

RESEARCH

Open Access



Built environment attributes and crime: an automated machine learning approach

Kyle Dakin¹, Weizhi Xie¹, Simon Parkinson^{1*} , Saad Khan¹, Leanne Monchuk² and Ken Pease³

Abstract

This paper presents the development of an automated machine learning approach to gain an understanding of the built environment and its relationship to crime. This involves the automatic capture of street-level photographs using Google Street View (GSV), followed by the use of supervised machine learning techniques (specifically image feature recognition) to recognise features of the built environment. In this exploratory proof-of-concept work, 8 key features (building, door, fence, streetlight, tree, window, hedge, and garage) are considered and a worked case-study is demonstrated for a small geographical area (8300 square kilometres) in Northern England. A total of 60,100 images were automatically collected and analysed across the area where 5288 crime incidents were reported over a twelve-month period. Dependency between features and crime incidents are measured; however, no strong correlation has been identified. This is unsurprisingly considering the high number of crime incidents in a small geographic region (8300 square kilometres), resulting in an overlap between specific features and multiple crime incidents. Furthermore, due to the unknown precise location of crime instances, an approximation technique is developed to survey a crime's local proximity. Despite the absence of a strong correlation, this paper presents a first-of-a-kind cross-discipline approach to attempt and use computation techniques to produce new empirical knowledge. There are many avenues of future research in this fertile and important area.

Keywords: Crime prevention, Supervised machine learning, Feature recognition, Crime analytics

Introduction

Criminologists have long since considered the relationship between the built environment and opportunities for crime and disorder. The study and manipulation of the physical, built environment to reduce the potential for crime is often referred to as Crime Prevention Through Environmental Design (CPTED) Crowe (2000a). CPTED draws upon theory from environmental criminology, architecture, urban design, and more recently, data science. CPTED is underpinned by the following principles: (1) physical security; (2) surveillance; (3) movement control; (4) management and maintenance and (5) defensible space Poyner (1983); Cozens et al. (2005); Armitage

(2013); Montoya et al. (2016); Armitage and Monchuk (2017). A commonality amongst these five principles is that prior consideration of a housing development can have positive impacts in reducing crime.

Attempts to reduce crime have historically focused upon changing the behaviour of people disposed to commit it. The rhetoric of diversion from crime and rehabilitation of those already established in criminal careers has dominated policy discourse, while evidence of general success in realising these aims has been at best modest (Pease 2010). Over the last forty years, attention has increasingly turned to situational crime prevention, whereby criminogenic features of the environment are designed or modified to make the risks attending criminality greater and the rewards less. The emphasis on situational change almost certainly represents the most fruitful approach to crime reduction. The approach aligns with the bulk of the relevant psychological literature. The

*Correspondence: s.parkinson@hud.ac.uk

¹ Department of Computer Science, University of Huddersfield, Queensgate HD1 3DH, UK

Full list of author information is available at the end of the article



© The Author(s) 2020. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

sub-disciplines of social and cognitive psychology substantially comprise demonstrations of how situations can be manipulated to influence behaviour. Tellingly, literature demonstrates the way in which people seem hard-wired to underestimate the power of situational change. The phenomenon is known as the ‘fundamental attribution error’ (Tetlock 1985). Recognition of this error leads directly to a refocusing of crime reduction policy towards changing situations, and to why this approach will face resistance, as the ubiquitous tendency to see the person rather than the situation as the major determinant of behaviour.

While the fundamental attribution error leads to the focus on person change, evidenced successes in crime reduction are mainly to be found in initiatives which change situations. Some of the most recent successes of this approach are to be found in Scott and Clarke (2020). Some have been popularised in the notion of ‘nudges’, identifying that apparently trivial situational changes resulting in non-trivial behaviour changes (Thaler and Sunstein 2009).

Some criminogenic features of situations are easy to remedy, such as the use of strong security mechanisms, whereas others are difficult and expensive, such as surveillance. The design of buildings, and the street networks in which they are located, together with the cost of remediation, provides the most obvious case where it is important to get the initial design right. Buildings and street networks carry their intrinsic crime risks for their lifetime, mitigated (but not removed) by security measures applied to individual homes (see (Thompson et al. 2018)). A single house burglary is currently costed at £5930 (Heeks et al. 2018), not taking account of opportunity costs imposed on police resources, together with increased population deviation in anticipation of or after experience of crime (Ellingworth and Pease 1998). Furthermore, area reputation will be harmed resulting in consequential depression in home values (Ihlanfeldt and Mayock 2010).

The reader may question whether the design of buildings and the configuration of street networks embed crime opportunities. Unsurprisingly, they do. Home and setting design was an early topic of interest for advocates of situational crime prevention. Pioneers of the approach were (Newman 1974; Jacobs 1961) and the approach came to be named as *Crime Prevention Through Environmental Design* (CPTED) (Crowe 2000b). Despite the evidence of considerable success in crime reduction in developments built according to CPTED principles (Armitage and Monchuk 2011), there remains limited consensus about which particular attributes of home and street network and in what combinations are optimal for reduced crime risk. In the UK CPTED, insofar

as it is applied, is delivered by a number of agencies (notably police, urban designers and planning authorities). CPTED advice is provided to planners by Designing out Crime Officers (DOCOs) who are employed within each of the forty-three territorial police forces of England and Wales. DOCOs review planning applications and assess the extent to which a development may pose opportunities for crime and disorder (Monchuk et al. 2018). On the basis of their assessment, remedial modifications to plans are advocated.

It is important to emphasise that it is in no way a criticism of DOCOs that their risk identification is largely untested. Skilled performance of all kinds depends upon the feedback of results. For example, medical judgements of treatment efficacy depend upon outcome data like rates of patient survival to recovery. DOCOs do not have systematic data on the crime experienced by developments with particular attributes. DOCOs have only the limited research literature and such inferences as they feel able to make through attendance at crime scenes as front-line police officers. There is therefore potential in exploring the addition of new quantitative information that DOCOs can use alongside the current body of knowledge and qualitative components.

The obvious (and possibly the only) way of systematically providing feedback on the crime consequences of design features is to examine the crime histories of developments built long enough ago for such histories to be meaningful. Examination of the original plans permits identification of building attributes and attribute combinations associated with subsequent crime. The ultimate aim and end-user application of the research programme of which this paper is an early part is to provide at the planning stage details of expected crime, by type, and to identify design adjustments to reduce this. Our vision is to provide a software tool, capable of analysing and learning patterns between crime and characteristics of the built environment to assist with offering a systematic approach to identifying crime risk, which is used alongside qualitative measures. It is also our vision that the tool should be self-updating to be able to learn new relationships, so as to reflect changes alongside aesthetic and other variations in home design. It would weight crime types by their associated harm.

The techniques to realise the vision outlined above are already available. The stages necessary to achieve this end are as follows:

1. Demonstration that the current DOCO based approach to the anticipation of crime from residential architectural plans varies across officers and provides on average modest predictive power. This has

already been done and is described in the next section of the paper.

2. Explore potential data sources and approaches to the identification of criminogenic features and feature combinations of homes and street networks. The present paper reports one element of this, risk predictability using Google Street View (GSV).
3. Application of supervised machine learning to the features of residential development plans and street networks extracted from GSV built at least a decade before. This is to yield risk assessment by crime type.
4. Devise a routine such that changes suggested to architectural places in terms of individual attributes will be ranked according to their expected crime reductive effect.
5. Routinely repeat the machine learning phase to identify new patterns, followed by checking previously analysed architectural plans to identify new and previously not known attribute relationships.

The story so far

The work in relation to stage 1 was published in Monchuk et al. (2018). Plans for an estate built and occupied a decade earlier were acquired, together with crime data for the lifetime of the development to date. Plans (not crime data) were shown to a sample of experienced DOCOs who were not familiar with the police force area in which the development concerned was to be found. They were invited to identify places where crime could be anticipated, and the type of crime likely to occur there. In brief, the results identified:

1. Individual officers varied widely in their identification of crime-prone locations.
2. There was substantial variation in the proportion of locations identified as crime prone. More specifically, there is a range of trade-offs between false negatives and false positives. Another way of expressing this is to say that the risk threshold varied between DOCOs.
3. The predictive accuracy of individual DOCOs varied, but was on average modest. Since the ceiling on possible accuracy is unknown (one of the justifications for the research programme outlined here) it may be that the best performing DOCOs' judgements are as good as it can get, with (for example) occupant characteristics accounting for the bulk of variation in crime experienced. Were that to turn out to be the case in the light of the research proposed here, investment in CPTED solutions should be limited to those identified, and other considerations e.g. aesthetic (see (Armitage et al. 2013)) would prevail

in home construction. However, if as we believe, an optimised CPTED is potentially powerfully crime reductive, funding commensurate with crime harm would be appropriate.

The original Monchuk et al. (2019) work evoked the writers' conviction that a machine learning approach provided the most promising route to identifying optimal crime reductive design. The work is a transitional paper seeking to identify an optimisation of prediction using those variables already used by DOCOs. It does this by applying automated deliberation techniques to automate repetitive rule-based logic. The results from this early work are promising and, while transitional, they do suggest a short term improvement to crime risk prediction. If the exercise were repeated for individual DOCOs it could identify the features used by the best performing DOCO, contrast this with features used by other DOCOs, and use that information as part of a DOCO training package.

Other recent work demonstrates the potential to use Machine Learning to automatically score the built environment using computer vision and GSV (Naik et al. 2014). In their research, a crowd-sourced approach is taken whereby participants score images based on how they perceive the safety, before machine learning algorithms try to learn the relationship between colour characteristics of the image and the participant's safety score. However, their approach is somewhat limited for crime reduction. Most significantly, it is scoring the built environment once it has been constructed, therefore minimising any opportunity to rectify through influencing design and planning—the key objective of CPTED. Furthermore, their research is constrained to identifying characteristics of the image (colour, etc.) and important crime contributing factors will be missed. For example, identifying a footpath with poor lighting, thus limiting opportunities for surveillance. Another limitation is that the Google Street View (GSV) image could have been taken with unfavourable lighting, making the image darker and thus resulting in receiving a lower safety rating.

In other recent and related research, authors have used GSV to acquire information on the built environment, with a particular emphasis on burglary (Langton and Steenbeek 2017). The researchers utilise GSV to replace the activity of making a physical site visit to determine attributes such as front door visibility, alarm, ease of access etc. The research is useful in using digital assets for performing assessments; however, there are still inherent limitations that motivate the research in this paper. The first is that a human is required to perform the extraction, which prevents scalability and introduces the

potential for a difference of opinion. The second is that the human analysts are searching for pre-determined features believed to have significance to crime reduction based on previous literature and subject knowledge. This restricts the potential to identify any new patterns that were previously unknown.

Vandeviver (2014) presents a survey on the user of GSV in criminological research. In their survey, an example of a relevant work using GSV to understand the built environment is that of performing neighbourhood audits (Kronkvist 2013). However, in their study, the authors are using GSV operated and interpreted by a human investigator. It is evident from their survey that although researchers have considered using GSV for analysing the built environment

Another recent study aims to gain an understanding of the effects of neighbourhood and house attributes on a burglar's selection (Vandeviver and Bernasco 2019). The research presents interesting findings that offenders prefer to target areas with a lower density of residential properties. Although this study provides a useful insight, it would be strengthened by exploring whether there are relationships features influencing the likelihood of burglary occurring. Furthermore, the focus of the paper is solely on burglary, and gaining an understanding of if neighbourhood and house attributes influence other types of crime is worthy of consideration.

A machine learning approach

The present paper explores the use of GSV to identify features visible in images acquired of home frontages. This is useful in two ways. First, it developed a method for the automated identification of individual home features from GSV images. Second, it presented the possibility of a detailed short-term study of individual home frontal features whose relevance to crime are contentious but would require lengthy and tedious research addressed by other means. Intruder alarms and street lighting are examples of frontal features with relevance to crime. The use of GSV in criminological research is by no means new; however, to the best of the authors' knowledge, this is the first research applying supervised machine learning to analyse the built environment.

The study reported here is deemed useful but transitional as a step towards the vision set out earlier. More specifically, additional data sources will be required at a later stage. The motivation for this is that according to the Crime Survey for England and Wales (2017)¹, only some 50% of burglaries involve front entry. Second, data

from individual homes were not available, so factors distinguishing individual home victimisation within an area are not captured. The current process for recording individual housing characteristics is performed by human experts using manual analysis techniques. However, there is a significant opportunity to consider the use of advancements in Artificial Intelligence (AI) to provide enhanced digital capabilities to upscale CPTED processing and improve consistency.

Preliminary work has involved extracting knowledge from 28 DOCOs within England and Wales and in the work reported here, the authors are particularly interested in two research questions of: (1) whether it is possible to automatically extract data on the features present within the built environment where a crime has occurred, as well as (2) taking the first steps towards using AI techniques to learn and identify key patterns flagging design features carrying crime risk.

This research provides the first empirical study of its kind known to the authors in working towards these two aims. In the pursuit of these aims, the following objectives are undertaken: (1) the development of a technique to extract street level images in the local proximity of a crime's location, taking into account that the exact location of the crime might not be available or is unknown; (2) the training of a machine learning algorithm to process acquired street level images to extract known features of the built environment; and finally, (3) use correlation techniques to develop a process to understand if there are strong relationships between environmental features and the location of a crime.

Outline method

This section presents the approach and how key technical challenges have been overcome. A brief summary is provided below to aid the reader in understanding the process undertaken in this research:

1. Crime data extraction: Data is downloaded and extracted from police.uk specifically focusing on one neighbourhood ward within Northern England.
2. Location generation: Due to the arbitrary assignment of location in open police data, a technique is proposed to generate locations close by the arbitrary location to gain an increased representation of the built environment within the local proximity.
3. Image collection: Once all the locations have been generated, street level images are then acquired using the GSV platform.
4. Feature selection: Each image is processed to identify key features within the built environment, using supervised machine learning that has been trained to recognise a series of features. For the purpose of this

¹ Crime in England and Wales: year ending June 2017 available at: <https://www.ons.gov.uk/releases/crimeinenglandandwalesyearendingjune2017>.

research, these consisted of: building, door, fence, streetlight, tree, window, hedge and garage.

5. Correlation: Using the acquired numeric data (number of features) for each crime location and then a statistical measure of dependency to determine if there are features that are strongly linked with different types of crime.

All computation performed in this research was on a high-performance computer with an Intel Xeon Platinum 8180 2.5 GHz processor, 128 GB RAM, and a NVidia GeForce RTX2080Ti. Each stage is now presented and discussed.

Crime data extraction

This research focused upon a single neighbourhood within Northern England. This ward was chosen for convenience and owing to its relatively high levels of recorded crime in the 12 month period between July 2018 and August 2019. The crime data was acquired via the open access police.uk website².

The data is categorised in to 14 different crime types and use of the miscellaneous ‘Other’ category. In this research, this category is omitted as we have no indication as to what type of crime has been committed. A total of 5795 instances were reported in the studied period, with a total of 507 appear in the Other category. This results in a total of 5288 analysed in this research. The number of instances of crime per category can be seen in the ‘Occurrence’ column in Table 2. It is evident that there is an uneven distribution between categories, with the lowest being bicycle theft with a count of 25 and the highest of violent and sexual offences with a count of 2033.

The different crime types are predetermined by the UK’s Single Online Home National Digital Team, whom collect and collate data on monthly basis from police forces throughout the UK. For readers unfamiliar with the definitions used in this research, the following list provides the definitions as provided on the police.uk website:

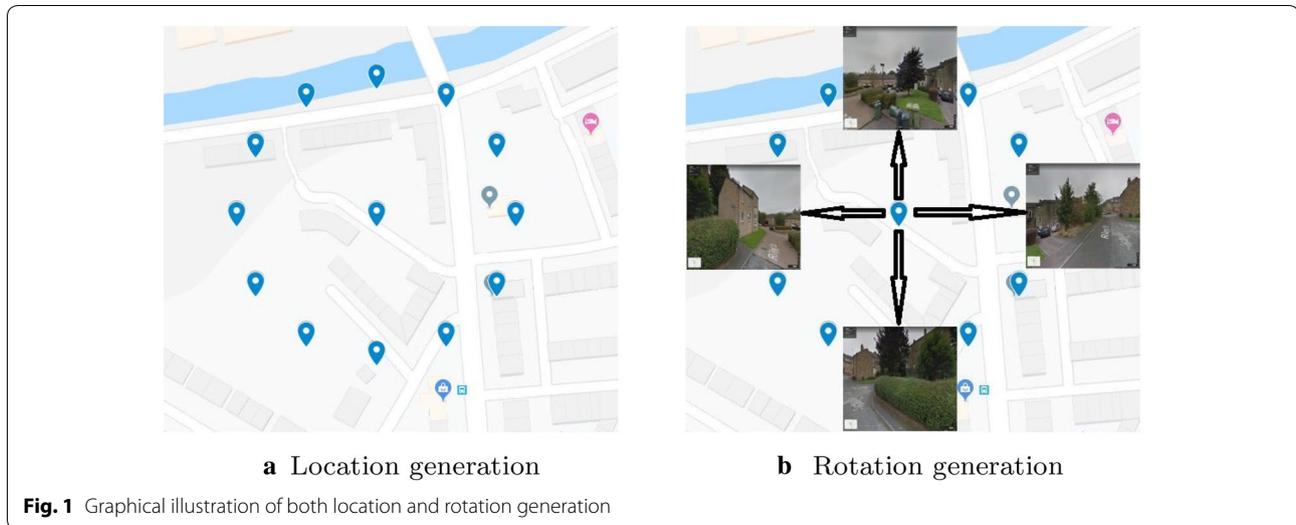
- Anti-social behaviour (ASB): Includes personal, environmental and nuisance anti-social behaviour. Note that ASB is not a crime but a civil offence.
- Bicycle theft: Includes the taking without consent or theft of a pedal cycle.
- Burglary: Includes offences where a person enters a house or other building with the intention of stealing.

- Criminal damage and arson: Includes damage to buildings and vehicles and deliberate damage by fire.
- Drugs: Includes offences related to possession, supply and production.
- Possession of weapons: Includes possession of a weapon, such as a firearm or knife.
- Public disorder: Includes offences which cause fear, alarm, distress or a possession of a weapon such as a firearm.
- Public order: Includes offences which cause fear, alarm or distress.
- Robbery: Includes offences where a person uses force or threat of force to steal.
- Shoplifting: Includes theft from shops or stalls.
- Theft from the person: Includes crimes that involve theft directly from the victim (including handbag, wallet, cash, mobile phones) but without the use or threat of physical force.
- Vehicle crime: Includes theft from or of a vehicle or interference with a vehicle.
- Violence and sexual offences: Includes offences against the person such as common assaults, Grievous Bodily Harm and sexual offences.

The acquired crime data does not contain a ward location within each address, but does however contain a location specified by longitude and latitude values. As the presented technique is focusing on the selected ward, it is necessary to determine if a crime’s location (longitude and latitude) is sited within the ward. This is achieved by querying the longitude and latitude values to return a ward using the postcode.io query service³. Each instance of crime is individually processed and, if the recorded latitude and longitude does not return a ward, points close to the location are generated and tested to see if they fall within a ward boundary. This process is repeated until the closest ward is found and returned. The same circular generation technique as presented in [Local generation](#) section with a radius of 50 m. The generated positions are processed incrementally until a ward is located and the process ends. In the experimental work undertaken in this paper, the method is only invoked in a few instances where there is incomplete information in the `postcodes.io` database. Once each crime instance has an associated ward, we then filter the entire data set to only contain crime instances within the ward of interest.

² data.police.uk open crime data available at: <https://data.police.uk/>.

³ Postcode.io is an open source project and based solely on open data, providing location-based information <https://postcodes.io/>.



Location generation

In this research, we use a location (a ward) within a town in the North of England with a residential population of approximately 17,000 over a geographical area of approximately 8300 square kilometres. The location includes a town centre location, as well as suburban residential housing estates. The ward was chosen due to this mixture of commercial and domestic properties and its relatively high crime statistics.

In this stage, we systematically generate distances close to the crime's longitude and latitude location as previously discussed to overcome inherent limitations of not having accurate and precise crime location data. The

technical approach for doing this is presented in Algorithm 1, which is essentially generating new longitude and latitude values within a circular pattern around the crime's location, using trigonometry. Algorithm 1 takes as input the crime's longitude and latitude and returns a set of new locations values. Figure 1a provides a graphical illustration whereby the location acquired from police data is in the centre and the generated points of interest are located around the central point in a circular pattern. In this work, a radius of 50 m (1 degree of latitude) is used, providing a distance of 26.1 ($2\pi r$, where r is 50 m) metres between newly generated points on the circumference.

Algorithm 1: Method to calculate points of interest

Input: A longitude position, $long$
Input: A latitudes position, lat
Input: The radius of the circle in metres, r
Output: A set of new longitude and latitudes pairs, $N = \{n_1, \dots, n_{12}\}$, where $n = \{long, lat\}$

```

1  $x = i/111111$ 
2 for  $j = 0; j < 360; j = j + 30$  do
3    $lat = radians(lat)$ 
4    $long = radians(long)$ 
5    $y = radians(j)$ 
6    $latRadians = asin(math.sin(lat) \times cos(x) + cos(lat) \times sin(x) \times cos(y))$ 
7    $longRadians =$ 
    $long + atan2(sin(y) \times sin(x) \times cos(lat), cos(x) - sin(lat) \times sin(latRadians))$ 
8    $newlong = degrees(latRadians)$ 
9    $newlat = degrees(longRadians)$ 
10   $N_j = \{newlong, newlat\}$ 

```

Furthermore, for each newly generated position, we generate rotations at 90 degree intervals (90, 180, 270, and 360) in order to survey different orientations of a physical environment, using GSV. An example is demonstrated in Fig. 1b whereby 4 images have been generated using Street View at one specific location. As evident in the images, there are different features that can be extracted from each image. For example, the top image contains a clear and obvious streetlight and tree, whereas the other three images contain dwellings and trees.

In order to gain an understanding of the intersection between the generated local proximity between two different crime instances, we have created a technique to measure the overlap between the area analysed for two crimes. The technique is presented in Algorithm 2. The algorithm tests for intersections using the square of the distance between the centres of the circles, generated using Algorithm 1. Algorithm 1 takes as input the following pre-established data items: longitude (x) and latitude (y) of a crime location, as well as the radius used to generate the proximity circle (r). Three sets (X, Y , and R) store these values and n is the crime instance number ($\|X\|, \|Y\|, \|R\| = n$). The algorithm calculates the distances from the two circle centres (d) and the sum of the two radii distances (rd). It is then possible to determine whether they intersect or border by checking the difference between d and rd . Running the Algorithm 2 on the 5,288 crimes provides a $NumMatching = 4300$. This demonstrates a strong cross-over between crime locations with around 75% of crime instances sharing intersecting circles. It is worth noting that in the data analysed, there are a total of 634 unique crime locations.

Table 1 Number of features identified in both training and testing data set

Feature	Training	Testing
Building	5017	827
Door	2755	489
Fence	3054	687
Streetlight	623	176
Tree	1019	191
Window	14768	2308
Hedge	1228	163
Garage	743	128

Image collection

Once all locations and orientations have been established for each crime instance, it is then necessary to retrieve the image using GSV. A web page application was written in the Python language to handle the image extraction tasks. The collection utilised the JavaScript Google Chrome extension to automate the process of loading GSV at a specific location and to acquire and store the image. Note that it would have been preferred to use the provided Google APIs to acquire street level images; however, the number of requests we need to make are beyond that of a free subscription.

As we have generated new locations around the provided location of crime, it is likely that some of the generated longitude and latitude locations do not fall on a road and therefore it will not be possible to acquire a Street

Algorithm 2: Method to calculate points of interest

```

Input: A set of longitude positions,  $X = \{x_i, x_i, \dots x_n\}$  where  $n$  is the number of
crime locations
Input: A set of latitudes positions,  $Y = \{y_i, y_i, \dots y_n\}$  where  $n$  is the number of
crime locations
Input: A set of radii measurements,  $R = \{r_i, r_i, \dots r_n\}$  where  $n$  is the number of
crime locations
Output: Number of crime circles intersecting or touching,  $NumMatching$ 
1 for  $i = 0; i < n; i = i + 1$  do
2   for  $j = 0; j < n; j = j + 1$  do
3      $d = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2}$ 
4      $rd = (r_i + r_j)$  if  $(rd == d) || (rd > d)$  then
5        $NumMatching ++$ 
6     
```



Fig. 2 Example of a street level image showing how the algorithm identified learnt features

View image. For example, as seen in Fig. 1a, many of the points are not located on a road. This process is easy to programmatically handle as an error response will be provided when the software attempts to acquire a Street View image for a location where it is not available. The output of this stage is a collection of images where the image name is recorded to match the crime, location and origin.

Feature selection

We trained the object detection method using 3356 images and tested with a further 496 images (15%). Features were manually labelled using an application that produces a separate XML file containing the bounding boxes surrounding a feature⁴. Table 1 provides details on how many features were identified in total across all training and testing images. In this exploratory research, we focus on selecting eight different attributes of the built environment, which can be seen in Table 1. Figure 2 illustrates an example whereby the algorithm has identified windows, doors, trees, and a fence in the image. As evident in the figure, the algorithm has identified the features within the image and assigned a confidence percentage score, for example, ‘window: 75%’. This score states that the algorithm has a confidence of 75% that it has identified a window based on those it has been trained to recognise. In this research, we utilised the TensorFlow Object Detection API⁵ due to its capabilities and ease-of-use. Alternatives, such as Google Vision are available; however, many features require a commercial

subscription. TensorFlow, due to its wide-scale use in many different scientific disciplines, has evolved to have a good range of functionality that is easy to use.

Limitations

To reiterate at this juncture, this paper reports on an exploratory piece to assess the feasibility and functionality of using an automated machine learning approach to the built environment. The list below summarises some of the inherent limitations and provides justifications as to why they do not detract from the research study.

- Crime location: Open source police data has an arbitrary assignment to be ‘on’ or ‘near’ a road location. This means that the actual location of the recorded crime could be significantly different from the location of the crime as it appears in the open source data. Police services in the UK do hold more precise location-based data, which the authors aim to acquire after the proof-of-concept set out in this paper; however, as we are considering crime types that are not just those involving a property (i.e., burglary), it will always be necessary to acquire image data in the local proximity of the crime to gain a wider understanding of the built environment. In other words, the street-level image at the location of the crime may not contain enough of the built environment. It is therefore the case that there will always be the need to survey around the crime’s location, given the known association between street networks and crime risk. In this work we are traversing around a potentially imprecise crime location; however, we increase the proximity of the area we survey to increase the likelihood of acquiring a more representative understanding of the built environment where the crime was committed.
- Case study ward size: In this research, we use a location (a ward) within a town in the North of England with a population of approximately 17,000 over a size of around 8300 square kilometres. However, the chosen destination is small and has characteristics that might not allow for meaningful findings. More specifically, it is a tightly packed urban area with relatively high crime statistics. This means that the cross-over between crime locations might not allow for a distinct set of images per crime location. However, the location was selected due to its mixture of commercial and domestic properties and high crime levels. The justification for using a location with these characteristics is that it provides a rich data set in terms of crime and environment characteristics. Selectively choosing a location with lower crime sta-

⁴ LabelImg was used in this research and it is a graphical image annotation tool and is available at: <https://github.com/tzutalin/labelimg>.

⁵ TensorFlow Object Detection API: https://github.com/tensorflow/models/tree/master/research/object_detection.

Table 2 Number of features per crime instance for each crime type in a 12 month period (07/2018 to 06/2019)

Crime type	Occurrence	Building			Door			Fence			Streetlight		
		Min	Avg	Max	Min	Avg	Max	Min	Avg	Max	Min	Avg	Max
Anti-social behaviour	728	0	6.8	18	0	1.8	16	0	1.8	26	0	0.5	6
Bicycle theft	25	0	7.1	16	0	2.1	13	0	1.3	13	0	0.3	2
Burglary	354	0	7.1	20	0	2	13	0	2	28	0	0.4	5
Criminal damage and arson	478	0	6.7	19	0	1.8	13	0	2	28	0	0.4	6
Drugs	228	0	7.2	16	0	1.9	10	0	1.8	21	0	0.5	6
Possession of weapons	65	0	6.8	17	0	2	11	0	1.4	11	0	0.4	3
Public order	545	0	6.1	17	0	1.5	11	0	1.4	15	0	0.3	6
Robbery	71	0	5.5	13	0	1.6	11	0	0.9	9	0	0.3	3
Shoplifting	326	0	5.7	16	0	1.3	11	0	1.2	16	0	0.4	5
Theft from the person	75	0	4.7	18	0	1.2	7	0	1.1	11	0	0.3	5
Vehicle Crime	360	0	1	9	0	1	15	0	0.3	3	0	1.5	16
Violence and sexual offences	2033	0	6.6	22	0	1.8	18	0	1.6	28	0	0.4	7
Average	440.7	0.0	5.9	16.8	0.0	1.7	12.4	0.0	1.4	17.4	0.0	0.5	5.8
Crime type	Occurrence	Tree			Window			Hedge			Garage		
		Min	Avg	Max	Min	Avg	Max	Min	Avg	Max	Min	Avg	Max
Anti-social behaviour	728	0	1.8	16	0	24.1	83	0	0.3	6	0	0.4	11
Bicycle theft	25	0	1.7	13	0	23.7	75	0	0.6	6	0	0.5	5
Burglary	354	0	2	20	0	23	93	0	0.4	7	0	0.4	11
Criminal damage and arson	478	0	1.9	19	0	22.3	80	0	0.3	7	0	0.4	6
Drugs	228	0	1.2	14	0	25.3	80	0	0.3	6	0	0.4	5
Possession of weapons	65	0	1.7	17	0	18.8	66	0	0.4	4	0	0.4	6
Public order	545	0	1.7	18	0	20.1	83	0	0.3	7	0	0.3	6
Robbery	71	0	2	12	0	16.5	62	0	0.3	4	0	0.2	5
Shoplifting	326	0	1.5	20	0	17.5	85	0	0.3	7	0	0.3	7
Theft from the person	75	0	1.4	13	0	15	77	0	0.2	3	0	0.2	3
Vehicle Crime	360	0	15.5	63	0	0.2	5	0	0.2	5	0	0	0
Violence and sexual offences	2033	0	1.7	25	0	21.4	93	0	0.3	7	0	0.4	11
Average	440.7	0.0	2.8	20.8	0.0	19.0	73.5	0.0	0.3	5.8	0.0	0.3	6.3

tistics and fewer properties would not be representative of the true problem.

- Number of features examined: The built environment comprises a number of different features such as buildings, street networks and street furniture (such as streetlights). In this study, the machine learning algorithm learnt how to automatically detect such features from GSV images. In this proof of concept, eight features are explored (Building, Door, Fence, Streetlight, Tree, Window, Hedge, and Garage). These features have been selected as starting point in this research and the final ambition is to significantly extend beyond these. Each feature requires extensive training and therefore has a high associated human time cost. The eight were selected for their natural alignment to the built environment and crime, but the authors recognise that there may be other

features that are important to include, but have not been incorporated into this proof-of-concept.

- Computational approach: The approach developed and presented in this paper is an exploratory proof-of-concept and as such there are numerous improvements that could be considered in future work. For example, alternative algorithms with different capabilities, considering varying numbers of impacts, and also in establishing and training the models of an increased feature set. The techniques used in this paper were selected because of the advantageous characteristics, such as ease-of-use and being open-source for free use.

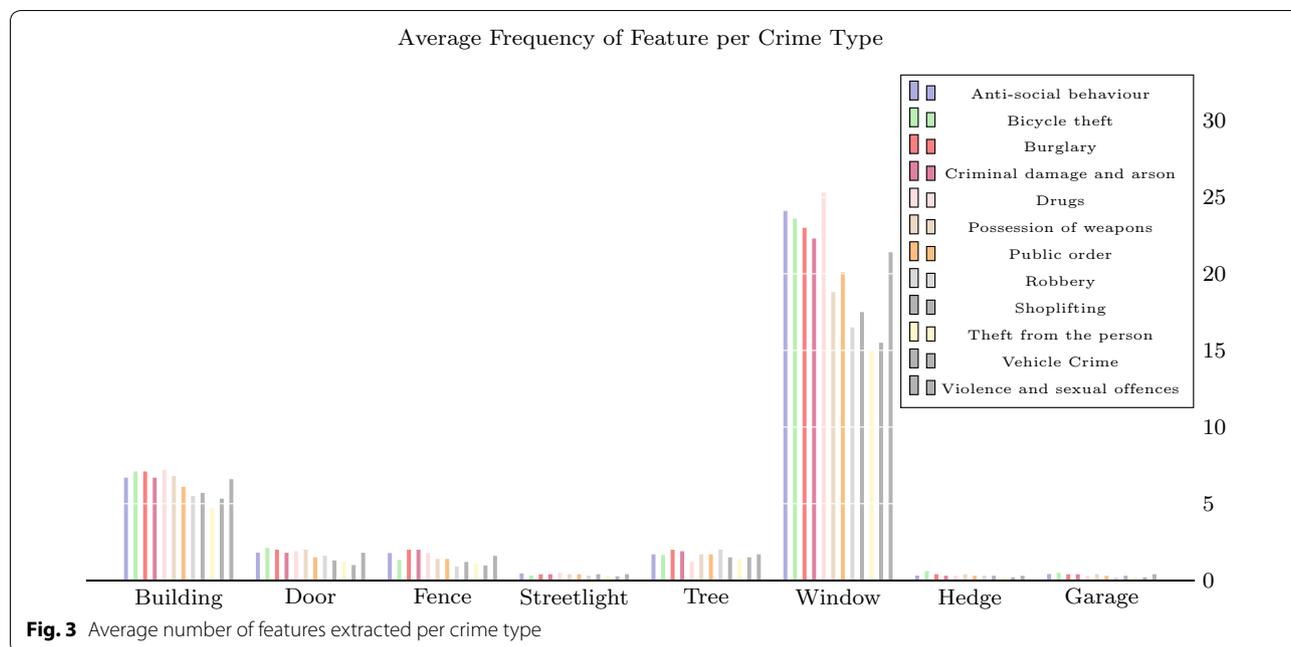


Fig. 3 Average number of features extracted per crime type

Results and discussion

In this section, the systematic analysis of all identified features is presented and discussed. In total, 60,100 images were extracted for the 5288 crimes in the ward of interest, providing an approximate average of 10 images per crime instance. This section is structured as follows: we first provide a descriptive analysis of the results, followed by the use of a statistical measure of dependence to determine if there are features that are strongly correlated to a specific crime type. The purpose of using a correlation technique is to determine if a feature of the built environment (e.g., streetlight) occurs more often than not with certain crime types (e.g., burglary). The statistical correlation technique used in this research is the χ^2 analysis technique.

Identified features

The data produced by the object detection algorithm was grouped for each crime type and then we performed various methods of statistical analysis. More specifically, we recorded the minimum, average, and maximum number of times the feature has been identified in an image of the same crime type. The data was combined by adding the results of images with the same crime ID and location as this would result in the data being from a 360 degree viewpoint of the crime location.

Table 2 presents the number of each crime type for a 12 month period between July 2018 to August 2019. The table presents the minimum, maximum and average number of each feature type per instance of crime

per crime category, except those in the ‘other’ category that are not included in this research. It is evident in the table that the minimum for each crime type and feature combination is 0, which is due to the fact that for each crime type there was at least one image with no identifiable features. Furthermore, it is evident that there is wide variation in the maximum and average number of features identified and per crime type. The difference (maximum–minimum) is more significant for certain features, demonstrating that for a specific crime instance there are fewer of those features to be identified. For example, as demonstrated through the average values in Table 2, there is a significant difference between the average number of buildings (5.9) and the average number of streetlights (0.5) and garages (0.3). However, this is to be expected as although the number of streetlights on a street might be high, their height means there is a greater chance that they will not be present in the image.

Figure 3 provides a graphical illustration of the data provided in Table 2. The figure presents the average values from the table, which represent the average number of features identified per crime instance for each crime category. From the figure, it is noticeable that the occurrence of each feature has some consistency, with window and building being the two highest. It is immediately evident that there is a proportional relationship between buildings, doors and windows as would be expected. This is because these features have been identified in a high portion of the images.

Table 3 Results for Chi square test

	Anti-social behaviour	Bicycle theft	Burglary	Criminal damage and arson	Drugs	Possession of weapons	Public order	Robbery	Shoplifting	Theft from the person	Vehicle Crime	Violence and sexual offences
Building	0.049171736	0.00419465	0.002945274	0.017695427	0.012567089	0.075277432	0.000659354	0.008140438	0.009503511	0.0000146	0.060032852	0.000323175
Door	0.009615258	0.020793548	0.006651332	0.000901065	0.010684809	0.103274205	0.007500919	0.027079278	0.017311229	0.004289079	0.060832707	0.000445047
Fence	0.007773285	0.075287479	0.07721061	0.100997268	0.001947025	0.001496339	0.000793808	0.11040128	0.001095057	0.001080291	0.023048432	0.001022461
Streetlight	0.003570158	0.025418916	0.01303054	0.00027064	0.000587713	0.002820387	0.000823301	0.002054649	0.009862088	6.52E-06	0.001494906	0.003103302
Tree	0.017455794	0.042665303	0.003518992	0.00027231	0.329965117	0.000790135	0.0004865	0.305091336	0.004736951	0.035206987	0.022780964	0.000388565
Window	0.06136163	0.013984676	0.012640516	0.00075709	0.156903242	0.194567982	0.006014738	0.057928357	0.001242845	0.000234946	0.001962376	2.03E-06
Hedge	0.009413935	0.152135475	0.001508649	0.004097326	0.021840133	0.021882287	0.000464755	0.000323893	0.00142401	0.006047027	0.007449936	0.002954804
Garage	0.001524997	0.026887385	0.001979047	0.006038218	0.003143268	0.021624719	0.000398972	0.014427641	0.005301108	0.039748623	0.009864861	1.63E-05
Max	0.06136163	0.152135475	0.07721061	0.100997268	0.329965117	0.194567982	0.007500919	0.305091336	0.017311229	0.039748623	0.060832707	0.003103302
Average	0.019985849	0.045170929	0.01493562	0.016378668	0.0672048	0.052716686	0.002142793	0.065680859	0.0063096	0.01082851	0.023433379	0.001031965
Difference	0.041375781	0.106964546	0.06227499	0.0846186	0.262760318	0.141851296	0.005358126	0.239410477	0.011001629	0.028920113	0.037399328	0.002071337

Table 4 Results ordered by highest Chi square result for each category

	Anti-social behaviour	Bicycle theft	Burglary	Criminal damage and arson	Drugs	Possession of weapons	Public order	Robbery	Shoplifting	Theft from the person	Vehicle Crime	Violence and sexual offences
Window						Window	Door	Tree	Door	Garage	Door	Streetlight
Building	Hedge	Fence	Streetlight	Fence	Window	Door	Window	Fence	Streetlight	Tree	Building	Hedge
Tree	Fence	Tree	Window	Garage	Hedge	Building	Streetlight	Window	Building	Hedge	Fence	Fence
Door	Tree	Garage	Door	Hedge	Building	Hedge	Fence	Door	Garage	Door	Tree	Door
Hedge	Garage	Streetlight	Tree	Door	Door	Garage	Building	Garage	Tree	Fence	Garage	Tree
Fence	Streetlight	Door	Building	Window	Garage	Streetlight	Tree	Building	Hedge	Window	Hedge	Building
Streetlight	Door	Window	Garage	Tree	Fence	Fence	Hedge	Streetlight	Window	Building	Window	Garage
Garage	Building	Building	Hedge	Streetlight	Streetlight	Tree	Garage	Hedge	Fence	Streetlight	Streetlight	Window

χ^2 analysis

The use of the χ^2 statistic measure has long since been used to measure the independence between terms and categories in text categorisation (Yang and Pedersen 1997). The challenge of determining independence and dependence between terms and categories in information retrieval systems shares many characteristics of measuring the relationship between crime type categories and features of the built environment. The χ^2 statistical measure has many successful applications in data mining and knowledge extraction tasks, particularly those in information security (Parkinson and Crampton 2016; Parkinson and Khan 2018). In this research, we utilise a two-way contingency table of feature f and crime type category c , where A is the number of times feature f and crime type c co-occur, B is the number of times f occurs without c , C is the number of times c occurs without f , D is the number of times neither f or c occur, and N is the total average number of objects detected. A measure of dependence is calculated by:

$$\chi^2(f, c) = \frac{N(AD - CB)^2}{(A + B)(A + C)(B + D)(C + D)} \quad (1)$$

Table 3 presents the values for each feature, f , and crime type category, c , with the strongest dependency values in italic. A strong dependency value means that the feature has been identified as being one that occurs the most frequently for that specific crime type. As evident in the table, the values are between 0 and 1, where 0 specifies independence between f and c and 1 specifies a strong dependency. The features for each crime type category with the highest value are highlighted, and it is evident that some crime types have the same feature as having the strongest dependency as other crime types. For example, burglary, criminal damage and arson share fence as being the feature with the strongest dependency. Following the calculation of χ^2 scores, it is then useful to compute the mean χ^2 for each crime type using the following equation where l is the number of features for each crime type:

$$\chi_{avg}^2(c) = \frac{1}{l} \sum_{j=1}^l \chi^2(a, f_j) \quad (2)$$

In Table 3, we include $\chi_{avg}^2(c)$ values, as well as a difference measure between the feature with the strongest dependency measure (c_{max}) and the average by calculating:

$$c_{diff} = c_{max} - \chi_{avg}^2(c) \quad (3)$$

In Table 3, c_{max} values are highlighted in italic. Furthermore, c_{diff} are illustrated in Fig. 4. Additionally, in Table 4, the features are in descending order $\chi^2(f, c)$,

with the top feature being the one that has the highest dependency score to the corresponding crime type. It is immediately evident from cross-referencing Table 3 and Fig. 4 that some crime categories have a better measure of dependency with a single feature; however, overall the $\chi^2(f, c)$ are low and do not go beyond 0.33, which is the dependency measure for the feature of tree to the crime category of drugs. This demonstrates that in general there is a weak dependency between the features and crime types. These results are not surprising when considering the necessary abstraction and generalisation required to overcome location challenges. However, in an attempt to understand these weak dependencies, further analysis and discussion are performed to consider the dependency scores for each feature and crime type pair.

In addition, Fig. 5 illustrates the $\chi^2(f, c)$ values in order of greatest to smallest for each feature within each crime category. The ordering of features is the same as in Table 4 but the individual graphs enable an easy understanding of the significance of each feature versus crime type. In each of the plots provided in Fig. 5, a best fit trend line generated by using linear regression is also added. The purpose of the trend line is to demonstrate how the the relationship between feature and crime type is increasing (in terms of $\chi^2(f, c)$ scores). The distance from the trend line can be used to state in comparative terms how strongly a feature relates to a crime category. From the use of these graphs and tables, we can determine that:

- Only bicycle theft, theft from the person, and violence and sexual offences categories have a unique (across all crime types) feature scoring the highest $\chi^2(f, c)$. More specifically, only bicycle theft with hedge, theft from the person with garage, and violence and sexual offences with streetlight. However, the c_{diff} for each category (difference between average $\chi^2(f, c)$ and $\chi_{avg}^2(c)$) is 0.10, 0.03, and 0.002, respectively. This demonstrates that significance of the top feature beyond the average is poor. Interestingly, from analysing Fig. 5b for bicycle theft, Fig. 5j for theft from the person, and Fig. 5l for violence and sexual offences, it is evident that there two features are either beyond or very close to the linear best fit line. This suggests that the top two features can be used to differentiate the crime category.
- There are many crime type categories that have the same feature with the highest $\chi^2(f, c)$ scores. The following observations can be made: (1) The window feature is identified as the best for both anti-social behaviour and Possession of weapons; (2) fence for both burglary and criminal damage and arson; (3) Tree for both drugs and robbery; and (4) Door for

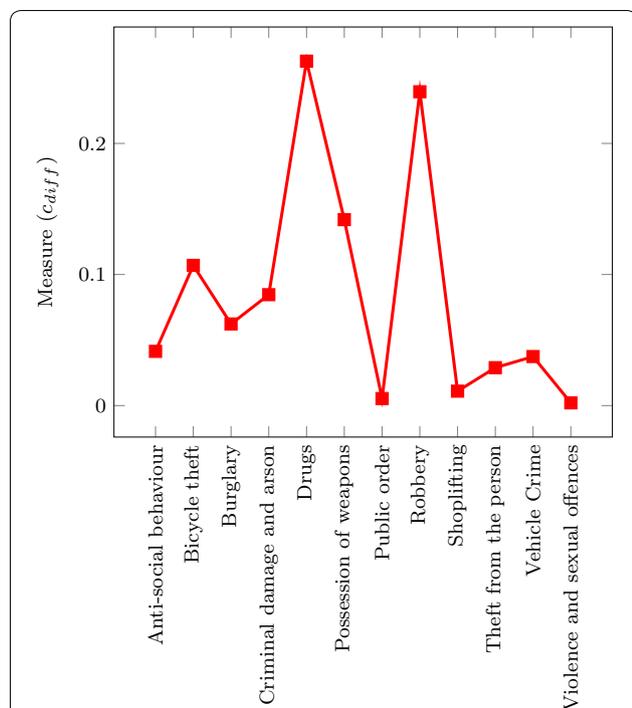


Fig. 4 Difference between the feature with the strongest dependency (c_{diff}) and the average $\chi^2_{avg}(c)$ for each crime type category c

public order, shoplifting and vehicle crime. Using Figs. 4 and 5 allows us to determine that for (1) the window feature has a lower (difference of 0.129) dependency score for anti-social behaviour when compared to possession of weapons. For observation (2), fence has a similar low score for both burglary and criminal damage and arson (difference of 0.03). For observation (3), tree has a similar relatively high score (greater than 0.3) for both drugs and robbery with a small difference of 0.03. For observation (4), door is common across three crime type categories, each having a low score (0.007, 0.017, and 0.0608, respectively)

- In terms of considering the top two features, no two crime types contain the same. This is significant as it means that those crime types sharing the same top feature do not share the next best feature. For example, considering observation (2) where fence has a similar low score for both burglary and criminal damage and arson, the next best features are street-light and building for burglary and criminal damage and arson, respectively. However, when considering the line of best fit in both Figs. 5c and d, it is evident that both these features fall below the linear best fit

line, indicating that they are a poor differentiating factor.

Based on the above observations, it can be established that it is possible to use the top two highest occurring features to differentiate images based on the average occurrence of that feature for each crime type. However, this discrimination would result in poor accuracy as the data is inconsistent, meaning that there are many instances where the top two crime features have either 0 or a significantly higher number of the identified top two. This will result in a large degree of incorrect classification. We can therefore state that the data does not have sufficient consistency to enable an automated classification approach. However, this finding is significant as it motivates the need for more accurate and precise data, especially involving a crime’s location.

Conclusion and future work

In this work a new approach to learning patterns between attributes of the built environment and crime is presented. The work is strongly motivated through the desire to improve the understanding of how characteristics of the built environment impact upon crime. As highlighted in the introduction, there is a wealth of research on the subject of CPTED. However, as yet, there is a lack of research utilising automated computing resources and intelligence to investigate patterns beyond what is currently known. This research set out to perform a first-of-a-kind exploratory study in this space. As such, there are many limitations to the study. Most notable, the small set of features, approximated location of crime, and limitation to studying one ward location. However, the approach presented in this paper is sufficient to motivate many future research directions.

The main finding of this research are that it is possible to train machine learning algorithms to recognise how to differentiate between different features in the built environment. Furthermore, due to limitations with publicly available datasets, a mechanism has been derived to acquire images from locations within the local proximity of the crime’s arbitrary assigned location. This mechanism clearly introduces uncertainty over whether or not the acquired images are of the crime’s actual location. The presented approach has good scalability and was able to process in-excess of 60,100 images from one small neighbourhood ward. A statistical dependency test was then used to establish if any features were particularly well correlated to a crime type. Only weak correlations were discovered, but this is not surprising considering the absence of precise crime locations and also that the location used for testing has a high volume of crime within a small geographic area, resulting in a high percentage

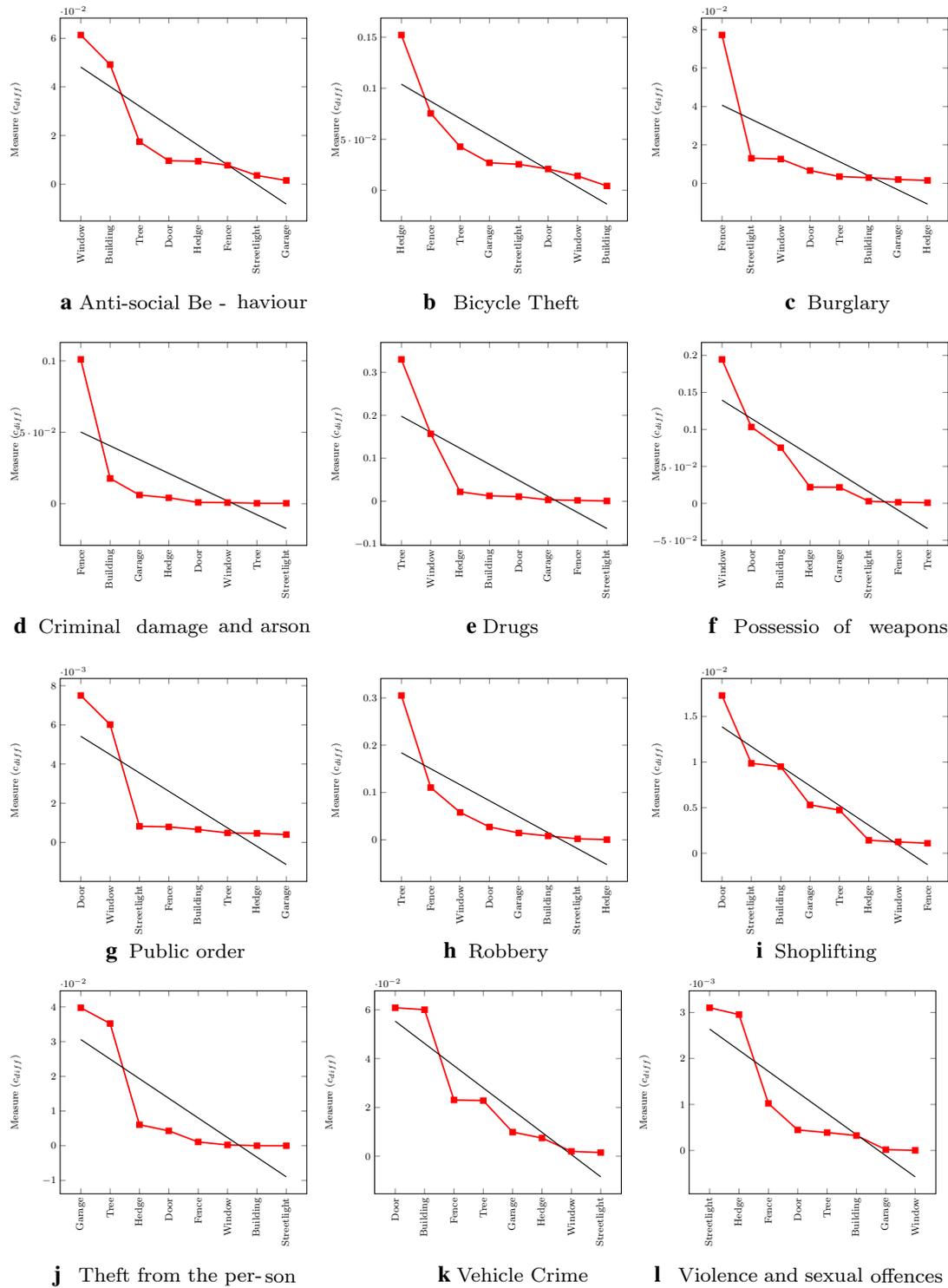


Fig. 5 $\chi^2(f, c)$ for each feature f and crime type c in decreasing order. The solid line is a best fit trend line used to illustrate features that have a stronger correlation

(75%) of cross-over between images and different crime types.

In terms of key findings between features of the built environment and crime, the following three points are summarised:

- Only bicycle theft, theft from the person, and violence and sexual offence have a unique feature as their highest dependency;
- There are many features identified as the best for multiple crime types. However, this is to be expected as there are fewer features than crime types. For example, a Fence is identified as having the strongest dependency for both the crime categories of burglary and criminal damage and arson; and
- When considering the two features that have the strongest dependency for each crime type, it is evident that each crime type has a different two features with occurring the highest. This is significant as it demonstrates that there are quantifiable differences between environmental characteristics and the locations where crime of the same type takes place.

Although this work has limitations, it presents an approach and lays foundations for future research in analysing the relationship between attributes of the built environment and crime. We see many avenues of future research activity within this area. The first is to overcome the approximation of location through acquiring crime data with precise location details. The second is the expansion of our feature set to identify many more characteristics of the built environment. The study of a larger geographic area should be undertaken to try and identify stronger patterns. The consideration of other data sources detailing features of the built environment, such as those available by local authorities and mapping agencies will be considered as an additional means of acquiring data. It is also important to mention that the authors recognise that this research has great potential throughout many different research arenas, and is not limited to crime reduction. We see the potential to automatically acquire and study fine-details of the built environment as a new source of information for research to use in conjunction with other crime data sources and systems to provide more useful insights. For example, focussing purely on acquiring streetlight locations on a large geographic scale, it might be possible to learn useful information as to their significance in the selection of crime location. The authors envisage these avenues of research resulting in techniques capable of complimenting current working practices that are largely qualitative, and not to serve as their replacement.

Acknowledgements

The authors thank in advance the anonymous reviewers for taking the time to read and review our manuscript.

Authors' contributions

KD and WX undertook the technical development under the supervision of SP and SK, who both led the data analysis and manuscript preparation. LM and KP provided valuable subject and research expertise throughout, as well as contributing towards the preparation of the manuscript. All authors read and approved the final manuscript.

Funding

No outside funding was used to support this work.

Availability of data and materials

All data and code used in this research are available from the authors upon request.

Competing interests

The authors declare that they have no competing interests.

Author details

¹ Department of Computer Science, University of Huddersfield, Queensgate HD1 3DH, UK. ² Applied Criminology & Policing Centre, University of Huddersfield, Queensgate HD1 3DH, UK. ³ University of Derby, Kedleston Road, Derby DE22 1GB, UK.

Received: 12 March 2020 Accepted: 15 June 2020

Published online: 08 July 2020

References

- Armitage, R. (2013). *Crime prevention through housing design: Policy and practice*. Berlin: Springer.
- Armitage, R., & Monchuk, L. (2011). Sustaining the crime reduction impact of designing out crime: Re-evaluating the secured by design scheme 10 years on. *Security Journal*, 24(4), 320–343.
- Armitage, R., & Monchuk, L. (2017). What is CPTED? Reconnecting theory with application in the words of users and abusers. *Policing: A Journal of Policy and Practice*, 13, 312–330.
- Armitage, R., Rogerson, M., & Pease, K. (2013). What is good about good design? Exploring the link between housing quality and crime. *Built Environment*, 39(1), 140–161.
- Cozens, P. M., Saville, G., & Hillier, D. (2005). Crime prevention through environmental design (CPTED): A review and modern bibliography. *Property Management*, 23(5), 328–356.
- Crowe, T. (2000). *Crime prevention through environmental design*. Oxford: Butterworth-Heinemann.
- Crowe, T., et al. (2000). *Crime prevention through environmental design*. Oxford: Butterworth-Heinemann.
- Ellingworth, D., & Pease, K. (1998). Movers and breakers: Household property crime against those moving home. *International Journal of Risk Security and Crime Prevention*, 3, 35–42.
- Heeks, M., Reed, S., Tafsiri, M., & Prince, S. (2018). The economic and social costs of crime second edition. Technical report, Research Report 99. London, UK: UK Home Office.
- Houghton, O. (1974). Defensible space-people and design in violent city-newman, o.
- Ihlanfeldt, K. & Mayock, T. (2010). Panel data estimates of the effects of different types of crime on housing prices. *Regional Science and Urban Economics*, 40(2):161 – 172, ISSN 0166-0462. <https://doi.org/10.1016/j.regsciurbeco.2010.02.005>. URL <http://www.sciencedirect.com/science/article/pii/S0166046210000086>.
- Jacobs, J. (1961). *The death and life of great American cities*. New York: Vintage.
- Kronkvist, K. (2013). Systematic social observation of physical disorder in inner-city urban neighborhoods through google street view: The correlation between virtually observed physical disorder, self-reported disorder and victimization of property crimes.

- Langton, S. H., & Steenbeek, W. (2017). Residential burglary target selection: An analysis at the property-level using Google Street View. *Applied Geography*, 86, 292–299.
- Monchuk, L., Pease, K., & Armitage, R. (2018). Is it just a guessing game? The application of crime prevention through environmental design (CPTED) to predict burglary. *Planning Practice & Research*, 33(4), 426–440.
- Monchuk, L., Parkinson, S., & Kitchen, J. (2019). Towards automating crime prevention through environmental design (CPTED) analysis to predict burglary. *Proceedings of the International Conference on Automated Planning and Scheduling*, 29, 539–547.
- Montoya, L., Junger, M., & Ongena, Y. (2016). The relation between residential property and its surroundings and day-and night-time residential burglary. *Environment and Behavior*, 48(4), 515–549.
- Naik, N., Philipoom, J., Raskar, R., & Hidalgo, C. (2014). Streetscore-predicting the perceived safety of one million streetscapes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, (pp. 779–785).
- Parkinson, S., & Crampton, A. (2016). Identification of irregularities and allocation suggestion of relative file system permissions. *Journal of Information Security and Applications*, 30, 27–39.
- Parkinson, S., & Khan, S. (2018). Identifying irregularities in security event logs through an object-based chi-squared test of independence. *Journal of Information Security and Applications*, 40, 52–62.
- Pease, K. (2010). Crime science. In *International Handbook of Criminology*, (pp. 29–48). CRC Press.
- Poyner, B. (1983). *Design against crime: Beyond defensible space*. London: Butterworths.
- Scott, M.S. & Clarke, R.V. (2020). Problem-oriented policing: Successful case studies.
- Tetlock, P.E. (1985). Accountability: A social check on the fundamental attribution error. *Social Psychology Quarterly*, pp. 227–236.
- Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: Improving decisions about health, wealth, and happiness*. City of Westminster: Penguin.
- Thompson, R., Tseloni, A., Tilley, N., Farrell, G., & Pease, K. (2018). *Which Security Devices Reduce Burglary?*, pages 77–105. 11 2018. ISBN 978-3-319-99941-8. https://doi.org/10.1007/978-3-319-99942-5_4.
- Vandeviver, C. (2014). Applying Google Maps and Google Street View in criminological research. *Crime Science*, 3(1), 13.
- Vandeviver, C. & Bernasco, W. (2019). Location, Location, Location: Effects of neighborhood and house attributes on burglars target selection. *Journal of Quantitative Criminology*, pp. 1–43.
- Yang, Y., & Pedersen, J. O. (1997). A comparative study on feature selection in text categorization. *ICML*, 97, 35.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

