *N* = 983, Kendall's rank correlation) and it performs better than the SUN model (R^2 = 0.27, SUN model; R^2 = 0.14). However, initial performance analysis suggests that, unlike the Places+SUN model, combining SUN features with the ImageNet features leads to a slightly poorer performance (R^2 = 0.27) compared to using solely the ImageNet features (R^2 = 0.28).

The reason for the performance decrease can be described by looking at the co-

efficients. The total of 552 (493 ImageNet plus 59 SUN features) out of 1 102 features are retained in the ImageNet+SUN model to compute the estimated income. Some of the ImageNet features might be correlated with the SUN features which may have caused more than half of the ImageNet features to be regulated by the elastic net. Whereas in the Places+SUN model, the regulated set of features were mainly the SUN features (only 29 out of 102 SUN features have a nonzero coefficient). This comparison also suggests that the SUN features work relatively more harmoniously together with the Places features.

Finally, we combine SUN attributes with ImageNet and Places categories, to create an extended vector with 1 467 features in total and once again build an elastic net model which we call the "combined model". Having computed better estimates (τ = 0.428 ,$p <$ 0.001, *N*=983, Kendall's rank correlation), the combined model outperforms the ImageNet, Places and SUN models built on the individual set of features (Combined model; R^2 = 0.33, ImageNet model; R^2 = 0.28, Places model; R^2 = 0.31, SUN model; R^2 = 0.14). Both the correlation between the estimates and actual household income and the R^2 values of the combined model and the ImageNet+Places model are very similar. However, the ImageNet+Places model has a slightly better performance making it the best performing model among the seven models we have generated. The performance measures calculated for each model are summarised under Table 5.2.

These results indicate that the SUN model is the worst performing model among the models we inspect. This might be explained by two factors. First of all, among the feature sets we build the models on, the SUN model has the least number of features. The other reason might be the nature of the dataset. The SUN attribute dataset was designed to complement scene categories in an attribute-space, which unlike category space has no distinct boundaries between the different attributes (Patterson et al., 2014). Considering all these dataset specific features, it is not surprising that the SUN model performs worse than the rest, whereas the Places model when combined with the SUN attributes benefits from a performance increase which supports the suggestion that the SUN attributes work well as a complement to category based feature vectors.

Although the Places+SUN model enhances the performance in comparison to the Places and the SUN models by broadening the place-based information, the model is missing object specific information since both Places and SUN features focus on characterising the scenes. The ImageNet+Places model which combines both object-specific and scene-specific information therefore computes income estimates that are better in line with the actual household income. On the other hand, the ImageNet and Places models individually perform reasonably well. However, they both are outperformed by the models with combined feature vectors. Our results therefore suggest that various aspects of diverse neighbourhoods can be captured better when we have a broader scene understanding which can be achieved by combining different image representations. However, we should also be cautious about adding too many features carrying similar information. This will add correlated predictors to the model and hence will be regulated by the elastic net. In short, the intuition that as we add more information, we will get better estimates is not always true

as we have observed in the combined model example.

In order to visually inspect whether these models are capable of capturing the income patterns in London, we visualise the estimated values per MSOA on a London map. Figure 5.10 depicts the spread of the median income estimates for the models we create in comparison to the actual household income. For easier comparison, we plot the ranked values for both the actual and estimated income on a scale $[1, 983]$ ranked in an increasing order where 1 corresponds to the lowest income value. Visual inspection highlights that despite the noisy appearance of the patterns in the maps, all models except the SUN model are successful in capturing the high income areas around central London as highlighted in Figure 5.3. The map with the actual income values is highlighted with a purple box and the best performing model is framed with a red box.

To better identify the strengths and weaknesses of these models, we take a closer look at the ImageNet+Places model, which is the best of the seven models in terms of the relationship between actual and estimated income values as well as the amount of the variance in the actual income values that it can capture. Figure 5.11 demonstrates actual and estimated income values as well as the difference between these values. For easier comparison, we again plot the ranked actual and estimated values. Having computed the correlation between the actual household income and the estimation error, we observe that the model tends to underestimate higher income values, whereas it overestimates the lower income values (τ = 0.56, $p < 0.001$, Kendall's rank correlation). Areas with overestimated income values are shaded in blue and the underestimated areas are represented by red, see Figure 5.11. Visual inspection suggests that despite capturing the general income characteristics, the estimations do not fully capture the values at both ends of the income spectrum; i.e. we fail to model the skewness of the target distribution. Underestimated income values follow a similar pattern as of high income areas while MSOAs with lower income and MSOAs which the model overestimated the household income are very much alike.

Finally, we investigate the effect of the individual features by analysing the coefficients of the ImageNet+Places model. Figure 5.12 shows ten largest positive and ten largest negative coefficients of the model. Twelve out of twenty of these coefficients are for the Places features. These coefficients suggest that pictures with high scores on categories as *beer garden*, *living room* or *church* tend to be taken in higher income areas while pictures with high scores of categories *loading dock*, *slum* and *industrial area* are more likely to be taken in parts of London with lower income. Considering these coefficients, we can conclude that ImageNet and Places categories which may come across as pleasant have higher coefficients, in contrast categories which appear to be less attractive get the largest negative coefficients.

However, CNNs trained for multi-class classification problems do not perform with the same accuracy across categories (Chatfield et al., 2014). Moreover, if they are used on a different dataset than the dataset they were initially trained, they tend to perform better. For instance, they detect objects which appear to be similar in both datasets. In Crowley

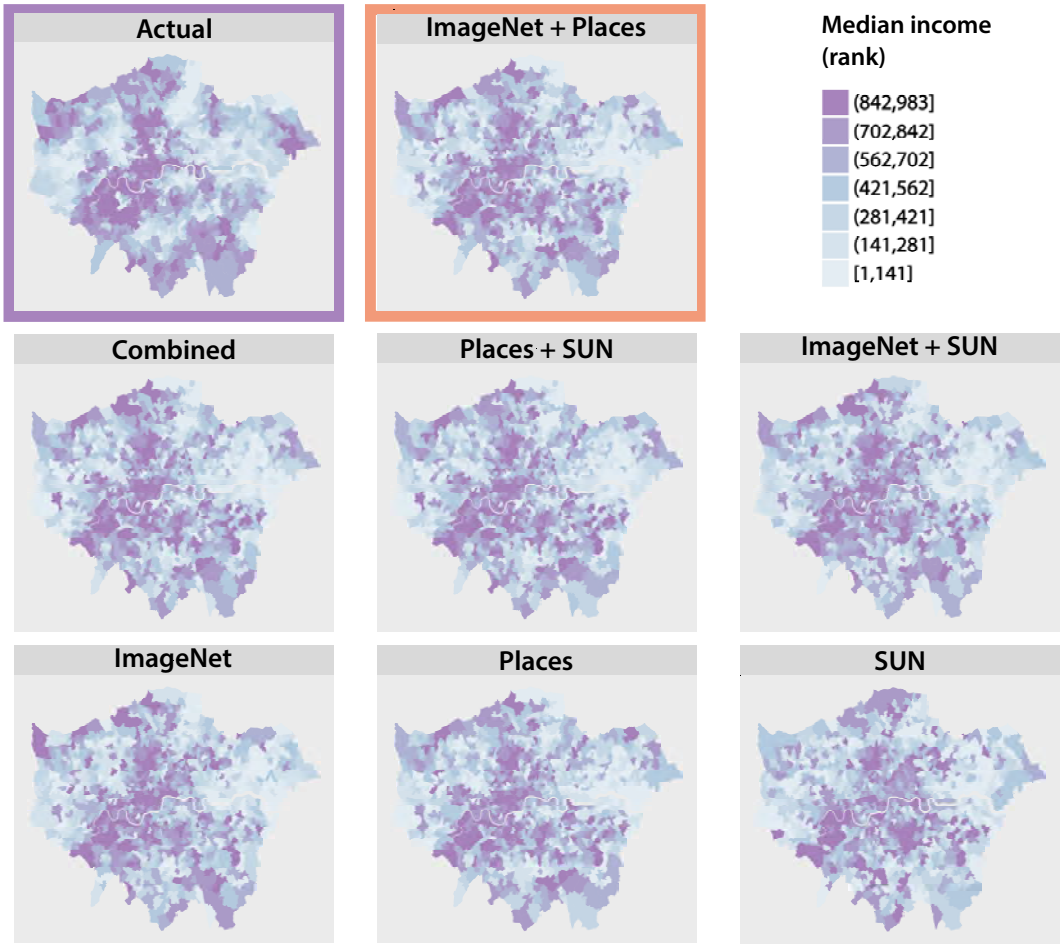

Figure 5.10: Actual and estimated income patterns across London.
Most of the models successfully capture the high income areas clustered around the city centre. The map showing the actual income values is framed with a purple box while the best performing model, the ImageNet+Places model, is highlighted with a red frame. Equal breaks are calculated for the ranked income values.

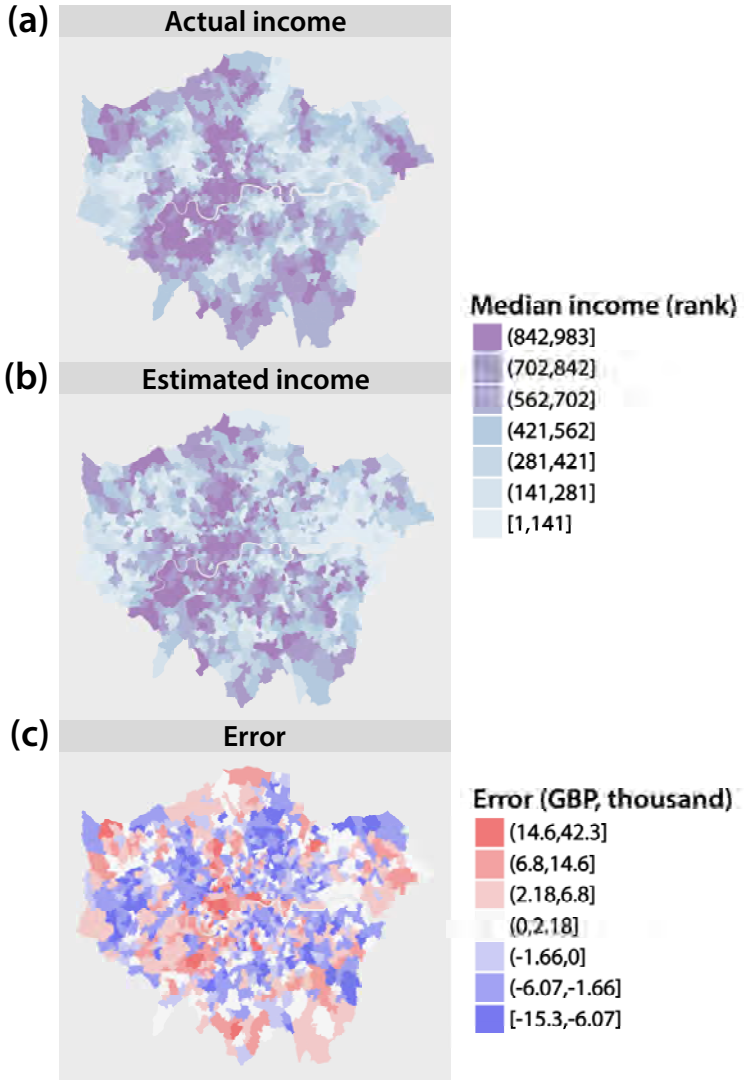

Figure 5.11: Comparison of the actual and estimated income values computed by the ImageNet+Places model.
(a) Map depicting the distribution of actual household income. **(b)** Map showing the distribution of the income values estimated by the ImageNet+Places model. The model captures the high income areas especially around central London stretching west along the River Thames. **(c)** Map highlighting the difference between actual and estimated income values per MSOA. Error values are calculated by subtracting the estimated values from the actual income values. The values underestimated by the model are shaded in red whereas the values overestimated by the model are filled with blue. Visual comparison of (a) and (b) suggests that the model is capable of capturing high-low income patterns when the values are ranked. We then investigate the actual-estimated income difference between the monetary values. Visual inspection suggests that although capturing the general income characteristics, the estimations are not great at both ends of the income spectrum, failing to model the skewness. Underestimated income values follow a similar pattern as of high income areas while MSOAs with lower income and MSOAs which the model overestimated the household income are very much alike. For figures (a) and (b), equal breaks are calculated for the ranked values, whereas for figure (c), we use the *k*-means clustering algorithm to create breaks for the error values.

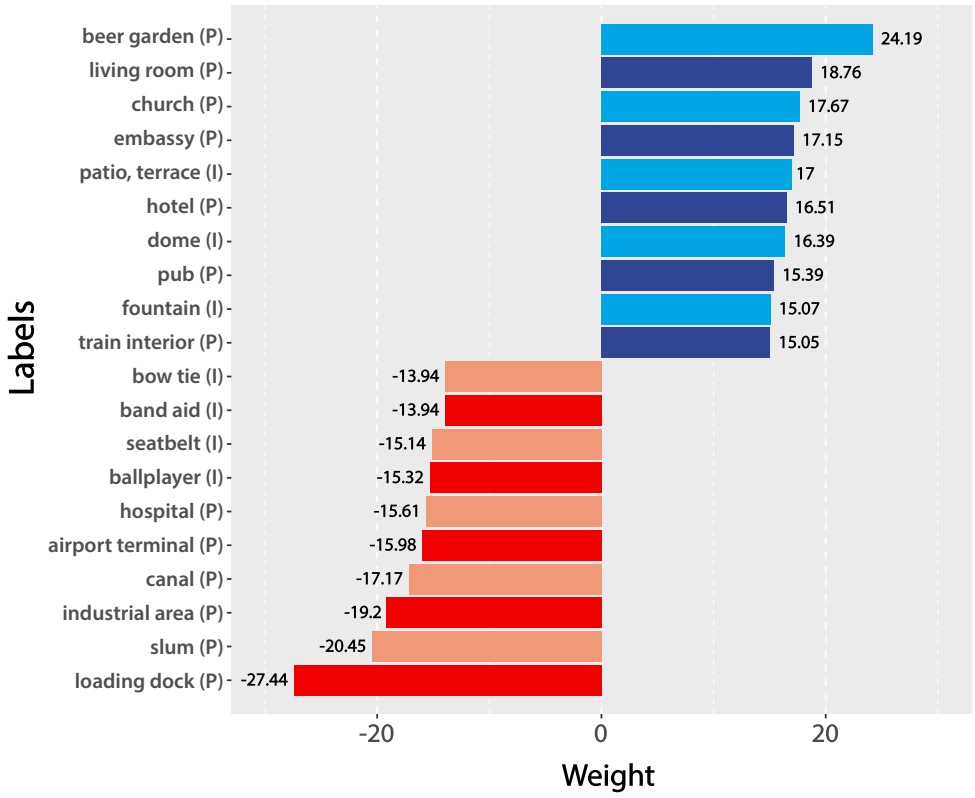

Figure 5.12: Ten largest positive and ten largest negative coefficients of the ImageNet+Places model.
Features with an "(I)" next to them indicate that they are ImageNet categories whilst features with a "(P)" are Places-365 categories. Twelve out of these twenty features are categories from the Places-365 dataset. Categories that may come across as pleasant have higher positive coefficients whereas categories which appear to be less attractive have higher negative coefficients. For instance, if an image has a higher score for categories such as *patio*, *hotel/outdoor* or *fountain*, then it is more likely that the image was taken in a higher income area. On the contrary, if categories such as *industrial area*, *slum* or *loading dock* have higher scores then those images are more likely to have been captured in areas with lower income.

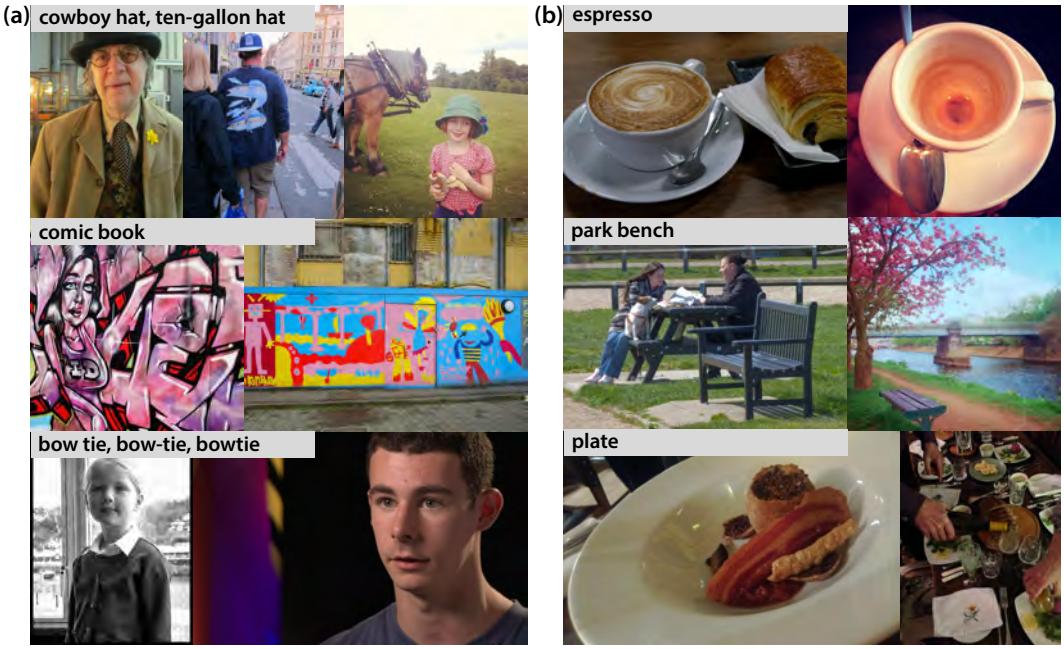

Figure 5.13: Sample *Flickr* pictures and their ImageNet labels generated by the pretrained CNN using the VGG-M-128 architecture.
(a) Sample *Flickr* pictures with unmatched labels. **(b)** Set of *Flickr* pictures with matching labels. Although the CNN fails to label pictures in (a) correctly, visual inspection suggests that pictures labelled under the same category share certain characteristics. For instance, the first set of pictures labelled as *cowboy hat* do not indeed contain a cowboy hat, however they all have a person wearing a hat of some sorts. Picture credits are provided in Appendix C.2.

and Zisserman (2014), the authors argued that the classifier was better at detecting people and horses in paintings since they look similar in natural images and in paintings however, it was more common for the classifier to make mistakes in detecting objects with simple shapes like buttons and wheels. Figure 5.13 shows sample *Flickr* pictures and their corresponding ImageNet labels returned by the the CNN built on the VGG-M-128 architecture. The CNN performs well when detecting *espresso*, *plate* and *park bench* whilst it wrongly classifies pictures as *cowboy hats*, *bow ties* and *comic books*.

Although the CNN fails to label these pictures correctly, it is apparent that photographs given the same label do share a number of characteristics. For instance, images labelled as *bow tie* tend to be portrait pictures, and the *cowboy hat* label has been given to images containing people wearing hats, even though the hat in question is not necessarily a cowboy hat. These examples indicate that we should be cautious when interpreting coefficients and their corresponding labels. Although the CNN is effective in grouping similar pictures into the same category, the labels do not always reflect the content of the category to the extent one might hope.

In this section, we analyse whether can expand our study on estimating restaurant

ratings using photographs of food shared on *Instagram*. By focusing on food-related *Instagram* pictures, we initially investigate whether we can estimate household income across London. Instead of focusing solely on pictures of food, we then further extend our investigation of estimating income to include the entire set of pictures from the *Instagram* dataset.

We build various models by exploiting the features created by using CNNs trained on three different training sets; ImageNet, Places-365 and SUN attribute databases. We demonstrate how using diverse features capturing different aspects of a given image can create different estimates of household income. Our results show that combining these distinct features can help us create better models however we need to be mindful which types of features we are combining. In short, our analyses suggest that *Instagram* pictures taken in London can help us estimate household income at MSOA level. In order to investigate whether we can generalise these results for other cities, in the next section we perform the same analysis on a set of *Instagram* pictures taken in New York City.

5.5 Using *Instagram* photographs to estimate household income in New York City

In the previous section, we have shown how *Instagram* pictures can be used to describe neighbourhood characteristics in order to estimate income in the Greater London area. In this section, we test whether our proposed method would hold for a different metropolitan with different dynamics. We therefore extend our study by analysing *Instagram* pictures taken in New York City to investigate if we can estimate median income at census tract level.

We start our analysis by creating feature vectors for each image by using the scores per category generated by a CNN trained on the ImageNet dataset. In order to create one feature vector per census tract, we then calculate the mean score per category grouping *Instagram* pictures with respect to the census tract they were taken at. Having generated a mean feature vector for each census tract, we again build an elastic net model by setting the logarithmic income values as the output variable.

For each census tract, we fit an elastic net model by leaving that census tract out. We then compute an income estimate for the discarded census tract by using the fitted model together with the feature vector of the census tract which we left out. This let us create income estimates across New York City. We call this model using ImageNet categories the “ImageNet model”. We repeat the same approach by replacing the feature vectors with the categories from the Places-365 Standard Database to create the “Places model”, as well as the SUN attribute database, to create the “SUN model”. As in the London analysis, we also experiment with combining different sets of categories in order to create broader image representations. Table 5.3 summarises the performance measures calculated by comparing the actual and estimated income values.

The models with the best performance are the ImageNet + Places model and the

Table 5.3: Performance scores for different models that aim to estimate the income of New York City at census tract level using information from *Instagram* photographs.

As in the London analysis, models which use feature vectors formed by combining two or more different set of categories perform better than models using only one set of categories. The ImageNet+Places model has the strongest correlation between the actual and estimated income values, whereas the elastic net model combining features from the ImageNet, Places and SUN models has the highest R^2 statistic, though performance does not differ greatly compared to the ImageNet+Places model. Both models with the best performance are highlighted in bold.

Model	R^2	τ	*p*-value	Nonzero-Coefficients
ImageNet	0.14	0.289	< 0.001	569
Places	0.16	0.295	< 0.001	340
ImageNet + Places	0.19	**0.314**	$\mathbf{< 0.001}$	**1007**
SUN	0.02	0.196	< 0.001	77
Places + SUN	0.17	0.296	< 0.001	357
ImageNet + SUN	0.15	0.294	< 0.001	628
Combined	**0.20**	0.313	< 0.001	**785**

combined model where feature vectors contained categories from all three datasets: the ImageNet, Places and SUN attribute databases. Both models capture around 20% of the variance in the median income at census tract level. These findings are in line with the results from the London analysis that differences in the key social measurements, namely income, can be captured better by combining the visual information extracted from online images by the CNNs trained on different datasets.

In order to inspect the similarities between the actual and estimated income patterns spread across New York City, we visualise the actual income per census tract and the income estimates generated by each model (Figure 5.14). We find that the majority of the models are able to identify places with higher household income situated around the Manhattan and Brooklyn area as well as capturing lower income areas located around Bronx. All models however miss higher income areas in Staten Island, with the best model, the combined model, underestimating income values across the census tracts with a median of almost 20 000 dollars.

It is crucial to understand the areas where the models provide better estimates and the areas where models fail to provide good income estimates. We examine one of our two best performing models, the combined model. In Figure 5.15, as a third map in addition to the other two maps depicting ranked actual and estimated household income, we therefore plot the error. We define error as the difference between the actual income and estimated income values for the combined model. Income values of the areas shaded in red are underestimated by the model while income estimates of the blue-shaded areas are higher than the original income values.

Finally, to have a better understanding of the contribution of the individual categories, we investigate the coefficients of the elastic net model. For visualisation, we again

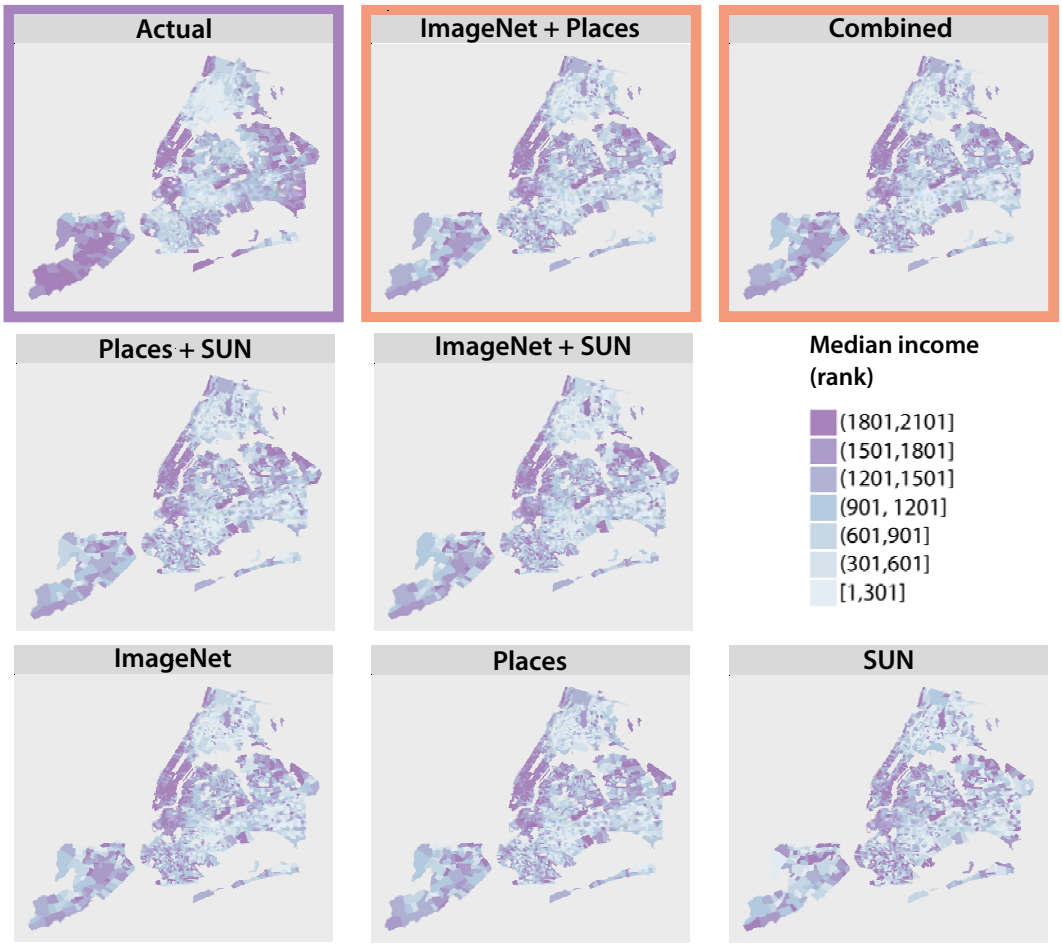

Figure 5.14: Actual and estimated income for census tracts in New York City. The majority of the models successfully capture the high income areas around Manhattan as well as the coastal part of Brooklyn facing Manhattan. On the other hand, all models fail the capture most of the high income areas in the Staten Island region. The map depicting the original income values is framed in purple while models with the best performance are highlighted with a red frame. Colour breaks are calculated with equal values using ranked income values.

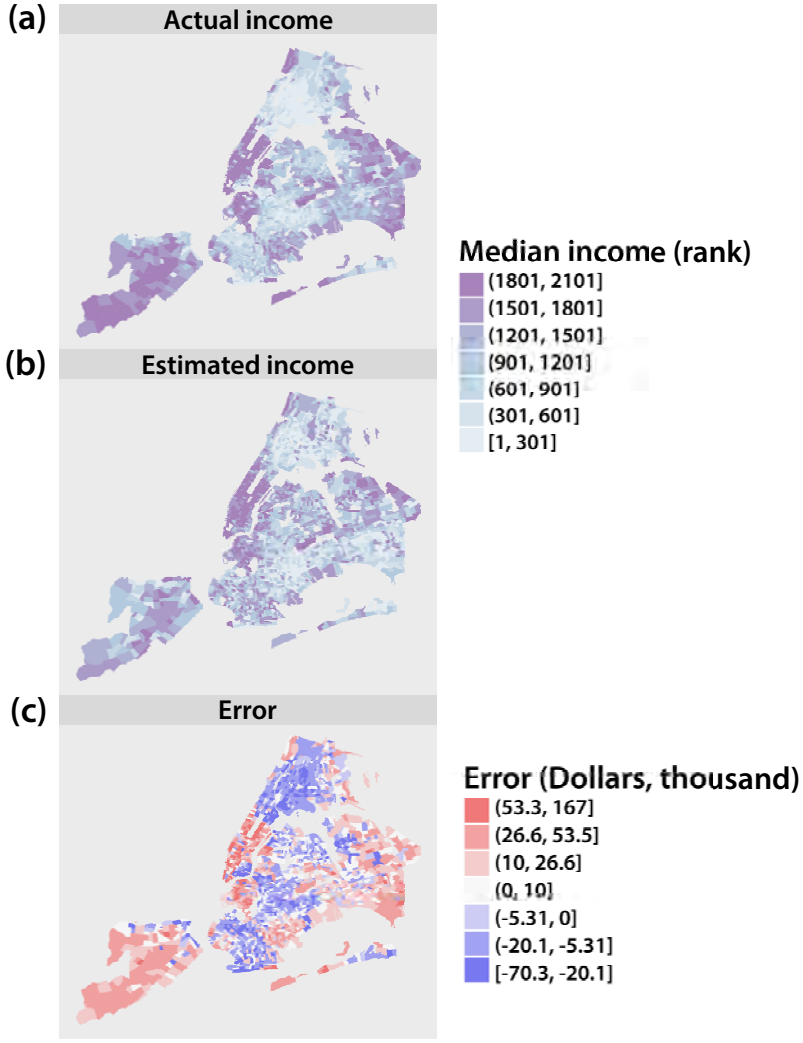

Figure 5.15: Visual comparison of actual and estimated income values computed by the Combined model across New York City.
(a) Distribution of the actual household income across the NYC. **(b)** Income values estimated by the Combined model. **(c)** Map depicting the difference between the actual and estimated household income. Visual inspection suggests the model captures the high income areas around Manhattan and Brooklyn area. However, income values across the Staten Island are mainly underestimated which might be due to the absence of a good set of *Instagram* pictures representing the neighbourhood as Staten Island have fewer pictures compared to the rest of the areas of the NYC.

focus on the combined model. Figure 5.16 depicts ten coefficients with the most positive and ten coefficients with the most negative effect in estimating income. As the combined model uses feature vectors created by using all categories from the ImageNet, Places-365 and SUN attribute datasets, the coefficients listed in the figure are a combination of all three. *Instagram* images with higher scores for the categories *restaurant*, *skyscraper* and *coffee shop* are more likely to be at taken at a higher income area compared to the pictures with a feature vector composed of higher scores for *traffic light* and *hospital room*. Compared to the coefficients of the ImageNet+Places model of London, categories with the largest coefficients extracted in the NYC analysis have certain differences. This could be explained by the distinct dynamics and characteristics of these two metropolitans. For instance, if a picture taken in the NYC has a higher score for the category *skyscraper*, then it is more likely that the picture was taken in an area with higher income. Similar relationship does not hold if the picture was taken in London. However, in London, other categories such as *beer_garden*, *patio* or *pub* suggest that a picture may be taken in a higher income area. The categories with the largest positive coefficients in the models built for London and NYC highlight differences of these two cities. While skyscrapers are one of the most popular and photographed landmarks in NYC, pubs and beer gardens can be more related with the city life in London. On the other hand, categories with negative coefficients have one similar category which is the *hospital* and *hospital room*. Three of the categories with negative coefficients are outfit related; *military uniform*, *sombrero* and *cowboy hat*. It is also worth mentioning again that although these categories group pictures with similar characteristics, as discussed in Section 5.4.2, these labels do not necessarily in line with what the category name suggests.

In this section, we investigate whether we obtain results that are in line with the results we discuss in Section 5.4.2 when we run the same analysis on a different city. Using New York City as our target location, we again build different models exploiting feature vectors created by the CNNs trained on three different image sets. We find that the relationship between the actual income and the income estimated using visual features of the *Instagram* images taken in New York City is significant yet weaker compared to the link between actual and estimated income values from London. This can be explained with the spatial granularity we pick to estimate income values. Specifically, although both *Instagram* datasets are fairly similar in terms of the number of photographs, the number of census tracts in New York City is twice as large as the number of MSOAs in London. This means that we have fewer pictures per spatial unit in NYC, hence we use less data to generalise features of a census tract. However, the lack of data may cause noisy and poor representations of neighbourhoods around New York City. A model using these representations would therefore generate poorer estimates in comparison to models using better generalised features. With these notes in mind, our results which are consistent with the findings from the previous section suggest that pictures uploaded to *Instagram* can be used to gain insights into the key socioeconomic attributes of a city.

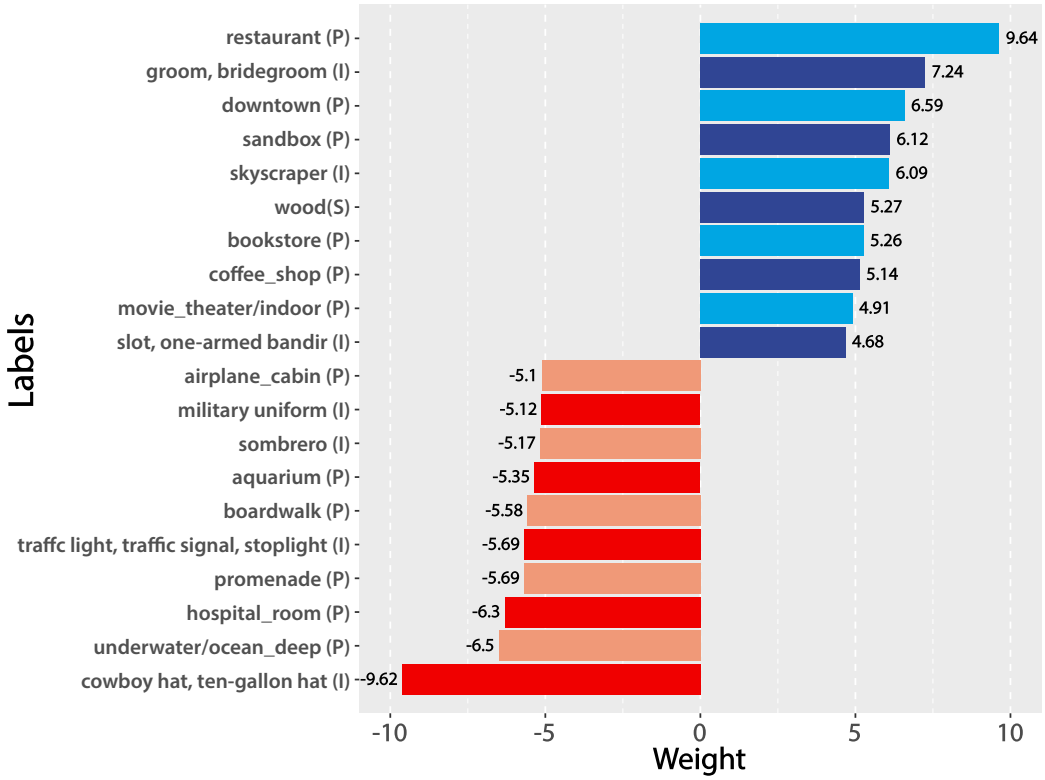

Figure 5.16: Ten largest positive and ten largest negative coefficients of the Combined model for New York City.
Features with an "(I)" next to them indicate that they are ImageNet categories, "(P)" represents Places-365 and "(S)" shows that categories are from the SUN attribute database. Compared to the coefficients of the ImageNet+Places model of London, the categories have certain differences reflecting the distinctions between London and New York City. *Instagram* images labelled with one of the categories represented in blue are more likely to be captured in a higher income area whereas *Instagram* images categorised as one of the red labels are more likely to be taken in an area with lower income.

5.6 Summary and discussion

In this chapter, we initially analyse whether photographs of food uploaded to the picture sharing platform *Instagram* can reveal information about the characteristics of the area they were taken. In order to investigate this hypothesis, we analyse a set of *Instagram* pictures taken in London over a six month period.

We start by building a classifier that can identify whether a picture is food-related or not by using pictures from the automatically created training set. Once we extract the food-related pictures taken in a restaurant, we investigate whether these pictures can be used as an indicator of the number of reviews the restaurant receives on *Yelp*. Our initial comparison suggests that the number of *Instagram* pictures taken at a restaurant is linked to the number of *Yelp* reviews about the restaurant. We then analyse the image content to investigate whether we can estimate the restaurant's rating. For each food-related picture taken at a restaurant, we create a feature vector by using a CNN trained on the ImageNet dataset. The feature vector contains scores for each ImageNet category indicating how likely a given image is from one of these categories. Our findings suggest that food-related *Instagram* pictures taken at a restaurant can, to a certain extent, help us estimate the restaurant's rating posted on *Yelp*.

We then extend these analyses to estimate a key socioeconomic attribute of a neighbourhood: income. By using the entire set of *Intagram* pictures classified as food-related, we estimate income of London at MSOA level. The comparison between estimated and actual income values suggests a significant yet fairly weak link, such that features extracted from pictures of food can only capture 8% of the variance in the income values.

Food pictures however constitute only a small portion of the entire *Instagram* dataset we analyse. We therefore seek to investigate whether using the entire set of pictures might further help us estimate income patterns across London. In addition to the ImageNet-based features extracted using a pretrained CNN, we also create sets of features using CNNs trained on different datasets such as the Places-365 and SUN attribute databases. Our findings suggest that models exploiting different categories as features can generate better income estimates which are in line with the actual household income across London.

We also investigate whether we can uncover a set of characteristics relating to the economic status of an area. Analysis of the model coefficients suggests that categories such as *embassy*, *patio* and *hotel*, which can be reminiscent of rich neighbourhoods get the highest positive scores. This indicates that if a picture has higher scores for any of these categories, then the picture was potentially taken in a higher income area. In contrast, categories such as *industrial area*, *slum* and *loading dock* that can be inherently related to poorer areas appear to have higher negative scores. This suggests that photographs that are automatically identified to be linked to any of these categories were presumably taken in areas with lower income.

Moreover, we show that categories do not always reflect what we expect them to. Although it might seem that CNNs provide wrong scores per category, they still tend

to group similar pictures to a certain extent. We therefore highlight that we need to be cautious when interpreting the coefficients of different categories.

Our results suggest that models exploiting more than one set of features are better at capturing changes in the income compared to the models that only use one set of feature vectors. However, we note that we need to be watchful when combining multiple features sets not to have too many correlated predictors in the final elastic net model.

Before moving to the New York City analysis, we should also underline that the representative power of features vectors can be enhanced by changing the underlying architecture of a CNN. There are various CNN architectures trained on the ImageNet, Places-365 and SUN attribute datasets which perform with different accuracy and precision in detecting objects. An architecture with a better detection performance across different categories can produce better feature vectors. Although models created with enhanced feature vectors may generate more accurate income estimates, it is less likely for such improvements to change the qualitative aspect of the results.

In the final part of this chapter, to investigate whether these results will be consistent for another city with different dynamics, we repeat the same approach for *Instagram* pictures taken over a six month period in New York City. For each census tract in New York City, we calculate estimated income by using seven different models. We find that income values estimated by using the information extracted from the pictures shared on *Instagram* is significantly correlated to the actual income values. These results are in line with the findings from the analysis focusing on the London area. Our analysis also reveal the difference between the categories associated with higher income areas in London and in New York City highlighting distinct characteristics of these metropolitans.

However, we found that, in comparison to the estimated values from London, income estimates from the New York City tend to be less in line with the actual household income across the New York City. This might be explained by the fact that New York City has less number of images per spatial unit which might have caused poorer generalisation once we created a mean feature vector. We also can't discard the inherent difference in the characteristics of these cities. As it is visible from the categories with highest positive and negative coefficients that London and New York City have distinct elements. We therefore cannot rule out the possibility that while features extracted from *Instagram* pictures can be at characterising traits of one city, they might not be true reflections of another city.

All in all, our results are consistent with the initial hypothesis that automatic analysis of the pictures uploaded to *Instagram* may help us estimate key socioeconomic attributes. We conclude that these findings illustrate the possibilities offered by online photographic data in gaining insights into key socioeconomic measures of cities around the world. Future studies drawing on our results and suggestions may explore whether change in these socioeconomic statistics can be monitored and captured with a finer time granularity.

CHAPTER 6

Forecasting 311 Complaints in New York City

6.1 Introduction

Fifty-four percent of the world's population live in urban areas, a figure which is forecast to reach 66 per cent by 2050 (UN, 2015). Due to the vast numbers of inhabitants, cities are faced with various complex problems which affect the daily lives of their residents in many ways. Some of these problems, such as fire, health emergencies, and crime are handled by emergency services. Yet, there are problems outside the remit of the emergency services that can have a serious impact on the efficient and harmonious operation of a city, such as illegal parking, broken traffic lights, sewers overflowing, and public areas falling into unsanitary conditions.

Monitoring such problems poses great challenges to urban management. In order to act rapidly in face of such problems, a number of local governments have introduced systems to allow citizens to report these incidents in near real time. A key example is New York City's 311 service. During 2016, 311 services were contacted almost 36 million times via calls, texts, mobile applications, online chat, and Twitter (NYC, 2017).

While rapid notification of such problems is useful, the ability to anticipate these incidents before they are reported would increase the capacity of local governmental services to act before problems became worse. As discussed in Chapter 2, a body of recent research has provided evidence that appropriate analysis of the massive datasets now generated by our everyday actions can support better forecasting of future behaviour, and thereby inform decision making and resource deployment (Conte et al., 2012; King, 2011; Lazer et al., 2009; Mitchell, 2009; Moat et al., 2014; Vespignani, 2009). In the area of crime, analysis of data collected by the police services has revealed that the occurrence of a burglary results in a short term increase in the probability that another burglary will occur on the same street (Bowers et al., 2004; Johnson and Bowers, 2004; Mohler et al., 2011).

In this chapter, we first exploit the vast amount of data on reports to New York City's 311 service to investigate whether we can forecast the location of incidents reported to the

311 services before the problems are reported. In particular, we determine whether we can anticipate the emergence of problem areas in a dynamic fashion, drawing on methodologies that have successfully been applied in the crime domain (Bowers et al., 2004).

However, well structured cross-platform services for reporting non-emergency incidents are not as common among cities worldwide as their emergency counterparts. Hence, alternative sources capturing everyday aspects of urban life might be useful in forecasting the location of the non-emergency incidents. One potential avenue offering fruitful insights into the daily dynamics of a city life are social media channels, of the kind analysed in earlier chapters of this thesis. In the second part of our analysis we therefore investigate whether photographs uploaded to social media channels can be used as an alternative source to forecast non-emergency incidents. Specifically, using New York City (NYC) as a calibration case, we analyse photographs shared on *Flickr* to create early warning signs for noise-related complaints reported to New York City's 311 services.

6.2 Data retrieval and preprocessing

Initially launched in 2003, New York City's 311 service helps residents of New York interact with more than 3 600 non-emergency government services. These range from reporting broken street lights to registering noise complaints relating to commercial, residential and non-residential properties. Tenants of rented properties can also report issues with their property that have not been adequately addressed by the landlord, such as a lack of heating in winter, or buildings that are inappropriately heated in summer. Systems then exist for the government of New York City to take action to resolve the problem reported by the tenant and subsequently bill the costs to the owner of the property. The service receives thousands of reports everyday via various channels, including social media platforms and mobile phone applications.

In this chapter, we analyse over five million 311 complaints recorded between 27th February 2012 and 31st December 2014. We retrieve this dataset from the New York City Open Data website which serves as an online repository for public data being generated by various departments, agencies and organisations in New York City. The data contains information on when and where 311 complaints were made, as well as information on what type of incident was reported, with categories such as "heating", "noise", "blocked driveway" and "unsanitary condition". In line with New York City Council legislation regarding public data (NYC, 2015), the location of the incident is recorded in the same format as provided by the submitter, such that the dataset discloses the full address of each incident down to the house number.

In total, the dataset we analyse contains 245 different incident types. We list all incident types in Appendix D in alphabetical order (Table D1). We find that some incident types are related. For example, noise-related incidents are recorded under eight different categories, including "Noise", "Noise-Commercial" and "Noise-Vehicle". We therefore merge such categories, and provide the complete list of merged incident types in Table D2.

We then focus on the 14 most frequently reported incident types. Calls falling into one of these 14 categories account for almost 70% of all incidents reported during the time period studied, or 3 568 285 complaints in total. Across the three years studied, this breaks down to 1 226 774 incidents in 2012, 1 129 394 incidents in 2013 and 1 212 117 incidents in 2014. Table D3 provides further information about each incident category. Figure D1 in Appendix D depicts incident counts for each of these categories.

For the second part of our study, we analyse 745 973 geotagged and publicly available *Flickr* pictures shared in the same week as they were captured in New York City between 1st January 2012 and 31st December 2014. We extracted *Flickr* data in JSON format annually via the *Flickr* API.

6.3 Methods

6.3.1 Creating NYC grid cells

We investigate whether we can generate weekly predictions of the areas of New York City in which different categories of 311 complaints will be made. To do this, we first divide the area of New York City into 500 $\times$ 500 m^2 grid cells, where vertical lines run from north to south, and horizontal lines run from east to west. We exclude cells whose centres fall outside the New York City boundaries. This produces a grid with a total of 3 672 grid cells. More than 80% of the grid cells (3 058 of the 3 672) have at least one incident reported over the three year period, with a median of four incidents reported per cell per week. Weekly counts show that each week, a median of 2 338 cells have at least one incident reported. This value is quite stable across different weeks, with a standard deviation of 65 over the entire three year period.

For each week and each grid cell, we extract the number of *Flickr* pictures taken and uploaded within the same week between 2012 and 2014. Over this three year period, more than 80% of the grid cells (3 016 of the 3 672) have at least one picture shared per cell per week. A median of 480 cells per week contain at least one *Flickr* picture with a median of 2 pictures per cell each week. Figure 6.1 depicts the spatial distribution of the pictures taken and uploaded to *Flickr* between 2012 and 2014. It is clearly visible that tourist attractions such as Central Park and the Statue of Liberty as well as John F Kennedy Airport are prominent photo sharing locations across New York City.

6.3.2 Calculating risk values using historical records

For the first part of our analysis, in order to decide whether there will be a non-emergency complaint reported from a grid cell by using historical 311 records, for each week and each grid cell, we calculate a risk value by using three different models: the spatiotemporal model, the static model and the seasonal model. Each model seeks to generate a risk surface for each incident type, where risk values are calculated for each cell. To assess

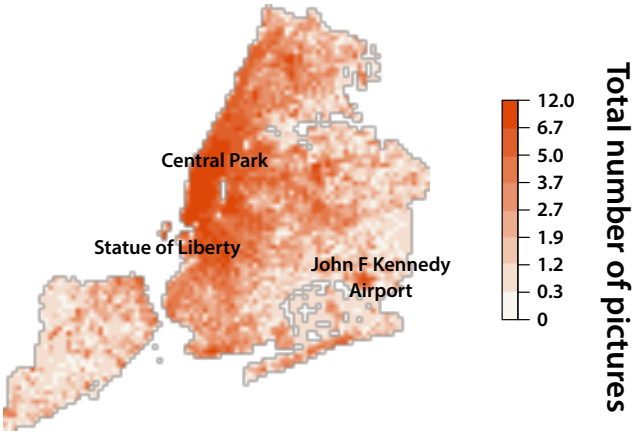

Figure 6.1: Total number of *Flickr* pictures taken and uploaded in the same week between 2012 and 2014.
Famous attractions including Central Park and Statue of Liberty as well as John F Kennedy Airport stand out as some of the photo sharing hotspots in New York City. Colour breaks were calculated using *k*-means clustering algorithm on logarithmically transformed numbers.

the models' performance, we use the risk surfaces to generate forecasts of cells in which we would expect 311 incidents to occur. We use data from 2012 to train the seasonal model, which requires 52 weeks history in order to generate predictions, and we use data from the first eight weeks of 2013 to train the spatiotemporal model, following the initial two month period used for training in Bowers et al. (2004). For this reason, we evaluate the performance of the models during the period from week 9 in 2013 to the final week of 2014.

6.3.2.1 Spatiotemporal model

The first model, the "spatiotemporal model", takes inspiration from the approach proposed in Bowers et al. (2004) for anticipating the location of future crimes. In this model, it is assumed that problems are most likely to occur in and around cells which have recently seen higher volumes of such problems. In other words, the location of previous events is of relevance, as is the recency with which they occurred. For each cell, we define a neighbourhood area *A*, with a radius of 5 km. We consider previous events in all weeks before week *t* from the first week of 2013 onwards, which we denote week 1, and begin to assess the quality of our forecasts in week 9. To calculate the risk value for a given incident type for each cell *i* in week *t*, we use the formula

$$\text{Risk Value}_i(t) = \sum_{a \in A} \sum_{\tau=1}^{t-1} \frac{1}{d\,(a,i)+1} \cdot \frac{1}{t-\tau} \cdot N_a(\tau), \tag{6.1}$$

where $N_a(t)$ is the number of 311 reports relating to the given incident type in cell *a* during week *t*, and where $d(a, i)$ is the distance between the centre of cell *a* and the centre of cell *i* measured in metres. We illustrate the implementation of this model in Figure 6.2.

We note that the influence of a previous incident occurring in cell a on the risk value for cell *i* is inversely proportional to the distance of the centre of cell *a* from the centre of cell *i*. For this reason, an incident which occurred in a cell on the boundary of the neighbourhood area, 5 km from cell *i*, would have 10% less influence than an incident which occurred in a neighbouring cell, 500 m from cell *i*. In comparison to an incident which occurred in cell *i* itself, an incident in a cell 5 km from cell *i* would have 0.02% of the influence on the final risk value for cell *i*. As this number is already very low, we set the neighbourhood area radius at 5 km and do not consider incidents in cells more than 5 km away in order to optimise the speed of risk value calculations.

To help determine whether data on how recently similar incidents have occurred nearby is of value in anticipating the future location of 311 incidents, we compare the spatiotemporal model to two further baseline models, the *static* model and the seasonal model.

6.3.2.2 Static model

In the "static model", it is assumed that the location at which similar incidents have occurred is of relevance, but the time at which they occurred is of no relevance. To implement this model, data on incidents which took place between the first week of 2013 and the final week of 2014 are used to calculate a static risk value for cell *i*. The calculated risk value for cell *i* therefore remains constant throughout the time period. While it is still affected by the proximity of other incidents, it is not affected by the time at which incidents occurred. For each cell *i*, we calculate the risk value,

$$\text{Risk Value}_i = \sum_{a \in A} \sum_{\tau=1}^{T} \frac{1}{d\,(a, i) + 1} \cdot N_a(\tau), \tag{6.2}$$

where $\tau = 1$ is the first week of 2013, and $\tau = T$ is the final week of 2014. Again, we illustrate the implementation of this model in Figure 6.2. Forecasts are assessed from week 9 of 2013 until the final week of 2014, as for the spatiotemporal model. By comparing the performance of the spatiotemporal model to the performance of the static model, we can investigate whether information on the recency of similar events nearby helps improve the quality of predictions. If this is the case, we would expect to see better predictions generated by the spatiotemporal model than the static model.

6.3.2.3 Seasonal model

If an incident type were to occur in a seasonal fashion, for example with more reports in winter, we might also expect to see better predictions generated by the spatiotemporal model than the static model, as a higher number of recent events may reflect that the season for a particular incident has begun. To distinguish between the possibilities of incidents

clustering in time because the problem is seasonal, and incidents clustering in time in a non-seasonal fashion which is better captured by the concept of recency, we create a second baseline model, the "seasonal model". In the seasonal model, it is assumed that the location at which similar incidents have occurred is of relevance, and that incidents occur with a seasonal pattern. To implement this model, data on incidents which took place during week $t - 52$ are used to calculate risk values for cells in week *t*. Again, the risk value is affected by the proximity of other incidents, but only those which took place at the same time of year in the previous year. To enable forecasts to be assessed from week 9 of 2013 until the final week of 2014 as for the previous two models, we draw on data from week 9 of 2012 onwards. For each cell *i*, we calculate the risk value,

$$\text{Risk Value}_i(t) = \sum_{a \in A} \frac{1}{d\,(a, i) + 1} \cdot N_a(t - 52). \tag{6.3}$$

Once again, we illustrate the implementation of this model in Figure 6.2. Once a risk surface consisting of risk values for all cells has been calculated, a risk threshold θ must be set so that predictions can be derived. A cell is considered to be at risk if its risk value is greater than θ (Figure 6.2). We evaluate the performance of all models for a wide range of values of θ, as described in more detail in the following section.

6.3.3 Calculating risk values using *Flickr* photographs

In order to create risk surfaces for noise complaints using photographic data shared on *Flickr*, we first need to inspect the content of the pictures. In the previous chapter, we showed that models using a combination of features from ImageNet and Places categories computed the best income estimates for both London and New York City. Here, to automatically analyse the image content, we therefore use VGG-M-128 which has been trained on the ImageNet dataset and VGG-16 which has been trained on the Places-365 database. Following a similar approach as in Chapter 5, for each photograph taken and uploaded in New York City in 2012, we create a 1 365 dimensional feature vector combining ImageNet labels generated by VGG-M-128 and Places-365 labels created by VGG-16. We then calculate the mean feature vector for each grid cell where at least one noise-related incident was reported across 2012.

To identify how much each feature contributes to the 311 report estimator for noise-related complaints, we use the mean feature vector. We fit an elastic net model by using each feature as a predictor and the total number of noise complaints reported per grid cell as the observed variable. We will refer to this model as the "*Flickr* model".

Before proceeding to the risk value calculation, we test the performance of the elastic net model in estimating the number of noise-related complaint reports. Leaving one grid cell out, we fit an elastic net model using the mean feature vectors from the remaining cells.

Using the mean feature vector of the grid cell that was initially left out, we then cal-

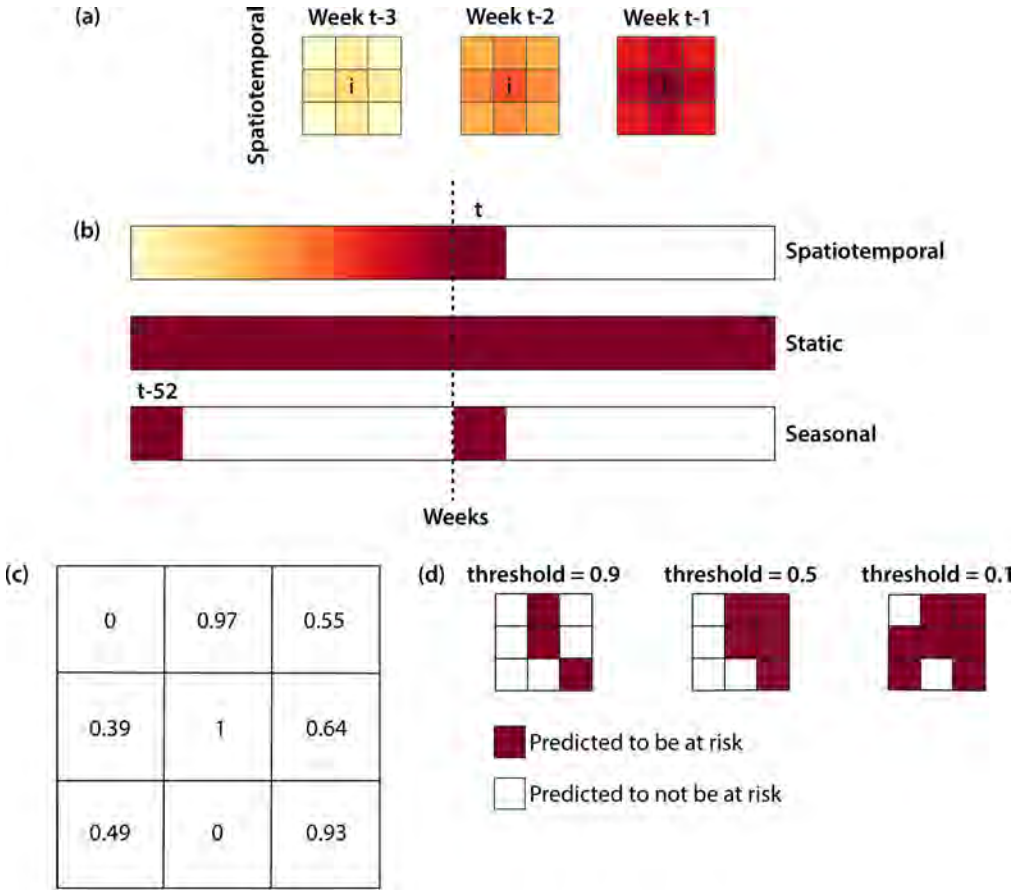

Figure 6.2: Three different models for identifying areas at risk.
(a) We create a 3 $\times$ 3 cell toy map to demonstrate how risk values for the central cell *i* are calculated using the spatiotemporal model. This model assumes that problems are most likely to occur in and around cells which have recently seen higher volumes of such problems. In other words, not only the proximity but the recency of these incidents is important. Here, we depict the 3 $\times$ 3 cell toy map across three weeks, Week $t-3$, Week $t-2$ and Week $t-1$. We colour cells according to the extent to which the problems occurring in that cell in the given week would increase the risk value estimated for the central cell *i* in Week t. Darker red indicates a greater influence. **(b)** We compare the spatiotemporal model to two further models. In the static model, data on incidents which took place during the whole 2013–2014 period are used to calculate a static risk value for cell *i*. The calculated risk value for cell *i* therefore remains constant throughout the time period. While it is still affected by the proximity of other incidents, it is not affected by the time at which incidents occurred. In the seasonal model, data on incidents which took place during week $t-52$ are used to calculate risk values for cells in week t. Again, the risk value is affected by the proximity of other incidents, but only those which took place at the same time of year in the previous year. **(c)** Once a risk surface has been calculated, predictions can be made. To explain how this prediction mechanism works, we create a hypothetical risk surface on a 3 $\times$ 3 cell toy map, using randomly generated risk values between 0 and 1. (d) We generate predictions from the hypothetical risk surface depicted in *(c)*. In red, we highlight the cells that would be determined to be "at risk" using three different thresholds of 0.9, 0.5 and 0.1. With a low risk threshold, nearly all cells are considered to be at risk. With a high risk threshold, very few cells are considered to be at risk.

culate the estimated number of noise-related complaints reported within a given cell. We repeat the same procedure for each grid cell with at least one noise-related complaint reported to the 311 services in 2012. Our analyses demonstrate a significant correlation between the actual number of reports of noise-related complaints and the estimated number of reports (τ = 0.32, $p < 0.001$, *N* = 2 015, Kendall's rank correlation). Our results suggest that visual features extracted from the *Flickr* images can capture 11% of the change in the number of noise-related complaints ($R^2 = 0.11$).

6.4 Analysis and results

6.4.1 Forecasting 311 complaints using historical records

For each week and for each of the 14 incident types in our dataset of 311 calls, we analyse the number of incidents reported within each cell. In Figure 6.3, we depict the spatial distribution of the four most frequently reported complaint categories during the years 2013 and 2014. Visual inspection suggests that the spatial distribution of incidents varies depending on the complaint type. For example, complaints relating to heating, noise and plumbing appear to cluster in certain areas of New York City, whilst complaints relating to street conditions are more widely spread across the city.

To gain further insight into the spatiotemporal structure of the 311 calls dataset, we visualise the times and locations of the four incident types that were most frequently reported to the New York City's 311 services between the first week of 2013 and the final week of 2014 (Figure 6.4; see Figure D2 in Appendix D for an alternative visualisation of this data using small multiples). Visual inspection suggests that reports of heating incidents appear to be more common in winter. The location of heating, noise and plumbing related incidents appears to remain relatively constant. In contrast, street condition problems appear to have a wider geographical spread in 2014 when in comparison to 2013.

We then investigate whether we can forecast the location of the future non-emergency incidents by utilising the previous records of the incidents reported to the 311 services. For each week and for each of the 14 complaint categories, we create risk surfaces by calculating the risk values using the spatiotemporal model, the static model, and the seasonal model introduced in 6.3.2, starting in week 9 of 2013 and working through to the final week of 2014. We rank the values of each risk surface, giving identical risk values a tied rank. The lowest risk value is allocated the rank 1. We determine the cells that would be considered as "at risk" by systematically changing the threshold values θ. Low values of θ result in most cells being considered as at risk, and high values of θ result in very few cells being considered as at risk. For a risk surface with *n* unique ranks of risk values, we generate a range of *n+1* values of θ, to enable us to test the full range of predictions which could be made by the generated risk surface, from no cells being considered as at risk to all cells being considered as at risk. In this way, the full range of possible values of θ, which are generated using the ranked risk values, are tested.

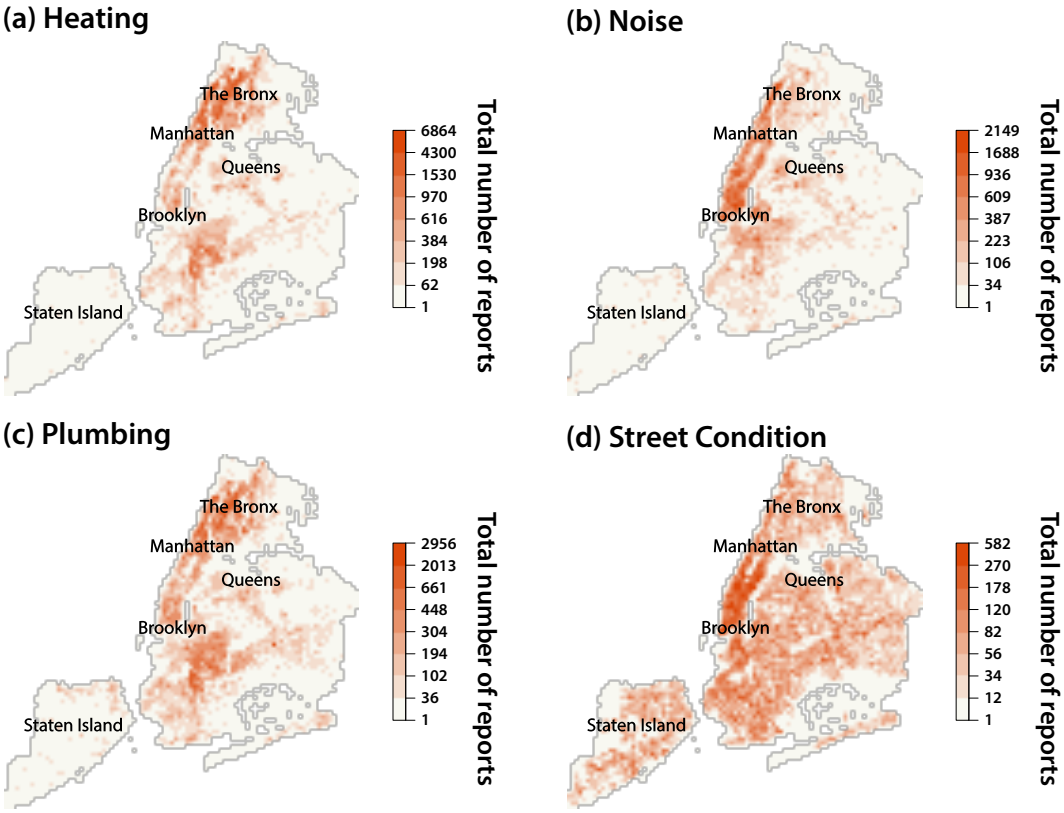

Figure 6.3: Location of incidents reported using the 311 service.
We depict the locations of the four most frequent categories of incidents reported to the New York City's 311 services during 2013 and 2014. These are **(a)** heating **(b)** noise **(c)** plumbing, and **(d)** street condition. We divide New York City into a grid of 500 $\times$ 500 m^2. The colour of each square indicates the volume of calls recorded for that location. Visual inspection suggests that reports of certain incident types, such as heating, noise, and plumbing have particularly high concentrations in certain areas of New York City, whereas reports of others, such as street condition, are more widely distributed across the city. Breaks were determined using the *k*-means clustering algorithm and rounded to the nearest integer.

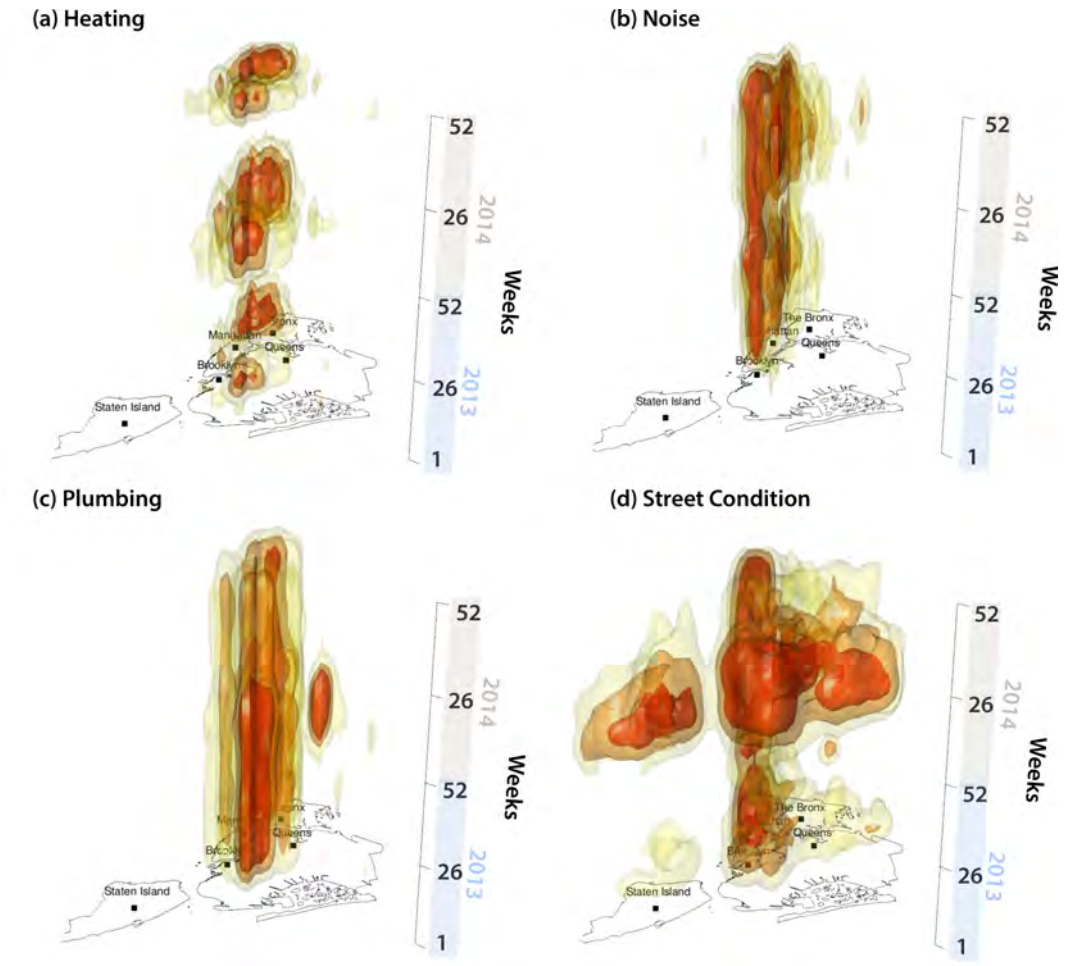

Figure 6.4: Time and location of incidents reported using the 311 service.
We visualise the times and locations of the four most frequent categories of incidents reported to the 311 service across 2013 and 2014. Again, these are **(a)** heating **(b)** noise **(c)** plumbing, and **(d)** street condition. The figure displays the contours of 25%, 50% and 75% of the distribution of incident reports, where the most opaque contour indicates the central 25% of the distribution. The distribution has been estimated using a 3D kernel density estimate. Visual inspection suggests that reports of heating incidents appear to be more common in winter. The location of heating, noise and plumbing related incidents appears to remain reasonable constant. In contrast, street condition problems appear to have a wider spread in 2014 in comparison to 2013. We provide an alternative visualisation of this data using small multiples in Figure D2 in Appendix D.

We examine whether incidents are reported in the cells we predicted would be at risk. For each model and each threshold, we calculate two metrics of the quality of the predictions, known as *sensitivity* and *specificity*. Sensitivity, or the *true positive rate*, refers to the proportion of cells marked as at risk in which incidents are then reported. Specificity, or the *true negative rate*, refers to the proportion of cells not marked as at risk, in which no incidents are then reported. To assess each models' relative performance, we compute the Receiver Operating Characteristic (ROC) curves obtained given the sensitivity and the specificity of the predictions at each value of θ. In Figure 6.5, we depict ROC curves for predictions of the locations of the four most frequent incident types reported in 2013 and 2014. For each ROC curve, we calculate the area under the curve (AUC). The AUC can range from 0 to 1, where AUC values of 1 indicate perfect predictions. If cells were randomly selected to be at risk, we would expect an AUC of 0.5 (depicted as "Random" in Figure 6.5).

Table 6.1 shows the AUC values with 95% confidence intervals using the spatiotemporal, static and seasonal models. It is clear from Table 6.1 and Figure 6.5 that the predictions generated by all three models are superior to randomly marking areas as at risk. To compare the performance of the spatiotemporal, static and seasonal models, we compare the ROC curves for each model using the method introduced by DeLong et al. (1988) which is based on comparing the AUC values (Table 6.2). In Table 6.1, we highlight those AUC values where the AUC for a model has been found to be significantly larger than other non-highlighted AUC values for other models for the same incident category. We find that the spatiotemporal model generates the best forecasts for 12 of the 14 complaint categories examined, with the exceptions of "Construction" and "Electric". This suggests that for most complaint categories, information on how recently similar incidents have occurred nearby is useful for improving forecasts of whether incidents will be reported in the near future. For complaints relating to csonstruction, we find that the seasonal model performs best, suggesting that information on previous seasonal patterns in reports of similar incidents nearby is of more relevance than information on the recency of similar local incidents. For the electric category, the static model provides the most accurate forecasts, suggesting that for this category, information on when previous incidents occurred does not benefit forecasts of the locations in which future incidents may be reported. However, as for all categories of complaint, predictions generated by all three models are clearly superior to randomly marking areas as at risk, for which an AUC of 0.5 would be expected.

6.4.2 Forecasting 311 complaints using *Flickr* photographs

In the previous section, focusing on the 14 most frequently reported incidents, we provide evidence that historical reports of such incidents can be used to identify locations of similar events before they occur. In the second part of our study, we analyse pictures uploaded to *Flickr* to uncover a signal which might help us predict locations of non-emergency complaints before they are reported to the 311 services. Considering that potential sources of

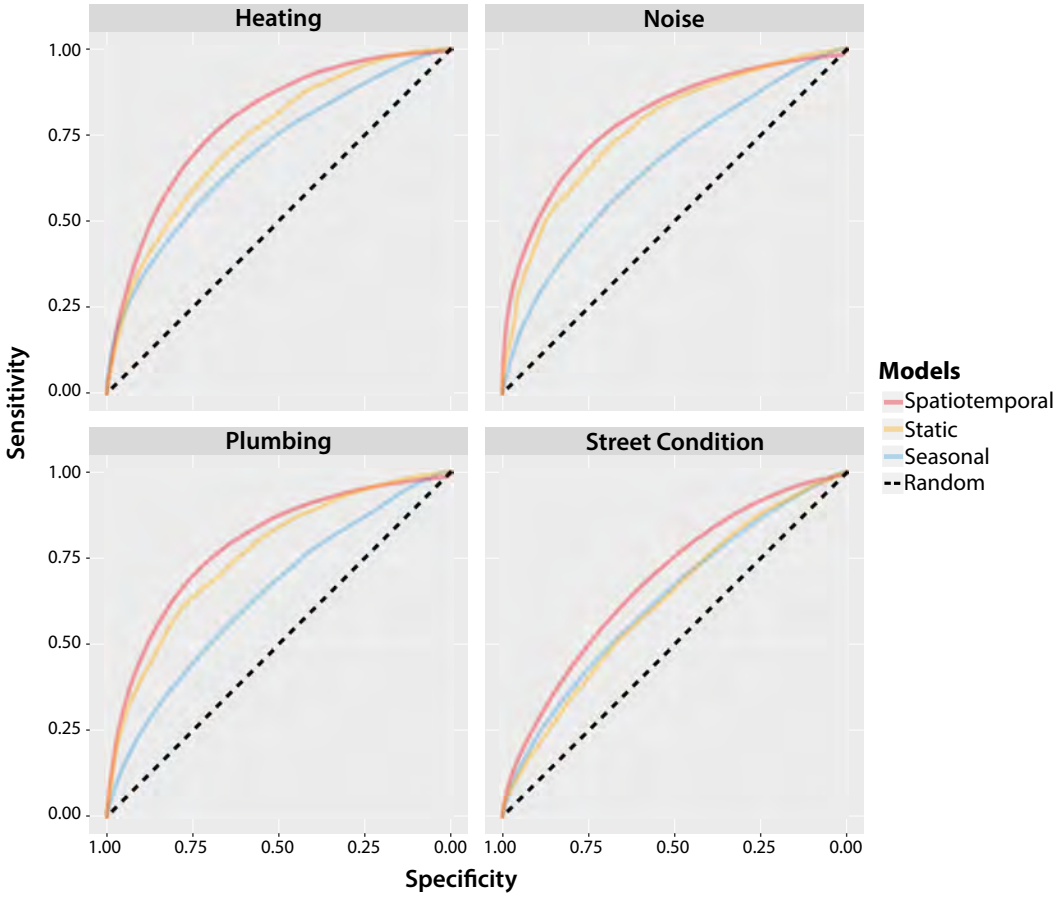

Figure 6.5: Evaluating different prediction models.
We evaluate the quality of predictions of the location of the four most frequent categories of complaints made to the 311 service during 2013 and 2014: Heating, Noise, Plumbing and Street Condition. We consider predictions made using the spatiotemporal (red), static (yellow), and seasonal (blue) models, and compare the performance of these models to the performance that would be expected if areas were randomly selected to be marked as at risk (black dashed line). For each week and each model, we generate forecasts using a wide range of thresholds, where low thresholds result in most cells being considered as at risk, and high thresholds result in very few cells being considered as at risk. For each model and each threshold, we calculate two metrics of the quality of the predictions, known as *sensitivity* and *specificity*. Sensitivity, or the *true positive rate*, refers to the proportion of cells marked as at risk in which incidents are then reported. Specificity, or the *true negative rate*, refers to the proportion of cells not marked as at risk in which no incidents are then reported. Using these two performance metrics, we can plot Receiver Operating Characteristic (ROC) curves and calculate the area under the ROC curves to evaluate the prediction performance. Visual inspection shows that all three models deliver better performances than would be expected if areas were randomly selected to be marked as at risk. Table 6.1 demonstrates that this conclusion holds for all 14 incident types we examine. For all of the four incident types depicted in this figure, the spatiotemporal model outperforms both the static and seasonal model, suggesting that information on how recently similar events have occurred can be used to improve predictions. This is true for 12 out of the 14 incident types we examine, with the exception of complaints in the categories *Construction* and *Electric* (Table 6.1).

Table 6.1: The area under the curve (AUC) values calculated for predictions generated by the spatiotemporal, static and seasonal models.

Values in parentheses depict the 95% confidence interval for the AUCs. The spatiotemporal model outperforms the static and seasonal models in forecasting the location of 311 reports, with the exception of incidents in the Construction and Electric categories. AUC values are highlighted in bold where the AUC for a model has been found to be significantly greater than other non-bold AUC values for other models for the same incident category, using the comparison method described in DeLong et al. (1988). See Table 6.2 for further details of these statistical tests.

Complaint	Spatiotemporal		Static		Seasonal	
Heating	**0.79**	**(0.788-0.792)**	0.743	(0.741-0.745)	0.697	(0.693-0.701)
Noise	**0.798**	**(0.796-0.8)**	0.772	(0.77-0.774)	0.663	(0.66-0.666)
Plumbing	**0.79**	**(0.788-0.792)**	0.76	(0.758-0.762)	0.644	(0.641-0.648)
Street Condition	**0.686**	**(0.684-0.688)**	0.624	(0.622-0.627)	0.631	(0.628-0.634)
Street Light Condition	**0.689**	**(0.686-0.691)**	0.643	(0.641-0.645)	0.614	(0.61-0.617)
Unsanitary Conditions	**0.744**	**(0.742-0.746)**	0.693	(0.691-0.695)	0.643	(0.639-0.646)
Paint	**0.785**	**(0.782-0.787)**	0.769	(0.766-0.771)	0.639	(0.634-0.644)
Construction	0.69	(0.687-0.693)	0.687	(0.683-0.69)	**0.716**	**(0.712-0.719)**
Blocked Driveway	**0.784**	**(0.782-0.786)**	0.737	(0.734-0.739)	0.617	(0.613-0.621)
Water System	**0.678**	**(0.675-0.681)**	0.64	(0.637-0.642)	0.624	(0.62-0.628)
Illegal Parking	**0.727**	**(0.725-0.729)**	0.676	(0.674-0.679)	0.596	(0.592-0.6)
Traffic Signal Condition	**0.711**	**(0.708-0.714)**	0.67	(0.666-0.673)	0.617	(0.612-0.622)
Sewer	**0.65**	**(0.647-0.652)**	0.598	(0.595-0.6)	0.601	(0.598-0.605)
Electric	0.743	(0.739-0.746)	**0.749**	**(0.746-0.752)**	0.619	(0.614-0.624)

Table 6.2: Paired comparisons of the AUC values for the spatiotemporal, static and seasonal models.

Statistical significance of the difference between the areas under the two ROC curves is calculated using DeLong's test for two ROC curves (DeLong et al., 1988). p values have been FDR corrected to account for the fact that multiple comparisons have been carried out.

	Spatiotemporal vs Static		Spatiotemporal vs Seasonal		Static vs Seasonal	
Complaint	Z	p	Z	p	Z	p
Heating	-31.10	< 0.001	-42.15	< 0.001	-20.34	< 0.001
Noise	-18.39	< 0.001	-72.20	< 0.001	-58.10	< 0.001
Plumbing	-19.69	< 0.001	-70.61	< 0.001	-56.07	< 0.001
Street Condition	-38.81	< 0.001	-27.46	< 0.001	3.22	< 0.001
Street Light Condition	-27.25	< 0.001	-36.09	< 0.001	-13.95	< 0.001
Unsanitary Conditions	-34.18	< 0.001	-51.79	< 0.001	-25.45	< 0.01
Paint	-8.29	< 0.001	-53.83	< 0.001	-47.94	< 0.001
Construction	-1.70	0.089	10.59	< 0.001	11.92	< 0.001
Blocked Driveway	-32.04	< 0.001	-77.54	< 0.001	-55.02	< 0.001
Water System	-19.43	< 0.001	-22.43	< 0.001	-6.45	< 0.001
Illegal Parking	-29.78	< 0.001	-56.55	< 0.001	-34.40	< 0.001
Traffic Signal Condition	-16.82	< 0.001	-31.16	< 0.001	-17.39	< 0.001
Sewer	-27.82	< 0.001	-21.31	< 0.001	1.56	0.118
Electric	2.91	< 0.001	-41.82	< 0.001	-44.72	< 0.001

urban noise can be identified using social media images such as pictures of a busy street, building works or even a party, as our case study, we focus on predicting the location of noise-related complaints.

We evaluate whether the *Flickr* model we have specified in section 6.3.3 can be used to predict the location of noise-related complaints before they are reported to the 311 services. For each *Flickr* photograph taken and uploaded within New York City across 2013 and 2014, we extract feature vectors using the pretrained CNNs we used to extract features of the pictures take 2012. We then create one mean feature vector reflecting the visual characteristics of the photographs taken and uploaded to *Flickr* on a given week in a given cell.

With the mean features as predictors, for each week, we create risk surfaces by using the elastic net fitted on the noise-related complaints reported in 2012. For easier comparison with the previous results, we calculate risk values starting in week 9 of 2013 through to the final week of 2014. We then rank the risk values of each risk surface, the lowest risk getting rank 1 and giving a tied average rank to identical risk values.

Having created ranked risk surfaces, we identify cells at risk again by systematically changing the threshold value θ. Using sensitivity and specificity metrics calculated for each threshold value, we compute the ROC curve for predictions on whether a given cell is likely to have reports on noise complaints or not. We then calculate the AUC for the ROC curve evaluating the predictive power of the *Flickr* model, AUC=0.609, 95% CI [0.606 – 0.611].

Figure 6.6 shows the performance of the predictions made using the *Flickr* model (green) against the performance of the predictions made using the spatiotemporal model (red). The performance obtained by randomly marking cells as at risk is represented by a black dashed line. Drawing on the method introduced in DeLong et al. (1988), we compare the performance of the *Flickr* and spatiotemporal models. Although the spatiotemporal model performs significantly better than the *Flickr* model in predicting the location of noise-related complaints (AUC for spatiotemporal model: 0.814; AUC for Flickr model: 0.622; Z = -146.44, $p < 0.001$, DeLong's test), the *Flickr* model is superior to random labelling of cells as risky or not, which has an AUC of 0.5.

Our results suggest that predictive performance of the *Flickr* model is significantly different from the spatiotemporal model, the latter proving to be more powerful. However, marking grid cells as at risk or not by solely using visual information extracted from *Flickr* still performs better than randomly picking cells to be at risk. Although using historical data is better at predicting the location of noise-related complaints in comparison to using visual information automatically extracted from *Flickr* pictures, our findings are in line with the suggestion that social media images can be utilised to create an indicator of where noise-related complaints might be reported, especially for cases where historical data is not available.

In order to investigate whether online images might add extra information to the historical data, we create a separate logistic regression model, where for each grid cell the predictors are weekly spatiotemporal risk values and weekly risk values generated by

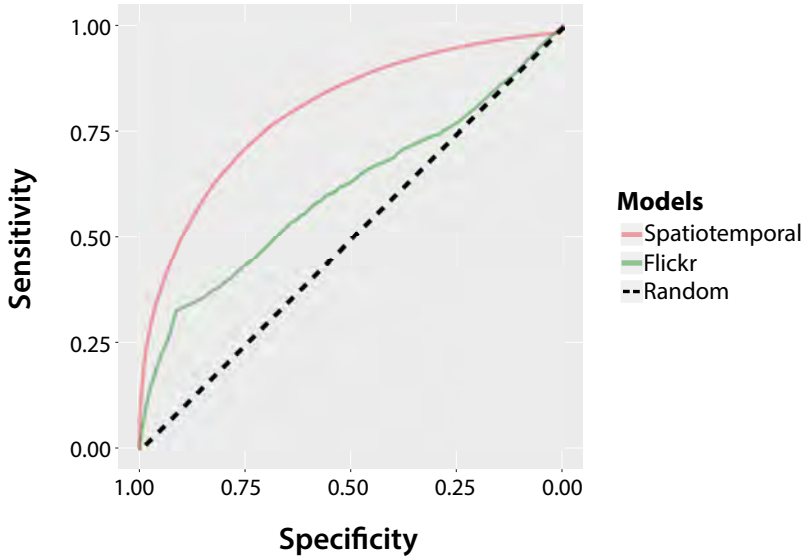

Figure 6.6: Evaluating different prediction models for noise-related complaints. We consider predictions made using the spatiotemporal (red) and *Flickr* (green) models, and compare the performance of these models to the performance that would be expected if areas were randomly selected to be marked as at risk (black dashed line). Using these two performance metrics sensitivity and specificity, we plot Receiver Operating Characteristic (ROC) curves and calculate the area under the ROC curves to evaluate the prediction performance. Visual inspection shows that both models deliver better performances than would be expected if areas were randomly selected to be marked as at risk.

using the *Flickr* model. We will refer to this as the *combined model*. We then compare this model to two other logistic regression model: one using risk surfaces generated by the spatiotemporal model and one using risk surfaces created by *Flickr* model.

As in section 6.3.2, we start evaluating these models in week 9 of 2013. We use the first eight weeks to fit the initial set of logistic regression models, which will then be used to predict the risk values for week 9. For each week from week 9 in 2013 to the final week of 2014, we create models by fitting logistic regression on the entire data from previous weeks to determine where a noise complaint will be reported on a given week.

We assess the performance of the logistic regression based models by computing the ROC curves. Table 6.3 lists AUC of the three logistic regression models with 95% confidence intervals. Figure 6.7 depicts the predictive performance of the three logistic regression based models. Table 6.4 provides performance comparison of these three models based on comparing their AUC values.

We find that when using logistic regression, spatiotemporal risk values still serve as the best predictors when identifying areas of noise complaints whereas the model combining risk values from the spatiotemporal and *Flickr* model predict risky cells with a similar performance. Among these three models, the model using information solely from *Flickr* has the least predictive power with the minimum AUC. Nevertheless, when forecasting the

Table 6.3: The area under the curve (AUC) values calculated for predictions generated by the logistic regression based *combined*, *Flickr* and spatiotemporal models.

Values in parentheses depict the 95% confidence interval for the AUCs. The spatiotemporal model outperforms the *Flickr* and *combined* models in forecasting the location of 311 reports on noise-related complaints, as confirmed using the comparison method described in DeLong et al. (1988). See Table 6.4 for further details of these statistical tests.

Complaint	Combined		Flickr		Spatiotemporal	
Noise	0.812	(0.81-0.814)	0.622	(0.62-0.623)	**0.814**	**(0.812-0.816)**

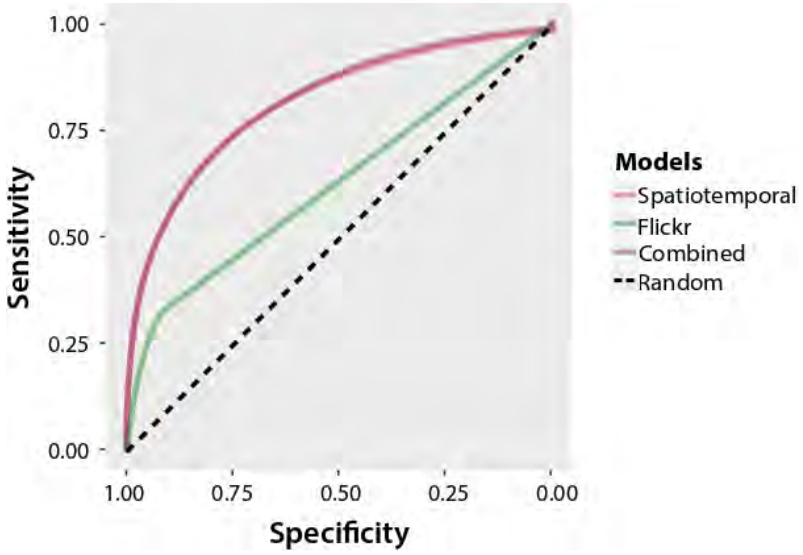

Figure 6.7: Evaluating predictive performance of logistic regression based models. We consider predictions made using the spatiotemporal (red) and *Flickr* (green) models as well as combining risk values from both models (purple). Visual inspection suggests that all three models deliver better performance than randomly labelling cells as risky or not (black dashed line). The spatiotemporal and *combined* models show a similar performance whereas they both outperform the *Flickr* model.

Table 6.4: Paired comparisons of the AUC values for the logistic regression based spatiotemporal, *Flickr* and combined models.

Statistical significance of the difference between the areas under the two ROC curves is calculated using DeLong's test for two ROC curves (DeLong et al., 1988). p values have been FDR corrected to account for the fact that multiple comparisons have been performed.

	Combined vs Flickr		Combined vs Spatiotemporal		Spatiotemporal vs Flickr	
Complaint	Z	p	Z	p	Z	p
Noise	-187.97	< 0.001	-8.94	< 0.001	-176.18	< 0.001

location of noise-related complaints, it is clear from Table 6.3 and Figure 6.7 that all three models perform better than randomly labelling cells as at risk. We should also note that despite using the same risk surfaces generated by the spatiotemporal model from historical reports, we find that the model using logistic regression performs better than the model using ranked risk values proposed under section 6.3.2 (AUC for logistic regression = 0.814; AUC for risk surfaces = 0.798; Z = -69.25, $p < 0.001$, DeLong's test).

6.5 Summary and discussion

Cities are faced with a myriad of problems every day, ranging from pot holes to noisy neighbours and broken traffic lights. To facilitate rapid monitoring of such issues, a number of local governments have recently developed systems to allow their citizens to report problems by phone or online. A key example is New York City's 311 service. However, while rapid reports of local problems might benefit policymakers, dynamic information on where problems might occur next would open up further opportunities to take effective action before problems become worse,

Here, we investigate whether models which were developed to anticipate criminal activity can be used to forecast the future location of urban problems outside the remit of the emergency services. We analyse a large dataset comprising three years of calls made to New York City's 311 service between the years 2012 and 2014. We find that, for the vast majority of incident categories, predictions of the future location of problems reported to the 311 service can be improved by considering how recently similar incidents have been reported nearby.

In our current methodology we divide New York City into grid cells for analysis and the generation of risk surfaces, but alternative approaches could be considered. In particular, while New York City census tracts are on average larger than the grid cells we have used here, their boundaries are likely to better reflect the structure of neighbourhoods. Future work could investigate whether the use of such alternative geometries might further improve the quality of predictions.

Further analyses could also investigate more complex approaches to analysing the time at which previous incidents were reported, for example by combining information on recent incidents, as taken into account by the spatiotemporal model, with information on incidents which occurred a year ago, as modelled by the seasonal model. Our current seasonal model incorporates seasonal information into its predictions following an approach that is common in time series analysis, generating risk surfaces for a given week using information on incidents that were reported during the same week in the previous year. However, future work could further examine the potential benefit of analysing a larger amount of data from the previous year, for example by generating risk surfaces for a given week using information on incidents that were reported over the course of a month during the previous year.

Future work could investigate in more detail to what extent information on nearby

incidents benefits predictions using historical reports on non-emergency incidents. At least some information on the location of previous incidents is required to generate the spatial risk surfaces we describe here, or it would be difficult to motivate allocating different risk values to different spatial locations. However, future analyses could manipulate the strength of influence of previous incidents in nearby grid cells on the risk value generated for a given grid cell. For example, should previous events in the surrounding grid cells have a stronger influence on the risk value generated, or perhaps no influence at all?

Not every city has been provided with a platform to report non-emergency incidents, whereas such problems might still be affecting the smooth functioning of a city. Motivated by the increasing number of online pictures capturing everyday aspects of an urban life, in the second part of our study, using New York City as a calibration case, we investigate whether online pictures can be used to forecast the location of non-emergency incidents before they are reported. Between the years 2012 and 2014, we analyse pictures uploaded to the photo sharing platform *Flickr* in the same week as they were captured. We extract risk surfaces for noise-related complaints by automatically analysing the visual content of the *Flickr* pictures. Our results suggest that the model incorporating the *Flickr* data can predict the location of noise-related incident reports better than randomly marking areas as at risk.

Nevertheless, once compared to the spatiotemporal model, predictive performance of the *Flickr* model is relatively poor. This can be explained with a closer look at the *Flickr* dataset. Although the size of this dataset might seem large, pictures are not evenly spread across New York City. As highlighted in Figure 6.1, picture hotspots in New York City are mainly dominated by historical landmarks and parks. Despite the large volume of images taken at these locations, there are indeed very few residents around some of these areas hence very few incident reports. Moreover, the grid surface we base our analysis has a very fine spatial granularity, meaning that the number of pictures per grid cell per week is not very large, ten pictures on average. Future studies might bring in photographic data from additional social media platforms to enrich the number of pictures in the dataset, as well as the diversity and spatial coverage.

We then investigate whether we can create a logistic regression model with better predictive power by combining information extracted from *Flickr* and data on previous noise-related complaints. We show that the *combined* model performs significantly better than the logistic regression based *Flickr* model. However, the model exploiting historical data is still better at predicting the location of noise-related complaints compared to the *combined* model. Although it might seem like the *combined* model includes more information, the signal embedded within risk surfaces created by the *Flickr* model is likely to be captured by the spatiotemporal model. As discussed above, *Flickr* data is not evenly spread and might incorporate noise. Since the meaningful signal is already captured by the spatiotemporal predictor, the remaining noise from the *Flickr* predictor brings the performance of the *combined* model down below the performance of the spatiotemporal model.

There is one more point to mention regarding the logistic regression analysis. In

predicting the location of noise-related incidents, the logistic regression model exploiting spatiotemporal risk values leads to a slightly higher AUC compared to the risk surfaces generated directly from the spatiotemporal risk values as described under section 6.4.1. We suggest this small difference is likely to be due to precision issues resulting from tied ranks when calculating the AUC for risk surfaces.

All in all, our findings provide evidence that the models we proposed can predict non-emergency incidents before they occur. However, not every non-emergency incident is reported to the 311 services or in contrast, higher number of complaints are received from some areas with certain demographics. An increasing body of research has shown that 311 reports can have inherent bias stemming from numerous factors. For instance, the usage of the non-emergency service varies across different neighbourhoods with different demographics (Eshleman and L Auerbach, 2015; Kontokosta et al., 2017). Furthermore, Legewie and Schaeffer (2016) also provided evidence that tension between the neighbours in an area can have an effect on the usage of the 311 services. The authors suggested more incidents are reported in areas with a conflict between the residents whereas in other areas residents initially try to solve the problem themselves without informing the authorities. When using 311 reports to inform decisions, it is therefore immensely important to consider the usage bias in order to avoid exacerbating any inequalities in reporting.

To summarise, city monitoring frameworks such as New York City's 311 system provide urban policy makers with rapid information on problems currently affecting the city. The first set of results we describe here suggest that appropriate analysis of the time and location of previously reported incidents could provide policy makers with additional early insight into the locations in which future incidents may be reported. We also seek to investigate alternative sources to identify risky locations. Our results suggest that visual information extracted through automatic analysis of photographic data uploaded to *Flickr* can give us insights into the location of noise-related complaints before they are reported to the 311 services. Our findings illustrate how the volumes of data now generated in urban environments may help us better manage the cities we live in.

CHAPTER 7

Conclusions

Measuring how people behave is of vital importance to both scientists and policy makers alike, who require this information to inform scientific theories and decisions regarding interventions. Traditionally, many measurements of core aspects of our daily lives have been drawn from surveys and interviews. While such data offer useful and rich insights into human behaviour, they also have certain drawbacks including the delay with which data can be collected, the resources required to collect these data, or the extent to which people are willing to or able to report on their behaviour.

Widespread usage of technological devices and the online services they connect us to generate large volumes of "digital traces" such as social media posts or search engine history, drawing a detailed picture of social behaviour. This online data tends to be available at high speed and low cost serving as an alternative source for measuring human behaviour at a national or even global scale.

Most of the studies under the fast growing discipline of computational social science (King, 2011; Lazer et al., 2009; Moat et al., 2014) have focused on analysing data from search engines (Choi and Varian, 2012; Ginsberg et al., 2009), online encyclopedia (Mestyán et al., 2013; Moat et al., 2013) or text based posts shared on social media channels such as *Twitter* (Bollen et al., 2011a; Ciulla et al., 2012). However, in recent years, social media platforms such as *Instagram* and *Flickr* which enable online users to share visual media have become ubiquitous. Ignited by the expanding volume of photographs uploaded to social media platforms, numerous studies have been conducted analysing the metadata as well as the textual data attached to these online photographs (Alanyali et al., 2016; Barchiesi et al., 2015a; Preis et al., 2013a; Wood et al., 2013)

In light of these previous studies, in this thesis, we exploit a less explored form of online data: the photographs themselves. Exploiting large quantities of online photographic data, the studies we have showcased here provide a series of examples of how globally shared photographs and metadata attached to them can help us study social processes as they unfold, identify behaviour patterns at a national or global scale, and offer alternative methods for measuring human behaviour.

In recent years, we have witnessed major protest outbreaks sweeping across countries and continents. One common behavioural pattern across these protests is the in-

creased usage of social media channels. This increased usage is generating large volumes of data offering new avenues for measuring and understanding protest activity. In Chapter 3, we presented the first part of our study on identifying global protest outbreaks by exploiting data attached to pictures uploaded to the photo sharing platform *Flickr*. Our findings illustrate that larger numbers of pictures shared with the word "protest" in 34 different languages corresponds to higher proportions of protest related newspaper articles.

Concentrating solely on text based data carries certain restraints. For instance, text analysis is highly dependent on language hence making it a major limitation especially when performing a global study that involves analysing different languages. Furthermore, not every photograph is uploaded with text attached, in which case we can't use these data points, and hence potentially miss out valuable information contained in the pictures themselves. However, in most cases location and time information are embedded automatically in photographs especially if they were captured with a mobile device. Owing to the advances in computational power as well as the abundance of training data, powerful algorithms such as deep neural networks have been adapted to analyse image content bringing computers a step closer to human-like visual perception. As discussed in Chapter 2, an increasing number of studies exploiting these architectures provide examples demonstrating their power in solving a number of problems arising in computer vision including object detection and scene classification. Drawing on these state-of-the-art image analysis methods, in Chapter 4 we presented how we can track protest outbreaks by analysing the visual content of photographs shared on *Flickr*. We created a convolutional neural network based framework to automatically identify protest-related pictures to uncover a quantifiable relationship between the protest pictures uploaded to *Flickr* across 244 countries and regions and the protest-related news articles on the newspaper *The Guardian*. By analysing this new form of language independent data, we provide evidence that photographs shared on social media may contain signs of protest outbreaks.

We also provided a comparison of the models introduced in Chapters 3 and 4 which shows that information extracted from online pictures captures the change in the proportion of protest related news articles in a similar fashion as the information extracted from the text attached to the pictures. However, we have demonstrated that a better indicator can be created by combining information from both text and image analysis. These findings together with the results from text and image analysis are in line with the suggestion that data on photographs shared online may facilitate monitoring of real world events as they unfold.

In Chapter 3 and Chapter 4, we showed how geotagged and timestamped pictures can be utilised to create near-real time indicators of global events. One of the main challenges of this study was the lack of reliable ground truth data. Using newspaper data as a proxy for ground truth is one of the most common approaches adopted by previous studies analysing social unrest (Braha, 2012; Compton et al., 2014; Steinert-Threlkeld et al., 2015). However, as discussed in the previous sections, this data might have a location bias.

For instance, articles published in a newspaper usually cover more news about

the country or region in which the newspaper is released though they do also include a certain amount of news coverage about other countries. The newspaper therefore might miss certain news or do not prioritise covering them which indeed may be important for the location where they originated. Building on our results, future studies might explore whether it is possible to identify cases where information extracted from online images can give us more information about local protest outbreaks compared to the newspaper chosen as a proxy for ground truth.

The difficulty in building a classifier varies considerably depending on the classification problem. For instance, for both humans and computers, mostly it is quite clear from the picture whether it is food-related or not as there are many distinctive features of food. On the contrary, protests can come in many forms, from silently standing protesters that can easily be confused with pedestrians. Some protests have protesters in fancy dress which can look quite similar to a street carnival or a festival. These all include additional levels of ambiguity to the definition of a protest scene hence making the classification problem harder for computers even for humans. In order to eliminate the vagueness in definition, one approach might be to break the problem into smaller pieces. That is, to create a separate classifier for each pattern that could represent a protest scene, such as protest signs, presence of police and megaphones, followed by an ensemble like combination of these individual classifiers. Another potential approach to improve the performance of the classifier might be to use a CNN trained on a different dataset. Extracting features by using a CNN initially trained on a dataset different to the ImageNet dataset, such as the Places-365 dataset, might help to capture different visual aspects of an image hence improving the overall accuracy of the final classifier. For instance, a classifier trained on features extracted with a CNN trained on the Places-365 dataset may be better at capturing scene related traits in an image. An even more powerful classifier can be trained by combining different sets of features extracted from different CNNs. Future studies can also exploit alternative methods to eliminate noise from the automatically created training set. For example, as discussed in Chatfield et al. (2015), a distance-based outlier removal can be used to remove noisy training images. However, this should be done with care so as not to violate the diversity of the training data.

Nevertheless, our findings from Chapters 3 and 4 form an example of how photographs shared online can be used to monitor social processes around the world as they unfold. The methods we have presented here can be generalised to track other global events that people tend to document and share on social media, such as natural disasters. By building an image analysis tool to automatically detect destruction, it might be possible to create rapid estimates of the level of damage and identify areas in need of urgent help by inspecting online pictures taken at the disaster area. Provided that they detect destruction with a plausible accuracy, these automatic measurements would potentially provide valuable indicators to policy makers.

Governments traditionally use survey-based methods to collect information on population demographics and the socioeconomic status of a country. In addition to being a

costly process, reports are usually published with a delay due to the laborious work required to collect and process data at scale. Alternative approaches therefore have been proposed including estimates calculated on historical data and other social statistics that are considered to be linked to the statistics in question. For instance, the Greater London Authority calculates income estimates for London area by using various data sources including historical income estimates released by the Office for National Statistics, household deprivation extracted from 2001 and 2011 census data, median house selling prices from the Land Registry dataset and Her Majesty's Revenue and Customs child poverty data as well as income data calculated from the Understanding Society dataset. Despite providing valuable information in a timely manner compared to nationwide censuses conducted every ten years in the UK, creating such estimates still relies heavily on survey-based data.

In Chapter 5, we hypothesised that pictures shared online can serve as an alternative source supplementing official statistics on key socioeconomic measurements. By analysing photographs uploaded to *Instagram* as well as using methods developed in computer vision, we showed that visual information extracted from online pictures may help us estimate income in London and New York City.

This new source of information can reduce the time and money required to gain further insights into socioeconomic statistics, complementing information extracted via traditional methods. Decision makers in the governmental and commercial arenas can therefore immensely benefit from this alternative source of information to create reports on the current status of major cities around the world. Future research drawing on our approach may explore whether change in these socioeconomic statistics can be monitored and captured with a finer time granularity. For instance, can we identify areas that have been undergoing gentrification by exploiting a wider longitudinal dataset of online photographs?

Modern cities are plagued by a myriad of problems, from noisy neighbours to illegal parking and failing street lights. In recent years, a number of cities have introduced systems to help citizens report such problems in order to facilitate their rapid resolution. In Chapter 6, we exploited data on complaints made to New York City's 311 service. We found that data from 311 complaints can be used not only to generate a real-time overview of problems currently faced in the city, but also to predict where related problems may be reported next.

However, could rapidly available data from online photographs improve these forecasts further? For example, might the content of certain pictures be associated with loud noises, and therefore allow us to forecast the location of noise complaints? In our final analysis, we investigate whether we can identify visual signals in photographs to predict the location of noise complaints before they are reported.

Our results suggested that information extracted from photographs posted on *Flickr* can help us predict the location of noise-related complaints reported to the 311 service at a level better than randomly predicting the location of noise incidents. Although predictions on the basis on photographic data are better than random predictions, *Flickr* data does not appear to enhance the baseline spatiotemporal model that exploits historical 311 records. Once compared to the model using historical data, our results indicate that the model

exploiting solely historical records perform better compared to using solely *Flickr* data or the model using a combination of *Flickr* data and historical data. One reason for this might be the uneven distribution of photographic data. The majority of *Flickr* pictures from New York City are clustered around tourist attractions or parks. Although such locations might be rich in terms of the number of pictures, some of these areas have very few or no residents, hence not many incident reports. Future studies building on these image analysis methods might utilise another image set with better coverage of the New York City area.

Overall, these studies represent the results of a programme of research seeking to quantify human behaviour by analysing photographic data shared online. Here, we have showcased examples from detecting global events to estimating socioeconomic statistics and predicting the occurrence of noise complaints. Our findings underline the potential of online images as a source of cheap and rapidly available measure of human behaviour around the world.

APPENDIX A

Table A1. List of the country and region names as used in the analysis in Chapters 3 and 4.

Afghanistan	Aland	Albania
Algeria	American Samoa	Andorra
Angola	Anguilla	Antarctica
Antigua and Barbuda	Argentina	Armenia
Aruba	Ashmore and Cartier Islands	Australia
Austria	Azerbaijan	Bahrain
Bangladesh	Barbados	Belarus
Belgium	Belize	Benin
Bermuda	Bhutan	Bolivia
Bosnia and Herzegovina	Botswana	Brazil
British Indian Ocean Territory	British Virgin Islands	Brunei
Bulgaria	Burkina Faso	Burundi
Cambodia	Cameroon	Canada
Cape Verde	Cayman Islands	Central African Republic
Chad	Chile	China
Colombia	Comoros	Cook Islands
Costa Rica	Croatia	Cuba
Curacao	Cyprus	Czech Republic
Democratic Republic of the Congo	Denmark	Djibouti
Dominica	Dominican Republic	East Timor
Ecuador	Egypt	El Salvador
Equatorial Guinea	Eritrea	Estonia

Ethiopia	Falkland Islands	Faroe Islands
Federated States of Micronesia	Fiji	Finland
France	French Guiana	French Polynesia
French Southern and Antarctic Lands	Gabon	Gambia
Gaza	Georgia	Germany
Ghana	Greece	Greenland
Grenada	Guam	Guatemala
Guernsey	Guinea	Guinea Bissau
Guyana	Haiti	Heard Island and McDonald Islands
Honduras	Hong Kong S.A.R.	Hungary
Iceland	India	Indian Ocean Territories
Indonesia	Iran	Iraq
Ireland	Isle of Man	Israel
Italy	Ivory Coast	Jamaica
Japan	Jersey	Jordan
Kazakhstan	Kenya	Kiribati
Kosovo	Kuwait	Kyrgyzstan
Laos	Latvia	Lebanon
Lesotho	Liberia	Libya
Liechtenstein	Lithuania	Luxembourg
Macau S.A.R	Macedonia	Madagascar
Malawi	Malaysia	Maldives
Mali	Malta	Marshall Islands
Mauritania	Mauritius	Mexico
Moldova	Monaco	Mongolia
Montenegro	Montserrat	Morocco
Mozambique	Myanmar	Namibia
Nauru	Nepal	Netherlands
New Caledonia	New Zealand	Nicaragua
Niger	Nigeria	Niue
Norfolk Island	Northern Cyprus	Northern Mariana Islands
North Korea	Norway	Oman
Pakistan	Palau	Panama
Papua New Guinea	Paraguay	Peru
Philippines	Pitcairn Islands	Poland
Portugal	Puerto Rico	Qatar
Republic of Serbia	Republic of the Congo	Romania
Russia	Rwanda	Saint Barthelemy

Saint Helena	Saint Kitts and Nevis	Saint Lucia
Saint Martin	Saint Pierre and Miquelon	Saint Vincent and the Grenadines
Samoa	San Marino	Sao Tome and Principe
Saudi Arabia	Senegal	Seychelles
Siachen Glacier	Sierra Leone	Singapore
Sint Maarten	Slovakia	Slovenia
Solomon Islands	Somalia	Somaliland
South Africa	South Georgia and South Sandwich Islands	South Korea
South Sudan	Spain	Sri Lanka
Sudan	Suriname	Swaziland
Sweden	Switzerland	Syria
Taiwan	Tajikistan	Thailand
The Bahamas	Togo	Tonga
Trinidad and Tobago	Tunisia	Turkey
Turkmenistan	Turks and Caicos Islands	Tuvalu
Uganda	Ukraine	United Arab Emirates
United Kingdom	United Republic of Tanzania	United States of America
United States Virgin Islands	Uruguay	Uzbekistan
Vanuatu	Vatican	Venezuela
Vietnam	Wallis and Futuna	West Bank
Western Sahara	Yemen	Zambia
Zimbabwe		

Table A2. List of translations of the word ‘protest’ in different languages.

Protest Keyword	Language
اجاجها	Arabic
Protesta	Austrian
Протест	Bulgarian
Protest	Czech
Protest	German
Protest	Estonian
Protesta	Basque
اعراض	Persian
Protesta	Galician
항의	Korean
Mótmæli	Icelandic
Contestazione	Italian
מחאה	Hebrew
ಪಪಭಟನನ	Kannada
Қарсылық қозғалысы	Kazakh
Protestas	Lithuanian
Протест	Macedonian
Bantahan	Malay
Protest	Dutch
反対運動	Japanese
Protesto	Portuguese
Protest	Romanian
Протест	Russian
Protest	Simple English
Protest	Slovak
Protest	Serbo-Croatian
Protesti	Finnish
Protest	Swedish
எதததபபரபபபரபபபமர	Tamil
การประทวง	Thai
Protesto	Turkish
Протест	Ukranian
פראטעסט	Yiddish
抗議	Chinese

Table A3. List of the total number of *Flickr* photographs analysed per country and region in Chapter 3.

	Location	Photographs
1	Afghanistan	734
2	Aland	436
3	Albania	2,121
4	Algeria	1,419
5	American Samoa	60
6	Andorra	1,577
7	Angola	290
8	Anguilla	76
9	Antarctica	701
10	Antigua and Barbuda	1,065
11	Argentina	49,670
12	Armenia	1,165
13	Aruba	346
14	Ashmore and Cartier Islands	0
15	Australia	324,903
16	Austria	103,291
17	Azerbaijan	1,039
18	Bahrain	2,278
19	Bangladesh	5,679
20	Barbados	678
21	Belarus	5,187
22	Belgium	104,520
23	Belize	887
24	Benin	164
25	Bermuda	524
26	Bhutan	206
27	Bolivia	4,288
28	Bosnia and Herzegovina	2,134
29	Botswana	860
30	Brazil	336,260
31	British Indian Ocean Territory	16
32	British Virgin Islands	123
33	Brunei	2,377
34	Bulgaria	8,580
35	Burkina Faso	292
36	Burundi	197
37	Cambodia	12,672

38	Cameroon	373
39	Canada	404,408
40	Cape Verde	620
41	Cayman Islands	986
42	Central African Republic	91
43	Chad	77
44	Chile	63,386
45	China	149,705
46	Colombia	34,793
47	Comoros	10
48	Cook Islands	40
49	Costa Rica	12,221
50	Croatia	9,979
51	Cuba	2,701
52	Curacao	792
53	Cyprus	5,366
54	Czech Republic	65,739
55	Democratic Republic of the Congo	562
56	Denmark	43,290
57	Djibouti	246
58	Dominica	75
59	Dominican Republic	7,477
60	East Timor	329
61	Ecuador	11,058
62	Egypt	12,849
63	El Salvador	2,412
64	Equatorial Guinea	20
65	Eritrea	23
66	Estonia	10,811
67	Ethiopia	1,474
68	Falkland Islands	269
69	Faroe Islands	592
70	Federated States of Micronesia	48
71	Fiji	163
72	Finland	55,050
73	France	493,813
74	French Guiana	880
75	French Polynesia	871
76	French Southern and Antarctic Lands	0
77	Gabon	128
78	Gambia	399

79	Gaza	702
80	Georgia	4,751
81	Germany	494,288
82	Ghana	754
83	Greece	42,367
84	Greenland	484
85	Grenada	84
86	Guam	1,777
87	Guatemala	5,772
88	Guernsey	2,090
89	Guinea	243
90	Guinea Bissau	7
91	Guyana	600
92	Haiti	1,016
93	Heard Island and McDonald Islands	0
94	Honduras	2,082
95	Hong Kong S.A.R.	58,471
96	Hungary	41,437
97	Iceland	15,565
98	India	67,407
99	Indian Ocean Territories	21
100	Indonesia	58,313
101	Iran	3,354
102	Iraq	5,505
103	Ireland	94,009
104	Isle of Man	3,874
105	Israel	35,847
106	Italy	492,311
107	Ivory Coast	117
108	Jamaica	1,879
109	Japan	480,205
110	Jersey	2,035
111	Jordan	8,619
112	Kazakhstan	2,757
113	Kenya	4,793
114	Kiribati	7
115	Kosovo	793
116	Kuwait	10,300
117	Kyrgyzstan	836
118	Laos	3,931
119	Latvia	9,534

120	Lebanon	3,243
121	Lesotho	76
122	Liberia	57
123	Libya	577
124	Liechtenstein	292
125	Lithuania	4,706
126	Luxembourg	4,138
127	Macau S.A.R	36
128	Macedonia	1,536
129	Madagascar	1,575
130	Malawi	465
131	Malaysia	65,237
132	Maldives	680
133	Mali	197
134	Malta	4,760
135	Marshall Islands	0
136	Mauritania	220
137	Mauritius	2,188
138	Mexico	118,988
139	Moldova	834
140	Monaco	80
141	Mongolia	1,909
142	Montenegro	2,310
143	Montserrat	37
144	Morocco	10,613
145	Mozambique	1,141
146	Myanmar	5,769
147	Namibia	783
148	Nauru	11
149	Nepal	4,525
150	Netherlands	262,659
151	New Caledonia	415
152	New Zealand	62,590
153	Nicaragua	2,319
154	Niger	80
155	Nigeria	552
156	Niue	6
157	Norfolk Island	235
158	North Korea	1,206
159	Northern Cyprus	1,550
160	Northern Mariana Islands	225

161	Norway	64,758
162	Oman	2,935
163	Pakistan	6,258
164	Palau	48
165	Panama	5,693
166	Papua New Guinea	279
167	Paraguay	1,897
168	Peru	21,277
169	Philippines	70,957
170	Pitcairn Islands	2
171	Poland	55,780
172	Portugal	57,766
173	Puerto Rico	8,591
174	Qatar	9,963
175	Republic of Serbia	9,368
176	Republic of the Congo	128
177	Romania	22,819
178	Russia	124,744
179	Rwanda	950
180	Saint Barthelemy	67
181	Saint Helena	170
182	Saint Kitts and Nevis	295
183	Saint Lucia	214
184	Saint Martin	258
185	Saint Pierre and Miquelon	35
186	Saint Vincent and the Grenadines	50
187	Samoa	25
188	San Marino	609
189	Sao Tome and Principe	105
190	Saudi Arabia	17,735
191	Senegal	978
192	Seychelles	261
193	Siachen Glacier	0
194	Sierra Leone	70
195	Singapore	64,376
196	Sint Maarten	1,131
197	Slovakia	13,811
198	Slovenia	8,673
199	Solomon Islands	143
200	Somalia	14
201	Somaliland	25

202	South Africa	28,101
203	South Georgia and South Sandwich Islands	33
204	South Korea	121,170
205	South Sudan	60
206	Spain	490,356
207	Sri Lanka	6,128
208	Sudan	950
209	Suriname	444
210	Swaziland	387
211	Sweden	82,758
212	Switzerland	127,250
213	Syria	2,016
214	Taiwan	394,698
215	Tajikistan	191
216	Thailand	90,141
217	The Bahamas	733
218	Togo	116
219	Tonga	28
220	Trinidad and Tobago	2,824
221	Tunisia	3,328
222	Turkey	38,437
223	Turkmenistan	25
224	Turks and Caicos Islands	355
225	Tuvalu	0
226	Uganda	995
227	Ukraine	27,198
228	United Arab Emirates	20,948
229	United Kingdom	1,890,670
230	United Republic of Tanzania	4,822
231	United States of America	3,812,116
232	United States Virgin Islands	1,275
233	Uruguay	6,617
234	Uzbekistan	881
235	Vanuatu	193
236	Vatican	0
237	Venezuela	16,865
238	Vietnam	53,333
239	Wallis and Futuna	1
240	West Bank	8,854
241	Western Sahara	41
242	Yemen	357

243	Zambia	1,311
244	Zimbabwe	478

Table A4. List of the total number of *The Guardian* news articles per country and region.

	Location	News
1	Afghanistan	2,212
2	Aland	0
3	Albania	142
4	Algeria	412
5	American Samoa	23
6	Andorra	60
7	Angola	169
8	Anguilla	24
9	Antarctica	268
10	Antigua and Barbuda	8
11	Argentina	1,255
12	Armenia	105
13	Aruba	16
14	Ashmore and Cartier Islands	0
15	Australia	8,430
16	Austria	729
17	Azerbaijan	126
18	Bahrain	322
19	Bangladesh	808
20	Barbados	115
21	Belarus	235
22	Belgium	1,070
23	Belize	86
24	Benin	89
25	Bermuda	120
26	Bhutan	68
27	Bolivia	282
28	Bosnia and Herzegovina	54
29	Botswana	124
30	Brazil	3,076
31	British Indian Ocean Territory	7
32	British Virgin Islands	95
33	Brunei	67
34	Bulgaria	452
35	Burkina Faso	176
36	Burundi	90
37	Cambodia	319
38	Cameroon	208

39	Canada	2,621
40	Cape Verde	55
41	Cayman Islands	94
42	Central African Republic	160
43	Chad	245
44	Chile	687
45	China	5,871
46	Colombia	503
47	Comoros	21
48	Cook Islands	34
49	Costa Rica	274
50	Croatia	438
51	Cuba	527
52	Curacao	9
53	Cyprus	829
54	Czech Republic	387
55	Democratic Republic of the Congo	215
56	Denmark	928
57	Djibouti	37
58	Dominica	17
59	Dominican Republic	97
60	East Timor	39
61	Ecuador	379
62	Egypt	1,770
63	El Salvador	133
64	Equatorial Guinea	54
65	Eritrea	65
66	Estonia	210
67	Ethiopia	470
68	Falkland Islands	120
69	Faroe Islands	55
70	Federated States of Micronesia	2
71	Fiji	194
72	Finland	613
73	France	7,242
74	French Guiana	24
75	French Polynesia	9
76	French Southern and Antarctic Lands	0
77	Gabon	86
78	Gambia	79
79	Gaza	520

80	Georgia	688
81	Germany	5,569
82	Ghana	546
83	Greece	1,620
84	Greenland	149
85	Grenada	48
86	Guam	42
87	Guatemala	212
88	Guernsey	95
89	Guinea	622
90	Guinea Bissau	37
91	Guyana	61
92	Haiti	270
93	Heard Island and McDonald Islands	2
94	Honduras	188
95	Hong Kong S.A.R.	1,215
96	Hungary	469
97	Iceland	634
98	India	4,354
99	Indian Ocean Territories	0
100	Indonesia	1,167
101	Iran	1,873
102	Iraq	2,398
103	Ireland	4,979
104	Isle of Man	136
105	Israel	2,151
106	Italy	3,970
107	Ivory Coast	279
108	Jamaica	378
109	Japan	2,796
110	Jersey	1,419
111	Jordan	1,675
112	Kazakhstan	297
113	Kenya	1,134
114	Kiribati	25
115	Kosovo	169
116	Kuwait	250
117	Kyrgyzstan	69
118	Laos	99
119	Latvia	221
120	Lebanon	641

121	Lesotho	58
122	Liberia	201
123	Libya	885
124	Liechtenstein	77
125	Lithuania	223
126	Luxembourg	378
127	Macau S.A.R	0
128	Macedonia	128
129	Madagascar	144
130	Malawi	209
131	Malaysia	565
132	Maldives	74
133	Mali	790
134	Malta	233
135	Marshall Islands	19
136	Mauritania	71
137	Mauritius	116
138	Mexico	1,743
139	Moldova	210
140	Monaco	501
141	Mongolia	138
142	Montenegro	322
143	Montserrat	26
144	Morocco	317
145	Mozambique	201
146	Myanmar	106
147	Namibia	119
148	Nauru	147
149	Nepal	365
150	Netherlands	1,376
151	New Caledonia	30
152	New Zealand	2,477
153	Nicaragua	154
154	Niger	249
155	Nigeria	935
156	Niue	6
157	Norfolk Island	4
158	Northern Cyprus	18
159	Northern Mariana Islands	3
160	North Korea	725
161	Norway	1,070

162	Oman	136
163	Pakistan	1,758
164	Palau	40
165	Panama	248
166	Papua New Guinea	296
167	Paraguay	126
168	Peru	442
169	Philippines	816
170	Pitcairn Islands	0
171	Poland	1,155
172	Portugal	1,153
173	Puerto Rico	65
174	Qatar	796
175	Republic of Serbia	328
176	Republic of the Congo	220
177	Romania	660
178	Russia	3,699
179	Rwanda	359
180	Saint Barthelemy	0
181	Saint Helena	6
182	Saint Kitts and Nevis	5
183	Saint Lucia	10
184	Saint Martin	18
185	Saint Pierre and Miquelon	2
186	Saint Vincent and the Grenadines	5
187	Samoa	136
188	San Marino	184
189	Sao Tome and Principe	4
190	Saudi Arabia	749
191	Senegal	254
192	Seychelles	43
193	Siachen Glacier	0
194	Sierra Leone	226
195	Singapore	801
196	Sint Maarten	7
197	Slovakia	194
198	Slovenia	239
199	Solomon Islands	41
200	Somalia	612
201	Somaliland	43
202	South Africa	2,766

203	South Georgia and South Sandwich Islands	1
204	South Korea	877
205	South Sudan	187
206	Spain	3,877
207	Sri Lanka	641
208	Sudan	414
209	Suriname	34
210	Swaziland	42
211	Sweden	1,467
212	Switzerland	1,216
213	Syria	2,860
214	Taiwan	293
215	Tajikistan	38
216	Thailand	689
217	The Bahamas	105
218	Togo	113
219	Tonga	103
220	Trinidad and Tobago	56
221	Tunisia	338
222	Turkey	1,953
223	Turkmenistan	46
224	Turks and Caicos Islands	22
225	Tuvalu	24
226	Uganda	532
227	Ukraine	747
228	United Arab Emirates	247
229	United Kingdom	29,106
230	United Republic of Tanzania	367
231	United States of America	3,148
232	United States Virgin Islands	0
233	Uruguay	463
234	Uzbekistan	112
235	Vanuatu	31
236	Vatican	632
237	Venezuela	454
238	Vietnam	911
239	Wallis and Futuna	3
240	West Bank	484
241	Western Sahara	20
242	Yemen	467
243	Zambia	249

244	Zimbabwe	413

APPENDIX B

Table B1. List of the country names as used when querying the *Bing* Image Search API.

Afghanistan	Albania	Algeria
Andorra	Angola	Antigua and Barbuda
Argentina	Armenia	Australia
Austria	Azerbaijan	Bahamas
Bahrain	Bangladesh	Barbados
Belarus	Belgium	Belize
Benin	Bhutan	Bolivia
Bosnia and Herzegovina	Botswana	Brazil
Brunei	Bulgaria	Burkina Faso
Burundi	Cambodia	Cameroon
Canada	Cape Verde	Central African Republic
Chad	Chile	China
Colombia	Comoros	Democratic Republic of the Congo
Congo	Costa Rica	Ivory Coast
Croatia	Cuba	Cyprus
Czech Republic	Denmark	Djibouti
Dominica	Dominican Republic	Ecuador
Egypt	El Salvador	Equatorial Guinea
Eritrea	Estonia	Ethiopia
Fiji	Finland	France
Gabon	Gambia	Georgia
Germany	Ghana	Greece
Grenada	Guatemala	Guinea
Guinea-Bissau	Guyana	Haiti
Honduras	Hungary	Iceland
India	Indonesia	Iran

Iraq	Ireland	Israel
Italy	Jamaica	Japan
Jordan	Kazakhstan	Kenya
Kiribati	North Korea	South Korea
Kuwait	Kyrgyzstan	Laos
Latvia	Lebanon	Lesotho
Liberia	Libya	Liechtenstein
Lithuania	Luxembourg	Macedonia
Madagascar	Malawi	Malaysia
Maldives	Mali	Malta
Marshall Islands	Mauritania	Mauritius
Mexico	F.S. Micronesia	Moldova
Monaco	Mongolia	Montenegro
Morocco	Mozambique	Myanmar
Namibia	Nauru	Nepal
Netherlands	New Zealand	Nicaragua
Niger	Nigeria	Norway
Oman	Pakistan	Palau
Palestine	Panama	Papua New Guinea
Paraguay	Peru	Philippines
Poland	Portugal	Qatar
Romania	Russia	Rwanda
Saint Kitts and Nevis	Saint Lucia	Saint Vincent and the Grenadines
Samoa	San Marino	So Tom and Prncipe
Saudi Arabia	Senegal	Serbia
Seychelles	Sierra Leone	Singapore
Slovakia	Slovenia	Solomon Islands
Somalia	South Africa	South Sudan
Spain	Sri Lanka	Sudan
Suriname	Swaziland	Sweden
Switzerland	Syria	Taiwan
Tajikistan	Tanzania	Thailand
Timor-Leste	Togo	Tonga
Trinidad and Tobago	Tunisia	Turkey
Turkmenistan	Tuvalu	Uganda
Ukraine	United Arab Emirates	United Kingdom
United States	Uruguay	Uzbekistan
Vanuatu	Venezuela	Vietnam
Western Sahara Sahrawi	Yemen	Zambia

Table B2. List of the total number of *Flickr* pictures analysed per country and region in Chapter 4.

	Location	Photographs
1	Afghanistan	665
2	Aland	379
3	Albania	1,661
4	Algeria	1,048
5	American Samoa	41
6	Andorra	1,354
7	Angola	220
8	Anguilla	126
9	Antarctica	565
10	Antigua and Barbuda	891
11	Argentina	39,318
12	Armenia	1,089
13	Aruba	333
14	Ashmore and Cartier Islands	0
15	Australia	275,765
16	Austria	69,351
17	Azerbaijan	808
18	Bahrain	1,844
19	Bangladesh	4,480
20	Barbados	549
21	Belarus	4,398
22	Belgium	80,781
23	Belize	793
24	Benin	82
25	Bermuda	372
26	Bhutan	194
27	Bolivia	3,478
28	Bosnia and Herzegovina	1,385
29	Botswana	671
30	Brazil	267,537
31	British Indian Ocean Territory	3
32	British Virgin Islands	71
33	Brunei	2,029
34	Bulgaria	6,443
35	Burkina Faso	204
36	Burundi	161
37	Cambodia	10,327

38	Cameroon	377
39	Canada	322,450
40	Cape Verde	417
41	Cayman Islands	727
42	Central African Republic	58
43	Chad	65
44	Chile	49,959
45	China	123,865
46	Colombia	29,077
47	Comoros	8
48	Cook Islands	41
49	Costa Rica	10,104
50	Croatia	7,814
51	Cuba	2,221
52	Curacao	472
53	Cyprus	3,443
54	Czech Republic	46,440
55	Democratic Republic of the Congo	428
56	Denmark	29,752
57	Djibouti	202
58	Dominica	44
59	Dominican Republic	5,890
60	East Timor	316
61	Ecuador	9,294
62	Egypt	10,230
63	El Salvador	2,207
64	Equatorial Guinea	23
65	Eritrea	20
66	Estonia	8,988
67	Ethiopia	1,342
68	Falkland Islands	217
69	Faroe Islands	489
70	Federated States of Micronesia	44
71	Fiji	321
72	Finland	39,763
73	France	382,639
74	French Guiana	1,263
75	French Polynesia	654
76	French Southern and Antarctic Lands	0
77	Gabon	124
78	Gambia	300

79	Gaza	601
80	Georgia	3,769
81	Germany	362,368
82	Ghana	769
83	Greece	34,426
84	Greenland	421
85	Grenada	66
86	Guam	1,132
87	Guatemala	5,473
88	Guernsey	947
89	Guinea	218
90	Guinea Bissau	4
91	Guyana	578
92	Haiti	683
93	Heard Island and McDonald Islands	0
94	Honduras	1,392
95	Hong Kong S.A.R.	46,465
96	Hungary	27,609
97	Iceland	12,655
98	India	57,353
99	Indian Ocean Territories	21
100	Indonesia	48,525
101	Iran	2,752
102	Iraq	4,058
103	Ireland	69,514
104	Isle of Man	2,793
105	Israel	28,048
106	Italy	409,448
107	Ivory Coast	92
108	Jamaica	1,259
109	Japan	421,180
110	Jersey	1,314
111	Jordan	7,481
112	Kazakhstan	2,289
113	Kenya	3,413
114	Kiribati	2
115	Kosovo	561
116	Kuwait	8,835
117	Kyrgyzstan	566
118	Laos	3,312
119	Latvia	8,812

120	Lebanon	2,494
121	Lesotho	59
122	Liberia	39
123	Libya	497
124	Liechtenstein	215
125	Lithuania	3,247
126	Luxembourg	3,283
127	Macau S.A.R	40
128	Macedonia	1,233
129	Madagascar	1,101
130	Malawi	353
131	Malaysia	53,507
132	Maldives	585
133	Mali	159
134	Malta	3,233
135	Marshall Islands	0
136	Mauritania	196
137	Mauritius	1,444
138	Mexico	96,216
139	Moldova	728
140	Monaco	72
141	Mongolia	1,616
142	Montenegro	1,688
143	Montserrat	36
144	Morocco	7,610
145	Mozambique	1,067
146	Myanmar	5,129
147	Namibia	655
148	Nauru	9
149	Nepal	3,590
150	Netherlands	193,646
151	New Caledonia	210
152	New Zealand	51,456
153	Nicaragua	1,941
154	Niger	68
155	Nigeria	464
156	Niue	0
157	Norfolk Island	201
158	North Korea	1,376
159	Northern Cyprus	155
160	Northern Mariana Islands	651

161	Norway	48,715
162	Oman	2,227
163	Pakistan	4,664
164	Palau	44
165	Panama	4,929
166	Papua New Guinea	247
167	Paraguay	1,450
168	Peru	18,198
169	Philippines	57,671
170	Pitcairn Islands	2
171	Poland	40,342
172	Portugal	45,878
173	Puerto Rico	6,913
174	Qatar	7,088
175	Republic of Serbia	7,462
176	Republic of the Congo	40
177	Romania	16,633
178	Russia	97,319
179	Rwanda	856
180	Saint Barthelemy	58
181	Saint Helena	166
182	Saint Kitts and Nevis	37
183	Saint Lucia	148
184	Saint Martin	130
185	Saint Pierre and Miquelon	24
186	Saint Vincent and the Grenadines	46
187	Samoa	15
188	San Marino	436
189	Sao Tome and Principe	91
190	Saudi Arabia	14,604
191	Senegal	761
192	Seychelles	193
193	Siachen Glacier	0
194	Sierra Leone	70
195	Singapore	50,926
196	Sint Maarten	1,376
197	Slovakia	8,893
198	Slovenia	6,769
199	Solomon Islands	102
200	Somalia	9
201	Somaliland	23

202	South Africa	21,305
203	South Georgia and South Sandwich Islands	16
204	South Korea	106,736
205	South Sudan	48
206	Spain	392,710
207	Sri Lanka	4,996
208	Sudan	601
209	Suriname	445
210	Swaziland	416
211	Sweden	60,916
212	Switzerland	95,385
213	Syria	1,821
214	Taiwan	321,110
215	Tajikistan	152
216	Thailand	71,136
217	The Bahamas	553
218	Togo	86
219	Tonga	29
220	Trinidad and Tobago	2,630
221	Tunisia	2,092
222	Turkey	28,924
223	Turkmenistan	15
224	Turks and Caicos Islands	295
225	Tuvalu	0
226	Uganda	751
227	Ukraine	19,762
228	United Arab Emirates	16,735
229	United Kingdom	1,412,387
230	United Republic of Tanzania	3,397
231	United States of America	3,031,219
232	United States Virgin Islands	884
233	Uruguay	4,534
234	Uzbekistan	700
235	Vanuatu	128
236	Vatican	0
237	Venezuela	13,602
238	Vietnam	43,310
239	Wallis and Futuna	1
240	West Bank	7,450
241	Western Sahara	13
242	Yemen	250

243	Zambia	723
244	Zimbabwe	331

B.1 Picture credits

Picture credits by *Flickr* user names from left to right: top: *JPetram* (CC BY-NC-ND); *paul_clarke* (CC BY-NC-ND); *ubiquit* (CC BY-SA), middle: *Steve Rhodes* (CC BY-NC-ND); *tmscblpz* (CC BY), bottom: *The All Nite Images* (CC BY-SA); *David Krawczyk* (CC BY-NC-SA); *Pictures of a pair of wandering travel mice* (CC BY-NC). We resized the pictures by keeping the aspect ratio and cropped them to adapt to the frames. To view a copy of these licenses, visit `https://creativecommons.org/licenses/by-nc-sa/4.0/` for CC BY-NC-SA; `https://creativecommons.org/licenses/by-nc/4.0/` for CC BY-NC; `https://creativecommons.org/licenses/by-nd/4.0/` for CC BY-NC-ND; `https://creativecommons.org/licenses/by/4.0/` for CC BY; `https://creativecommons.org/licenses/by-sa/2.0/` for CC BY-SA; `https://creativecommons.org/licenses/by-nd/4.0/` for CC BY-ND.

APPENDIX C

Table C1. Subcategories of restaurants on *Yelp*.

Afghan (afghani)	Indian (indpak)
African (african)	Indonesian (indonesian)
American (New) (newamerican)	International (international)
American (Traditional) (tradamerican)	Irish (irish)
Arabian (arabian)	Italian (italian)
Argentine (argentine)	Japanese (japanese)
Armenian (armenian)	Ramen (ramen)
Asian Fusion (asianfusion)	Kebab (kebab)
Australian (australian)	Korean (korean)
Austrian (austrian)	Kosher (kosher)
BBQ & Barbecue (bbq)	Laotian (laotian)
Bangladeshi (bangladeshi)	Latin American (latin)
Basque (basque)	Live & Raw Food (raw_food)
Belgian (belgian)	Malaysian (malaysian)
Bistros (bistros)	Mediterranean (mediterranean)
Brasserie (brasseries)	Falafel (falafel)
Brazilian (brazilian)	Mexican (mexican)
Breakfast & Brunch (breakfast_brunch)	Middle Eastern (mideastern)
British (british)	Lebanese (lebanese)
Buffet (buffets)	Modern European (modern_european)
Bulgarian (bulgarian)	Mongolian (mongolian)
Burgers (burgers)	Moroccan (moroccan)
Burmese (burmese)	Nicaraguan (nicaraguan)
Cafes (cafes)	Noodles (noodles)
Cafeterias (cafeteria)	Pakistani (pakistani)
Cajun/Creole (cajun)	Pan Asian (panasian)
Cambodian (cambodian)	Persian/Iranian (persian)

Caribbean (caribbean)
Cheesesteaks (cheesesteaks)
Chicken Shop (chickenshop)
Chicken Wings (chicken_wings)
Chinese (chinese)
Cantonese (cantonese)
Dim Sum (dimsum)
Szechuan (szechuan)
Crepes (creperies)
Cuban (cuban)
Czech (czech)
Delis (delis)
Diners (diners)
Dinner Theater (dinnertheater)
Ethiopian (ethiopian)
Filipino (filipino)
Fish & Chips (fishnchips)
Fondue (fondue)
Food Courts (food_court)
Food Stands (foodstands)
French (french)
Mauritius (mauritius)
Reunion (reunion)
Game Meat (gamemeat)
Gastro Pubs (gastropubs)
Georgian (georgian)
German (german)
Gluten Free (gluten_free)
Greek (greek)
Guamanian (guamanian)
Halal (halal)
Hawaiian (hawaiian)
Himalayan/Nepalese (himalayan)
Honduran (honduran)
Hot Dogs (hotdog)
Hungarian (hungarian)
Peruvian (peruvian)
Pizza (pizza)
Polish (polish)
Pop-Up Restaurants (popuprestaurants)
Portuguese (portuguese)
Russian (russian)
Salad (salad)
Sandwiches (sandwiches)
Scandinavian (scandinavian)
Scottish (scottish)
Seafood (seafood)
Singaporean (singaporean)
Slovakian (slovakian)
Soul Food (soulfood)
Soup (soup)
Southern (southern)
Spanish (spanish)
Sri Lankan (srilankan)
Steakhouses (steak)
Supper Clubs (supperclubs)
Sushi (sushi)
Syrian (syrian)
Taiwanese (taiwanese)
Takeaway & Fast Food (hotdogs)
Tapas & Small Plates (tapasmallplates)
Tapas Bars (tapas)
Tex-Mex (tex-mex)
Thai (thai)
Turkish (turkish)
Ukranian (ukrainian)
Vegan (vegan)
Vegetarian (vegetarian)
Venison (venison)
Vietnamese (vietnamese)
Waffles (waffles)

C.1 Picture Credits

Picture credits by *Flickr* user names from top to bottom: *andyaldridge* (CC BY-NC-SA); *alalsacienne* (CC BY-NC-ND); *El Villano* (CC BY-NC-SA); *tompagenet* (CC BY-SA). We resized the pictures by keeping the aspect ratio fixed. To view a copy of these licenses, visit `https://creativecommons.org/licenses/by-nc-sa/4.0/` for CC BY-NC-SA; `https://creativecommons.org/licenses/by-nd/4.0/` for CC BY-NC-ND; `https://creativecommons.org/licenses/by-sa/2.0/` for CC BY-SA.

C.2 Picture Credits

Picture credits by *Flickr* user names from left to right: **(a)** top: *Ozzy Delaney* (CC BY); *DG Jones* (CC BY-NC-SA); *mockduck* (CC BY-NC), middle: *steve.wilde* (CC BY-ND); *Martin Pettitte* (CC BY), bottom: *Rooney Dog* (CC BY-ND); *FixersUK* (CC BY-ND). **(b)** top: *miyagawa* (CC BY-SA); *partiallyblind* (CC BY-NC-SA), middle: *Much Rambling* (CC BY-ND); *Htchoi_430* (CC BY-NC-ND), bottom: *whatleydude* (CC BY); *Shakespearesmonkey* (CC BY-NC). Apart from resizing by keeping the aspect ratio fixed, we have not made changes on the photographs. To view a copy of these licenses, visit `https://creativecommons.org/licenses/by-nc-sa/4.0/` for CC BY-NC-SA; `https://creativecommons.org/licenses/by-nc/4.0/` for CC BY-NC; `https://creativecommons.org/licenses/by-nd/4.0/` for CC BY-NC-ND; `https://creativecommons.org/licenses/by/4.0/` for CC BY; `https://creativecommons.org/licenses/by-sa/2.0/` for CC BY-SA; `https://creativecommons.org/licenses/by-nd/4.0/` for CC BY-ND.

APPENDIX D

Table D1. All incident categories as they are recorded by the New York City's 311 services in alphabetical order. Due to the case sensitive nature of NYC's 311 records, there are multiple categories with the same name.

Adopt-A-Basket	Illegal Parking
AGENCY	Illegal Tree Damage
Agency Issues	Indoor Air Quality
Air Quality	Indoor Sewage
Alzheimer's Care	Industrial Waste
Animal Abuse	Investigations and Discipline (IAD)
Animal Facility - No Permit	Invitation
Animal in a Park	Lead
APPLIANCE	Legal Services Provider Complaint
Asbestos	Lifeguard
Beach/Pool/Sauna Complaint	Literature Request
Benefit Card Replacement	Litter Basket / Request
Bereavement Support Group	Maintenance or Facility
BEST/Site Safety	Misc. Comments
Bike Rack Condition	Miscellaneous Categories
Bike/Roller/Skate Chronic	Missed Collection (All Materials)
Blocked Driveway	Mold
Boilers	Municipal Parking Facility
Bottled Water	New Tree Request
Bridge Condition	Noise
Broken Muni Meter	Noise - Commercial
Broken Parking Meter	Noise - Helicopter
Building Condition	Noise - House of Worship
Building/Use	Noise - Park
Bus Stop Shelter Placement	Noise - Residential
Calorie Labeling	Noise - Street/Sidewalk

Case Management Agency Complaint
City Vehicle Placard Complaint
Collection Truck Noise
Comment
Complaint
Compliment
Construction
CONSTRUCTION
Consumer Complaint
Cranes and Derricks
Curb Condition
Damaged Tree
DCA / DOH New License Application Request
DCA Literature Request
Dead Tree
DEP Literature Request
Derelict Bicycle
Derelict Vehicle
Derelict Vehicles
DFTA Literature Request
DHS Advantage -Landlord/Broker
DHS Advantage - Tenant
DHS Advantage - Third Party
DHS Income Savings Requirement
Dirty Conditions
Disorderly Youth
DOF Literature Request
DOF Parking - Address Update
DOF Parking - DMV Clearance
DOF Parking - Payment Issue
DOF Parking - Request Copy
DOF Parking - Request Status
DOF Parking - Tax Exemption
DOF Property - City Rebate
DOF Property - Owner Issue
DOF Property - Payment Issue
DOF Property - Property Value
DOF Property - Reduction Issue
DOF Property - Request Copy
DOF Property - RPIE Issue
Noise Survey
Noise - Vehicle
NONCONST
Non-Emergency Police Matter
Non-Residential Heat
NORC Complaint
OEM Literature Request
Open Flame Permit
Opinion for the Mayor
Other Enforcement
OUTSIDE BUILDING
Overflowing Litter Baskets
Overflowing Recycling Baskets
Overgrown Tree/Branches
PAINT - PLASTER
PAINT/PLASTER
Panhandling
Parking Card
Plant
Plumbing
PLUMBING
Poison Ivy
Portable Toilet
Posting Advertisement
Public Assembly
Public Assembly - Temporary
Public Payphone Complaint
Public Toilet
Radioactive Material
Rangehood
Recycling Enforcement
Request for Information
Request Xmas Tree Collection
Rodent
Root/Sewer/Sidewalk Condition
SAFETY
Sanitation Condition
Scaffold Safety
School Maintenance
SCRIE

DOOR/WINDOW
DOT Literature Request
DPR Internal
DPR Literature Request
Drinking
Drinking Water
EAP Inspection - F59
Elder Abuse
ELECTRIC
Electrical
Elevator
ELEVATOR
Emergency Response Team (ERT)
Ferry Complaint
Ferry Inquiry
Ferry Permit
Fire Alarm - Addition
Fire Alarm - Modification
Fire Alarm - New System
Fire Alarm - Reinspection
Fire Alarm - Replacement
Fire Safety Director - F58
FLOORING/STAIRS
Food Establishment
Food Poisoning
Forensic Engineering
For Hire Vehicle Complaint
For Hire Vehicle Report
Forms
Found Property
Gas Station Discharge Lines
GENERAL
GENERAL CONSTRUCTION
General Construction/Plumbing
Graffiti
Harboring Bees/Wasps
Hazardous Materials
Hazmat Storage/Use
HEAP Assistance
HEAT/HOT WATER
HEATING
Senior Center Complaint
Sewer
SG-98
SG-99
Sidewalk Condition
Smoking
Snow
SNW
Special Enforcement
Special Natural Area District (SNAD)
Special Projects Inspection Team (SPIT)
Sprinkler - Mechanical
Squeegee
Stalled Sites
Standing Water
Standpipe - Mechanical
Street Condition
Street Light Condition
Street Sign - Damaged
Street Sign - Dangling
Street Sign - Missing
STRUCTURAL
Summer Camp
Sweeping/Inadequate
Sweeping/Missed
Sweeping/Missed-Inadequate
Tanning
Tattooing
Taxi Complaint
Taxi Compliment
Taxi Report
Teaching/Learning/Instruction
Traffic
Traffic Signal Condition
Trans Fat
Transportation Provider Complaint
Tunnel Condition
Unleashed Dog
Unlicensed Dog
Unsanitary Animal Facility
Unsanitary Animal Pvt Property

Highway Condition	UNSANITARY CONDITION
Highway Sign - Damaged	Unsanitary Pigeon Condition
Highway Sign - Dangling	Urinating in Public
Highway Sign - Missing	Utility Program
Home Care Provider Complaint	VACANT APARTMENT
Home Delivered Meal Complaint	Vacant Lot
Home Delivered Meal - Missed Delivery	Vending
Homeless Encampment	Violation of Park Rules
Homeless Person Assistance	Water Conservation
Home Repair	WATER LEAK
Housing - Low Income Senior	Water Quality
Housing Options	Water System
HPD Literature Request	Weatherization
Illegal Animal Kept as Pet	Window Guard
Illegal Animal Sold	X-Ray Machine/Equipment
Illegal Fireworks	

Table D2. Incident categories which were grouped together during analysis in Chapter 6.

Original Category	New Category
Noise	Noise
Noise-Residential	
Noise-Commercial	
Noise-Street/Sidewalk	
Noise-Vehicle	
Noise-Park	
Noise-Helicopter	
Noise-House of Worship	
Noise Survey	
Collection Truck Noise	
Sanitation Condition	Unsanitary Conditions
Dirty Conditions	
Unsanitary Condition	
Unsanitary Animal Facility	
Unsanitary Pigeon Condition	
Unsanitary Animal pvt Property	
Heating	Heating
Heat/Hot Water	
Non-Residential Heat	
Water System	Water System
Water Leak	
Water Conservation	
Standing Water	
Water Quality	
Drinking Water	
Bottled Water	
Electric	Electric
Electrical	
Plumbing	Plumbing
General Construction/Plumbing	
Sewer	Sewer
Root/Sewer/Sidewalk Condition	
Construction	Construction
General Construction	
Paint-Plaster	Paint
Paint/Plaster	

Table D3. Descriptions of incidents falling within the 14 incident categories analysed.

Category	Description
Heating	Residents of New York City can make complaints about a lack of heating or hot water in residential buildings between October and May. They can also report if the heating is left on during summer months. This service can be used by tenants of rented property, or owners of apartments who are experiencing problems due to issues with maintenance of the building in which the apartment is situated. The service is also aimed at users of commercial and non-residential buildings such as senior centres and day care centres. Residents are asked to try to resolve such issues with landlords, managing agents or superintendents in the first instance, but mechanisms exist for New York City to act at the building owner's expense if no solution is forthcoming (City of New York, 2018f).
Noise	Complaints about various sources of noise as listed in Table D2, such as noise from neighbours (City of New York, 2018i) or noise from the street City of New York (2018j).
Plumbing	Complaints about plumbing work carried out by a licensed plumber but without a valid permit (City of New York, 2018l), or about a licensed plumber carrying out work incorrectly (City of New York, 2018h).
Street Condition	Complaints about the condition of a street, covering problems such as potholes, cave-ins and street surface damage caused by utility companies (City of New York, 2018m).
Street Light Condition	Reports of broken, defective or fallen street lights as well as requests for new street light installations (City of New York, 2018q).

Unsanitary Conditions	Under the provisions of the process described for heating complaints, residents of New York City can complain about various maintenance problems affecting an apartment which they rent, or an entire residential building, if these problems have not been satisfactorily addressed by the landlord or the building owner. This includes unsanitary conditions such as mould, sewage, or pests. Legal action can be taken against landlords or building owners who do not resolve such issues (City of New York, 2018n). In addition, residents can also report complaints about unsanitary conditions caused by animals. For instance, residents can report pigeon droppings on window ledges, sidewalks and exteriors of commercial and residential properties, as property owners are required to clean up unsanitary pigeon conditions such as excessive droppings on their properties City of New York (2018k). They are also required to clean up animal waste on their property, including adjoining sidewalks and gutters, even if it is not caused by their own pet (City of New York, 2018d). Dog walkers are also required to pick up waste deposited by their dogs while walking and breaches of these rules can be reported to the 311 services (City of New York, 2018d). The 311 services also accept complaints about dirty animal facilities, including animal shelters, groomers and petting zoos (City of New York, 2018a).
Paint	Problems with peeling paint on walls and ceilings in residential buildings can be reported as a maintenance issue under the same provisions. City of New York (2018n)
Construction	Complaints relating to issues with construction in the city, including construction taking place after hours or against approved plans City of New York (2018c).
Blocked Driveway	Complaints about cars partially or completely blocking residential driveways (City of New York, 2018s).

Water System	Residents in apartment buildings can report issues with their water system that have not been resolved by the landlord, such as a lack of water, or low water pressure. Residents in private and commercial buildings can also report such problems where these have been deemed not to be due to internal problems in the building (City of New York, 2018t). Other water related problems in this category include complaints about the quality and safety of purchased bottled water (City of New York, 2018b), concerns about excessive water usage in the context of the city's water conservation initiatives (City of New York, 2018u) and water that a property owner has left standing for over 5 days in which mosquitoes might be able to breed (City of New York, 2018p).
Illegal Parking	Complaints about non-emergency vehicles that are illegally parked and commercial vehicles parked on a residential road (City of New York, 2018g).
Traffic Signal Condition	Complaints about the conditions of traffic lights such as lights changing out-of-sequence. Requests can also be made for new light installations or a review of the timing of a traffic signal (City of New York, 2018r).
Sewer	Complaints about sewer related problems such as damaged sewers, leaks or sewer odour (City of New York, 2018o).
Electric	Residents in apartment buildings can report issues with their electrics that have not been resolved by the landlord. This category also contains reports of broader electrical issues such as power outages (City of New York, 2018e).

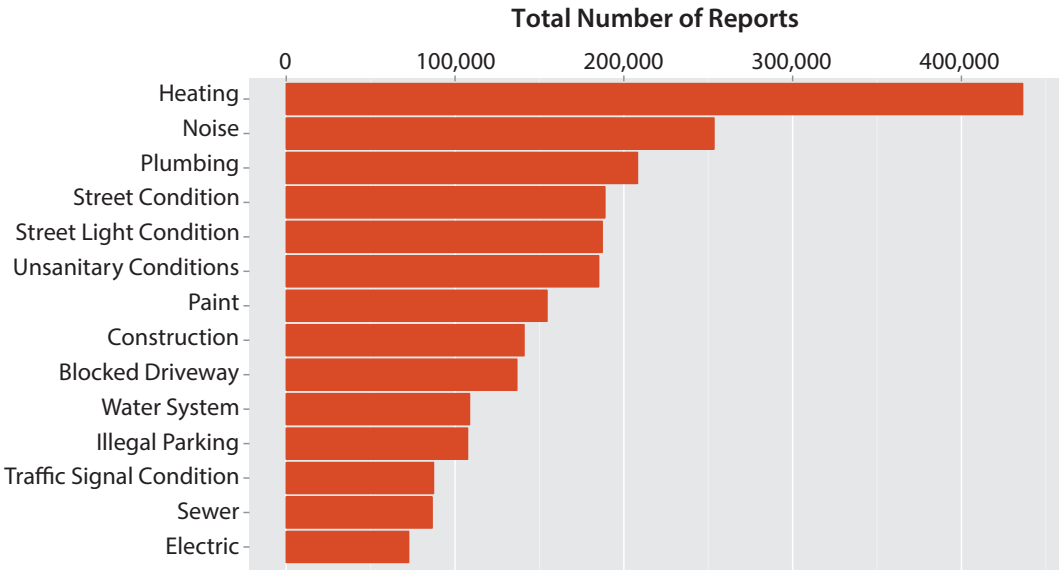

Figure D1. Total number of 311 reports per incident category. The total number of reports made to New York City's 311 services during 2013 and 2014 for the 14 incident categories covered in this study. Incident categories referring to related issues are merged as described under Chapter 6. The final subset of incidents covers over 65% of all incidents reported during 2013 and 2014.

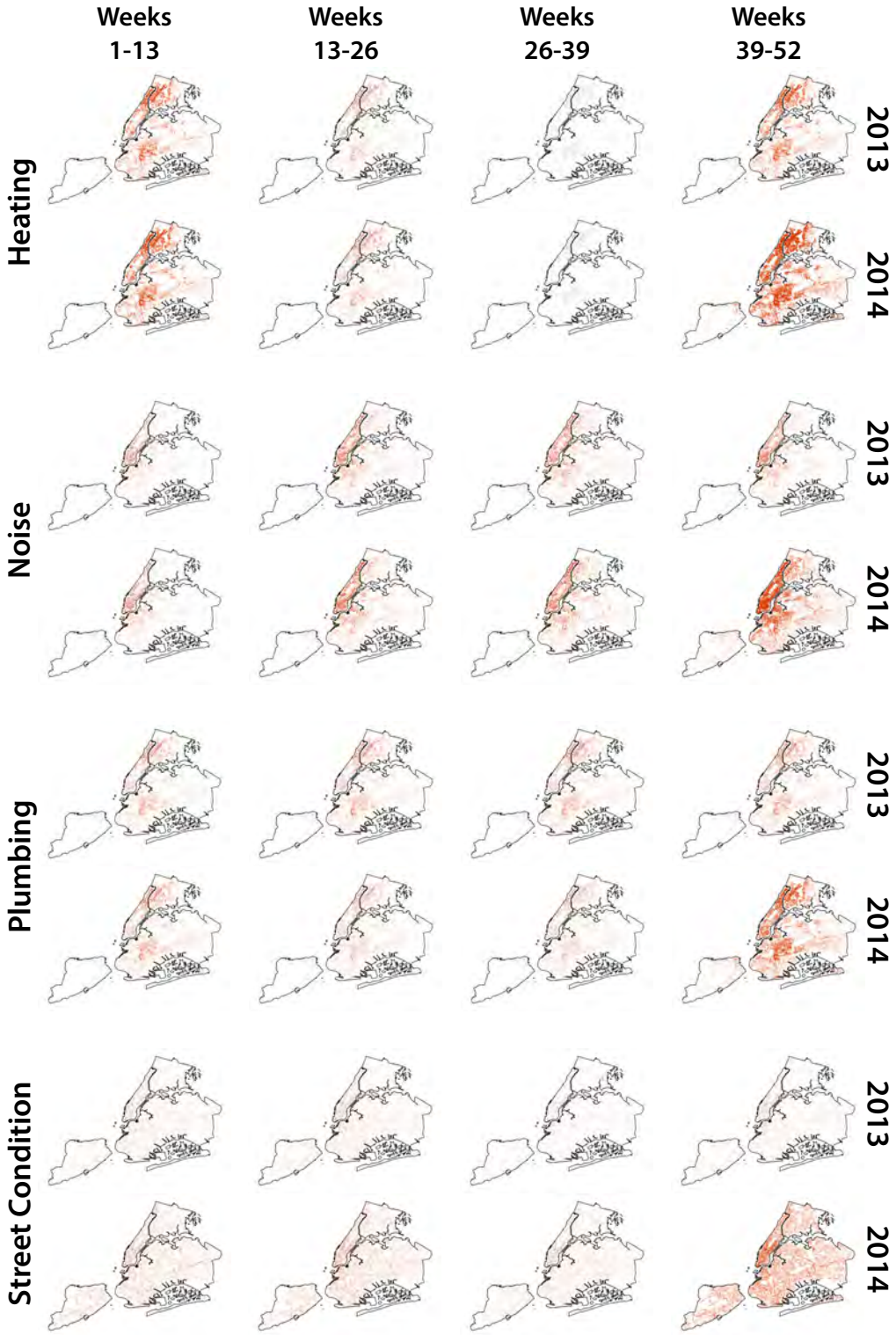

Figure D2. Time and location of incidents reported using the 311 service, visualised using small multiples. We visualise the times and locations of the four most frequent categories of incidents reported to the 311 service across 2013 and 2014. Again, these are Heating, Noise, Plumbing, and Street Condition.

References

Aiello, L. M., Schifanella, R., Quercia, D., and Aletta, F. (2016). Chatty maps: constructing sound maps of urban areas from social media data. *Royal Society Open Science*, 3(3):150690.

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

Alanyali, M., Preis, T., and Moat, H. S. (2016). Tracking protests using geotagged Flickr photographs. *PLOS ONE*, 11(3):e0150466.

Alashkar, T., Jiang, S., Wang, S., and Fu, Y. (2017). Examples-rules guided deep neural network for makeup recommendation. In *Proceedings of the Thirty-first AAAI Conference on Artificial Intelligence*, pages 941–947. AAAI.

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

Angus, E., Thelwall, M., and Stuart, D. (2008). General patterns of tag usage among university groups in Flickr. *Online Information Review*, 32(1):89–101.

Antenucci, D., Cafarella, M., Levenstein, M., Ré, C., and Shapiro, M. D. (2014). Using social media to measure labor market flows. Working paper 20010, National Bureau of Economic Research.

Aral, S. and Walker, D. (2012). Identifying influential and susceptible members of social networks. *Science*, page 1215842.

Arietta, S. M., Efros, A. A., Ramamoorthi, R., and 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.

Arthur, C. (2011). Egypt blocks social media websites in attempted clampdown on unrest. Available: `http://www.theguardian.com/world/2011/jan/26/egypt-blocks-social-media-websites`. Accessed: 2015-01-29.

Askitas, N. and Zimmermann, K. F. (2009). Google econometrics and unemployment forecasting. *Applied Economics Quarterly*, 55(2):107–120.

Bakhshi, S., Shamma, D. A., and Gilbert, E. (2014). Faces engage us: Photos with faces attract more likes and comments on Instagram. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 965–974. ACM.

Ballard, D. H. and Brown, C. M. (1982). *Computer Vision*. Prentice-Hall, Inc.

Barchiesi, D., Moat, H. S., Alis, C., Bishop, S., and Preis, T. (2015a). Quantifying international travel flows using Flickr. *PLOS ONE*, 10(7):e0128470.

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

Bentley, R. A., O'Brien, M. J., and Brock, W. A. (2014). Mapping collective behavior in the big-data era. *Behavioral and Brain Sciences*, 37(1):63–76.

Blumenstock, J. and Eagle, N. (2010). Mobile divides: gender, socioeconomic status, and mobile phone use in Rwanda. In *Proceedings of the 4th ACM/IEEE International Conference on Information and Communication Technologies and Development*, page 6. ACM.

Boichak, O. (2017). Battlefront volunteers: Mapping and deconstructing civilian resilience networks in Ukraine. In *Proceedings of the 8th International Conference on Social Media & Society*, page 3. ACM.

Bollen, J., Mao, H., and Pepe, A. (2011a). Modeling public mood and emotion : Twitter sentiment and socio-economic phenomena. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, pages 450–453. AAAI.

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

Bordino, I., Battiston, S., Caldarelli, G., Cristelli, M., Ukkonen, A., and Weber, I. (2012). Web search queries can predict stock market volumes. *PLOS ONE*, 7(7):e40014.

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

Bottou, L. and Bousquet, O. (2008). The tradeoffs of large scale learning. In *Advances in Neural Information Processing Systems*, pages 161–168.

Bourlard, H. and Wellekens, C. J. (1989). Links between Markov models and multilayer perceptrons. In *Advances in Neural Information Processing Systems*, pages 502–510.

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

Braha, D. (2012). Global civil unrest: contagion, self-organization, and prediction. *PLOS ONE*, 7(10):e48596.

Branigan, T. (2009). China blocks Twitter, Flickr and Hotmail ahead of Tiananmen Anniversary. Available: `http://www.theguardian.com/technology/2009/jun/02/twitter-china`. Accessed: 2015-01-29.

Budak, C. and Watts, D. J. (2015). Dissecting the spirit of Gezi: Influence vs. selection in the Occupy Gezi movement. *Sociological Science*, 2:370–397.

Cadieu, C. F., Hong, H., Yamins, D. L., Pinto, N., Ardila, D., Solomon, E. A., Majaj, N. J., and DiCarlo, J. J. (2014). Deep neural networks rival the representation of primate IT cortex for core visual object recognition. *PLOS Computational Biology*, 10(12):e1003963.

Chatfield, K., Arandjelović, R., Parkhi, O., and Zisserman, A. (2015). On-the-fly learning for visual search of large-scale image and video datasets. *International Journal of*

Multimedia Information Retrieval, 4(2):75–93.

Chatfield, K., Simonyan, K., Vedaldi, A., and Zisserman, A. (2014). Return of the devil in the details: Delving deep into convolutional nets. In *British Machine Vision Conference*.

Chen, R., Mohammed, N., Fung, B. C., Desai, B. C., and Xiong, L. (2011). Publishing set-valued data via differential privacy. *Proceedings of the VLDB Endowment*, 4(11):1087–1098.

Choi, H. and Varian, H. (2012). Predicting the present with Google Trends. *Economic Record*, 88(s1):2–9.

Choromanska, A., Henaff, M., Mathieu, M., Arous, G. B., and LeCun, Y. (2015). The loss surfaces of multilayer networks. In *Artificial Intelligence and Statistics*, pages 192–204.

Christie-Miller, A. (2014). Erdogan bans Twitter as corruption claims spread. Available: `http://www.thetimes.co.uk/tto/news/world/europe/article4040322.ece`. Accessed: 2015-01-29.

Chung, J. S. and Zisserman, A. (2018). Learning to lip read words by watching videos. *Computer Vision and Image Understanding*.

City of New York (2018a). Animal Facility Complaint. Available: `http://www1.nyc.gov/nyc-resources/service/1067/animal-facility-complaint`. Accessed: 2018-03-16.

City of New York (2018b). Bottled Water Complaint. Available: `http://www1.nyc.gov/nyc-resources/service/3580/bottled-water-complaint`. Accessed: 2018-03-16.

City of New York (2018c). Building Construction Complaint. Available: `http://www1.nyc.gov/nyc-resources/service/1270/building-construction-complaint`. Accessed: 2018-03-16.

City of New York (2018d). Dog or Animal Waste Complaint. Available: `http://www1.nyc.gov/nyc-resources/service/1535/dog-or-animal-waste-complaint`. Accessed: 2018-03-16.

City of New York (2018e). Electrical Complaint or Power Outage. Available: `http://www1.nyc.gov/nyc-resources/service/2246/electrical-complaint-or-power-outage`. Accessed: 2018-03-16.

City of New York (2018f). Heat or Hot Water Complaint. Available: `http://www1.nyc.gov/nyc-resources/service/1813/heat-or-hot-water-complaint`. Accessed: 2018-03-16.

City of New York (2018g). Illegal Parking Complaint. Available: `http://www1.nyc.gov/nyc-resources/service/1894/illegal-parking-complaint`. Accessed: 2018-03-16.

City of New York (2018h). Licensed Plumber or Electrician Complaint. Available: `http://www1.nyc.gov/nyc-resources/service/2223/licensed-plumber-or-electrician-complaint`. Accessed: 2018-03-16.

City of New York (2018i). Noise from Neighbor. Available: `http://www1.nyc.gov/nyc-resources/service/1197/noise-from-neighbor`. Accessed: 2018-03-16.

City of New York (2018j). Noise from Street or Sidewalk. Available: http://www1.nyc.gov/nyc-resources/service/1203/noise-from-street-or-sidewalk. Accessed: 2018-03-16.

City of New York (2018k). Pigeon Droppings or Odor Complaint. Available: http://www1.nyc.gov/nyc-resources/service/2219/pigeon-droppings-or-odor-complaint. Accessed: 2018-03-16.

City of New York (2018l). Plumbing Work Without Permit Complaint. Available: http://www1.nyc.gov/nyc-resources/service/2231/plumbing-work-without-permit-complaint. Accessed: 2018-03-16.

City of New York (2018m). Pothole or Street Surface Complaint. Available: http://www1.nyc.gov/nyc-resources/service/2238/pothole-or-street-surface-complaint. Accessed: 2018-03-16.

City of New York (2018n). Residential Maintenance Complaint. Available: http://www1.nyc.gov/nyc-resources/service/1950/residential-maintenance-complaint. Accessed: 2018-03-16.

City of New York (2018o). Sewer Odour or Leak. Available: http://www1.nyc.gov/nyc-resources/service/2440/sewer-odor-or-leak. Accessed: 2018-03-16.

City of New York (2018p). Standing Water Complaint. Available: http://www1.nyc.gov/nyc-resources/service/2510/standing-water-complaint. Accessed: 2018-03-16.

City of New York (2018q). Streetlight Request or Complaint. Available: http://www1.nyc.gov/nyc-resources/service/2533/streetlight-request-or-complaint. Accessed: 2018-03-16.

City of New York (2018r). Traffic Signal (Vehicle Stoplight) Complaint or Request. Available: http://www1.nyc.gov/nyc-resources/service/4641/traffic-signal-vehicle-stoplight-complaint-or-request. Accessed: 2018-03-16.

City of New York (2018s). Vehicle Blocking Driveway Complaint. Available: http://www1.nyc.gov/nyc-resources/service/1217/vehicle-blocking-driveway-complaint. Accessed: 2018-03-16.

City of New York (2018t). Water Complaint. Available: http://www1.nyc.gov/nyc-resources/service/2713/water-complaint. Accessed: 2018-03-16.

City of New York (2018u). Water Wasting Complaint. Available: http://www1.nyc.gov/nyc-resources/service/2733/water-wasting-complaint. Accessed: 2018-03-16.

Ciulla, F., Mocanu, D., Baronchelli, A., Gonçalves, B., Perra, N., and Vespignani, A. (2012). Beating the news using social media: the case study of American Idol. *EPJ Data Science*, 1:8.

Compton, R., Lee, C., Xu, J., Artieda-Moncada, L., Lu, T.-C., De Silva, L., and Macy, M. (2014). Using publicly visible social media to build detailed forecasts of civil unrest. *Security Informatics*, 3(1):1.

Connor, J. T., Martin, R. D., and Atlas, L. E. (1994). Recurrent neural networks and robust time series prediction. *IEEE transactions on neural networks*, 5(2):240–254.

Conover, M., Ratkiewicz, J., Francisco, M. R., Gonçalves, B., Menczer, F., and Flammini, A. (2011). Political polarization on Twitter. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, pages 89–96. AAAI.

Conover, M. D., Ferrara, E., Menczer, F., and Flammini, A. (2013). The digital evolution of Occupy Wall Street. *PLOS ONE*, 8(5):e64679.

Conte, R., Gilbert, N., Bonelli, G., Cioffi-Revilla, C., Deffuant, G., Kertesz, J., Loreto, V., Moat, S., Nadal, J.-P., Sanchez, A., et al. (2012). Manifesto of computational social science. *The European Physical Journal Special Topics*, 214(1):325–346.

Cormode, G. (2011). Personal privacy vs population privacy: learning to attack anonymization. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1253–1261. ACM.

Crowley, E. J. and Zisserman, A. (2014). In search of art. In *Workshop at the European Conference on Computer Vision*, pages 54–70. Springer.

Curme, C., Preis, T., Stanley, H. E., and 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.

De Montjoye, Y.-A., Hidalgo, C. A., Verleysen, M., and Blondel, V. D. (2013). Unique in the crowd: The privacy bounds of human mobility. *Scientific Reports*, 3:1376.

De Montjoye, Y.-A., Radaelli, L., Singh, V. K., et al. (2015). Unique in the shopping mall: On the reidentifiability of credit card metadata. *Science*, 347(6221):536–539.

DeLong, E. R., DeLong, D. M., and Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, pages 837–845.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255. IEEE.

Doersch, C., Singh, S., Gupta, A., Sivic, J., and Efros, A. (2012). What makes Paris look like Paris? *ACM Transactions on Graphics*, 31(4):1–9.

Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., and Darrell, T. (2014). Decaf: A deep convolutional activation feature for generic visual recognition. In *International Conference on Machine Learning*, pages 647–655.

Dos Santos, R., Shah, S., Chen, F., Boedihardjo, A., Lu, C.-T., and Ramakrishnan, N. (2014). Forecasting location-based events with spatio-temporal storytelling. In *Proceedings of the 7th ACM SIGSPATIAL International Workshop on Location-Based Social Networks*, pages 13–22. ACM.

Edwards, A., Housley, W., Williams, M., Sloan, L., and Williams, M. (2013). Digital social research, social media and the sociological imagination: Surrogacy, augmentation and re-orientation. *International Journal of Social Research Methodology*, 16(3):245–260.

Eshleman, C. and L Auerbach, J. (2015). Identifying and Evaluating Predictors of New York City Service Requests. Available: `https://www.researchgate.net/`

publication/283506452/Identifying-and-Evaluating-Predictors-of-New-York-City-Service-Requests. Accessed: 2018-03-16.

Ettredge, M., Gerdes, J., and Karuga, G. (2005). Using web-based search data to predict macroeconomic statistics. *Communications of the ACM*, 48(11):87–92.

Garcia, C. and Delakis, M. (2004). Convolutional face finder: A neural architecture for fast and robust face detection. *IEEE Transactions on pattern analysis and machine intelligence*, 26(11):1408–1423.

Gayo Avello, D., Metaxas, P. T., and Mustafaraj, E. (2011). Limits of electoral predictions using Twitter. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*. AAAI.

Gebru, T., Krause, J., Wang, Y., Chen, D., Deng, J., Aiden, E. L., and 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*, page 201700035.

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., and Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, 457(7232):1012.

Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587.

GLA (2015). Methodology of household income data model. Available: https://files.datapress.com/london/dataset/household-income-estimates-small-areas/gla-household-income-estimates-method-paper-Update%2008-2015.pdf. Accessed: 2018-01-31.

Glorot, X., Bordes, A., and Bengio, Y. (2011). Deep sparse rectifier neural networks. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pages 315–323.

Goel, S., Hofman, J. M., Lahaie, S., Pennock, D. M., and Watts, D. J. (2010). Predicting consumer behavior with Web search. *Proceedings of the National Academy of Sciences*, 107(41):17486–17490.

Gong, Y. and Zhang, Q. (2016). Hashtag recommendation using attention-based convolutional neural network. In *International Joint Conference on Artificial Intelligence*, pages 2782–2788.

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

Goodfellow, I., Bengio, Y., Courville, A., and Bengio, Y. (2016). *Deep learning*, volume 1. MIT press Cambridge.

Graves, A., Mohamed, A.-r., and Hinton, G. (2013). Speech recognition with deep recurrent neural networks. In *Proceedings of the IEEE International Conference on acoustics, speech and signal processing*, pages 6645–6649. IEEE.

Gruzd, A. and Tsyganova, K. (2015). Information wars and online activism during the

2013/2014 crisis in Ukraine: Examining the social structures of pro-and anti-Maidan groups. *Policy & Internet*, 7(2):121–158.

Gutiérrez-Roig, M., Preis, T., Seresinhe, C. I., Letchford, A., and Moat, H. S. (preprint). Inferring urban income statistics from Google Street View.

Hale, S. A. (2014). Multilinguals and Wikipedia editing. In *Proceedings of the 2014 ACM conference on Web science*, pages 99–108. ACM.

Hijmans, R. J., Williams, E., and Vennes, C. (2015). Geosphere: spherical trigonometry. R package version 1.3-11. Available: `https://CRAN.R-project.org/package=geosphere`.

Hochman, N. and Manovich, L. (2013). Zooming into an Instagram city: Reading the local through social media. *First Monday*, 18(7).

Hochman, N. and Schwartz, R. (2012). Visualizing Instagram: Tracing cultural visual rhythms. In *Proceedings of the workshop on Social Media Visualization in conjunction with the sixth international AAAI conference on Weblogs and Social Media*, pages 6–9.

Huang, W., Wu, Z., Chen, L., Mitra, P., and Giles, C. L. (2015). A neural probabilistic model for context based citation recommendation. In *Proceedings of the Twenty-ninth AAAI Conference on Artificial Intelligence*, pages 2404–2410. AAAI.

Hubel, D. H. and Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, 160(1):106–154.

Hulth, A., Rydevik, G., and Linde, A. (2009). Web queries as a source for syndromic surveillance. *PLOS ONE*, 4(2):e4378.

Instagram (2017). Instagram's 2017 year in review. Available: `https://instagram-press.com/blog/2017/11/29/instagrams-2017-year-in-review/`. Accessed: 2018-02-05.

Johnson, S. D. and Bowers, K. J. (2004). The burglary as clue to the future. *European Journal of Criminology*, 1(2):237–255.

Karpathy, A. and Fei-Fei, L. (2015). Deep visual-semantic alignments for generating image descriptions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3128–3137. IEEE.

Kennedy, L., Naaman, M., Ahern, S., Nair, R., and Rattenbury, T. (2007). How Flickr helps us make sense of the world: context and content in community-contributed media collections. In *Proceedings of the 15th ACM international conference on Multimedia*, pages 631–640. ACM.

Kilgarriff, A. and Fellbaum, C. (2000). Wordnet: An electronic lexical database. *Language*, 76(3):706.

Kim, D., Park, C., Oh, J., Lee, S., and Yu, H. (2016). Convolutional matrix factorization for document context-aware recommendation. In *Proceedings of the 10th ACM Conference on Recommender Systems*, pages 233–240. ACM.

Kim, Y. (2014). Convolutional neural networks for sentence classification. *CoRR*, abs/1408.5882.

King, G. (2011). Ensuring the data-rich future of the social sciences. *Science*, 331:719–721.

Kittur, A., Chi, E. H., and Suh, B. (2008). Crowdsourcing user studies with Mechanical Turk. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 453–456. ACM.

Koltsova, O. and Selivanova, G. (2015). Explaining offline participation in a social movement with online data: The case of observers for fair elections.

Kontokosta, C. E., Hong, B., and Korsberg, K. (2017). Equity in 311 reporting: Understanding socio-spatial differentials in the propensity to complain. *CoRR*, abs/1710.02452.

Kramer, A. D., Guillory, J. E., and 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.

Kristoufek, L. (2013a). BitCoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the Internet era. *Scientific Reports*, 3:3415.

Kristoufek, L. (2013b). Can Google Trends search queries contribute to risk diversification? *Scientific Reports*, 3:2713.

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

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 1097–1105.

Lazer, D., Kennedy, R., King, G., and Vespignani, A. (2014). The parable of Google Flu: traps in big data analysis. *Science*, 343(6176):1203–1205.

Lazer, D., Pentland, A. S., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., et al. (2009). Computational social science. *Science*, 323:721–723.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.

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

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

Lee, P. (2016). Learning from Tays introduction. Available: `https://blogs.microsoft.com/blog/2016/03/25/learning-tays-introduction/`. Accessed: 2018-04-05.

Legewie, J. and Schaeffer, M. (2016). Contested boundaries: Explaining where ethnoracial diversity provokes neighborhood conflict. *American Journal of Sociology*, 122(1):125–161.

Letchford, A., Preis, T., and Moat, H. S. (2016). Quantifying the search behaviour of different demographics using Google Correlate. *PLOS ONE*, 11:e0149025.

Liu, L., Qu, B., Chen, B., Hanjalic, A., and Wang, H. (2018). Modelling of information diffusion on social networks with applications to WeChat. *Physica A: Statistical Mechanics*

and its Applications, 496:318–329.

Luca, M. (2016). Reviews, reputation and revenue: The case of Yelp. Available: `http://www.ssrn.com/abstract=1928601`. SSRN Electronic Journal.

Luca, M. and Zervas, G. (2016). Fake it till you make it: Reputation, competition, and Yelp review fraud. *Management Science*, 62(12):3412–3427.

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

McAuley, J. and Leskovec, J. (2013). Hidden factors and hidden topics: understanding rating dimensions with review text. In *Proceedings of the 7th ACM conference on Recommender systems*, pages 165–172. ACM.

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

Mikolov, T., Karafiát, M., Burget, L., Černockỳ, J., and Khudanpur, S. (2010). Recurrent neural network based language model. In *Proceedings of the Eleventh Annual Conference of the International Speech Communication Association*, pages 1045–1048.

Mitchell, T. M. (2009). Mining our reality. *Science*, 326(5960):1644–1645.

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

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

Moat, H. S., Preis, T., Olivola, C. Y., Liu, C., and 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., and Tita, G. E. (2011). Self-exciting point process modeling of crime. *Journal of the American Statistical Association*, 106(493):100–108.

Naik, N., Kominers, S. D., Raskar, R., Glaeser, E. L., and Hidalgo, C. A. (2017). Computer vision uncovers predictors of physical urban change. *Proceedings of the National Academy of Sciences*, 114(29):7571–7576.

Ng, J. Y.-H., Hausknecht, M., Vijayanarasimhan, S., Vinyals, O., Monga, R., and Toderici, G. (2015). Beyond short snippets: Deep networks for video classification. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 4694–4702. IEEE.

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

NYC (2015). A Local Law to amend the administrative code of the city of New York, in relation to the standardization of address and geospatial information on the open data portal. Available: `http://legistar.council.nyc.gov/LegislationDetail.aspx?ID=2460428&GUID=A23883BC-E875-44C9-B5C4-C475626044D0`. Accessed: 2018-03-16.

NYC (2017). 311 Sets New Record with Nearly 36 Million Interactions in 2016. Available: http://www1.nyc.gov/office-of-the-mayor/news/033-17/311-sets-new-record-nearly-36-million-interactions-2016. Accessed: 2018-02-26.

NYC City Planning (2015). New York City census FactFinder. Available: http://www1.nyc.gov/assets/planning/download/pdf/data-maps/maps-geography/census-factfinder/cff-userguide.pdf. Accessed: 2018-01-18.

Okabe, D. (2004). Emergent social practices, situations and relations through everyday camera phone use. In *International Conference on Mobile Communication and Social Change*.

ONS (2012). Available: https://www.ons.gov.uk/methodology/geography/ukgeographies/censusgeography. Accessed: 2018-02-07.

Patterson, G., Xu, C., Su, H., and Hays, J. (2014). The sun attribute database: Beyond categories for deeper scene understanding. *International Journal of Computer Vision*, 108(1-2):59–81.

Pinkovskiy, M. and Sala-i Martin, X. (2014). Lights, camera,... income!: Estimating poverty using national accounts, survey means, and lights. Technical Report 19831, National Bureau of Economic Research.

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

Preis, T. and Moat, H. S. (2015). Early signs of financial market moves reflected by Google searches. In Gonçalves, B. and Perra, N., editors, *Social Phenomena: From Data Analysis to Models*, pages 85–97. Springer International Publishing, Switzerland.

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

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

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

Quercia, D., Schifanella, R., Aiello, L. M., and McLean, K. (2015). Smelly maps: The digital life of urban smellscapes. In *Proceedings of the Nineth AAAI International Conference on Weblogs and Social Media*. AAAI.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088):533.

Sainath, T. N., Mohamed, A.-r., Kingsbury, B., and Ramabhadran, B. (2013). Deep convolutional neural networks for LVCSR. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 8614–8618. IEEE.

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

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

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

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

Sermanet, P., Kavukcuoglu, K., Chintala, S., and LeCun, Y. (2013). Pedestrian detection with unsupervised multi-stage feature learning. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 3626–3633. IEEE.

Sharif Razavian, A., Azizpour, H., Sullivan, J., and Carlsson, S. (2014). Cnn features off-the-shelf: an astounding baseline for recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 806–813.

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, 69(1):99–118.

Simonyan, K. and Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations*, pages 1–14.

Sloan, L., Morgan, J., Burnap, P., and Williams, M. (2015). Who tweets? Deriving the demographic characteristics of age, occupation and social class from Twitter user meta-data. *PLOS ONE*, 10(3):e0115545.

Smith, A. and Monica, A. (2018). Social media use in 2018. Available: `http://www.pewinternet.org/2018/03/01/social-media-use-in-2018/`. Accessed: 2018-04-10.

Springenberg, J. T., Dosovitskiy, A., Brox, T., and Riedmiller, M. A. (2014). Striving for simplicity: The all convolutional net. *CoRR*, abs/1412.6806.

Steinert-Threlkeld, Z. C., Mocanu, D., Vespignani, A., and Fowler, J. (2015). Online social networks and offline protest. *EPJ Data Science*, 4(1):19.

Sung, K.-K. and Poggio, T. (1998). Example-based learning for view-based human face detection. *IEEE Transactions on pattern analysis and machine intelligence*, 20(1):39–51.

Traud, A. L., Mucha, P. J., and Porter, M. A. (2012). Social structure of facebook networks. *Physica A: Statistical Mechanics and its Applications*, 391(16):4165–4180.

Tumasjan, A., Sprenger, T., Sandner, P., and Welpe, I. (2010). Predicting elections with Twitter: What 140 characters reveal about political sentiment. In *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, pages 178–185. AAAI.

UN (2015). World Urbanization Prospects: The 2014 Revision, (ST/ESA/SER.A/366).

US Census Bureau (2010). Geographic terms and concepts - census tract. Available: `https://www.census.gov/geo/reference/gtc/gtc_ct.html`. Accessed: 2018-01-18.

Vaillant, R., Monrocq, C., and Le Cun, Y. (1994). Original approach for the localisation of ob-

jects in images. *IEE Proceedings-Vision, Image and Signal Processing*, 141(4):245–250.

Vázquez, A., Pastor-Satorras, R., and Vespignani, A. (2002). Large-scale topological and dynamical properties of the internet. *Physical Review E*, 65(6):066130.

Vespignani, A. (2009). Predicting the behavior of techno-social systems. *Science*, 325(5939):425.

Wachter, S. (2018). GDPR and the Internet of Things: Guidelines to protect users identity and privacy. Available: `http://www.ssrn.com/abstract=3130392`. SSRN Electronic Journal.

Wachter, S., Mittelstadt, B., and Russell, C. (2017). Counterfactual explanations without opening the black box: Automated decisions and the GDPR. Available: `http://www.ssrn.com/abstract=3063289`. SSRN Electronic Journal.

Wagenmakers, E.-J. and Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, 11(1):192–196.

Waibel, A., Hanazawa, T., Hinton, G., Shikano, K., and Lang, K. J. (1990). Phoneme recognition using time-delay neural networks. In *Readings in speech recognition*, pages 393–404. Elsevier.

Waitrose (2017). The Waitrose food and drink report 2016. Available: `https://www.johnlewispartnership.co.uk/content/dam/cws/pdfs/Resources/the-waitrose-food-and-drink-report-2016.pdf`. Accessed: 2018-01-18.

Weilenmann, A., Hillman, T., and Jungselius, B. (2013). Instagram at the museum: communicating the museum experience through social photo sharing. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 1843–1852. ACM.

Wikimedia (2018). Total page view. Available: `https://stats.wikimedia.org/v2/`. Accessed: 2018-04-10.

Wood, S. A., Guerry, A. D., Silver, J. M., and Lacayo, M. (2013). Using social media to quantify nature-based tourism and recreation. *Scientific Reports*, 3:2976.

Xiao, J., Hays, J., Ehinger, K. A., Oliva, A., and Torralba, A. (2010). Sun database: Large-scale scene recognition from abbey to zoo. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 3485–3492. IEEE.

Yelp (2018). About yelp. Available: https://www.yelp.co.uk/about. Accessed: 2018-04-05.

Zhang, F., Tanupabrungsun, S., Hemsley, J., Robinson, J. L., Semaan, B., Bryant, L., Stromer-Galley, J., Boichak, O., and Hegde, Y. (2017). Strategic temporality on social media during the general election of the 2016 US presidential campaign. In *Proceedings of the 8th International Conference on Social Media & Society*, page 25. ACM.

Zhang, Z., Lyons, M., Schuster, M., and Akamatsu, S. (1998). Comparison between geometry-based and gabor-wavelets-based facial expression recognition using multi-layer perceptron. In *Proceedings of the Third IEEE International Conference on Au-*

tomatic Face and Gesture Recognition, pages 454–459. IEEE.

Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., and Oliva, A. (2014). Learning deep features for scene recognition using Places database. In *Advances in Neural Information Processing Systems*, pages 487–495.

Zhu, Y., Xiong, L., and Verdery, C. (2010). Anonymizing user profiles for personalized web search. In *Proceedings of the 19th international conference on World wide web*, pages 1225–1226. ACM.