Street-Frontage-Net: urban image classification using deep convolutional neural networks

Quantifying aspects of urban design on a massive scale is crucial to help develop a deeper understanding of urban designs elements that contribute to the success of a public space. In this study, we further develop the Street-Frontage-Net (SFN), a convolutional neural network (CNN) that can successfully evaluate the quality of street frontage as either being active (frontage containing windows and doors) or blank (frontage containing walls, fences and garages). Small scale studies have indicated that the more active the frontage, the livelier and safer a street feels. However, collecting the city-level data necessary to evaluate street frontage quality is costly. The SFN model uses a deep CNN to classify the frontage of a street. This study expands on the previous research via five experiments. We find robust results in classifying frontage quality for an out-of-sample test set that achieves an accuracy of up to 92.0%. We also find active frontages in a neighbourhood has a significant link with increased house prices. Lastly, we find that active frontage is associated with more scenicness compared to blank frontage. While further research is needed, the results indicate the great potential for using deep learning methods in geographic information extraction and urban design.

Keywords: urban design, deep learning, convolutional neural network, machine vision, London, Google Street View, object recognition

1.0 INTRODUCTION

Advances in computer vision using artificial neural networks have allowed aspects of urban design to be quantified on a massive scale that was previously not possible. This is crucial to help develop a deeper understanding of urban design elements that inform domains of urban planning such as economic progress, transportation and residential well-being. In this study, we specifically focus on the urban design concept of active versus blank frontage, where active frontage is defined as ground floor buildings having windows and doors as opposed to blank walls, fences and garages (ODPM 2005). Quantitatively, the concept of active frontage has been expressed through frontage

classes (Law et al. 2017), on which this study focuses, or through ordinal indicators such as the façade evaluation scale (Heffernan et al. 2014). The quality of street frontages is an important factor in urban design, as it represents the space where the built environment and pedestrians interact most closely (Gehl 1971; 2010; Heffernan et al. 2014). For example, the greater the number of doors and windows facing a street, the more "eyes-on-the-streets", which contributes to the perception of a safe, convivial and sociable street (Jacobs 1961; Alexander et al. 1977). There are also health benefits such as improved walkability, and economic benefits such as increased land value (Heffernan et al. 2014; Nase et al. 2013). Despite being advocated in planning guidance (Llewelyn Davies Yeang and Homes Communities Agency 2013), previous research on this topic had been limited, firstly in terms of scale, as the majority of previous research used small scale questionnaires, and secondly in terms of establishing what benefits active frontages provide (Heffernan et al. 2014; Kickert 2016; Nase et al. 2013). The collection of such data to evaluate the quality of street frontage is both costly and time-consuming.

One approach to collecting this data more efficiently is to cast this as an image classification problem (Krizhevsky et al. 2012). In previous research, the use of deep learning methods in urban frontage classification was examined (Law et al. 2017) with the creation of the Street-Frontage-Net model (SFN) to classify active versus blank frontages. The aim of this research is to expand on this previous research by refining and explaining the SFN model and also validating the importance of this urban design indicator through statistical analyses with both house price and scenicness.

In this study, we first evaluate the accuracy of the SFN model in predicting active versus blank street frontage using different deep convolutional neural network

architectures. We also evaluate the association of street frontage quality with increases in house prices, and with differences in scenicness. Finally, we explore the usefulness of our SFN model by visualising the neural network model and examining its predictions along the street segment.

2.0 RELATED WORKS

Despite the importance of active frontages in the urban design literature and the ubiquity of urban image data, limited urban computational research has been conducted on the classification of street frontages using street image data (Law et al. 2017; Liu et al. 2017). Previous computational research on street image data involved manual feature extraction to detect edges for classification. Image feature descriptors such as scale-invariant feature transform (SIFT) and histograms of oriented gradients (HOG) are used, for instance, to detect distinctive architectural features in the Parisian cityscapes (Doersch et al. 2012), to estimate the perceived safety via online crowd-source games (Streetscore, 2014; Naik et al. 2014), and to identify visual discriminative elements in predicting city attributes such as theft rates, graffiti presence and house prices (Arietta et al. 2016).

Recent advances in deep learning research (Bengio et al. 2015) have opened up new research opportunities in various domains including urban studies (Seresinhe et al. 2017). Of particular interest are Convolutional Neural Networks (CNNs), initially popularised by LeCun et al. (1998), which are well adapted to classifying images and extracting image features in machine vision. CNNs regained popularity in 2012 when the network architecture AlexNet (Krizhevsky et al. 2012) significantly outperformed previous architectures on the ImageNet database by reducing the top-five error from 26.0% to 15.3%. The ImageNet database is one of the most widely used visual

databases in the world for object recognition research, with over 14 million annotated images (Deng et al. 2009). Researchers continue to introduce several advances to CNN architectures to improve model accuracy and to extend them to other machine vision problems such as object detection and image segmentation (Girshick 2015; Ren et al. 2015). A few of the important architectures used in this study are illustrated below.

The VGG16 neural network architecture was introduced by Simonyan and Zisserman (2014) and has become one of the baseline CNN architectures used in object recognition due to its performance and simplicity. This network is the most straightforward of the four we use in this study, having thirteen convolutional layers, five pooling layers and three fully connected layers. This model achieves an accuracy of over 92.0% on the ImageNet database.

The ResNet-50 network was put forward by He et al. (2015) and uses extremely deep neural networks that are strung together with residual blocks. Instead of learning the mapping between inputs and outputs, the residual blocks simply learn the difference between the two, hence the name. The architecture we use in this study has a total of 50 layers.

The Inception-V3 network, which was introduced by Szegedy et al. (2015), extracts features from the Inception modules that use small convolutions within each module simultaneously. This approach reduces the number of parameters significantly within each module. The Inception-V3 network uses a global average pooling layer instead of multiple fully-connected-layers.

Lastly, the Xception architecture (Chollet 2017) introduced Depthwise Separable Convolutions into the Inception architecture. The Xception model learns the three

colour channels separately rather than simultaneously, hence the name Separable Convolutions. The Xception architecture can outperform Inception-V3 networks with a similar numbers of parameters.

CNNs are now being used to understand city characteristics, including place categories (Zhou et al. 2014a; Zhou et al. 2014b), the beauty of outdoor places (Seresinhe et al. 2017), demographic characteristics (Gebru et al. 2017) and in predicting house prices (Law et al. 2018). More recently, these techniques have been used to estimate greenery and street enclosures, which were found to have an effect on walkability (Li et al. 2018). More closely related, CNNs have also been used to predict the continuity of a street facade as rated by domain-experts (Liu et al. 2017) and in developing the Street Frontage Net model (SFN), a street frontage classifier, using both Google Street View images and 3D-model synthetic street images (Law et al. 2017).

3.0 METHOD AND MATERIALS

The aim of this study is to refine and validate the SFN model, which classifies the ground floor of a front-facing street view image into four frontage categories: blank, single-side active (single-active), both-sides active (both-active) and non-urban. In order to train the classifier, we use Greater London (Figure 1) as the case study from where two ground truth street view image datasets are collected. These ground truth images are subsequently used in the five experiments detailed in section 4.0.

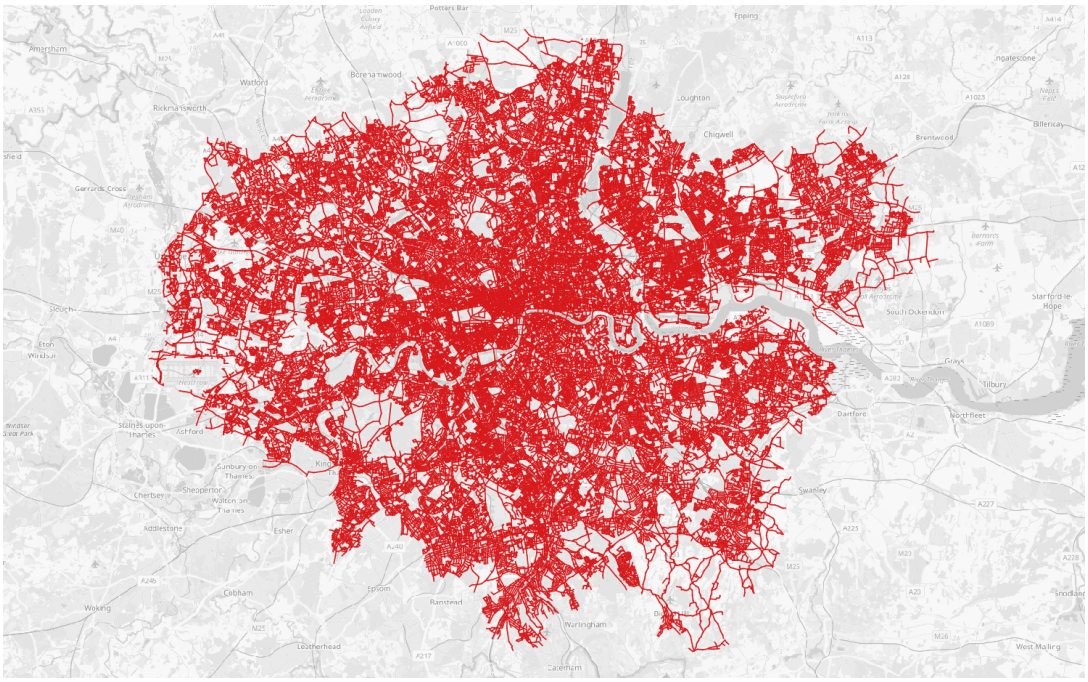

Figure 1: The objective of this study is to refine and validate the Street-Frontage-Net Model. The Greater London study area is used as the case study. Contains Ordnance Survey data. Crown copyright and database right 2017.

3.1 Materials

This study uses the same datasets that were used in Law et al. (2017), which include street images collected from the Google Street View API (Google 2017) and street images generated from the CityEngine software (ESRI 2013). To collect the Google Street View dataset, we first construct a graph from the street network of London, using the OS Meridian line2 dataset (Figure 2). More specifically, we take the geographic median and the azimuth of the street edge between two junctions to determine both the location (longitude, latitude) and the bearing (degree) of each street view image location. We then download a total of 112,650 front-facing street images in London, and pre-process them by removing invalid images, such as those taken inside a building, those that are too dark and those that are not available on Google Street View. We then resize the valid images to a set dimension (256 x 256 pixels), and hand label

approximately 10% of the total images (10,004) by frontage class, of which 2,632 are blank, 2,001 are single-active, 2,814 are both-active and 2,557 are non-urban. For the Google Street View images, four street frontage classes are adopted, namely: 0 - blank frontage on both sides of the street, 1 - active frontages on one side of the street, 2 - active frontage on both sides of the street and 3 - non-urban images (Figure 3).

The second dataset comprises street images generated from a 3D model of an abstract city in ESRI CityEngine, which is a parametric engine for city building (ESRI 2013). Creating synthetic image data using a 3D virtual environment can greatly enhance the efficiency of collecting urban image data. In total, 4,800 images are generated with this process. We remove invalid images, such as repeated images, images that are near an intersection and those at the end of a road. This process results in 1,029 filtered images, which are then similarly resized to a set dimension (256 x 256 pixels). The ground truth labelling is performed automatically, and three sets of images are produced: 0 - blank frontages on both sides of the street, 1 - active frontages on one side of the street and 2 - active frontages on both sides of the street (Figure 4). The non-urban frontage class is not included, as it is not realistic to synthetically create non-urban scenes using the parametric software.

Figure 2: This figure shows the steps involved in retrieving a front-facing Google Street View image. From left to right: the street network graph, get the location and bearing of

each street, download the street image and manually label the image. 2017 Google Inc. Google and the Google logo are registered trademarks of Google Inc.

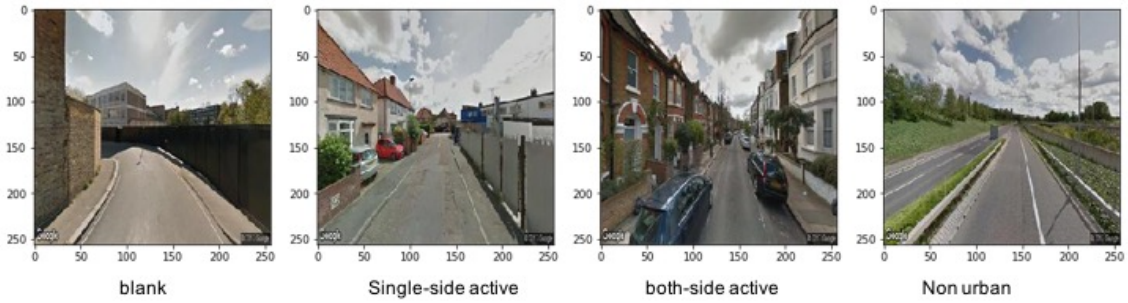

Figure 3: This figure shows four typical Google Street View urban frontage images. Active frontage is defined as the ground floor building frontage having windows and doors, as opposed to blank walls, fences and garages. From left to right: blank frontage, single-active frontage, both-active frontage, and non-urban frontage. 2017 Google Inc. Google and the Google logo are registered trademarks of Google Inc.

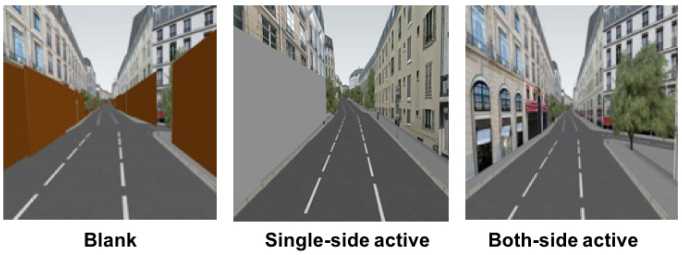

Figure 4: This figure shows the 3D model urban frontage images. From left to right: blank frontage, single-active frontage and both-active frontage. 2017 ESRI. All rights reserved.

4.0 EXPERIMENTS

We use the image datasets, as detailed in *the method and materials section,* in the following experiments. In the first experiment, the sensitivity and robustness of the SFN model are tested by comparing different CNN architectures. In the second experiment, the frontage variables are used as inputs in a house price regression model. The relationship between scenicness and urban frontage quality are examined in the third experiment. In the fourth experiment, predictions about model generalisations along

different sections of the street are tested. In the final experiment, a CNN visualisation method is applied to interpret the results of the model.

4.1 Experiment One – CNN model selection and tests

Law et al.'s (2017) study showed that using a simple AlexNet CNN model to classify the frontage of a street image results in a high accuracy (75%). However, with recent advances in CNN architecture, Law's results need to be re-examined using state-of-the art CNN architecture. This first experiment is divided into three parts. First, we test two transfer learning CNN models using the VGG16 architecture, whereby one model uses the weights learnt from the ImageNet database (Deng et al. 2009) and another model uses the weights learnt from the Places365 database (Zhou et al. 2014a). While the ImageNet database focuses on object recognition, Places365 concerns the recognition of place categories, such as "residential neighbourhood" and "train station platform", and thus should be more relevant for this research. In the second part of this experiment, we compare the VGG-16 model (Simonyan and Zisserman 2014), the ResNet-50 model (He et al. 2015), the Inception-V3 model (Szegedy et al. 2015) and the Xception model (Chollet 2017). All these models were pre-trained using the ImageNet dataset and subsequently fine-tuned for the study. In the third part of this experiment, we will compare running these four models with and without the 3D model images. The objective is to study the robustness of the results. Furthermore, we will perform a spatial out-of-sample test, where the trained model will make inferences for another city – Paris – to study the extent to which the London-SFN model generalises to a different geography.

4.2 Experiment Two – estimating the effect of urban frontage on house price through the hedonic price approach

The ultimate aim of this research is to contribute a valuable method to urban studies for analysing city life, from traffic flows to the provision of housing. Example research with street frontage might include to what extent street frontage quality might affect well-being or house prices. Previous studies using urban characteristics such as street frontage have been limited to small areas due to the cost and time required to collect such data (Nase et al. 2013). However, deep learning methods can retrieve large-scale results of estimated effects on urban performance. The aim of this second experiment is to test the effects of different frontage classes on house prices. We frame the second experiment as a simplified house price regression problem, whereby house prices are broken down into their utility-bearing components, including frontage quality, using the hedonic price approach (Rosen 1974; Cheshire and Sheppard 1995). This concept is analogous to comparing two properties, each with nearly identical features, except that one property has one bedroom and the other has two bedrooms. The price differential between the two is equal to the implicit price of the extra bedroom. This approach often includes structural features such as the size of the house and its type. It can also include location features such as job accessibility, or neighbourhood features such as land use diversity. Since its introduction, the hedonic price approach has become an established real estate method for pricing environmental goods and constructing housing price indices, and serves as evidence in the development of welfare policies (Palmquist 1984; Ridker and Henning 1967). To date, the effects of urban frontage on property values has largely been supported by a limited number of quantitative studies (Nase et al. 2013). Testing the effects of street frontage quality on house prices can reveal the economic value of urban frontage design.

Table 1: Descriptive statistics for the house price model. The table includes the house price in 2011, number of bedrooms, size, age, levels of accessibility, land use diversity and the neighbourhood active frontage score.

Statistic	N	Mean	St.Dev.	Min	Max
lnprice	6,110	12.6	0.5	11.3	15.3
size	6,110	97.8	40	29	278
age	6,110	81.4	37.5	0	311
bedrooms	6,110	2.6	1	1	8
Accessibility	6,110	9.5	0.4	7.9	10.1
Land use div.	6,110	0.1	0.1	0	0.8
Active front R800	6,110	0.3	0.1	0.001	0.7

Table 2: Cross-tabulation for the frontage classes of the house price model.

Class	n	%
Blank	1189	19.46%
Single-active	1489	24.37%
Both-active	3153	51.60%
non-urban	279	4.57%

The dataset used in this experiment was compiled from the UK Land Registry (2017) with attributes from the Nationwide Housing Society (2012). The dataset for this experiment includes traditional housing attributes such as structural, neighbourhood and location features. The structural features for each property transaction include the price paid in 2011, the type of property, the number of bedrooms, the size and the age of the property. Location features include employment accessibility, which was computed as a gravity-based accessibility measure that takes the sum of jobs accessed within 60 minutes divided by their travel time. Neighbourhood features include distance to the nearest parks and land use diversity (entropy) within 800 meters. The datasets used to calculate these location features originate from the Ordnance Survey (2017), the Office for National Statistics (2017), Valuation Office Agency (2015) and Historic England (2017).

Two urban frontage indicators are used for the housing price model. The first is the urban frontage class, which is predicted for each street using the best performing SFN model from the first experiment. The second indicator is the neighbourhood frontage score, which is computed by taking, for each property, the average active frontage probability within 800m.

The overall hedonic price of the property is represented by a function ***H(·)***, parameterized by ***B*** that takes its housing attributes ***X***, its urban frontage class ***f*** and the neighbourhood active frontage score ***a***. The frontage class ***f*** is cast as a one hot encoding, where a binary dummy variable is created for each frontage class. We consider a baseline model ***H(X),*** which only depends on housing attributes, and a baseline + frontage model ***H(X, f, a),*** which depends on the housing attributes, the frontage class and the neighbourhood active frontage variable simultaneously. These equations yield the following multiple variable regression models:

$$H(X) = \beta_0 + \sum \beta_1 * X + \epsilon$$

(1)

$$H(X, f, a) = \beta_0 + \sum \beta_1 * X + \sum \beta_2 * f + \beta_3 * a + \epsilon$$

(2)

We then take model 2 to yield two sub-models. The first sub-model is estimated for Inner London and the second is estimated for Outer London. These sub-models aim to test the significance of frontage quality effects on house prices geographically. The boundary for the sub-models are shown in Figure 5. All the models are estimated by minimising the mean squared error loss function using the ordinary least square (OLS) method where standard regression statistics and significance are reported.

Figure 5: The Inner and Outer London boundary is used for the second experiment. We want to study the extent to which the association between frontage quality and house prices varies geographically.

Cross sectional OLS models have certain limitations, one of which is spatial autocorrelation, which can result in unreliable estimates for the frontage variables. As a result, a standard Moran's I test is computed for the model (Moran 1948). If significant, corresponding spatial regression models such as the spatial lag and spatial error models are then estimated as a robustness test (Anselin 1988).

4.3 Experiment Three – Scenicness and frontage quality comparison

In the previous experiment, we examined the extent to which urban frontage quality is associated with house prices. A related question is: how does urban frontage quality contribute to the visual aesthetics or scenicness of an environment, and how does this scenicness relate to specific architectural and environmental features? For example, do

people perceive classical architectural features with Corinthian columns as being scenic?

The aim of this experiment is to examine the extent to which frontage quality relates to perceptions of street view aesthetics. A logical conjecture is that a blank frontage not only negatively influences the safety perception of an area, but it also limits views of the surroundings, and thus reduces the environment's visual interest and therefore its scenicness. In order to test this idea, we conduct a small-scale exploratory descriptive study using a modified version of the Scenic-or-Not CNN model from Seresinhe et al. (2017).

This modified Scenic-or-Not CNN, henceforth referred to as the Street-View-Scenic CNN, uses ground truth data on scenicness from two different online games, *Scenic-Or-Not* and *Scenic-London.* The *Scenic-Or-Not* dataset comprises nearly 217,000 images, sourced from *Geograph (http://http://www.geograph.org.uk/)*, covering nearly 95% of the 1km grid squares in Great Britain, and was originally used to train the CNN presented in Seresinhe et al. (2017). In order to improve the performance for Google Street View images, this CNN is further trained on the 6,946 Google Street View images that comprise the *Scenic-London* dataset. We use the *Google Street View API* to randomly sample four images per Inner London Lower Layer Super Output Area (LSOA). LSOAs are defined by the Office for National Statistics for statistical analyses, with geographic areas ranging from 0.018 to 684 square km and between 983 and 8,300 residents (1,500 on average). We specifically choose to focus on Inner London to increase the sample of urban imagery our CNN is trained on, as we are most interested in understanding the characteristics of urban design. We use systematic unaligned sampling to generate coordinates for the images, whereby the sample space (i.e. the

entire LSOA) is split into four equally sized sub-areas, and a random [x, y] coordinate is generated for each sub-area.

For our *Scenic-London* dataset, 34,955 ratings were gathered, following a similar procedure to *Scenic-Or-Not* (Seresinhe et al. 2017), whereby we presented our images in a web interface to be rated on a scale of 1-10. Only images rated at least three times have been included in the CNN training dataset.

This Street-View-Scenic CNN currently achieves an accuracy score of 0.44 (Kendall's Rank correlation between the predicted scenic scores and the actual scenic scores) for Street View images. We then use the Street-View-Scenic CNN to predict scenic ratings on a higher resolution dataset of Google Street View images – around 500,000 images at a 100-square-meter-resolution grid for London.

For this experiment, we associate the predictions from the Street-View-Scenic CNN model with the predictions from the Street-Frontage-Net CNN model for the study area of Barnsbury (Figure 6). We make geographical visualisations to compare the predictions. We also conduct a one-way analysis of variance (ANOVA) test, which compares the mean scenic ratings between the different frontage classes and determines whether any of these means are statistically significantly different from each other.

Figure 6: The Barnsbury Case Study is used for the third and fourth experiment.

4.4 Experiment Four – within-street prediction tests

A shortcoming of the previous study (Law et al 2017) was the use of just a single sample image for each street. This type of sampling might be adequate at the city-scale when comparing the average distribution between neighbourhoods, reflecting the continuity of active frontages along a street. However, this sampling is inadequate for a study within a neighbourhood, as street-level differences can be important. As a result, in this fourth experiment, we test the extent to which the predictions of the SFN model generalise at different street segment resolutions, specifically how a single image model differs from predictions taken every 20m, 40m, 80m, 120m, 160m, 200m and 240m within the Barnsbury neighbourhood in London, as illustrated in Figure 6.

4.5 Experiment Five – model visualisation and interpretation

Recent advancements in deep learning have allowed computers to achieve human-like performance for many visual tasks. However, the application of CNNs can be limited due to the black-box nature of neural network models. In recent years, researchers have

developed models and methods to interpret these models. These studies have introduced methods such as visualising the layers of CNN models, diagnosing CNN representations, disentangling complex convolutional-layers, and building explainable models (Zhang et al. 2018). For this research, we use the most direct method in visualising CNN representation by using the gradient-based methods. Envisaging what the CNN sees can help us better trust a network prediction. More specifically, we apply Gradient-weighted Class Activation Mapping (Grad-CAM), which uses the gradient for a particular class that flows into the last convolutional layer of an input image to highlight activated regions for a particular class (Selvaraju et al. 2016). In particular, we apply the Grad-CAM method to four representative images of the four classes: blank frontage, single-side active frontage, both-side active frontage and non-urban image (Figure 3). The results can provide an indication of whether the model is acting upon the actual frontage area of an image rather than possible covariates within the image. Lastly, we visualise the predictions of the SFN model in order to interpret its geographical distribution.

4.6 Training Setup

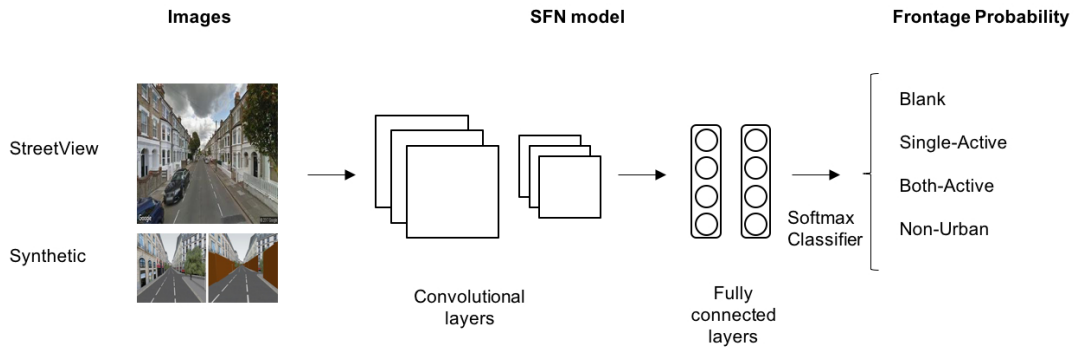

Figure 7: This figure shows the Street-Frontage-Net model, where a street view image is the input and the probability of each frontage class is the output.

The SFN model adopts a standard CNN architecture setup for image classification,

where a street view image is the input and the probability of each frontage class is the output (Figure 7). We use transfer learning to leverage the knowledge of a pre-trained neural network, which is a useful approach when millions of labelled images are not available to train a deep learning model from scratch. To train the transfer learning model for a different architecture, we first freeze the weights of the penultimate convolutional layer of each model, and then train the last convolutional layer and the subsequent layers for the new classification task. As each architecture studied in this paper is inherently different, future research should include architecture-specific sensitivity tests to optimise the accuracy for each CNN architecture. In this case, we try to maintain the settings as close to each other as possible for comparison purposes. We divide the dataset into training (70%), validation (15%) and testing (15%). We train the CNN using stochastic gradient descent with Nesterov momentum to minimise the categorical cross entropy loss function, as defined below, where **p** is the true distribution and **q** is the predicted probability distribution. The cross entropy loss function is typically used in a multi-category classification problem.

$$H(p,q) = -\sum p(x)\log\left(q(x)\right) \tag{3}$$

We train for 60 epochs with an initial learning rate of 10e-5. A Softmax activation function is used in the final layer to estimate the probability distribution of an image class, as detailed below, and the frontage class with the highest probability is selected as the predicted label.

$$p_{nk} = \frac{\exp(\mathrm{x_{nk}})}{\sum \exp\left(x_{nk}\right)}$$

(4)

5.0 RESULTS

5.1 Experiment One - CNN model selection

To identify the most appropriate setup and to test the sensitivity of the model, we first compare the Google Street View images' ground truth label with the most likely frontage class predicted in Model 1, which uses the ImageNet weights and Model 2, which uses the Places365 weights. The results in Table 3 indicate that while both models achieve a high accuracy, the model that uses the ImageNet weights (87.5%) is more accurate than the model that uses the Places365 weights (78.9%). We introduce here the confusion matrix to measure the performance of the classification model. In a confusion matrix, the i-th and j-th entries show the number of occasions where the network predicted category i corresponds to the observed category j. Hence, off-diagonal elements indicate mis-classifications. The confusion matrix (Figure 8) illustrates that the fine-tuned ImageNet model is significantly more accurate at predicting both the blank class and the single-active class than the fine-tuned Places365 model. This result is surprising considering the Places365 database has been built to more accurately reflect the features of environmental images. One reason for this result may be that a general image database such as ImageNet can capture more diverse sets of general image features than the Places365 database.

Table 3: Transfer learning accuracy comparing a VGG-16 ImageNet and a VGG-16 Places 365 model.

Transfer	accuracy
VGG16-Imagenet	87.50%
VGG16-Places365	78.90%

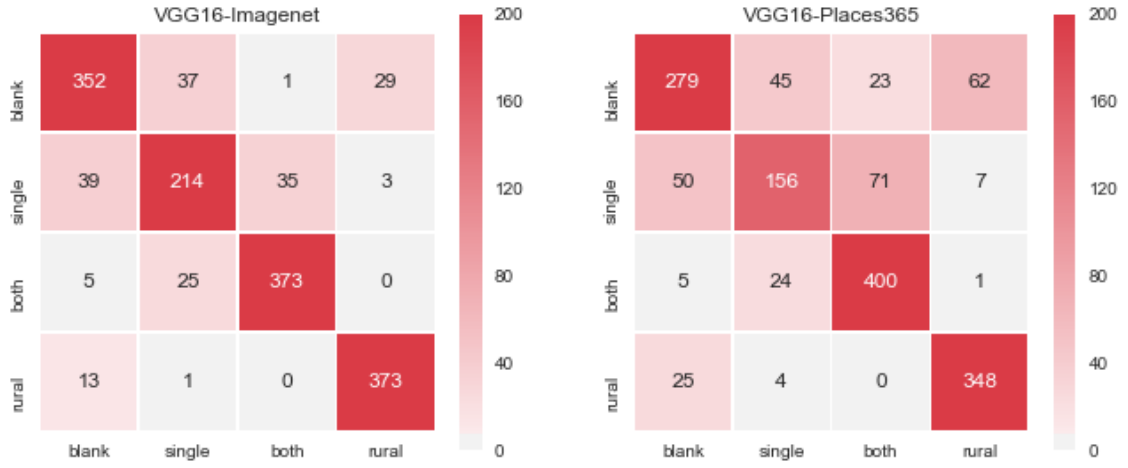

Figure 8: Confusion matrix for the transfer learning experiment comparing a VGG-16 ImageNet model and a VGG-16 Places365 model. The result shows that the model trained with the ImageNet database has a higher overall accuracy.

We then compare the same ImageNet-based model for four different architectures: VGG16, ResNet-50, Inception and Xception. The results (Table 4) show that VGG16 yields the most accurate results, followed by Xception, ResNet-50 and Inception-V3. This outcome is not surprising considering the simplicity of the classification task. As each neural network architecture differs significantly from the others, more research is needed to optimise for architecture-specific transfer learning strategies.

Table 4: Google Street Views prediction accuracy comparing the different CNN architecture.

baseline	accuracy
VGG16	87.50%
Res-50	77.90%
Inception	77.00%
Xception	81.30%

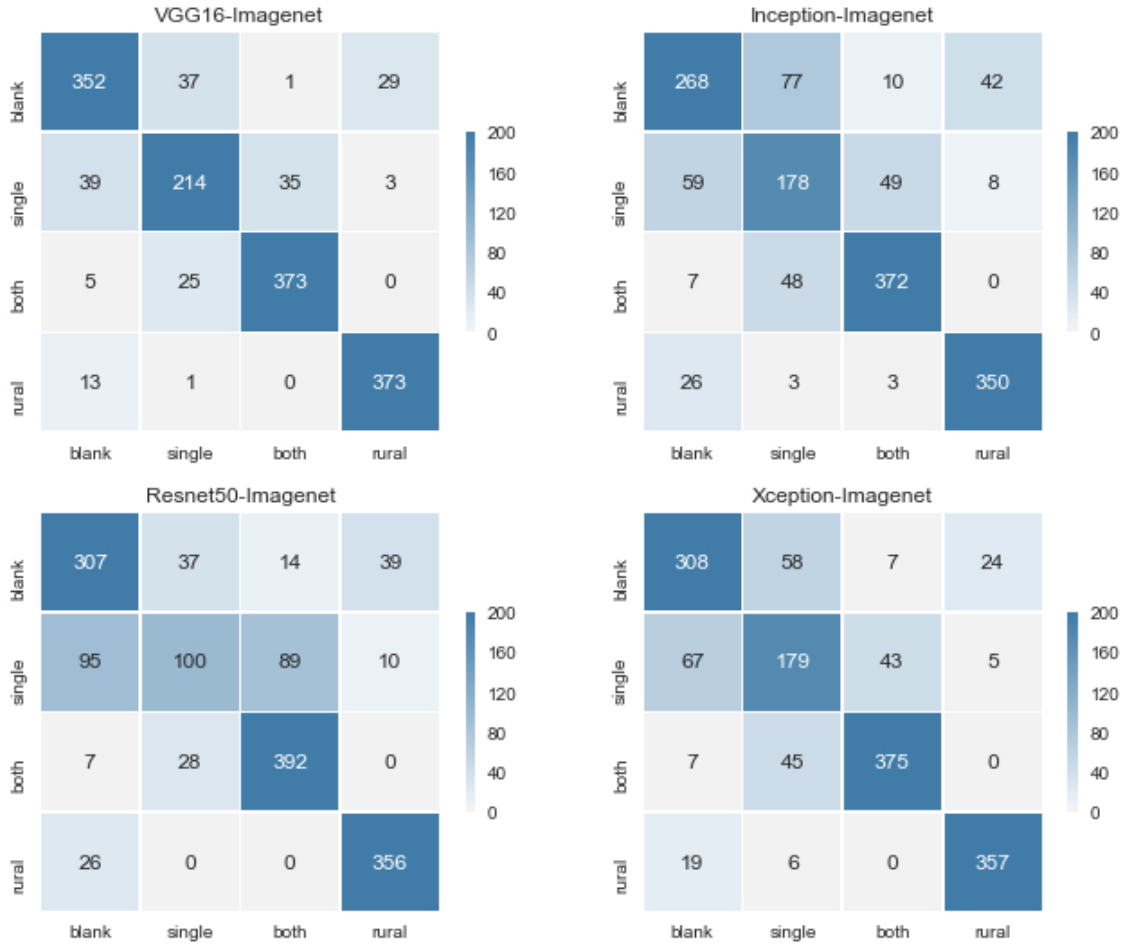

Figure 9: Confusion matrix results comparing different CNN architectures. The result show that VGG16 generally achieves the highest accuracy, followed by Xception, ResNet-50 and Inception-V3. The results show some confusion between the single-active class, the blank class and the both-active class.

The confusion matrix (Figure 9) illustrates that the VGG16 model generally predicts with a higher accuracy than the other three models. The single-active class displays slight confusion with both the blank frontage class and the both-active class. The non-

urban class, on the other-hand, has slight confusion with the blank frontage class. These results are not surprising, as there are clear overlaps between the single-active class with both the blank frontage class and both-active class. Future research in object detection is required to overcome these uncertainties. We then compare the model trained on Google Street View images with a model trained on both Google Street View images and 3D model street view images. The results (Table 5 and Figure 10) show that the model augmented with 3D-model images performs slightly better than the same model without 3D-model images, across all CNN architectures. This outcome improves previous SFN research (Law et al. 2017) and is robust across different CNN architectures.

Table 5: Google Street View + 3D model street view prediction accuracy

base+3D	accuracy
VGG16	92.70%
Res-50	78.10%
Inception	79.00%
Xception	81.30%

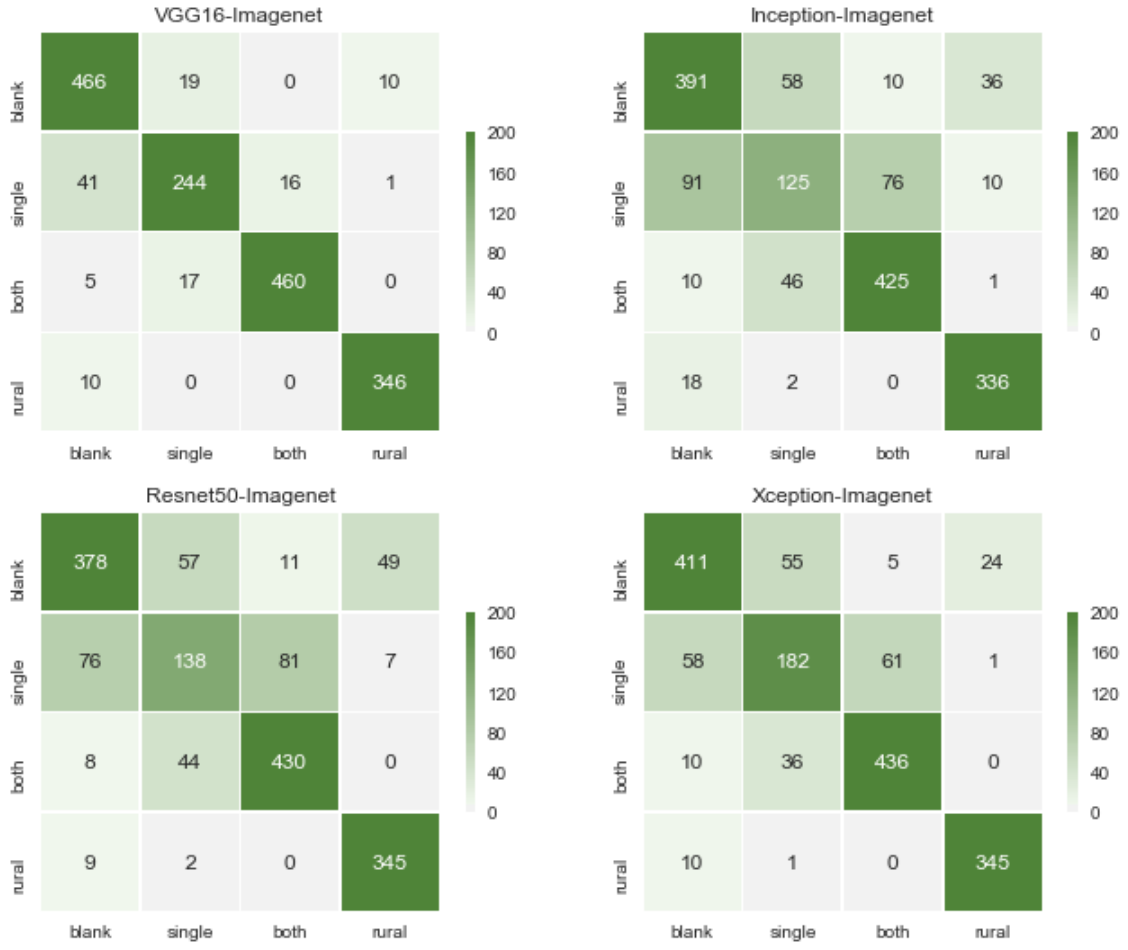

Figure 10: Confusion matrix results for the four CNN architecture using the Google Street View + 3D-model dataset. The results are similar to previous research, whereby the model augmented with 3D-model images performs better than the model without these synthetic images.

In order to further validate the result, a spatial out-of-sample test is conducted for Paris, France. 281 images located in Central Paris are randomly downloaded using the Google

Street View API and manually labelled according to the four frontage classes. The London-SFN model is then used to predict the labels of these images, which are then compared to the observed labels. The prediction achieves a high accuracy of 77.26% on these novel images, which demonstrates the strong generalisability of the model. The confusion matrix is shown in Figure 11.

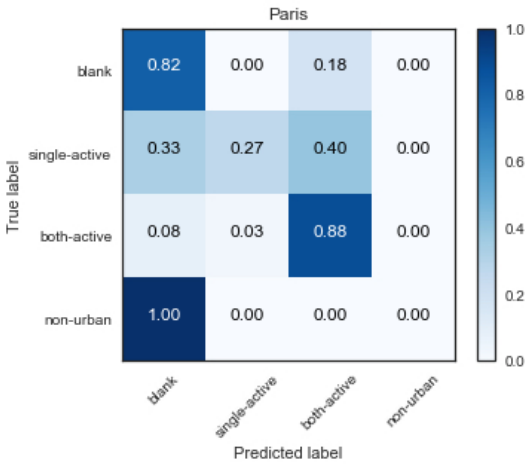

Figure 11: Confusion Matrix results for Paris show that the London-SFN model generalises well geographically. However, the result is particularly poor for the single-active class.

5.2 Experiment Two - estimating the effect of urban frontage on house price through the hedonic price approach

To test the validity or the usefulness of the urban frontage indicator, we construct a house price model and regress London house prices with frontage classes and neighbourhood frontage scores, holding various housing attributes constant. We find that the frontage class for a particular street segment containing a home is an insignificant factor. However, we observe that the neighbourhood active frontage score is a significant factor (Figure 12). These results indicate that the amount of active frontage immediately surrounding a particular house might not be a crucial factor in isolation, but that the frontage across an entire neighbourhood can be important. This

result is reasonable, as isolated cases of blank frontages here and there will likely not be too important, but a cluster of blank frontages will undoubtedly have more of an adverse effect on overall safety perception of an area. We also find that the neighbourhood active frontage effect is significant in Inner London and insignificant in Outer London. These results may indicate the effect of walkability: active frontages have greater value in areas where pedestrians tend to walk more, as the active frontages might be more salient.

	Dependent variable:			
	lnprice			
	Baseline	Frontage-Model	Inner London	Outer London
Structural	Included	Included	Included	Included
Location	Included	Included	Included	Included
Neighbourhood	Included	Included	Included	Included
Single-Active		-0.004	0.02	-0.018
		-0.01	-0.021	-0.012
Both-Active		0.007	0.042**	-0.011
		-0.01	-0.017	-0.011
Non-Urban		0.027	0.101**	0.009
		-0.018	-0.043	-0.019
Active_R800		0.208***	0.701***	0.049
		-0.035	-0.057	-0.044
Constant	11.037***	11.274***	7.922***	11.987***
	-0.135	-0.14	-0.316	-0.159
Observations	6,110	6,110	2,078	4,032
R^2	0.675	0.678	0.713	0.667
Adjusted R^2	0.675	0.677	0.711	0.666
Residual Std. Error	0.266 (df = 6101)	0.265 (df = 6097)	0.263 (df = 2065)	0.252 (df = 4019)
F Statistic	1,587.134*** (df = 8; 6101)	1,068.646*** (df = 12; 6097)	427.531*** (df = 12; 2065)	671.749*** (df = 12; 4019)
Note:	*p**p***p<0.01			

Figure 12: House price regression model results. Model 1 is the baseline model without any frontage attributes. Model 2 includes both the frontage class variable and the neighbourhood frontage score variable. Model 3 is a sub-model for Inner London and Model 4 is a sub-model for Outer London. The results show no significance on the frontage class variable but significance on the neighbourhood active frontage score variable. These results indicate the amount of active frontage may not be a crucial factor in isolation but is important at the neighbourhood level.

The spatial dependence tests (Moran's I) show that the OLS model has significant spatial autocorrelation (Table 6). In response, a spatial lag and a spatial error model are subsequently estimated. The spatial lag model, on the one hand, still shows significant spatial autocorrelation. The spatial error model, on the other hand, shows no significant spatial autocorrelation while achieving the lowest AIC when compare to both the OLS and the spatial lag model. Figure 13 shows the neighbourhood frontage score variable is significant after controlling for spatial effects (spatial error model). Its effects, however, decrease when compare to the OLS model.

Table 6: Moran's I statistics for the OLS model, the spatial error model and the spatial lag model.

	Moran's I	P-Value
OLS	0.502	2.20E-16
Spatial-Lag	0.15	2.20E-16
Spatial-Error	-0.003	0.771

	Dependent variable:	
	lnprice	
	spatial lag	*spatial error*
	5	6
Structural	Included	Included
Location	Included	Included
Neighbourhood	Included	Included
Single-Active	-0.001	0.008
	-0.008	-0.007
both-Active	0.007	0.011*
	-0.007	-0.007
Non-Urban	0.004	0.008
	-0.013	-0.012
active R800	-0.022	0.114**
	-0.026	-0.049
Constant	1.492***	12.182***
	-0.149	-0.407
Observations	6,110	6,110
Log Likelihood	1,102.86	1,639.40
$sigma^2$	0.039	0.032
Akaike Inf. Crit.	-2,175.71	-3,248.79
AIC (OLS)	1,131.40	1,131.40
Wald Test (df = 1)	8,254.251***	13,745.180***
LR Test (df = 1)	3,311.220***	4,384.302***
Note:	*p **p ***p<0.01	

Figure 13: Spatial regression model results. The results show that the neighbourhood score variable is significant for the spatial error model which controls for spatial effects.

5.3 Experiment Three - Scenicness and frontage quality comparison

We conduct an exploratory descriptive analysis of the frontage quality of a Google Street View image and its perceived scenicness in the Barnsbury study area. Figure 14 shows the geographical distribution of the frontage classification (on the left) along with the urban scenicness of the same area (on the right). The results suggest there is a visual correspondence between blank frontages (blue streets) and low scenicness (blue dots). The results from the ANOVA tests (Table 7) confirm these observations, suggesting that scenicness scores are significantly different between frontage classes. The resulting box plot (Figure 15) points to greater differences between the blank frontage class and the both-active frontage class, whereby the blank frontage class is found to be less scenic and the active frontage class more scenic. Further research with a larger sample size is required to confirm these results.

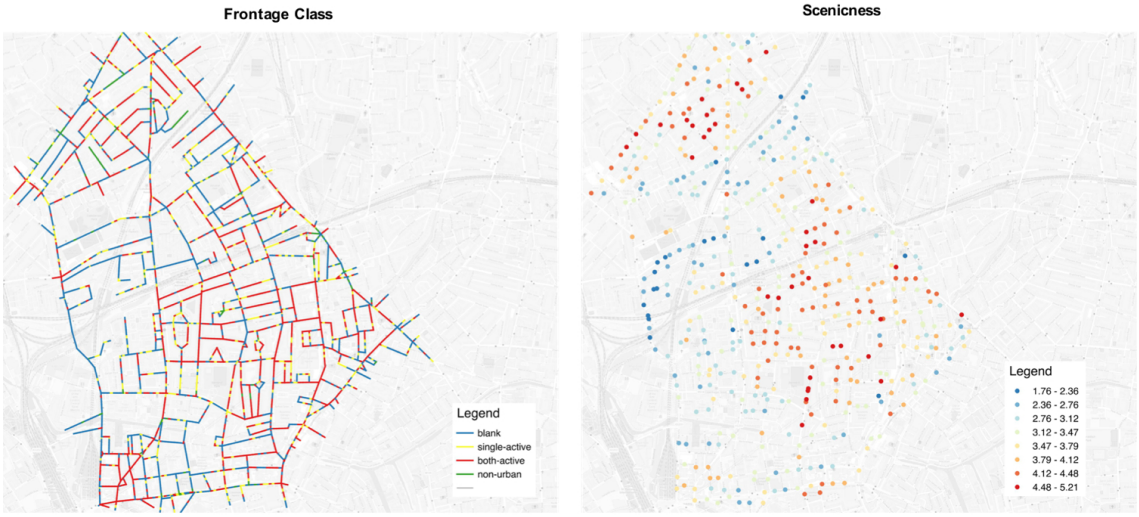

Figure 14: Comparing urban frontage class on the left and urban scenicness on the right. The results show clear associations between the absence of active frontages and low levels of scenicness. Contains Ordnance Survey data Crown copyright and database right 2017.

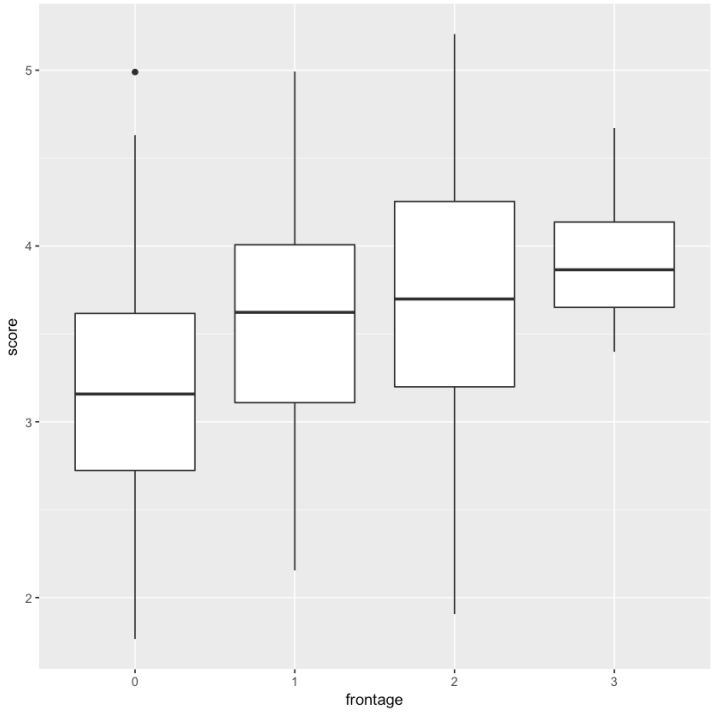

Figure 15: Box plot illustrating the extent to which scenicness varies for each frontage class in the Barnsbury study area. The difference is especially significant with the blank frontage class and the both-active class.

Table 7: ANOVA table illustrating scenicness is statistically significantly different between the classes.

	Df	Sum Sq	Mean Sq	F-value	Pr(>F)
Frontage	1	29.9	29.9	76.95	<2e-16 ***
Residuals	486	188.8	0.389		
Signif.	*** 0.001	** 0.01	* 0.05		

5.4 Experiment Four - within-street prediction tests

To study the extent to which the SFN model predictions generalise across a street segment, we compare the predictions for the Barnsbury area at different street resolutions. The cross-tab report (Figure 16) presents the results. First, the single image model performs similarly to the 160m, 200m and 240m model. The single image model's accuracy significantly falls at 120m and reaches a low at 20m. This outcome is

not surprising, as an entire street is not necessarily homogeneous. Obvious heterogeneity exists between frontage classes.

The accuracy is worst for the non-urban frontage class, with an accuracy of only 29% when comparing the single-image model with the 20m-image model. The both-active frontage class and the blank frontage class is most accurate, at 65% to 70% when comparing the single-image model to both the 20m-image model and the 40m-image model respectively. These results suggest that blank class and both-active class are predominately homogeneous along a street in London, while single-active class and non-urban class are predominately heterogeneous along a street segment. In other words, an entire street of blank frontages or both-active frontages are more likely to occur than a whole street of non-urban frontages or single-active frontages. Figure 17 illustrates the comparison of the low-resolution single-image model with the medium-resolution 120m-model and the high-resolution 20m model. This figure shows remarkable visual similarity between the three maps. Figure 17 also demonstrates that longer streets have greater heterogeneity in terms of frontage types. These results suggest that the use of low-resolution sampling is adequate on a city scale, but high-resolution sampling is needed at a neighbourhood scale, so that the distribution effects of the interconnections between frontages can be more appropriately addressed.

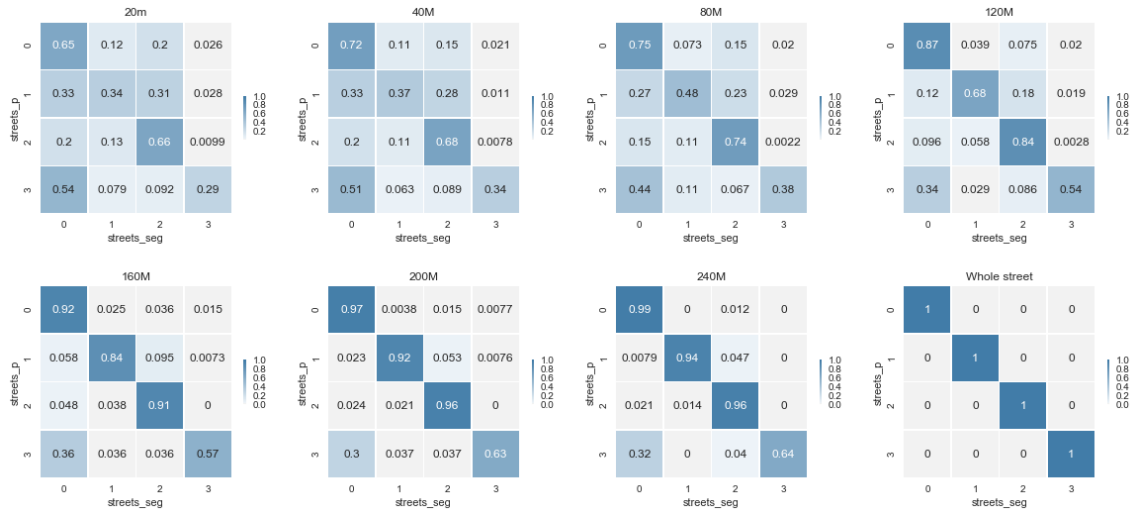

Figure 16: Cross-tabulations for predictions of different street resolutions. The results show that the single-image model is most similar to the 160m and 200m model, and that the accuracy reduces significantly from 120m. The accuracy is far better for the both-active and blank frontage class.

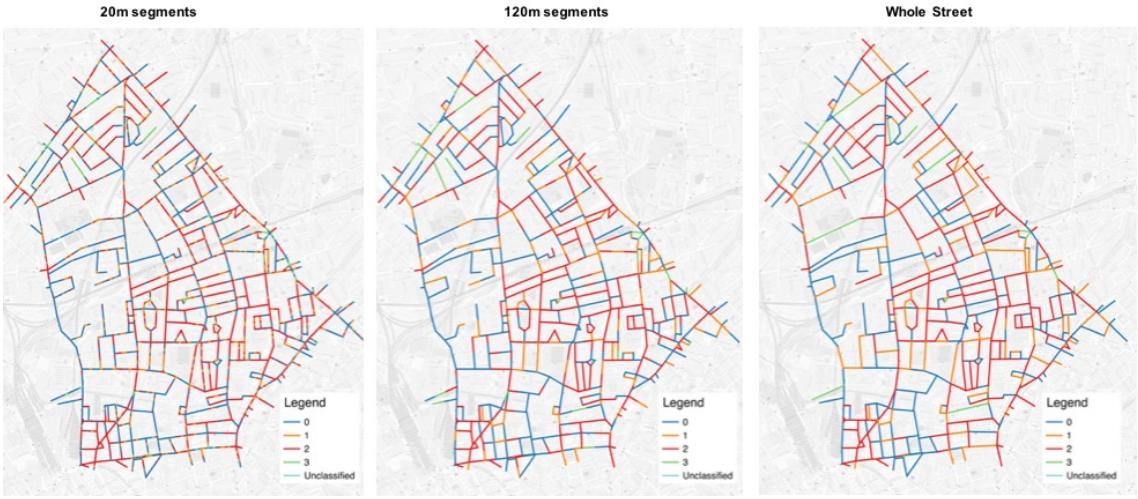

Figure 17: Prediction comparison for different resolutions of street segments. The figure shows visual correspondence across the three resolutions.

5.5 Experiment Five - model visualisation and interpretation

For interpretation purposes, we visualise the SFN model using the Grad-CAM method to better understand the visual pattern of a neural unit. Figure 18 presents the results in a visualisation matrix, whereby images from the four classes – as identified in the

methods section – are situated along the x-axis. The visualisation illustrates, from left to right: blank frontage, single-active frontage, both-active frontage and non-urban frontage. The y-axis shows the location of a particular frontage class's activation in the image: blank frontage activation, single-frontage activation, both-active frontage activation and non-urban frontage activation. The results suggest that each frontage class image is highlighted by its subsequent frontage class activation. This outcome is illustrated by the activation along the diagonal of the gradient visualisation matrix. This outcome suggests that the model accurately identifies the frontages of a Google Street View image rather than other correlated areas of the images. These results also clearly show that the single-active class is inherently a combination of blank class and both-active classes. These outcomes are also reflected in the poorer accuracy of the single-active class. We have included a few more successful and failed examples in the supplementary information. This visualisation technique is useful in identifying model activation errors.

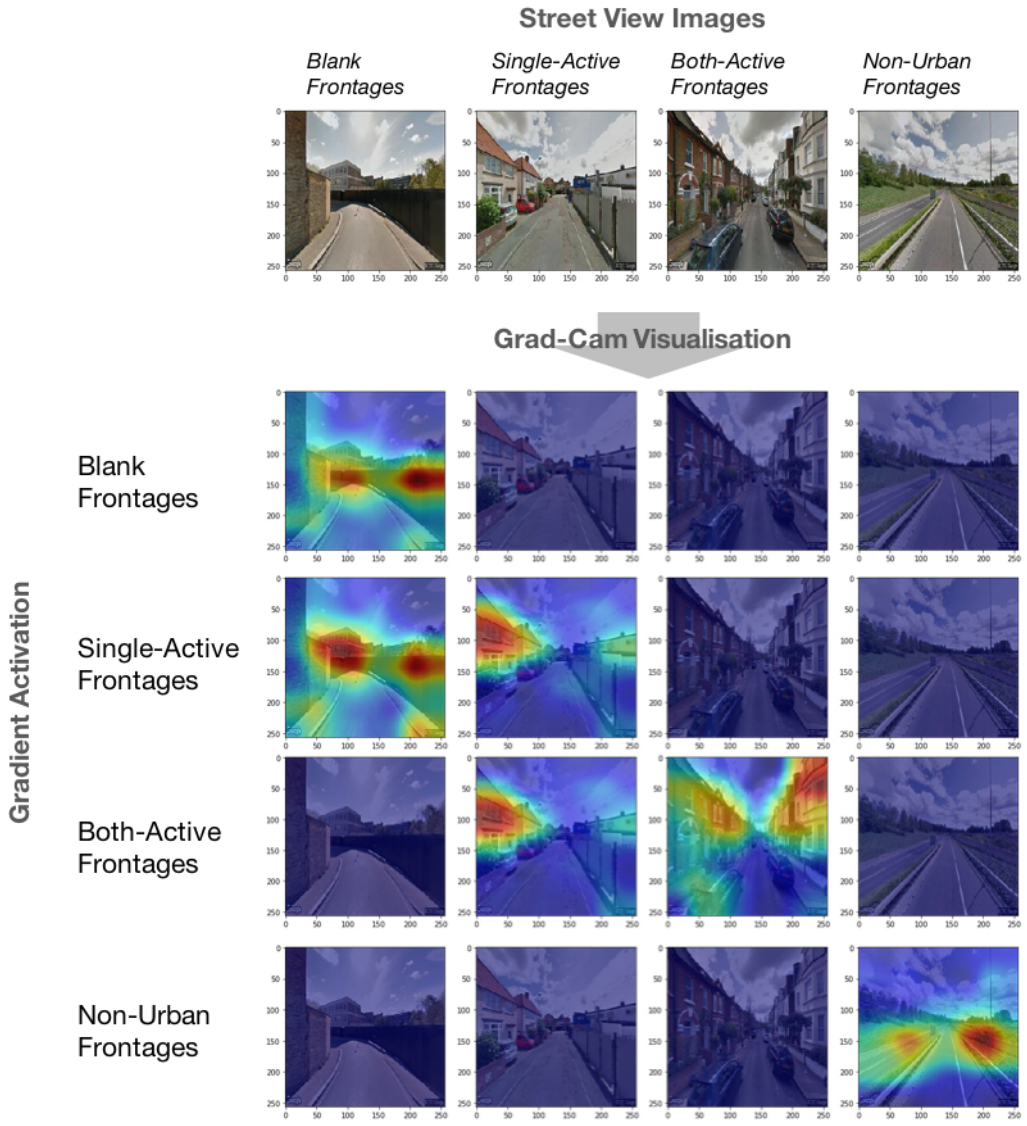

Figure 18: Grad-CAM visualisation of the four frontage classes with respect to the input image. The visualisation successfully highlights the frontage area of the street view image.

Finally, we use the best performing SFN model to predict both the frontage classes and the probability of an active frontage on every street segment in London. Figure 19 shows the most likely frontage class for each street image of London, where red denotes both-active, yellow represents single-active, blue indicates blank and green signifies non-urban. Examples of the four classes are highlighted in Figure 19, including Regent Street with both-active frontages, Hyde Park with non-urban frontages and streets in

Canary Wharf with blank frontages. The results indicate that Inner London has, as expected, a higher quantity of both-active frontages than Outer London. Single-active frontages are geographically more scattered compared to the other frontage classes, while blank frontages are found in Outer London. Non-urban frontages are found along parks and motorways. These results are consistent with the hypothesis that newer areas are generally less active and are also less pedestrian friendly.

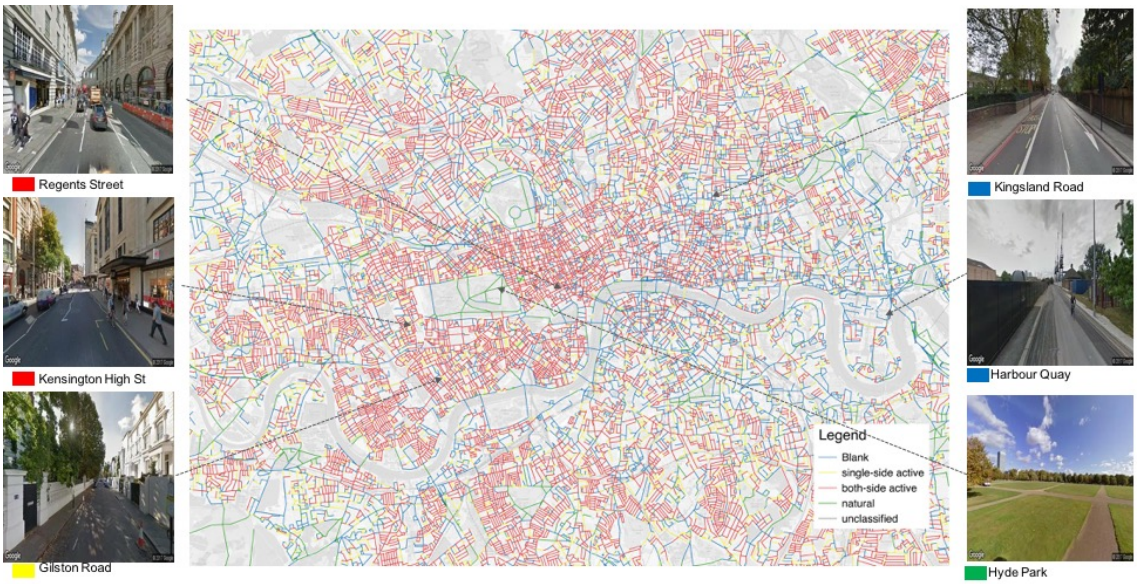

Figure 19: Prediction from the SFN classifier for every street in London where red signifies both-active, yellow signifies single-active, blue signifies blank and green signifies non-urban. Contains Ordnance Survey data. Crown copyright and database right 2017.

Figure 20 shows the probability of active frontages in London, in which red represents a greater probability of active frontages in Inner London and blue indicates a lower probability of active frontages in Outer London. Both of these maps will be put online as an interactive web-map for demonstration purposes.

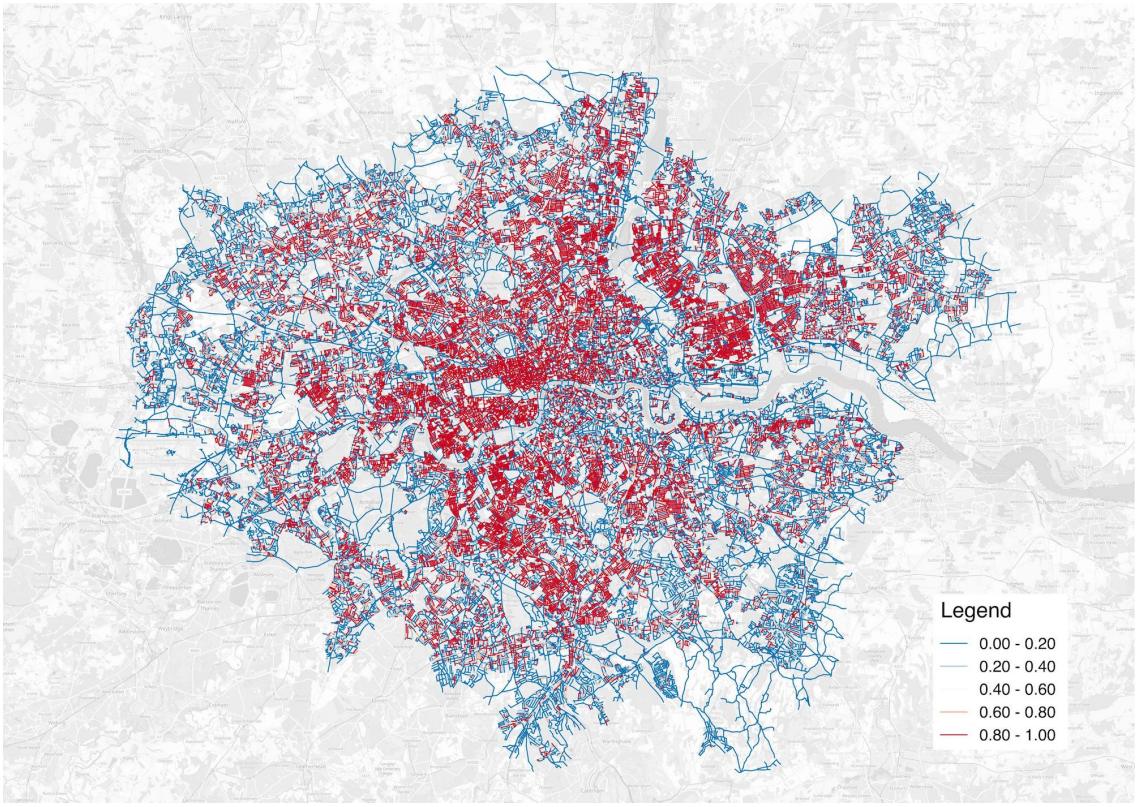

Figure 20: Active frontage probability for every street in London where red signifies higher probability for active frontages and blue signifies lower probability for active frontages. Contains Ordnance Survey data. Crown copyright and database right 2017.

6.0 Conclusion

To conclude, this study successfully demonstrates the classification of urban street frontage quality using deep learning methods. We have also shown that the quality of urban street frontage can be an important component to help understand aspects of urban planning, such as house prices and perceived scenicness.

This research extends from previous research (Law et al. 2017) by validating the SFN Model, and finds that the VGG16-ImageNet model performs better than more complex architecture with a similar specification. The results are consistent for both the model with only street view images as well as for the model with both street view and 3D model images. The results from the second experiment show that having more active

frontages at a neighbourhood level is correlated with higher house prices. One interpretation of this result is that neighbourhoods with a greater probability of active frontages are perceived as being safer (Jacobs 1961), leading to higher house prices. This effect is also greater for Inner London than for Outer London, suggesting that frontage quality is more valuable where pedestrians walk more. These results are consistent even after accounting for spatial dependency in the data. The results from the third experiment show that scenicness is significantly different between different frontage classes, and that active frontages have higher scenicness scores than blank frontages. One possible interpretation of this result is that blank frontages limit the variability of the scene, thereby reducing the visual interest of the environment. The results from the fourth experiment show that both-active frontages and blank frontages are largely homogenous (70%) along a single street segment and that single-active frontages and non-urban frontages are largely heterogeneous (30%) along a street segment. These results show that higher resolution sampling is necessary when studying frontages within a neighbourhood. Finally, the fifth experiment shows encouraging results when visualising the CNN model to determine which parts of a typical street image gets activated. The geographical visualisation also supports the general consensus that Inner London has a higher frequency of active frontages than Outer London.

7.0 Discussion

This research successfully demonstrates how the quality of urban street frontage can be an important factor for studies that explore different aspects of urban planning, such as the economic valuation of a city and the well-being of its citizens, however there are several limitations that remain. Firstly, there exists misclassification between single-active frontages and both blank frontages and both-active frontages. To overcome this

concern, initial research has been conducted to develop a street frontage object detector that uses the Faster R-CNN model (Ren et al. 2015) to detect and localise building frontages, as seen in Figure 21. This result shows a promising research direction in retrieving more information from the image dataset.

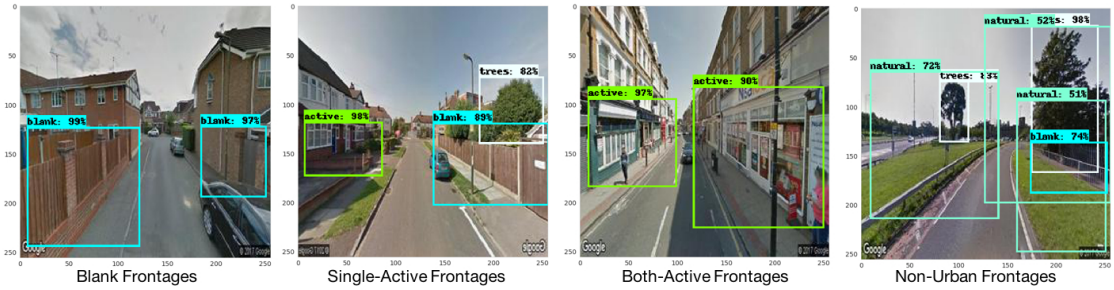

Figure 21: Predictions from a Faster R-CNN model highlighting the urban frontages of a front-facing Street View image.

Future research is also needed to study the extent to which generated 3D images can replace real images in the training data used by the CNN. The success of replicating the training data in a 3D virtual environment can greatly enhance the efficiency of collecting urban image data. Secondly, the focus on a single case study reduces the extent to which the results can be generalised. The small experiment with the Paris Street View images represents an early attempt in this direction, however further research is needed to develop a universal model for urban frontage assessment. Despite the above-mentioned limitations, the results of this study are highly encouraging, and confirm that deep learning methods can be successfully used for the purpose of categorising urban street frontage on a large scale. We also explore, through a preliminary statistical analysis, how our predictions of urban street frontage can be related to scenicness, as well as to factors important to urban studies such as house prices. These results show clear potential in the use of deep learning methods for geographical knowledge discovery and for scientific studies in urban design.

ACKNOWLEDGMENTS

We thank Ben Bird for comments that greatly improved the manuscript. This work was supported by The Alan Turing Institute under the UK Engineering and Physical Sciences Research Council (EPSRC) grant no. EP/N510129/1. (Alan Turing Institute: TU/A/000016 and TU/D/000019).

REFERENCES

Alexander, C., Ishikawa, S., Silverstein, M., Jacobson, M.,Fiksdahl-King, I., and Angel, S., 1977. A Pattern Language. New York: Oxford University Press.

Anselin, L. 1988. Spatial Econometrics: Methods and Models. Kluwer Academic, Dordrecht.

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

Cheshire, J., and Sheppard, S., 1995. On the Price of Land and the Value of Amenities. Economica , 62(246): 247-267.

Chollet, F., 2017. Xception: Deep Learning with Depthwise Separable Convolutions. arXiv preprint arXiv:1610.02357.

Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., and Li, F.F., 2009. ImageNet: A Large-Scale Hierarchical Image Database. Conference on Computer Vision and Pattern Recognition (CVPR).

Doersch, C., Singh, S., Wu, C., and Hui, W., 2012. What makes Paris look like Paris. ACM Transactions on Graphics.

ESRI, 2013. City Engine. http://www.esri.com/software/cityengine/. (accessed on 2013).

Gebru, T., Krause, J., Wang, Y., Chen, D., Deng, J., Aiden, E., and Li, F., 2017. Using Deep Learning and Google Street View to Estimate the Demographic Makeup of Neighbourhoods across the United States. PNAS.

Gehl, J. 1971. Life Between Buildings: Using Public Space. The Danish Architectural Press.

Gehl, J. 2010. Cities for People. Island Press.

Girshick, R., 2015. Fast R-CNN. IEEE International Conference on Computer Vision (ICCV).

Google., 2017. Google Maps. https://www.maps.google.com/. (accessed on 2017).

He, K., Zhang, X., Ren, S., and Sun, J., 2015. Deep Residual Learning for Image Recognition. arXiv preprint arXiv:1512.03385.

Heffernan, E., Heffernan, T., and Pan, W., 2014. The Relationship between the Quality of Active Frontages and Public Perceptions of Public Spaces. Urban Design International. 19(1): 92-102.

Historic England, 2017. Historic England Listed Parks and Gardens dataset. https://historicengland.org.uk/listing/the-list/data-downloads/. (accessed on 2017).

Jacobs, J., 1961. The Death and Life of Great American Cities. Random House Inc.

Kickert, C. 2016. Active Centers - Interactive Edges: the rise and fall of ground floor frontages. Urban Design International. 21(1): 55-77.

Krizhevsky, A., Sutskever, I. and Hinton, G.E., 2012. Image Net Classification with Deep Convolutional Neural Networks. Advances in neural information processing. (NIPS)

Land Registry, 2017. House Price Transaction Dataset. https://www.gov.uk/search-house-prices. (accessed on 2017).

Law, S., Shen, Y., and Seresinhe, C., 2017. An Application of Convolutional Neural Network in Street Image Classification: the case study of London. ACM GeoAI '17 Workshop on Artificial Intelligence and Deep Learning for Geographic Knowledge Discovery.

Law, S., Paige, B., and Russell, C., 2018. Take a Look Around: Using Street View and Satellite Images to Estimate House Prices. arXiv preprint arXiv: 1807.07155

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.

Li, L., Tompkin, J., Michalatos, P., and Pfister, H., 2017. Hierarchical Visual Feature Analysis for City Street View Datasets. IEEE VIS 2017 Workshop on Visual Analytics for Deep Learning.

Li, Y., Paluri, M., Rehg, J., and Dollar, P. 2016. Unsupervised learning of edges. Conference on Computer Vision and Pattern Recognition (CVPR).

Li, X.J. Santi, P., Courtney, T.K., Verma, S.K., and Ratti, C. 2018. Investigating the Association between Streetscapes and Human Walking Activities using Google Street View and Human Trajectory Data. Transactions in GIS. 22(4):1-16.

Liu, L., Silva, E., Wu, C. and Hui, W., 2017. A Machine Learning-based Method for the Large-scale Evaluation of the urban environment. Computers, Environment and Urban Systems. 65: 113-125.

Llewelyn Davies Yeang and Homes Communities Agency, 2013. Urban Design Compendium. English Partnerships and The Housing Corporation.

Mirowski, P., Grimes, M.K., Malinowski, M., Hermann, K.M., Anderson, K., Teplyashin, D., Simonyan, K., Kavukcuoglu, K., Zisserman, A., and Hadsell, R., 2018. Learning to Navigate in Cities Without a Map. arXiv preprint arXiv:1804.00168.

Moran, P., 1948. The Interpretation of Statistical Maps. Biometrika. 35: 255-60.

Naik, N., Philipoom, J., Raskar, R., and Hidalgo, C.A., 2014. StreetScore – Predicting the Perceived Safety of One Million Streetscapes. CVPR Workshop on Web-scale Vision and Social Media.

Nase, I., Berry, J. and Adair, A. 2013. Hedonic Modelling of High Street Retail Properties: A Quality Design Perspective, Journal of Property Investment & Finance, 31(2): 160-178.

Nationwide, 2012. House Price Dataset. Permission granted from LSE.

Office of the Deputy Prime Minister (ODPM), 2005. Safer Places: The Planning System and Crime Prevention. Home Office.

Office for National Statistics, 2017. Office for National Statistics datasets https://www.ons.gov.uk. (accessed on 2017).

Ordnance Survey. 2017. Ordnance Survey datasets. https://www.ordnancesurvey.co.uk/opendatadownload/products.html. (accessed on 2017).

Palmquist, R.B., 1984. Estimating the Demand for the Characteristics of Housing. Review of Economics and Statistics. 66(3), 394-404.

Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster R-CNN: Towards Real-time object detection with region proposal networks. arXiv preprint. arxiv:1506.01497.

Ridker, R.G. and Henning, J.A., 1967. The Determinants of Residential Property Values with Special Reference to Air Pollution. Review of Economics and Statistics. 49(2): 246-257

Rosen, S., 1974. Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. Journal of Political Economy. 82(1): 34-55.

Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D., 2016. Grad-CAM: Why did you say that? Visual Explanations from Deep Networks via Gradient-based Localization. arXiv preprint arXiv:1610.02391v3.

Seresinhe, C., Preis, T., and Moat, S., 2017. Using Deep Learning to Quantify the Beauty of Outdoor Places. Royal Society Open Science. 4(7): 170170.

Simonyan, K., and Zisserman, A., 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv preprint arXiv:1409.1556.

Streetscore, 2014. Streetscore. http://streetscore.media.mit.edu. (accessed on 2014)

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A., 2015. Going Deeper with Convolutions. Conference on Computer Vision and Pattern Recognition (CVPR).

Valuation Office Agency, 2015. Valuation Office Agency dataset. (accessed on 2015)

Zhang, Q.S. and Zhu, S.C., 2018. Visual Interpretability for Deep Learning: a survey. arXiv preprint arXiv:1802.00614v2.

Zhou, B., Lapedriza, J., Xiao, A., Torralba, A., and Oliva, A., 2014a. Learning Deep Features for Scene Recognition using Places Database. Advances in neural information processing (NIPS).

Zhou, B., Liu, L., Oliva, A., and Torralba, A., 2014b. Recognizing City Identity via Attribute Analysis of Geo-tagged Images. European Conference on Computer Vision (ECCV).