remote sensing

Article

A Hierarchical Urban Forest Index Using Street-Level Imagery and Deep Learning

Philip Stubbings [1], Joe Peskett [1], Francisco Rowe [2,*] and Dani Arribas-Bel [2]

[1] Data Science Campus, Office for National Statistics, Newport NP10 8XG, UK; philip.stubbings@ons.gov.uk (P.S.); joseph.peskett@ons.gov.uk (J.P.)

[2] Geographic Data Science Lab, Department of Geography & Planning, University of Liverpool, Liverpool L69 7ZT, UK; D.Arribas-Bel@liverpool.ac.uk

* Correspondence: F.Rowe-Gonzalez@liverpool.ac.uk; Tel.: +44-1517-94-2845

Received: 29 April 2019; Accepted: 5 June 2019; Published: 12 June 2019

Abstract: We develop a method based on computer vision and a hierarchical multilevel model to derive an Urban Street Tree Vegetation Index which aims to quantify the amount of vegetation visible from the point of view of a pedestrian. Our approach unfolds in two steps. First, areas of vegetation are detected within street-level imagery using a state-of-the-art deep neural network model. Second, information is combined from several images to derive an aggregated indicator at the area level using a hierarchical multilevel model. The comparative performance of our proposed approach is demonstrated against a widely used image segmentation technique based on a pre-labelled dataset. The approach is deployed to a real-world scenario for the city of Cardiff, Wales, using Google Street View imagery. Based on more than 200,000 street-level images, an urban tree street-level indicator is derived to measure the spatial distribution of tree cover, accounting for the presence of obstructing objects present in images at the Lower Layer Super Output Area (LSOA) level, corresponding to the most commonly used administrative areas for policy-making in the United Kingdom. The results show a high degree of correspondence between our tree street-level score and aerial tree cover estimates. They also evidence more accurate estimates at a pedestrian perspective from our tree score by more appropriately capturing tree cover in areas with large burial, woodland, formal open and informal open spaces where shallow trees are abundant, in high density residential areas with backyard trees, and along street networks with high density of high trees. The proposed approach is scalable and automatable. It can be applied to cities across the world and provides robust estimates of urban trees to advance our understanding of the link between mental health, well-being, green space and air pollution.

Keywords: urban forestry; green space; street-level imagery; deep learning; image segmentation

1. Introduction

Urban trees provide key social, environmental and economic benefits. As global urban population expands, and congestion and pollution levels rise, access to urban green space becomes increasingly important [1]. Urban green space, including parks, green roofs, community gardens and street trees, provide critical ecosystem services. These spaces promote physical activity, boost psychological well-being, filter air, remove pollution, attenuate noise, cool temperatures, mitigate climate change and improve urban public health [2,3].

Urban trees play a major role in removing environmental air pollution and improving human health. In the United Kingdom, they are estimated to annually remove 1.4 million tonnes of air pollutants; that is, equivalent to £1 billion in health damage costs [4]. By providing shade and cooling, urban trees also moderate temperatures, helping to reduce the risk of heat-related illnesses by an estimated 4–6 degrees [5,6]. Urban green space also has key health benefits. Trees are aesthetically pleasing features, inviting physical activity, which reduces obesity [7] and mental stress levels [8]. Visual accessibility to trees enhances contemplation and offers a sense of peace and tranquillity [9,10]. At the same time, lack of urban green space is associated with increased risk of respiratory diseases, obesity and mortality [11,12]. Lack of access to urban green spaces has also been associated with attention deficit disorder during childhood [13]. Ensuring widespread access to urban green space is thus key to rendering these environmental and health benefits.

Estimating the size, location and distribution of urban forest, however, remains a challenge. Traditional approaches to producing tree and vegetation inventories have relied on manual data collection processes via trained surveyors and community-based crowdsourcing [14,15]. These approaches can entail a considerable gap between data collection and data release times, are expensive, infrequent, rely on small samples, are prone to sampling errors and provide a very coarse measure of green space [16,17]. These challenges have thus undermined the scalability, automatisation and consistency of these approaches.

To address these issues, remote sensed satellite and aerial imagery-based approaches have also been developed to estimate green spaces [18,19]. While automatable, satellite-based approaches present challenges in urban contexts as vegetation cover measures are typically based on moderate-resolution satellite imagery e.g., 30 m per pixel, which represents a very coarse spatial scale for cities. Efforts employing high-resolution light detection and ranging (LiDAR) data to map tree canopy have proven well suited for urban contexts [19], but high data acquisition costs and specialised proprietary software hamper their implementation [15]. Additionally, a key limitation is that while satellite and aerial imagery may provide a relatively accurate quantification of urban greenery, they do not offer a good representation of street-level vegetation [20]. The landscape of urban vegetation which humans perceive and experience at the street-level may differ significantly from that remotely sensed from above.

Coupled with publicly available street-level imagery, advances in computer vision and computing capacity are enabling reliable detection and measurement of urban environments to capture how humans experience these environments. Street-level imagery data, such as Google Street View (GSV), provide extensive geographical coverage, and standardized, geocoded and high resolution images of the urban environment. Computer vision algorithms have been developed to process street-level imagery, measuring perceived urban safety [21], urban change [22], wealth [23], infrastructure [24], demographics [25] and building type classification [26]. In the context of urban trees, a small but growing number of studies have sought to develop computer vision approaches to address three key areas: (1) quantify the shade provision of urban canopy [27–29]; (2) catalog the location of urban trees [30,31]; and, (3) estimate the percent of urban street-level tree cover [15,20,32,33]. Our paper seeks to contribute to the third line of inquiry.

Estimating urban street-level tree cover generally involves two key steps: (1) identification and classification of vegetation pixels in individual street-level images; and, (2) amalgamation of these pixels at a street segment or area level to generate an index of visible vegetation cover at a relevant geographical level. To identify vegetation pixels, Li et al. [32] provide a first systematic approach to measure the percent of total pixels in a GSV image employing an unsupervised mean shift segmentation. While this method has the advantage of not needing data training, the resulting segmentation outcomes are influenced by shadows and illumination, and green non-vegetation pixels are misclassified as urban tree canopy [33]. Progressing this work, Seiferling et al. [15] applied a supervised geometric segmentation algorithm to identify vegetation pixels and measure vegetation cover in GSV images. This algorithm classifies image pixels into geometric classes of an image i.e., ground plane, sky plane and vertical surfaces, but it does not take into account the overall semantic context of images, misclassifying image pixels [34], and requiring pre-computation of image features [33]. An alternative is the Pyramid Scene Parsing Network (PSPNet) algorithm. PSPNet has been consistently demonstrated to yield improved pixel classification in urban environments [33,34], and has been trained and designed to specifically work in urban settings [34].

Additionally, street-level imagery based studies on the quantification of urban trees have used Yang's et al. [20] index of visible green cover to provide a summary measure of percentage of urban street-level tree cover at area level. However, the index is the average of the number of identified vegetation pixels in images at a same street location, and as a result, it suffers from three key limitations. First, the index does not account for systematic variation in the percent of urban street-level tree cover across images due to "obstructing" urban features, such as vehicles, street light poles and pedestrians. Second, it does not correct for the variation in the size of the sample of images collected at particular locations. Third, it does not measure the uncertainty resulting from image-to-image variation in the percent of tree cover and the variation in sample size by the aggregation of areas.

To address these shortcomings, the present paper aims to develop a scalable, automatable and consistent approach based on recent advances in Deep Convolutional Neural Networks (DCNN) and multilevel regression modelling to estimate a hierarchical area-level score of urban street-level trees. Semantic image segmentation is applied using the PSPNet [34] to classify image pixels and estimate the percentage of vegetation cover in street-level images. Based on these individual image estimates, a hierarchical two-level modelling approach is used to derive an area score which corrects for image-to-image variability due to the presence of "obstructing" urban features; accounts for sample variability; and incorporates measures of uncertainty. Before deploying our approach to GSV images of the city of Cardiff, United Kingdom, the efficiency of the DCNN segmentation against two alternative approaches is tested following widely used methods: a pixel-wise vegetation classification algorithm used in controlled crop environments; and a generalisation of that technique that builds a more flexible image mask. Although our application focuses on Cardiff, the proposed approach is scalable and can be applied in different urban contexts, complementing metrics of urban tree cover based on above-the-ground imagery (e.g., satellite and aerial).

The proposed approach makes two key contributions on the identification and quantification of urban forest. First, it contributes evidence supporting the superiority of PSPNet in identifying and classifying image vegetation pixels in a UK-based urban setting. Related work has largely drawn on US cities, making it difficult to establish the accuracy of PSPNet in non-US urban environments. European cities tend to be more compact, denser and busier than American cities. Second, this paper makes a major methodological contribution by applying a hierarchical two-level modelling approach to derive a sophisticated street-level tree score. Unlike indices used in previous studies, the proposed score effectively accounts for image-to-image variation in the percent of urban tree cover due to obstructing objects; adjusts for sample variation; and, provides a measure of uncertainty.

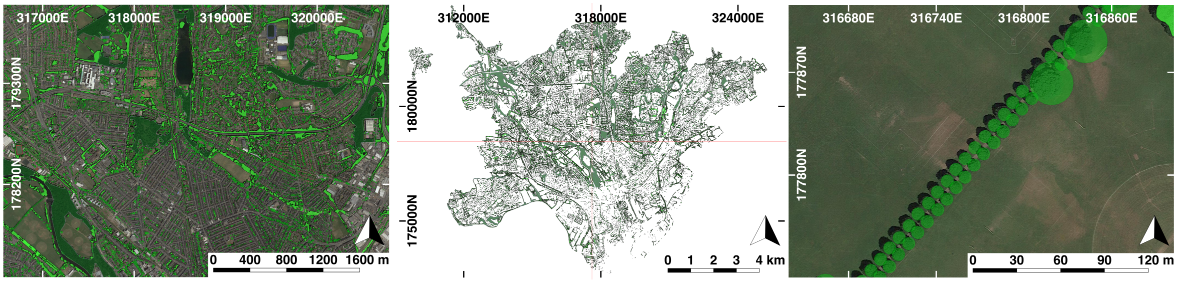
The remainder of this paper is structured as follows. The next section introduces the study area, Cardiff, and data used in the study. Section 3 explains our proposed deep learning image segmentation algorithm and the two alternative approaches used for comparative assessment. Section 4 presents the results of the comparative assessment, introduces our proposal of a hierarchical score, and applies it to the case of Cardiff. Section 5 concludes by discussing the implications of our results and avenues for further research using our proposed urban tree score.

2. Study Area and Data

2.1. Cardiff

The urban forest in Cardiff, United Kingdom, is the study context (Figure 1). Cardiff is the capital of Wales, the most populous city in the country, and the 11th largest city in the United Kingdom, with a population of over 330 thousand according to the latest 2011 census. Cardiff is recognised for its extensive green urban fabric, housing over 330 parks and gardens and one of the UK's largest parks, Bute, expanding 56 hectares and over 2000 trees [35]. In 2013, urban trees were estimated to cover 16% of the total area of the city [14]. In addition, extensive aerial surveys have been conducted by Natural Resources Wales

(NRW) [14] to measure and understand tree canopy in the city. Data drawn from these surveys provide a useful framework to assess our approach. These data are described in Section 2.3.



(**a**) Center of Cardiff (**b**) Cardiff's woodland and trees (**c**) Crown diameter illustration

Figure 1. Natural Resources Wales (NRW) Dataset—Distribution and Coverage. Note: Satellite imagery copyright Google.

2.2. Imagery Data

We rely on two primary datasets to test and deploy our image segmentation algorithms. First, the Mapillary dataset [36] is used to benchmark the performance of our proposal; second, GSV is used to deploy our approach to the case of Cardiff. While GSV has been used here, it should be noted that this approach is not dependent on GSV data and may be applied to arbitrary images captured at street level from multiple sources.

Mapillary data. The Mapillary dataset consists of 25,000 street-level images captured using a variety of cameras from around the world. Pixels in each image have been labelled as belonging to one of 100 possible categories describing various components present in urban scenes. This is considered a high quality labelled dataset, which has undergone a two-stage quality assurance process. To ensure consistency with the standardly shaped images used for Cardiff, only Mapillary images with a standard "4 to 3" aspect ratio are used, resulting in a dataset of approximately 10,000 street-level images.

GSV data. The empirical application relies on detailed imagery for every segment of the street network. As such, OpenStreetMap [37] is used to derive street segments, and the GSV API as a source of street level imagery. To conduct this study, 220,068 640 × 640 pixel street-view images are sampled from the left and right-hand side of the road at 10-metre intervals along the entire Cardiff road network. The set of GSV images used were sampled at mixed times of the year, excluding winter, between 2012 and 2017. The most recent images tend to belong to main/arterial roads, whereas the oldest images belong to low traffic side-streets.

2.3. Tree Data

To empirically assess the performance of our proposed approach, we ideally require ground truth data providing full geographical coverage information on the location and density of existing trees. While such ground truth data does not exist, a large-scale survey of urban trees can arguably provide a good approximation. For our purpose, we use the world's first nationwide urban tree mapping survey [14], conducted in Wales by NRW. Data collection took place in three phases during 2006, 2011, and 2013, using aerial photography. Individual trees were identified manually using a "desk-based analysis" and originally represented as points. The study reports tree crown cover for three sizes (in diameter), categorised as small (3–6 m), medium (6–12 m) and large (12 m or more). This approach is applied in a variety of contexts including, but not limited to: green open space, transport corridors, commercial areas and woodland. Data from the most recent phase of the survey conducted in 2013 is used and each point is turned into a circular polygon. The tree survey is complemented with data on green space areas from the National Forest Inventory (NFI), provided by NRW and the Welsh Government Lle geo-portal.

The final result is a set of polygons representing individual and small groups of amenitytrees and larger areas of urban woodland. To our knowledge, this is the only detailed geospatial inventory of urban trees in Cardiff. In our analysis of the proposed urban forest score, these datasets are complemented with additional information on land cover and land uses in Cardiff that are also obtained from the original report by NRW [38].

Despite all of its advantages and the extent of geographical coverage as illustrated in Figure 1, the Cardiff trees dataset contains some inaccuracies. These inaccuracies likely derive from the tree labelling process and aerial image quality, which the report indicates was not sufficient to detect tree canopy of less than three metres in circumference, a common case for trees growing along transport corridors. Nonetheless, the data represent the current state-of-the-art and we believe meet the quality requirements for the present study.

3. Methods

This section presents a methodological framework to enable reliable, automated identification of urban trees from street-level images, and quantification of street-level tree cover at an aggregate geographical scale (Figure 2). The framework involves two stages: (1) identification and classification of vegetation pixels; and, (2) aggregation of these pixels to generate a tree score. As indicated above, GSV imagery is used for our application. This section presents the methods used in these two stages as follows. First, methods for vegetation identification in street-level images are introduced (Section 3.1). One of the most common approaches in both plant phenotyping and street-level studies of green space, the so-called "green pixel $L*a*b*$ threshold", is presented (Section 3.1.1), followed by a novel generalisation of the $L*a*b*$ method (Section 3.1.2), before our preferred segmentation algorithm based on the PSPNet neural network (Section 3.1.3). The aim of including the two first methods is to use them as a benchmark, to demonstrate the PSPNet's ability to successfully identify green space in images, and effectiveness in terms to appropriately predict pixel classes. Then, the hierarchical modelling approach to aggregate vegetation pixels and estimate the density of street-level tree cover at a relevant geographical level is discussed (Section 3.2).

3.1. Identification and Classification of Vegetation in Street-Level Images

In this section, three alternative approaches of vegetation identification and classification in street-level images are presented. Their goal is conceptually simple. Given a street-level image, these algorithms determine the amount of vegetation present in the scene by identifying and classifying each image pixel as belonging or not to the vegetation class. A percentage summary of visible density can then be defined as the proportion of pixels labelled as vegetation in a single image or set of images.

3.1.1. Image Segmentation Using a Green Pixel $L*a*b*$ Threshold

Our first approach to pixel-wise vegetation identification and classification is based on the observation that vegetation tends to be green, at least in spring and summer seasons. This idea is widely accepted and used in the plant phenotyping domain [39], in which it is possible to segment plant leaves according to shade of green, and has also been used to identify green space in street-level imagery [32].

The intuition behind the green pixel $L*a*b*$ threshold technique is as follows. An image is initially expressed as a three-dimensional tensor of dimensions $W \times H \times 3$; where W corresponds to the width of the image; H to its height; and, 3 relates to the RGB triplets that describe a color. This input is projected to the $L*a*b*$ space, which in contrast to RGB, consists of three dimensions defined as lightness (L, or luminosity), alpha ($a*$) and beta ($b*$). The $L*$ parameter represents the luminosity ranging from 0 (black), through 50 (grey) and to 100 (white). In $L*a*b*$, the colour space is represented by the $a*$ and $b*$ parameters. An $a*$ of 0 corresponds to grey, whereas an increasingly negative $a*$ value corresponds to a higher saturation of green, and increasingly positive $a*$ value corresponds to a higher saturation of red. Similarly, the $b*$ channel encodes the saturation level for blue (negative $b*$) and yellow (positive $b*$), respectively.

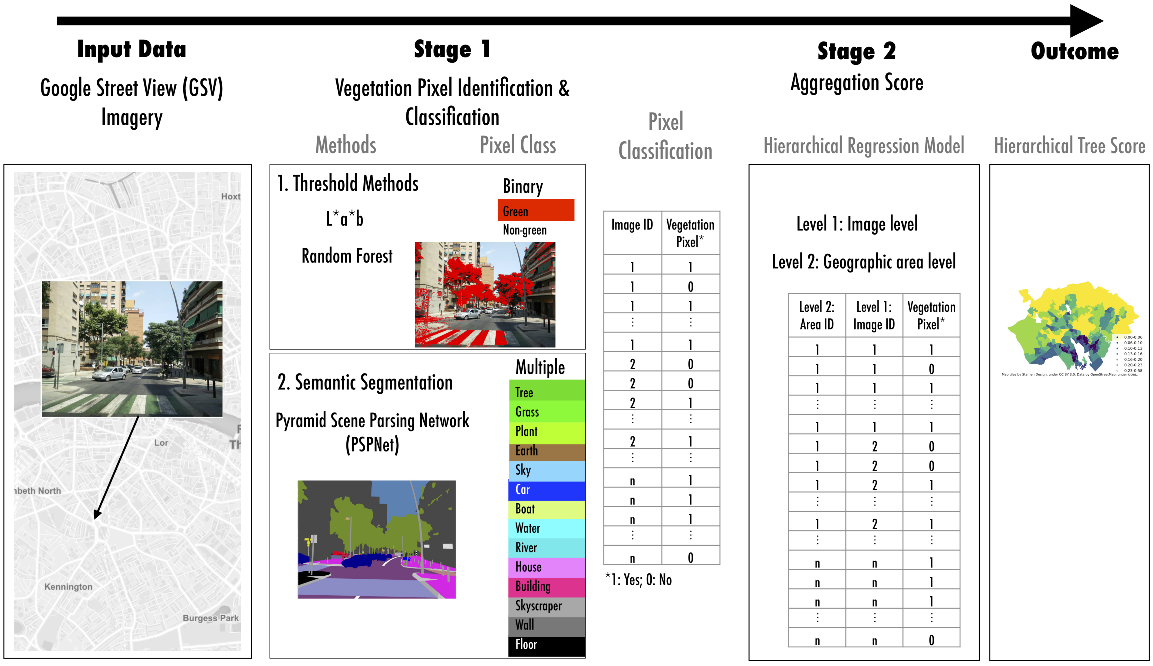


Figure 2. Methodological framework. Input data are obtained from Google Street View (GSV). These data are used in Stage 1 to identify and classify GSV image pixels. The three image segmentation methods are used and presented in Section 3.1: first, two threshold binary classification methods: The $L * a * b*$ based method (Section 3.1.1) and a novel Random Forest improvement (Section 3.1.2); and then, a semantic image segmentation method (i.e., PSPNet) (Section 3.1.3) is proposed as an improvement. A dataset of vegetation pixel classification is derived for each image and used in Stage 2 in a hierarchical modelling framework (Section 3.2) to derive a novel tree score, to estimate street-level tree cover at an aggregate geographical level.

The advantage of the $L * a * b*$ approach, as compared to direct color comparison, is that color classification (captured in $a*$ and $b*$) is invariant to changes in luminosity and results in a linearly separable space. This can be seen in Figure 3, where different cross-sections from the L*a*b* colour space are displayed for three levels of increasing lightness. Any region in the top-left quadrant ($a* < 0$, $b* < 0$) is considered to be green. The shade of green will remain relatively static with respect to the lighting level, which makes it possible to identify "greenness" in a way that is robust to varying lighting conditions, as is the case in real-world images.

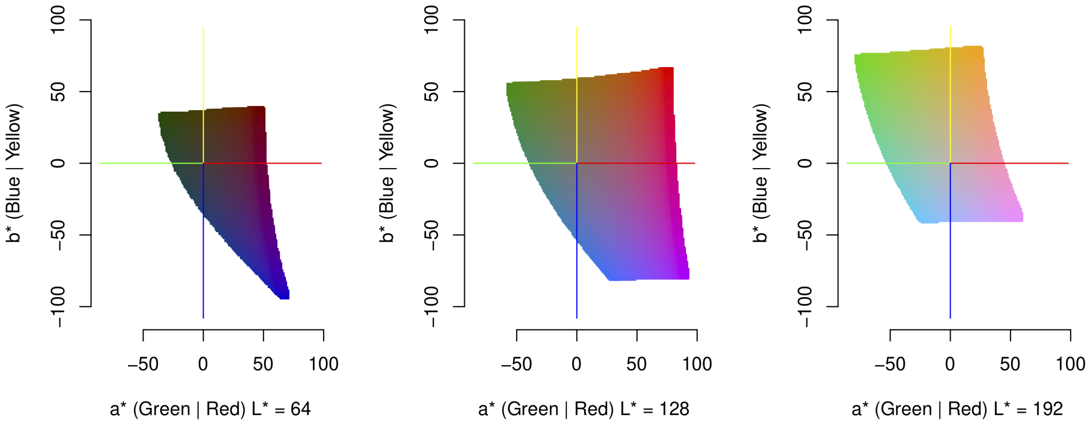


Figure 3. RGB to $L * a * b*$ colour space luminosity intersections—Varying degrees of luminosity (different $L*$ value) do not affect the identification of green-space (upper-left quadrant).

Within the colour spectrum represented in Figure 3, each quadrant is defined as a range of $a*$ $(A_1 \leq a \leq A_2)$ and $b*$ $(B_1 \leq b \leq B_2)$ values. These ranges can then be used to segment an image by labelling an individual pixel as green (vegetation) if its corresponding $a*$ value is within the threshold range. In the plant phenotyping domain [39], researchers have reported varying threshold parameters, which can be used for leaf segmentation in lighting-controlled images of plants. For example, Scharr et al. [39] find a range of $(-25 \leq a \leq -15)$ to be effective for extracting vegetation foreground in images of tobacco plants. A similar methodology is followed here to filter green pixels in street-level imagery.

To identify an appropriate $L * a * b*$ threshold, Bayesian parameter optimisation [40] is used, relying on the Matthews Correlation Coefficient ([41], MCC) as an objective function to find an optimal set of (A_1, A_2, B_1, B_2) parameters which minimise the pixel-by-pixel classification error. Each set of parameters enumerated by the optimisation method are evaluated using the mean MCC validation score after two-fold cross validation over the Mapillary training data. Figure 4 displays the range of $a*$ and $b*$ values in the Mapillary dataset along with the optimal parameters obtained from Bayesian parameter optimisation. These optimal parameters correspond to $-31 \leq a \leq -6$ and $5 \leq b \leq 57$ respectively.

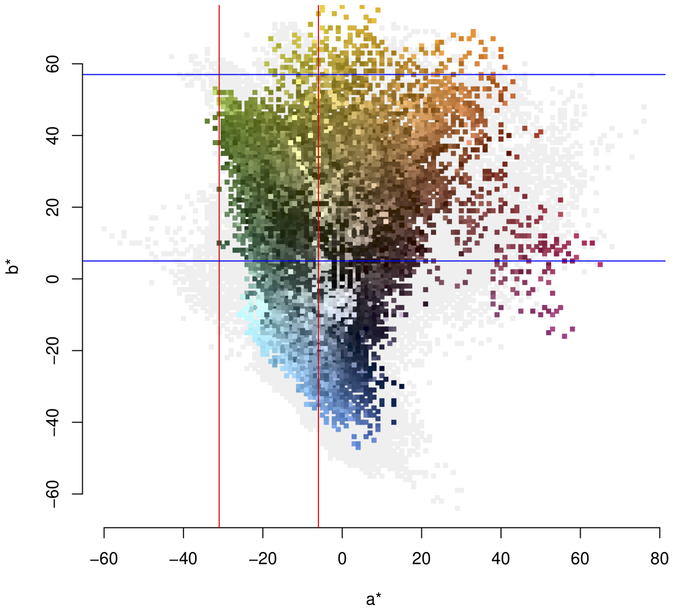


Figure 4. Vegetation (coloured region) and non-vegetation pixels (grey) in the Mapillary dataset. Each point represents an individual pixel from a randomly sampled set of images from the dataset. Pixels are coded according to the colour represented by the $a * b*$ pair. The sample includes values with varying patterns of $L*$ (lightness), and as such exhibits wide variability. Only pixels belonging to the vegetation class are retained, with the remaining classes colour coded as grey on the background to illustrate the overall feature space. The vertical lines extending from the $a*$ axis correspond to the optimal A_1 (left), A_2 (right) parameters, while the horizontal lines extending from $b*$ represent the optimal B_1 (bottom), B_2 (top) parameters.

3.1.2. Image Segmentation Using a Random Forest

Although widely used, the $L * a * b*$ method is rather limited. A more flexible generalisation is thus proposed. We have seen that a pixel is classified as vegetation if its $(a*, b*)$ values are contained within a rectangular threshold range. However, the optimal threshold is likely to conform to a different shape. So greater flexibility to define this range is needed as evidenced in Figure 4. It shows a non-rectangular range for the green colour space in the $a*$ and $b*$ scale for the Mapillary dataset.

Given as input two $(a*, b*)$ features, a Random Forest model [42] is trained on the Mapillary data to classify pixels into vegetation and non-vegetation classes. This approach is used because of its simplicity and minimal tuning of parameters [43]. As for the $L*a*b*$ method, we use Bayesian parameter optimisation and the MCC objective function to select the optimal parameters which correspond to the *number of estimators* and the *estimator maximum depth*.

The optimal random forest model has 11 estimators restricted to a depth of 14 leaves. As with the $L*a*b*$ based method, the model was trained to identify vegetation and non-vegetation class pixels. This process involves enumerating all possible $(a*, b*)$ combinations to produce a matrix containing the model's confidence of a pixel belonging to the vegetation class. Figure 5 (left panel) visualises this matrix, displaying the pattern of class confidence, with lighter (darker) colours indicating higher (lower) model confidence. The right panel graph displays a bitmap decision mask representing image pixels above the optimal model confidence threshold of 0.32 which is obtained by maximising the MCC, recall and R^2 scores based on the Mapillary labelled images (ground-truth data). The model is thus generalised to an elliptic region of the colour-space as opposed to rectangular as conceptualised in the Green pixel $L*a*b*$ threshold approach.

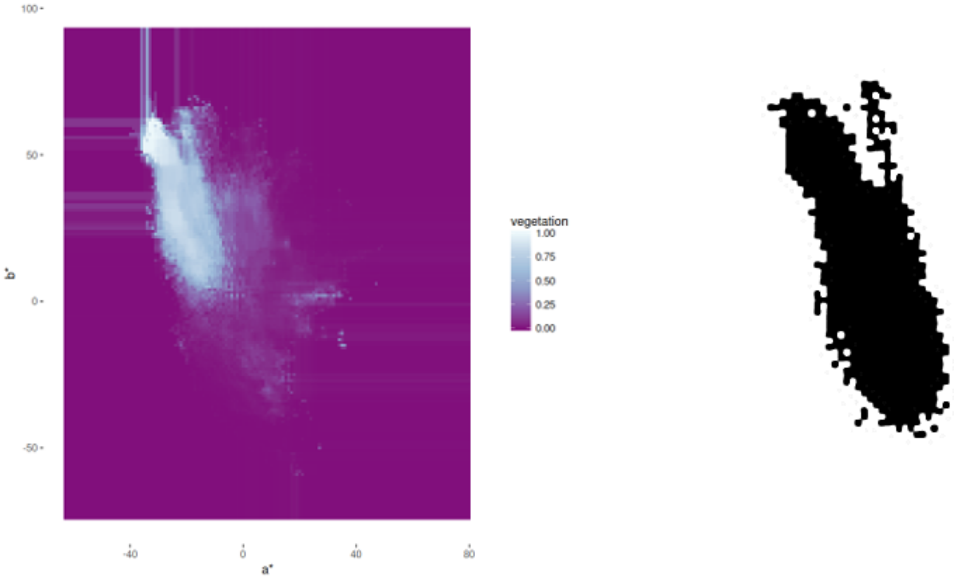


Figure 5. Random Forest segmentation. Using a $a*b*$ space (**left**), a random forest classifier is trained to predict green space (area above 0.32), generating a bitmap mask (**right**). A pixel is then classified as vegetation if its $a*b*$ colour coordinates are contained within the mask.

3.1.3. Scene Parsing via Pyramid Scene Parsing Network (PSPNet)

The objective of the $L*a*b*$ threshold and the random forest extension described in the previous section is to classify image pixels belonging to a vegetation class. As such, the image segmentation approach can be considered a pixel-wise binary classification problem. The vegetation detection task can be further generalised as a scene parsing problem in which given an image, the objective is to assign a label to each and every pixel in a coordinated way. Having assigned labels to all pixels in a scene, we can then extract only those pixels belonging to a vegetation class.

Scene parsing is an extension of image segmentation in which the objective is to holistically describe the entire scene. This contrasts with traditional image segmentation, where only pixels belonging to a specific class are singled out. Both scene parsing and image segmentation can be viewed as extensions of object detection. Its aim is to identify the location of specific objects in an image, and classifies image pixels into a class from a range of discrete categories describing the image.

Scene parsing has gained increased recognition and relevance in recent years due to its important role in applications such as autonomous driving.

Abstracting street tree identification to a scene parsing problem introduces a new set of challenges but also offers potential advantages over an image segmentation approach. By labelling pixels as belonging to one-of-many different classes, a scene may be described as a rich hierarchy in which objects may consist of component parts. For example, windows can be linked to buildings, or leafs to trees. Describing the context of the entire scene may therefore improve the classification accuracy of individual classes.

The current state-of-the-art in scene parsing resides in deep learning [44], particularly in DCNN [45,46]. Recent research has resulted in a number of sophisticated architectures, including the Fully Convolutional Network (FCN) [47], Segmentation Network (SegNet) [48], Deep Labelling for Semantic Image Segmentation (DeepLabV3) [49], Pyramid Scene Parsing Network (PSPNet) [34] and a more recent iteration of DeepLabV3, DeepLabV3+ [50]. These models build on earlier work in the related field of image object detection and classification, where advances such as "AlexNet" [46], Very Deep Convolutional Networks (VGG16) [51], Inception Network (GoogleLeNet) [52] and Deep Residual Learning for Image Recognition (ResNet) [53] have improved notably our ability to detect and classify objects in general terms.

We adopt a scene parsing approach based on PSPNet using a Chainer implementation provided by Tokui et al. [54]. This decision is made on the basis of the following three main reasons. First, PSPNet is specifically designed to parse urban scenes so it represents a good fit for the identification of vegetation on street-level images. Second, PSPNet features a unique pyramid parsing architecture that uses both local and global contextual information in images to classify pixels. This allows objects to be classified in the context of other objects (e.g., branch within tree). Third, PSPNet has been shown to outperform several of the most popular deep learning algorithms, including FCN, SegNet, DeepLab and DilatedNET, in major performance evaluation competitions such as the 2012 PASCAL VOC benchmark [55], the 2016 ImageNet scene parsing challenge [56], and the ongoing Cityscapes benchmark [57]. Cityscapes is a particularly relevant achievement in the context of this paper because it is a benchmark for urban scene image segmentation. PSPNet relies on a Fully Convolutional Neural Network (FCNN) for pixel prediction and a pyramid parsing module for harvesting sub-region image representations. The PSPNet architecture is illustrated in Figure 6.

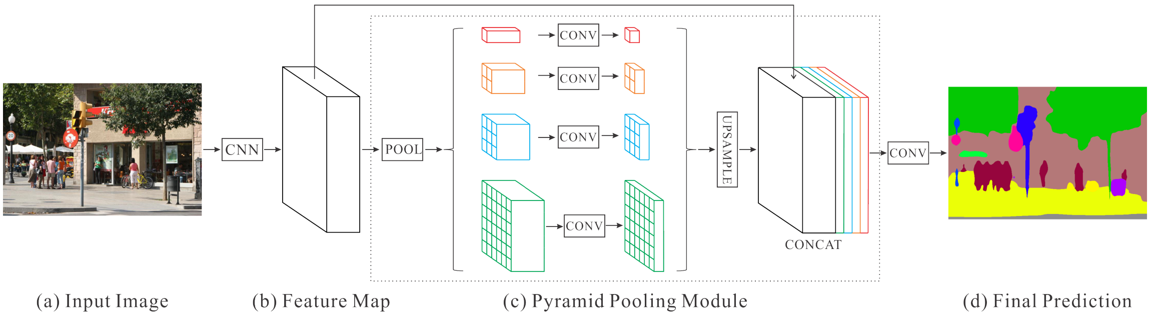


Figure 6. The PSPNet architecture (Figure reproduced here with kind permission from the PSPNet author [34])—PSPNet first employs a pre-trained ResNet [53] CNN to extract a feature map for the input image (**a**). This feature map (**b**) is then fed into a unique *pyramid pooling module* (**c**) which is a 4-layer configuration of average pooling kernels of different sizes which are intended to capture varying regions within the image. The first kernel covers the entire image, the second half of the image, while the third and fourth kernels cover smaller regions of the input image. The feature maps resulting from the pooling kernels are then up-sampled and concatenated to form a *global prior* representation of the different scales and regions in the input image. This global prior is then concatenated with the original ResNet feature map (**b**) and finally fed into a convolution layer (**d**) to produce a pixel-by-pixel class prediction.

PSPNet addresses one of the most challenging aspects of scene parsing by considering high-level contextual information when predicting pixel-by-pixel class categories. Specifically, it reduces the mismatched relationship effect [34], in which objects are classified based on appearance alone. An example of this case is a boat misclassified as a car, which would not be a possibility, if the information that cars do not usually exist on water were incorporated into the model. It has been shown that many errors in scene parsing are either partially, or completely related to the issue of mismatched relationships [34]. Solving this problem thus yields substantially higher accuracy compared to models that ignore overall scene context. This ability to incorporate context into scene understanding is of particular relevance in the context of vegetation detection. Trees and vegetation typically exist within a variety of contexts. In urban environments, these may take the form of road side trees, parks, gardens, balconies and even green walls. Contextual understanding is therefore expected to result in higher prediction accuracy compared to an approach which aims to identify vegetation based on appearance alone.

Rather than training the PSPNet architecture from scratch, we rely on pre-trained weights. Such an alternative is adopted because this architecture has already been successfully trained for similar purposes, and this can be leveraged by using the final weights. Our contribution in this context is to compare the predictive performance of two sets of weights to specifically identify urban vegetation: a pre-trained PSPNet model previously used to obtain first place in the Cityscapes benchmark [57]; and a model trained on the ADE20K benchmark data [58], which came first place in the ImageNet scene parsing challenge 2016 [56].

3.2. A Hierarchical Street-Level Tree Score

In this section, a methodology is proposed to develop a hierarchical score that aggregates individual vegetation estimates, as generated for every street-level image by PSPNet into a single score for a geographical area. Aggregated indicators have key advantages. They are easy to interpret, allow characterization of areas, and are more likely to be adopted in less technical arenas, such as policy-making. Methodologically, a certain degree of geographical aggregation reduces non-systematic error in our vegetation estimates derived from individual images.

Our approach requires every observation in the initial dataset to have assigned exogenously a second level. This level needs to aggregate all the individual observations into a smaller number of groups in a way that every first level observation (e.g., an individual image) belongs to one, and only one group in the second level (e.g., an administrative definition of neighbourhood). Once this hierarchy is established, the estimation of the scores unfolds in two main stages, one at each so-called level. First, a semantic scene parsing algorithm (e.g., PSPNet) to extract the percentage of tree cover from each GSV image in our sample. This step yields a set of individual estimates but also, as a by-product of this process, PSPNet automatically detects a range of objects within the image that characterise the urban environment, including cars, buses, pedestrians and roads. Second, a LSOA-level score of street-level trees is derived. Areas differ in geographical size and street layout, resulting in a different number of images to be aggregated. Generally this is the case, particularly when administrative areas are used. Amalgamating street-level images at a particular geography thus results in some areas containing a larger number of images while others feature a smaller share. This imbalance may affect the variance of the aggregate estimates. Additionally, street-level images are captured at different times during the day, and this usually translates into objects, such as cars, pedestrians and buses being captured. This is in addition to the built environment and natural attributes. These objects may obstruct the visibility of existing vegetation from camera sight, resulting in artificially lower levels of vegetation identified in an image.

To correct for differences in the percentage of image coverage of these attributes, a hierarchical regression model of two levels is adopted: level 1 at the image level, and level 2 at the area level. In mathematical notation, this can be expressed as:

$$\begin{array}{rcl} V_i & = & \alpha_l + \beta X_i + \epsilon_i \\ \alpha_l & \sim & \mathcal{N}(\alpha, \sigma_\alpha) \\ \epsilon_i & \sim & \mathcal{N}(0, \sigma_\epsilon) \end{array} \tag{1}$$

where V_i represents the proportion of pixels in image i that are classified as vegetation by PSPNet, α_l is a random effect at the area l level, β are the parameters relating to the proportion of coverage of the range of attributes identified from an image X_i, and ϵ_i is an i.i.d. error term, assumed to be normally distributed with a standard deviation of σ_ϵ. Every area-specific random effect α_l is hierarchically connected through an α hyper-parameter. Assuming X_i are demeaned (i.e., scaled so the average is set to zero), α_l can be interpreted as the average percentage of vegetation for all the images within a given area l, controlling for the obstructing effect of the range of image features represented by X_i. The estimated average percentage of vegetation is our proposed hierarchical LSOA street-level urban forest score.

It is important to contextualise the derivation of this score within the broader framework of multilevel models. The particular interpretation of the model in Equation (1) as a vegetation score, its application in the context of aggregating estimates individually derived from images, as well as the interpretation of X_i as a correction for obstructing objects, are novel contributions of this paper. However, the more general approach of multilevel modelling on which the scores rely is a robust and widely used technique in statistics and, more specifically, in regression analysis. For that reason, it is beyond the scope of this paper to cover its internal mechanisms in detail—see [59,60].

A hierarchical score as derived from Equation (1) has three main advantages. First, it allows to naturally account for image-to-image variation in the percentage of image covered by "blocking objects" such as buses, pedestrians and cars in the estimation of the average percentage of vegetation at the LSOA level. Second, because the LSOA random effects α_l are hierarchically connected through a hyper-parameter, the estimates are robust to over-fitting in cases where too few observations are available for a given area. This is the so-called *shrinkage effect* [59], one of the main reasons these modes are widely popular. Third, our scores are derived within a probabilistic model and estimates of the uncertainty can be obtained. Because each α_l is assumed to follow a normal distribution with modelled parameters, fitting the model also returns estimates for such measures of uncertainty (i.e., σ_α). This can be particularly useful when the α_l scores are used to inform comparisons between areas, as point estimates may be misleading, creating the illusion of differences that may be statistically insignificant.

4. Results and Discussion

This section first discusses the results of a comparative assessment between our proposed PSPNet approach and the pixel-based methods described in the previous section. The relative performance of these methods to classify image pixels into vegetation and non-vegetation classes is assessed. Our hierarchical model to derive scores of urban street-level trees is then introduced. Scores for the city of Cardiff are computed and their systematic variation is explored through regression analysis.

4.1. Comparative Assessment

For each image in the Mapillary test (*ground-truth*) dataset, the performance of PSPNet is evaluated, along with the green pixel $L*a*b*$ threshold and novel $L*a*b*$ random forest mask extension described previously. To this end, a range of statistical metrics are used to assess prediction accuracy as shown in Table 1. These metrics are based on statistical measures of sensitivity and specificity. These measures are commonly used to assess the accuracy of binary classification outcomes, and comprise: True Positive (TP), False Positive (FP); True Negative (TN) and False Negative (FN). In the context of our application, they refer to:

- TP: correctly identified vegetation pixels
- FP: incorrectly identified vegetation pixels

- TN: correctly rejected vegetation pixels
- FN: incorrectly rejected vegetation pixels

Table 1. Evaluation metrics.

Metric	Description	Equation
BACC	Class balanced accuracy	(TP/P) + (TN/N)/2
Pre	Precision	TP/TP + FP
Rec	Recall	TP/FN + TP
IoU	Intersection over union	TP/(TP + FP + FN)
MCC	Matthews correlation coefficient	$\frac{(TP*TN-FP*FN)}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$
R^2	Goodness of fit	$1-(SS_{res}/SS_{total})$
τ	Kendall's Tau	$(n_c-n_d)/(n*(n-1)/2)$

SS_{res}: regression sum of squares; SS_{total} total sum of squares; n_c: number of concordant pairs; n_d: number of discordant pairs; n: number of observations.

A commonly-used measure of accuracy for class balanced binary classification (BACC) is used. This indicator is adjusted to account for the unbalanced nature of our dataset, which contains higher numbers of pixels in the non-tree class than in the tree class. It can be interpreted as the average *recall* over both tree and non-tree classes. Two common precision (Pre) and recall (Rec) measures and indicators of precision/recall tradeoff, namely the Intersection over Union (IoU) [61] and Matthews correlation coefficient (MCC) [41] scores, are also utilised. While the first two indicators provide independent measures of precision and recall, the latter two indicators assess the tradeoff between precision and recall. IoU is typically referred to as the Jaccard Index, and can be interpreted as the average percentage overlap between the predicted and actual image segment areas. As a complimentary measure, the MCC score can be interpreted as a (class balanced) correlation between the predicted and actual pixel classes.

In addition to these summary metrics to measure pixel-by-pixel classification accuracy, the overall model capacity to predict the percentage of visible tree in each image is also assessed using two measures. First, the model goodness of fit is summarised by assessing predicted vs actual percentage tree present in each image within the Mapillary ground-truth data based the R^2 metric. Second, the Kendall's rank correlation coefficient is computed as a non-parametric measure of ordinal similarity, which is interpreted as a comparison between the predicted and actual percentage rankings of each image.

To begin our comparison, a simplified version of the previously described $L*a*b*$ method has been included. In this version only the (green to red) $a*$ parameter is used to threshold images. As with the full $a*b*$ model, a set of optimal parameters is derived. However, this time with an exhaustive grid-search yields an optimal set of values ($A_1=-31, A_2=-11$). Using the optimal $-31 \leq a* \leq -11$ parameters, this model ($a*$) can only reach a BACC of 55%.

Adding back the (blue to yellow) $b*$ parameter and using the optimal $-31 \leq a* \leq -6$ and $5 \leq b* \leq 57$ model obtained using Bayesian optimisation as described in the previous section, the $a*b$ BACC of 62% and accompanying summary statistics reported in Table 2 indicate that the use of both ($a*, b*$) parameters results in more accurate predictions. Yet, nearly 40% of the predictions are inaccurate. Since the algorithm is purely based on colour pixel patterns, it is likely this hinders its ability to differentiate between green objects, such as a green tree and a green car.

Next, the *random forest mask approach* ($RF(a*, b*)$) is then applied to relax the rectangular space. Although this approach produces some improvements in terms of the precision, recall and correlation metrics in Table 2, these improvements are marginal resulting in a BACC accuracy score of 62%.

Table 2. Performance comparison of competing approaches to classify vegetation pixels in Mapillary dataset.

Model	BACC	Pre	Rec	IoU	MCC	R^2	τ
$a*$	55%	0.33	0.14	0.10	0.15	0.04	0.15
$a * b*$	62%	0.47	0.28	0.21	0.29	0.20	0.28
Random Forest mask	62%	0.48	0.29	0.22	0.31	0.25	0.32
PSPNet (ADE20K)	85%	**0.82**	0.73	**0.63**	**0.74**	**0.83**	0.76
PSPNet (Cityscapes)	**90%**	0.66	**0.87**	0.60	0.72	**0.83**	**0.77**

BACC = Balanced accuracy, Pre/Rec = Precision or recall, IoU = Intersection-over-Union (Jaccard index), MCC = Matthews correlation coefficient, τ = Kendall's tau. Higher scores in each column in bold.

Finally, the application of *PSPNet* to Mapillary images is presented. Performance results from applying this algorithm are available in the two bottom rows of Table 2. The various performance metrics consistently show a significant improvement in prediction quality compared to the two previously discussed approaches. For example, the BACC accuracy score is at least 20% higher and the R^2 is over three times higher than that of the $RF(a*, b*)$ approach (0.83 vs 0.25). Models using two sets of weights are compared. The model based on the Cityscapes weights displays slightly better results over the ADE20K model in terms of accuracy (BACC) and recall. These results may reflect the fact that Cityscapes was specifically trained to identify features in urban environments, while ADE20K has a more general purpose.

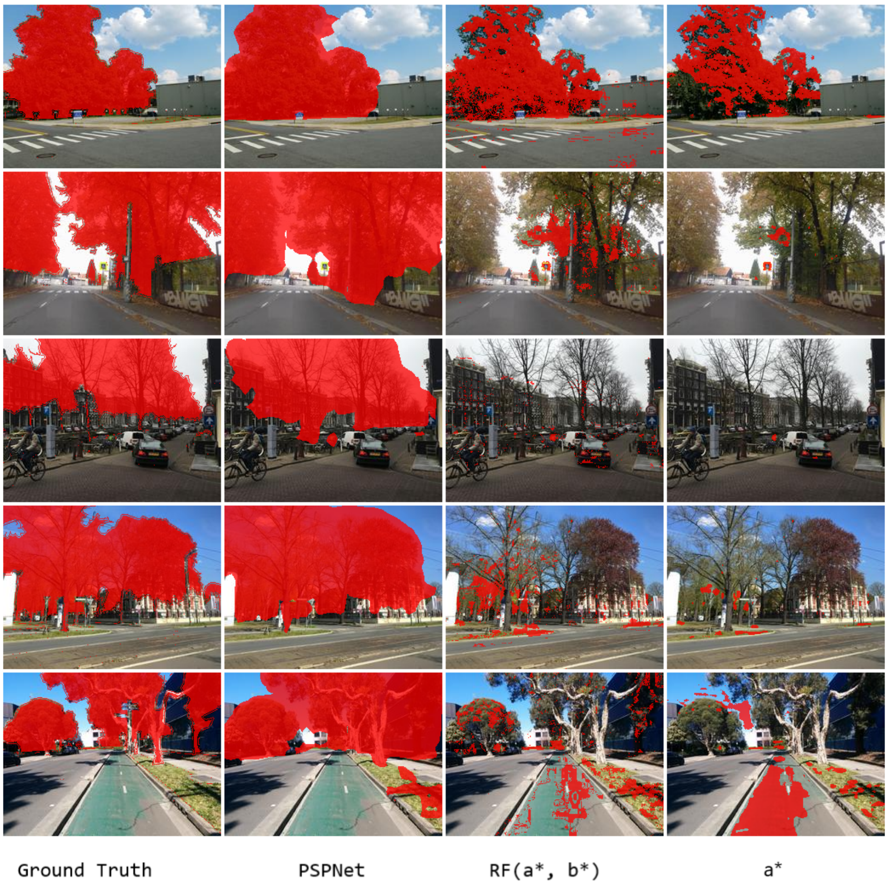


Figure 7. Illustration of competing vegetation segmentation approaches. Every column corresponds to the manually annotated output and our three competing approaches. Every row corresponds to a scene illustrating different types of challenges, respectively: "clearcut" segmentation of vegetation; vegetation in autumn where green leaves are not present; vegetation in winter with no leaves; vegetation with no green leaves; and vegetation with lane painted in green. Pixels labelled as vegetation are highlighted in red.

We also assess the computational performance of the various algorithms. We report two metrics of processing time per image: Millisecond per image and images per second (Table 3). All our analyses were performed on a Linux machine with a NVIDIA GeForce GTX 1060 6GB GPU and Intel(R) Core(TM) i7-8700 CPU @ 3.20 GHz. The results show that while the $RF(a*, b*)$ approach is inferior to PSPNet in terms of segmentation accuracy, it is significantly more efficient in terms of processing time. This is due to the nature of the bitmap mask, as described in Section 3.1.2, which was implemented based on highly efficient vectorised code. The $RF(a*, b*)$ approach is also more efficient than the simple $L*a*b*$ method based on $a*$ and $a*b*$ parameters, demonstrating that our generalised version of the $L*a*b*$ method is more flexible and efficient. Note that the $a*$ and $a*b*$ approaches depend on the same implementation resulting in equal processing times.

Table 3. Computing performance comparison of competing approaches to classify vegetation pixels in Mapillary dataset.

Model	Millisecond per Image	Images per Second
$a*$	0.2663 < 1 ms	3755.16
$a*b*$	0.2663 < 1 ms	3755.16
Random Forest mask	0.1959 < 1 ms	5104.65
PSPNetGPU	87.8 ms	11.39
PSPNetCPU	1921 ms	0.52

To better understand the comparative performance of the three competing approaches, Figure 7 displays the segmentation labels produced by each method and those contained in the ground-truth Mapillary dataset. The $a*b*$ method has been omitted due to its similarity with $RF(a*, b*)$ which is a marginal improvement over $a*b*$. The first row consists of easily identifiable trees, which were labelled nearly perfectly by the PSPNet model and with some success by the $RF(a*, b*)$ and $a*$ threshold methods. The second and third rows illustrate one of the main issues with the $L*a*b*$ based methods. Green vegetation is less prevalent in autumn and winter months, and color-based only approaches perform poorly identifying un-leafy trees. The forth and fifth rows highlight another key limitation of the $L*a*b*$ methods. This is in distinguishing that not all tree species are green, and not all green objects are trees. In contrast, the PSPNet model is robust to all these situations, likely due to its ability to learn beyond the characteristics of an individual pixel and into the structure of the overall arrangement of pixels as discussed in Section 3.1.3. Further, Figure 8 illustrates the relationship between the *actual* percentage vegetation and our *predicted* percentage vegetation over approximately 10,000 images in the Mapillary dataset.

Taken together, these results evidence the robustness and accuracy of PSPNet loaded with the Cityscapes pre-trained weights to identify and classify vegetation pixels from street-level images, although they also show PSPNet is computationally demanding for high throughput applications. Our experiments reveal the superiority of the neural network approach over the pixel-based techniques derives from the ability of the neural network to incorporate information on the configuration of values *throughout* the image. This feature permits the network to include patterns and shapes across pixels that provide important information when it comes to predicting vegetation presence. In contrast pixel-based methods, such as those used as a benchmark in this exercise, cannot incorporate more information than that contained in a single pixel to generate its prediction. As our results show, this drawback has important predictive consequences. Additionally, the performance of weights from Cityscapes is also understandable as the original intention and training focus was on urban scenes. Having used very similar input data to those used in the context of this experiment, the Cityscapes weights offer an excellent fit-for-purpose in the context of this paper, as the predictive performance results suggest. This approach is next applied to derive a hierarchical score to characterize the spatial distribution of vegetation in urban environments.

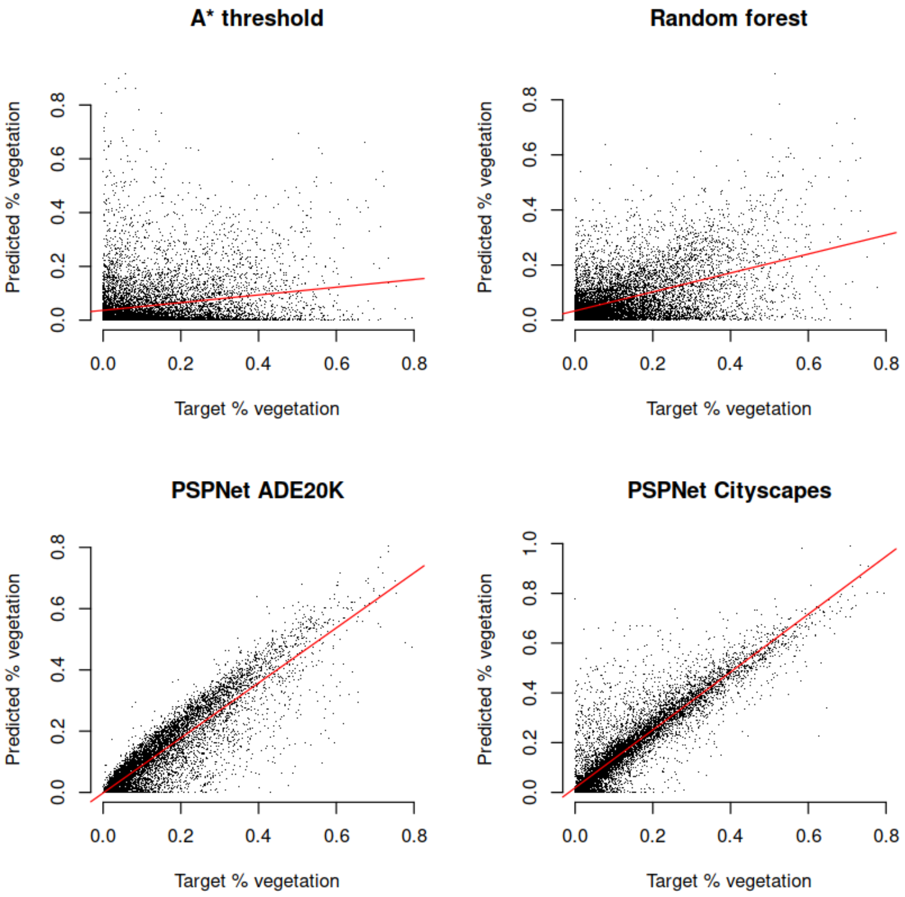
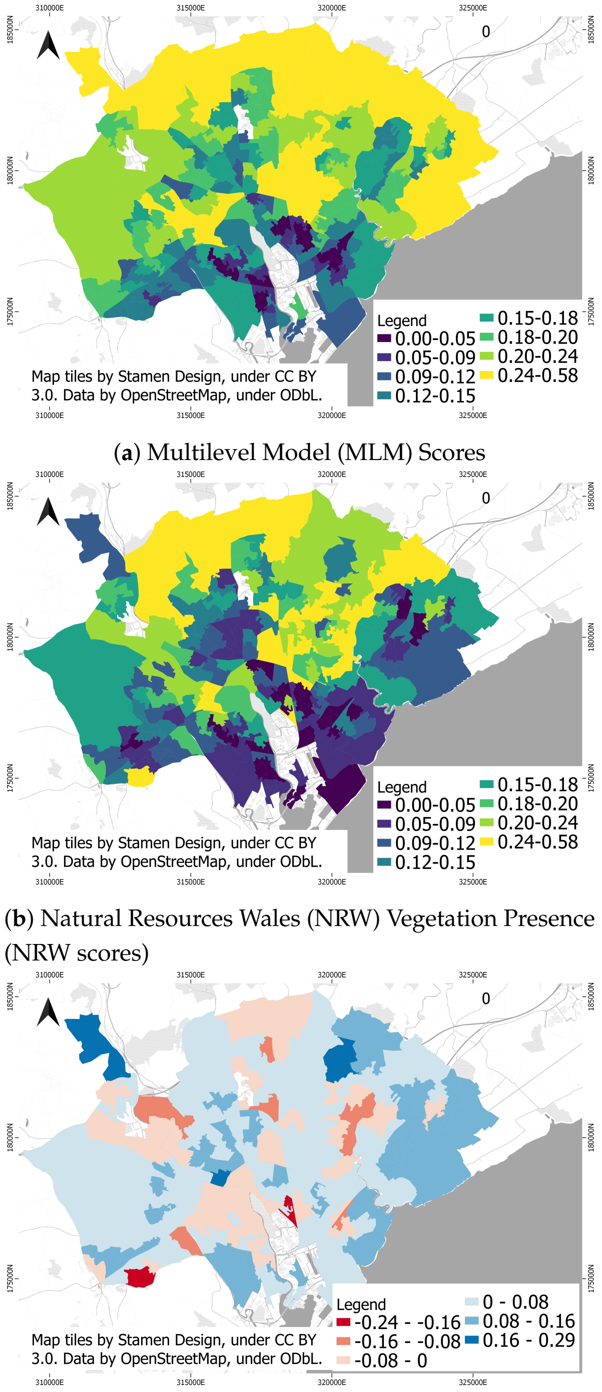


Figure 8. Comparison of actual levels of vegetation with predictions from different approaches. Horizontal axes display the actual percentage of vegetation; vertical axes contain predicted values by each approach. Best linear fit (bivariate correlation) is displayed as red lines.

4.2. Area Scores

Once vegetation pixels at the street-level are identified and classified based on PSPNet, hierarchical scores as described in Section 3.2 can be calculated. To this end, three aspects are considered. First, a higher geographical level to aggregate pixels from individual image estimates is needed. Lower layer Super Output Areas (LSOAs) are adopted. They represent the middle layer in the set of statistical geographies developed by the ONS [62]. LSOAs offer a good compromise between tractability and spatial resolution. The second aspect involves defining features which may obstruct visibility of vegetation in GSV images. Extracted using PSPNet, the following categories of pixels are employed: roads, sky, persons, riders, cars, trucks, buses, trains, motorcycles, and bicycles. Third, an estimation method is required to recover the α_l parameters in the model of Equation (1). For this task, the model is estimated using restricted maximum likelihood (REML) based on the `lme4` R package [63].

Figure 9a presents the distribution of hierarchical scores for LSOAs across Cardiff. As expected, it shows a marked core-periphery geographical pattern. More densely populated LSOAs in the city centre display the lowest scores of street-level vegetation ranging from 0 to 0.13, while more peripheral LSOAs at the city fringe, particularly at the north and northeast, exhibit the highest presence of street-level vegetation increasing from 0.23 to 0.58. Middle-range scores indicate moderate abundance of street-level vegetation in intermediate areas.



(**a**) Multilevel Model (MLM) Scores

(**b**) Natural Resources Wales (NRW) Vegetation Presence (NRW scores)

(**c**) MLM - NRW score

Figure 9. Comparison of Urban Forest scores. (**a**): Multilevel Model (MLM) Scores. (**b**): NRW vegetation percentage cover scores. (**c**): the difference between the MLM score and the NRW vegetation estimates.

To understand systematic differences in the patterns of vegetation coverage across LSOAs based on our hierarchical score, an ordinary least squared regression model is estimated to assess the relationship between average street-level coverage and a range of land use classes at the LSOA level. Our hierarchical score is regressed over the percentage of a LSOA area used for 12 different land use classes: commercial areas, education, hospitals, burial locations, remnant countryside, formal open space, informal open space, woodland, high density residential areas, low density residential areas and unclassified land. Data for the land use classes were obtained from the 2013 NRW urban tree cover study [38].

The left panel in Table 4 displays the estimates of our regression model. Statistically significant (p-value < 10%) and positive regression coefficients indicate that LSOAs with large shares of areas allocated to commercial, education, formal open space, informal open space, low density residential and unclassified land uses tend to have larger percentages of street-level vegetation cover. Particularly, coefficients above one reveal that unclassified open land with no development, and low density residential areas are associated with the largest average shares of street tree coverage. These coefficients also indicate that a one point increase in the percentage of the LSOA area used for unclassified land and low density residential buildings is associated with an increase of vegetation coverage of 1.28% and 1.19% respectively. The nature of vegetation in these areas tends to differ. While uncontrolled vegetation tends to be abundant in LSOAs with extensive areas of unclassified open land and scarce building infrastructure, well maintained trees are comparatively more prevalent in LSOAs largely dedicated to low density residential development. Unsurprisingly, LSOAs primarily used for high density residential buildings represent the only feature negatively associated with the average of tree coverage. The corresponding regression coefficient is statistically significant and negative, indicating that a one point increase in the share of LSOA dedicated to high density residential buildings is related to a reduced percentage of vegetation cover by 2.12%. In the context of the Cardiff study area, this is likely due to the prevalence of terraced housing of which building facade is likely to be the most dominant feature in each street-level image.

Table 4. Regression results.

Variables	**MLM Score**			**MLM—NRW Score**		
	Coef	**SE**		**Coef**	**SE**	
Intercept	0.081	0.011	***	0.082	0.011	***
Commercial Areas	0.932	0.391	**	−0.066	0.391	
Education	0.855	0.407	**	−0.144	0.407	
Hospitals	0.498	0.868		−0.500	0.868	
Burial Locations	0.010	0.190		−0.990	0.191	***
Remnant Countryside	0.357	2.134		−0.646	2.135	
Formal Open Space	0.302	0.083	***	−0.697	0.083	***
Informal Open space	0.364	0.114	***	−0.635	0.114	***
Woodland	0.411	0.354		−0.588	0.354	***
High Density Residential	−2.616	0.722	***	−3.619	0.722	***
Low Density Residential	1.191	0.145	***	0.189	0.145	
Transport Corridors	0.718	0.825		−0.280	0.826	
Unclassified	1.277	0.262	***	0.273	0.262	
R^2	0.587			0.448		
Adj. R^2	0.561			0.412		

Significance level (p-value): *** 1%, ** 5%, * 10%.

As a complement to the study of systematic variation in our scores, they are also compared to the percentage of vegetation cover extracted from NRW described in Section 2.3. We believe such analysis can bring an additional layer of reliability to our scores, and improve the understanding of our model performance analysing areas where differences exist. While our approach measures vegetation that is visible from *a pedestrian's perspective*, the NRW captures vegetation as observed from *the sky*. Thus, vegetation measures extracted for particular road segments are likely to differ, as such their geographical correspondence at the LSOA level is assessed.

Figure 9b shows the spatial distribution of the NRW vegetation percentage cover, and Figure 9c displays the difference between our hierarchical score and the NRW vegetation estimates. Positive values (blue) correspond to areas where our hierarchical score displays a higher proportion of vegetation than the NRW estimate, with negative values (red) indicating the reverse pattern. The overall pattern is remarkably consistent displaying marginal differences in vegetation scores. Yet, differences exist, showing two very prominent outliers. These outliers correspond to over-estimations of vegetation of our hierarchical score over the NRW vegetation percent cover for two LSOAs in the north-west of Cardiff. These differences in estimates can be explained by the existence of a small number of long roads within these wards passing through high-density road-side woodland. Viewed at street-level, as captured by our score, almost 60% of the visual field is vegetation. In contrast, the NRW vegetation score measures less vegetation as only part of the aerial images used for calculation is covered by roads.

Similarly, our score significantly under-estimates vegetation cover compared to the NRW score in a LSOA in central Cardiff, in the area of Cathays, reflecting high density terraced housing. At street level, little vegetation is observable and therefore our score measures scarce levels of vegetation, whereas the NRW percent vegetation cover is higher as it captures backyard garden trees. Our score also significantly under-estimates vegetation cover in the south-west in the area of Caerau. This difference highlights an aspect of our estimates that warrants caution, as this area has an abundance of street-side grass and small trees. While this is captured by street-level imagery, this form of vegetation covers a small share of overall imagery. By contrast, aerial imagery does a better job measuring floor-level and miniature trees, suggesting that complementing street-level and aerial-based approaches would produce a more accurate measure of urban green space in some areas.

To better understand systematic patterns in differences between our hierarchical score and the NRW vegetation percent cover, an OLS regression model is estimated using these differences as a dependent variable and the range of land use classes described above. The right panel in Table 4 reports the results. Negative coefficients reveal that our hierarchical score tends to produce significantly smaller estimates of vegetation cover in LSOAs with large shares of land used for burial, formal open, informal open, woodland and high density residential spaces. These differences between vegetation scores are particularly large for LSOAs comprising extensive high density residential areas. Differences are amplified by 3.2% with 1% increases in the share of LSOA covered by high density residential areas. As exemplified by Cathays above, vegetation in LSOAs accommodating large high density residential areas is generally in backyard gardens and thus is not captured by our street-level pedestrian view score.

5. Conclusions

Measuring the size, location and distribution of the stock of urban forest is challenging. Data at a very high degree of spatial granularity is required to accurately achieve this task. This paper proposes a novel, scalable, automatable and consistent approach to quantify urban trees at the street level in cities. Our method is based on a semantic image segmentation approach (PSPNet) combined with a hierarchical multilevel model and street-level imagery. We demonstrated the accuracy of our approach against a widely-used pixel threshold and a more flexible extension. Our approach is also validated by estimating a hierarchical tree score to measure tree percentage cover and measuring its correspondence with high-resolution aerial tree crown cover data. The results show a remarkable degree of correspondence capturing similar percentages of urban trees per areas.

We argue our street tree vegetation index represents a robust measure of the amount of nature humans perceive and experience at the street level. It does so by more appropriately capturing tree cover in areas with large burial, woodland, formal open and informal open spaces where shallow trees are abundant, in high density residential areas with backyard trees, and along street networks with high density of high trees, compared to aerial tree cover estimates. We thus argue the methodology proposed in this paper could form the base to develop research and data products that can inform and further our understanding of the link between mental health, well-being, green space and air pollution. We hope future research progresses in this direction as new and open sources of street view imagery, such as mapstreetview and OpenStreetCam proliferate and improve their geographical coverage.

Author Contributions: Conceptualization, P.S., J.P., F.R. and D.A.-B.; methodology, P.S., J.P., F.R. and D.A.-B.; software, P.S., J.P. and D.A.-B.; validation, P.S., J.P., F.R. and D.A.-B.; formal analysis, P.S., J.P., F.R. and D.A.-B.; investigation, P.S., J.P., F.R. and D.A.-B.; resources, P.S., J.P.; data curation, P.S., J.P.; writing—original draft preparation, P.S., F.R. and D.A.-B.; writing—review and editing, P.S., F.R. and D.A.-B.; visualization, P.S., F.R. and D.A.-B.; supervision, P.S., F.R. and D.A.-B.; project administration, P.S.; funding acquisition, P.S.

Funding: This research received no external funding.

Acknowledgments: This project has been supported by the Office for National Statistics (ONS) Natural Capital team. We wish to thank Emily Connors for guidance and feedback during this project. This work has also benefitted from various discussions and feedback from several researchers in the ONS Data Science Campus. In particular, Alex Noyvirt for implementing a VGG-16 image class prediction model during the early stages of this project, Gareth Clews for advice on SQL schema, and Bernard Peat for discussions and initial help in background research. We also wish to thank Sonia Williams and Tom Smith for providing feedback on this article.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

BACC	Class balanced accuracy
GSV	Google Street View
LSOA	Lower Layer Super Output Area
MCC	Matthews Correlation Coefficient
NRW	Natural Resources Wales
OLS	Ordinary Least Squares
ONS	Office of National Statistics
PSPNet	Pyramid Scene Parsing Network
RF	Random Forest
REML	Restricted Maximum Likelihood
FCNN	Fully Convolutional Neural Network
CCN	Convolutional Neural Network
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
LiDAR	high-resolution light detection and ranging
DCNN	Deep Convolutional Neural Networks
NFI	National Forest Inventory
FCN	Fully Convolutional Network
SegNet	Segmentation Network
DeepLab	Deep Labelling for Semantic Image Segmentation
VGG16	Very Deep Convolutional Networks
GoogleLeNet	Inception Network
ResNet	Deep Residual Learning for Image Recognition
IoU	Intersection over Union
Pre	Precision
Rec	Recall
SS	Sum of Squares

References

1. Blanco, H.; Alberti, M.; Forsyth, A.; Krizek, K.J.; Rodriguez, D.A.; Talen, E.; Ellis, C. Hot, congested, crowded and diverse: Emerging research agendas in planning. *Prog. Plann.* **2009**, *71*, 153–205. [CrossRef]
2. Escobedo, F.J.; Kroeger, T.; Wagner, J.E. Urban forests and pollution mitigation: Analyzing ecosystem services and disservices. *Environ. Pollut.* **2011**, *159*, 2078–2087. [CrossRef] [PubMed]
3. Wolch, J.R.; Byrne, J.; Newell, J.P. Urban green space, public health, and environmental justice: The challenge of making cities 'just green enough'. *Landsc. Urban Plan.* **2014**, *125*, 234–244. [CrossRef]

4. Office of National Statistics. *UK Air Pollution Removal: How Much Pollution Does Vegetation Remove in Your Area?* Technical Report; ONS: Newport, Wales, UK, 2018.
5. Nowak, D.J.; McHale, P.J.; Ibarra, M.; Crane, D.; Stevens, J.C.; Luley, C.J. Modeling the effects of urban vegetation on air pollution. In *Air Pollution Modeling and Its Application XII*; Springer: New York, NY, USA, 1998; pp. 399–407.
6. Wenting, W.; Yi, R.; Hengyu, Z. Investigation on temperature dropping effect of urban green space in summer in Hangzhou. *Energy Procedia* **2012**, *14*, 217–222. [CrossRef]
7. Giles-Corti, B.; Macintyre, S.; Clarkson, J.P.; Pikora, T.; Donovan, R.J. Environmental and lifestyle factors associated with overweight and obesity in Perth, Australia. *Am. J. Health Promot.* **2003**, *18*, 93–102. [CrossRef]
8. Woo, J.; Tang, N.; Suen, E.; Leung, J.; Wong, M. Green space, psychological restoration, and telomere length. *Lancet* **2009**, *373*, 299–300. [CrossRef]
9. Kaplan, S.; Kaplan, R. Health, supportive environments, and the reasonable person model. *Am. J. Public Health* **2003**, *93*, 1484–1489. [CrossRef]
10. Song, Y.; Gee, G.C.; Fan, Y.; Takeuchi, D.T. Do physical neighborhood characteristics matter in predicting traffic stress and health outcomes? *Transp. Res. Part F Traffic Psychol. Behav.* **2007**, *10*, 164–176. [CrossRef]
11. Woodcock, J.; Edwards, P.; Tonne, C.; Armstrong, B.G.; Ashiru, O.; Banister, D.; Beevers, S.; Chalabi, Z.; Chowdhury, Z.; Cohen, A.; et al. Public health benefits of strategies to reduce greenhouse-gas emissions: Urban land transport. *Lancet* **2009**, *374*, 1930–1943. [CrossRef]
12. James, P.; Hart, J.; Banay, R.; Laden, F. Exposure to greenness and mortality in a nationwide prospective cohort study of women. *Environ. Health Perspect.* **2016**, *124*, 1344–1352. [CrossRef]
13. Louv, R. *Last Child in the Woods: Saving Our Children from Nature-Deficit Disorder*; Algonquin Books: Chapel Hill, NC, USA, 2008.
14. Natural Resources Wales . *Tree Cover in Wales' Towns and Cities*; Technical Report; Natural Resources Wales: Cardiff, UK, 2016.
15. Seiferling, I.; Naik, N.; Ratti, C.; Proulx, R. Green streets- Quantifying and mapping urban trees with street-level imagery and computer vision. *Landsc. Urban Plan.* **2017**, *165*, 93–101. [CrossRef]
16. Fuller, R.A.; Gaston, K.J. The scaling of green space coverage in European cities. *Biol. Lett.* **2009**, *5*, 352–355. [CrossRef] [PubMed]
17. Dickinson, J.L.; Zuckerberg, B.; Bonter, D.N. Citizen science as an ecological research tool: Challenges and benefits. *Annu. Rev. Ecol. Evolut. Syst.* **2010**, *41*, 149–172. [CrossRef]
18. Homer, C.; Dewitz, J.; Fry, J.; Coan, M.; Hossain, N.; Larson, C.; Herold, N.; McKerrow, A.; VanDriel, J.N.; Wickham, J.; et al. Completion of the 2001 national land cover database for the counterminous United States. *Photogramm. Eng. Remote Sens.* **2007**, *73*, 337.
19. MacFaden, S.W.; O'Neil-Dunne, J.P.; Royar, A.R.; Lu, J.W.; Rundle, A.G. High-resolution tree canopy mapping for New York City using LIDAR and object-based image analysis. *J. Appl. Remote Sens.* **2012**, *6*, 063567. [CrossRef]
20. Yang, J.; Zhao, L.; Mcbride, J.; Gong, P. Can you see green? Assessing the visibility of urban forests in cities. *Landsc. Urban Plann.* **2009**, *91*, 97–104. [CrossRef]
21. Naik, N.; Philipoom, J.; Raskar, R.; Hidalgo, C. Streetscore-predicting the perceived safety of one million streetscapes. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Columbus, OH, USA, 24–27 June 2014; pp. 779–785.
22. Naik, N.; Kominers, S.D.; Raskar, R.; Glaeser, E.L.; Hidalgo, C.A. *Do People Shape Cities, or Do Cities Shape People? The Co-Evolution of Physical, Social, and Economic Change in Five Major US Cities*; Technical Report; National Bureau of Economic Research: Cambridge, MA, USA, 2015.
23. Glaeser, E.L.; Kominers, S.D.; Luca, M.; Naik, N. Big data and big cities: The promises and limitations of improved measures of urban life. *Econ. Inq.* **2018**, *56*, 114–137. [CrossRef]
24. Zhang, W.; Witharana, C.; Li, W.; Zhang, C.; Li, X.; Parent, J. Using Deep Learning to Identify Utility Poles with Crossarms and Estimate Their Locations from Google Street View Images. *Sensors* **2018**, *18*, 2484, doi:10.3390/s18082484. [CrossRef]
25. Gebru, T.; Krause, J.; Wang, Y.; Chen, D.; Deng, J.; Aiden, E.L.; Fei-Fei, L. Using deep learning and Google Street View to estimate the demographic makeup of neighborhoods across the United States. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 13108–13113, doi:10.1073/pnas.1700035114. [CrossRef] [PubMed]

26. Kang, J.; Körner, M.; Wang, Y.; Taubenböck, H.; Zhu, X.X. Building Instance Classification Using Street View Images. *ISPRS J. Photogramm. Remote Sens.* **2018**, *145 Pt A*, 44–59
27. Li, X.J.; Ratti, C.; Seiferling, I. Quantifying the shade provision of street trees in urban landscape: A case study in Boston, USA, using Google Street View. *Landsc. Plan.* **2017**, *169*, 81–91. [CrossRef]
28. Li, X.J.; Ratti, C. Mapping the spatial distribution of shade provision of street trees in Boston using Google Street View panoramas. *Urban For. Urban Green.* **2018**, *31*, 109–119, doi:10.1016/j.ufug.2018.02.013. [CrossRef]
29. Li, X.; Ratti, C. Using Google Street View for Street-Level Urban Form Analysis, a Case Study in Cambridge, Massachusetts. In *The Mathematics of Urban Morphology*; D'Acci, L., Ed.; Springer: Cham, Switzerland, 2019; pp. 457–470, doi:10.1007/978-3-030-12381-9_20.
30. Wegner, J.D.; Branson, S.; Hall, D.; Schindler, K.; Perona, P. Cataloging Public Objects Using Aerial and Street-Level Images and Urban Trees. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016; pp. 6014–6023. doi:10.1109/CVPR.2016.647. [CrossRef]
31. Branson, S.; Wegner, J.D.; Hall, D.; Lang, N.; Schindler, K.; Perona, P. From Google Maps to a fine-grained catalog of street trees. *ISPRS J. Photogramm. Remote Sens.* **2018**, *135*, 13–30. [CrossRef]
32. Li, X.; Zhang, C.; Li, W.; Ricard, R.; Meng, Q.; Zhang, W. Assessing street-level urban greenery using Google Street View and a modified green view index. *Urban For. Urban Green.* **2015**, *14*, 675–685. [CrossRef]
33. Cai, B.Y.; Li, X.; Seiferling, I.; Ratti, C. Treepedia 2.0: Applying Deep Learning for Large-scale Quantification of Urban Tree Cover. *arXiv* **2018**, arXiv:1808.04754.
34. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid scene parsing network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2881–2890.
35. Cardiff Research Centre. *Bute Park Restoration Project—Public Consultation Spring 2013*; Technical Report; Cardiff Research Centre: Cardiff, UK, 2013.
36. Neuhold, G.; Ollmann, T.; Bulò, S.R.; Kontschieder, P. The Mapillary Vistas Dataset for Semantic Understanding of Street Scenes. In Proceedings of the International Conference on Computer Vision (ICCV), Venice, Italy, 22–27 October 2017; pp. 5000–5009.
37. OpenStreetMap Contributors. Planet Dump. 2017. Available online: https://planet.osm.org (accessed on 11 June 2019).
38. Natural Resources Wales. *Town Tree Cover in the City and County of Cardiff*; Technical Report; Natural Resources Wales: Cardiff, UK, 2016.
39. Scharr, H.; Minervini, M.; French, A.P.; Klukas, C.; Kramer, D.M.; Liu, X.; Luengo, I.; Pape, J.M.; Polder, G.; Vukadinovic, D.; et al. Leaf segmentation in plant phenotyping: A collation study. *Mach. Vis. Appl.* **2016**, *27*, 585–606. [CrossRef]
40. Snoek, J.; Larochelle, H.; Adams, R.P. Practical Bayesian Optimization of Machine Learning Algorithms. In *Advances in Neural Information Processing Systems 25*; Pereira, F., Burges, C.J.C., Bottou, L., Weinberger, K.Q., Eds.; Curran Associates, Inc.: Red Hook, NY, USA , 2012; pp. 2951–2959.
41. Matthews, B.W. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim. Biophys. Acta (BBA) Protein Struct.* **1975**, *405*, 442–451. [CrossRef]
42. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
43. Friedman, J.; Hastie, T.; Tibshirani, R. *The Elements of Statistical Learning*; Springer: New York, NY, USA, 2001; Volume 1.
44. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436. [CrossRef] [PubMed]
45. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. doi:10.1109/5.726791. [CrossRef]
46. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems 25*; Pereira, F., Burges, C.J.C., Bottou, L., Weinberger, K.Q., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2012; pp. 1097–1105.
47. Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. In Proceedings of the 28th IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015.
48. Badrinarayanan, V.; Kendall, A.; Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *arXiv* **2015**, arXiv:1511.00561.

49. Chen, L.C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking atrous convolution for semantic image segmentation. *arXiv* **2017**, arXiv:1706.05587.
50. Chen, L.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.
51. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556.
52. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.E.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going Deeper with Convolutions. In Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015.
53. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016, pp. 770–778.
54. Tokui, S.; Oono, K.; Hido, S.; Clayton, J. Chainer: A next-generation open source framework for deep learning. In Proceedings of the Workshop on Machine Learning Systems (LearningSys) in the Twenty-Ninth Annual Conference on Neural Information Processing Systems (NIPS), Montreal, QC, Canada, 7–12 December 2015; Volume 5, pp. 1–6.
55. Everingham, M.; Van Gool, L.; Williams, C.K.; Winn, J.; Zisserman, A. The pascal visual object classes (voc) challenge. *Int. J. Comput. Vis.* **2010**, *88*, 303–338. [CrossRef]
56. Zhou, B.; Zhao, H.; Puig, X.; Fidler, S.; Barriuso, A.; Torralba, A. Semantic understanding of scenes through the ADE20K dataset. *arXiv* **2016**, arXiv:1608.05442.
57. Cordts, M.; Omran, M.; Ramos, S.; Rehfeld, T.; Enzweiler, M.; Benenson, R.; Franke, U.; Roth, S.; Schiele, B. The cityscapes dataset for semantic urban scene understanding. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 3213–3223.
58. Zhou, B.; Zhao, H.; Puig, X.; Fidler, S.; Barriuso, A.; Torralba, A. Scene parsing through ade20k dataset. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; Volume 1, p. 4.
59. Gelman, A.; Hill, J. *Data Analysis Using Regression and Multilevel/Hierarchical Models*; Cambridge University Press: Cambridge, UK, 2006.
60. Goldstein, H. *Multilevel Statistical Models*; John Wiley & Sons: Hoboken, NJ, USA, 2011; Volume 922.
61. Everingham, M.; Eslami, S.M.A.; Van Gool, L.; Williams, C.K.I.; Winn, J.; Zisserman, A. The Pascal Visual Object Classes Challenge: A Retrospective. *Int. J. Comput. Vis.* **2015**, *111*, 98–136.10.1007/s11263-014-0733-5. [CrossRef]
62. ONS. Office for National Statistics: Census Geography. 2019. Available online: https://www.ons.gov.uk/methodology/geography/ukgeographies/censusgeography (accessed on 11 June 2019).
63. Bates, D.; Mächler, M.; Bolker, B.M.; Walker, S.C. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **2015**, *67*. [CrossRef]