Mapping the urban forest at street level

Philip Stubbings, Joe Peskett
Data Science Campus - Office for National Statistics

September 2018

Abstract

Urban trees provide a wide range of environmental, social and economic benefits [29], such as improving air quality and are known to be associated with lower crime levels and greater community cohesion. In collaboration with ONS Natural Capital [40], we have developed an experimental method for estimating the density of trees and vegetation present at 10 metre intervals along the road network for 112 major towns and cities [30] in England and Wales.

Our approach uses images sampled from Google Street View as input to an image segmentation algorithm as to derive a percentage vegetation density map for the road network of an entire city. The developed system is built on recent advancements in the field of deep learning for semantic image segmentation.

This article reports on the effectiveness of our approach for deriving a city- wide geospatial vegetation indicator, starting with the robustness of our initial attempts at identifying green vegetation in arbitrary scenes, through to evaluating models of increasing complexity and finally the use and validation of deep image segmentation neural networks for visual scene understanding.

Contents

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, Tom Smith, Amelia Jones and Gareth Pryce for providing feedback on this article.

Introduction

Urban trees provide numerous social, environmental and economic benefits. In a recent study produced by the Centre for Ecology and Hydrology for the Office for National Statistics (ONS) Natural Capital Accounts [35], the UK's trees were estimated to remove 1.4 million tonnes of air pollutants in a single year, resulting in an annual saving of £1 billion in avoided health damage costs [43]. In another study, London's 8.42 million trees have been estimated [32] to remove 2,241 tonnes of pollution per year, which in addition to other services, is estimated to provide £132.7 million in annual benefits.

Various tree valuation methods [34] have been devised to consider the social and economic benefits trees provide to a community, the environment and economy as a whole. The objective of these methods is to attempt to derive a value that goes beyond replacement cost so that trees are considered as assets rather than liabilities. Although tree valuation methods may be crude and vary in the type and number of benefits they attempt to quantify, recognising the positive impact of trees is important for policy-making and urban planning.

Before attempting to quantify the benefits of street trees in urban areas, it is of course necessary to understand exactly where trees and vegetation exist. When considering street trees alone, one would need to consider the entire road network, which is a daunting task. Our project attempts to solve this problem by making use of automated tree detection procedure coupled with street-level image data.

The result of this work is a consistent methodology that can be used to augment existing tree valuation approaches, with the main benefit being the capability to assess urban vegetation from a remote location. In addition, our approach may be used in combination with more established remote sensing or earth observation techniques such as the use of satellite image data.

To estimate the benefits of trees in an area, it is first necessary to build an inventory, which can then be used for geospatial analysis. Building a tree and vegetation inventory can be achieved in several ways, including traditional surveying and community-based crowdsourcing through to the use of satellite data to build wide-area vegetation indices and local-area automated tree crown detection, which may also include the use of aerial photography.

We are specifically interested in a scalable, automated and consistent methodology, which can be used for the generation of a geospatial vegetation dataset. Furthermore, the methodology should be robust to the seasonality of trees, their species-specific characteristics and the features of the urban environments in which they grow.

In collaboration with the ONS Natural Capital team, we set out to explore the question: *Can a national urban vegetation index be generated using computer vision and machine learning techniques?*

This project has been developed in three phases, starting with the development of an image processing pipeline, further development and improvements to vegetation detection methods and finally an evaluation of the developed method, which in turn comprises of three sub-studies designed to evaluate the approach in different settings. This article focuses on the evaluation of our methods.

When starting this project, we considered several different approaches. One of our main requirements is to be able to generate an urban vegetation dataset in an automated, scalable and consistent way. Therefore, an obvious avenue of research would likely include some form of satellite image processing, from which we would focus on developing object detection methods to identify individual trees and image segmentation methods for large areas of urban woodland. However, there already exist (commercial) tree datasets, which have been derived from aerial sources. Most notably, Bluesky's National Tree Map (tm) [18] provides a detailed national tree survey and has been used as the basis for numerous studies to date.

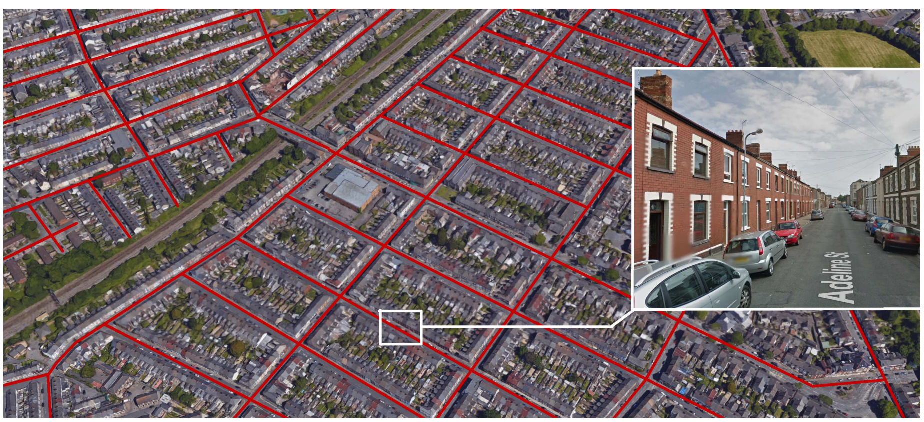

Figure 1: A high density residential area comprised of predominantly non-publicly accessible (or visible) vegetation. Images copyright Google.

Instead, we decided to focus on the detection of amenity (street) trees and vegetation from the point of view of a pedestrian. Our aim is to account for urban vegetation that is (visibly) accessible from ground level, excluding trees obscured from view in private areas. Furthermore, we have focused on vegetation that surrounds the urban road network since road traffic is a major source of air pollution.

Based on this direction, we implemented a proof of concept using images obtained from Google Street View API [16], Open street map road network data [17] and a simple colour thresholding method to identify areas of green in each street-view image. This initial proof of concept was designed to demonstrate our idea, which we later discovered to be similar to research behind the MIT Treepedia project [41, 10, 11]. This similarity is purely coincidental: our choice of Google Street View imagery, road network data and initial image thresholding method was primarily selected with rapid-prototyping in mind, based on existing practical experience of Google's APIs, open street map data and prior-experience of vegetation detection methods in the horticulture domain.

To assess the performance and feasibility, three separate studies have been conducted, from which the aim is to demonstrate the effectiveness of the vegetation detection algorithm in different contexts.

In the first study, the performance of several different models has been evaluated in the context of a pixel-wise image segmentation task.

Evaluating the ability of the model to classify pixels as belonging to vegetation or not in the context of a pixel-wise image segmentation task will demonstrate the performance of the approach as a binary classifier. The intention with this part of the evaluation will be to compare and optimise alternative models.

The second and third studies in this article evaluate the performance of the selected model in a geospatial context, covering the model's higher-level ability to estimate and rank LSOA regions by vegetation density and finally the model's ability to quantify the presence of vegetation and trees at street level.

Evaluation 1: Street-scene image segmentation

Given as input an arbitrary street-level image, the objective of the model is to determine the quantity of vegetation present in the scene. Specifically, the task is to determine the number of pixels belonging to the vegetation class. A summary of visible vegetation density can then be defined as the ratio of pixels belonging to the vegetation class.

In light of the current research interest in autonomous cars, there are a number of street-level image segmentation benchmark datasets. In particular, the Cityscapes [01], ADE20K [02], Mapillary Vistas [12] and most recently, Apolloscape [13] datasets are of particular interest, since they provide high-quality ground-truth labels for urban scenes, covering a range of categories including trees and vegetation, cars, people, buildings and so on. For this study, the Mapillary Vistas dataset has been selected to evaluate model classification performance.

Figure 2: An example training instance from the Mapillary Vistas dataset [12] overlaid with tree, sidewalk and car labels. Image copyright Mapillary.

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 1 of 100 possible categories describing the various components present in urban scenes. The dataset has been selected for use due to its high level of quality derived from a two-stage quality assurance (QA) process [12].

In the context of our research, the primary objective is to identify the vegetation class, of which the Mapillary dataset contains a diverse range of instances covering numerous species obtained in each of the four seasons.

In this study, only Mapillary images with a standard 4 to 3 aspect ratio have been used, resulting in a dataset of approximately 10,000 street-level images.

First approach: Green pixel L*a*b* threshold

Our first approach at pixel-wise vegetation classification is based on the observation that vegetation tends to be green (at least in the spring and summer) and is based on findings from the plant phenotyping domain [03], in which it is possible to segment plant leaves according to shade of green.

The input (RGB) image is first converted to L*a*b* [36] colour space. By projecting an image to L*a*b* space, it is possible to separate colour into a three-dimensional plane, in which L* corresponds to the luminosity or lightness of the image, a* corresponds to green to red, and finally, b* corresponds to blue to yellow.

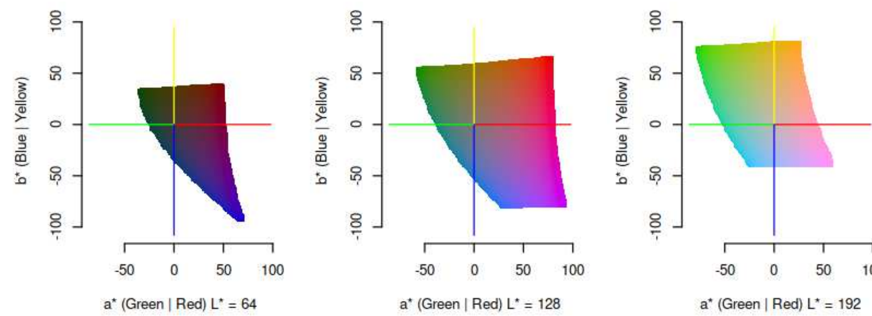

Figure 3: RGB to L*a*b* colour space luminosity intersections. The colour space is linearly separable and stable with respect to changes in lightness (L*)

Figure 3 illustrates different cross sections from the L*a*b* colour space from three levels of increasing lightness. Any region in the top-left quadrant ($a* < 0 < b*$) is considered to be green. Note that the shade of green will remain relatively static with respect to the lighting level. As such, it should be possible to identify "greenness" in a way that is robust to varying lighting conditions.

By restricting the a* parameter to lie within a threshold, $A_1 \leqslant a \leqslant A_2$ it is possible 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 [03], researchers have reported varying threshold parameters, which can be used for leaf segmentation in lighting-controlled images of plants. In the plant phenotyping domain [03], a threshold of $-25 \leqslant a \leqslant -15$ is reported as an effective range for extracting vegetation foreground in images of tobacco plants.

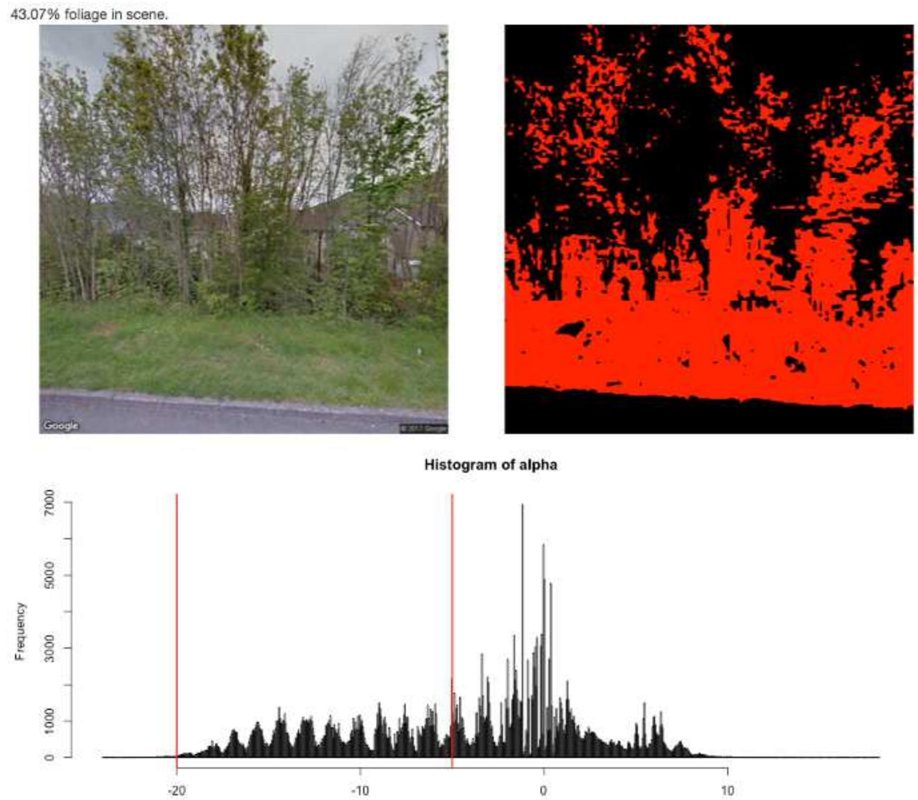

Figure 4: Filtering green pixels from a Google Street View image with a* threshold: $-20 \leqslant a \leqslant -15$. Street View image copyright, Google.

Following the same methodology, it is possible to isolate or filter green pixels in street-level imagery. In the absence of ground-truth data, the early proof of concept of our method used a static threshold derived empirically by means of an interactive tool as illustrated in Figure 4.

The Mapillary labelled image dataset allows for the possibility to explore the threshold parameters in greater detail.

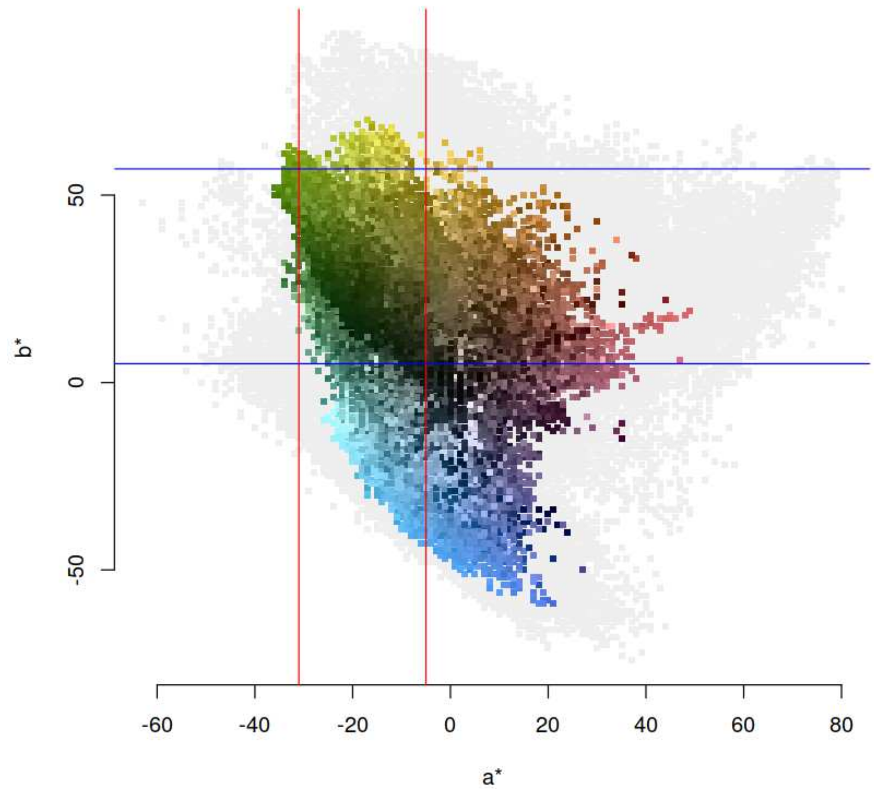

Figure 5: Visualising vegetation pixels (coloured region) and non-vegetation pixels (grey) in the Mapillary dataset. There exists an optimal set of threshold parameters $A_1 \leqslant a \leqslant A_2 \wedge B_1 \leqslant b \leqslant B_2$ with respect to the pixel classification error.

Figure 5 illustrates the range of a* and b* pixel colour values present in the Mapillary dataset. Each point represents an individual pixel from a random subset of the image dataset coded as the actual colour represented by the a*b* pair. Note that this sample includes values with different l* (lightness) and as such, exhibits slight variations. Only pixels belonging to the vegetation class have been retained here, with the remaining classes colour coded as grey to illustrate the overall feature space. The vertical lines correspond to the optimal A_1 (left), A_2 (right) and B_1 (bottom), B_2 (top) parameters.

Of particular interest is the range of colour belonging to the vegetation class.

While most pixels are green (not shown here), there is a large number of non-green pixels. For example, yellow or red pixels may be attributed to seasonality and tree species, brown may be attributed to autumn and tree trunk, while blue may be attributed to sky behind leaves and trunk captured in less fine-grained segment labelling.

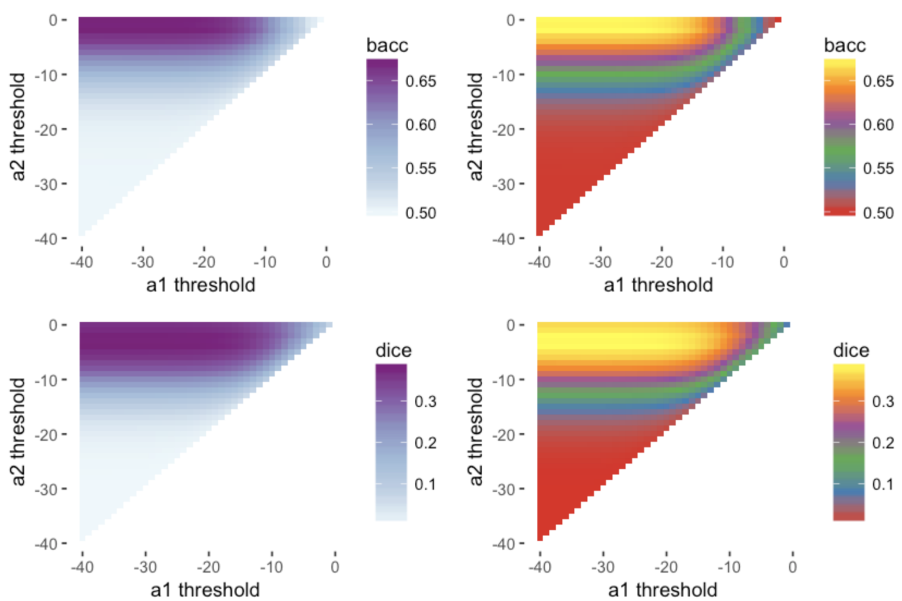

Figure 6: Balanced accuracy surface (top), Sørensen–Dice score (bottom) for fixed $B_1 \leqslant b \leqslant B_2$ and variable $A_1 \leqslant a \leqslant A_2$

Altering the A_1 and A_2 thresholds has an effect on the accuracy of the model. Specifically, increasing A_2 will capture more red pixels, potentially increasing the false positive rate, whilst decreasing A_2 will increase the false negative rate. Therefore, there is an opportunity to optimise the threshold values to maximise classifier performance.

Figure 6 shows the resulting gradient expressed as class balanced accuracy (top) and Sørensen–Dice coefficient (F1-score) (bottom) from a grid-search of all possible A_1, A_2 parameter pairs where $A_1 \leqslant A_2$. The right-hand side images show the same result in another colour scheme as to emphasise the gradient. B_1 and B_2 (blue to yellow) have been clamped to `0` and `12` respectively.

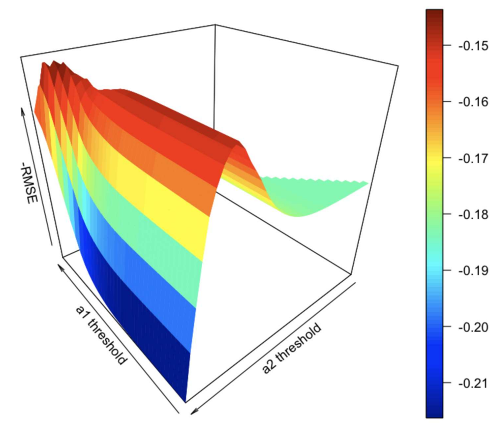

Figure 7: $A_1 \leqslant a \leqslant A_2$ inverted RMSE surface: error increases after $A_2 > 0$

It is clear from the classification accuracy surface in Figure 6 that there is an optimal point before A_1 approaches 0. Altering the A_2 threshold has little effect, until it approaches 0 and the threshold region narrows. This can be seen with greater clarity when observing the (inverted) RMSE error surface from the same grid search as shown in Figure 7. The performance of the model rapidly decreases with $A_2 > 0$. In fact, the optimal A_1, A_2 parameters in this context are `-31, -6` respectively, which is close to the range reported for detecting tobacco plants in plant phenotyping domain [03].

In addition to the A_1, A_2 (green to red) thresholds, the model can be extended by including B_1, B_2 (blue to yellow) thresholds, such that it is possible to exclude turquoise blue and green.

We have made use of Bayesian parameter optimisation [19], using the Matthews Correlation Coefficient [04] (MCC) as an objective function as to find an optimal set of A_1, A_2, B_1, B_2 parameters with respect to pixel-by-pixel classification error. Each set of parameters enumerated by the optimisation method have been evaluated using the mean MCC validation score having performed two-fold cross validation over the training data. Our resulting optimal L*a*b* threshold model used to derive percentage vegetation for a single image is as follows:

$$\frac{\sum_{n=1}^{N} vegetation_n}{N}$$

where N = Number of pixels in image, and:

$$vegetation_n = \begin{cases} 1, & \text{if } -31 \leqslant a \leqslant -6 \land 5 \leqslant b \leqslant 57 \\ 0, & \text{otherwise} \end{cases}$$

This model considers both a* and b* values for each pixel, resulting in a class balanced accuracy of 62% over the ground-truth data. Using the a* alone, our best model ($A_1 = -31$ $A_2 = -11$), found from an exhaustive grid search, could only reach 55% accuracy. Performance of all models will be summarised later in this article.

Second approach: Random forest vegetation mask

The a*b* threshold method can be generalised to the problem of locating a binary mask such that a pixel is classified as vegetation if its a*b* values are contained within the mask region. In the case of the a*b* method discussed previously, this region is rectangular in nature. Therefore, it is possible that the a*b* model can be further improved by allowing for a greater degree of expressiveness: the optimal mask may be elliptic, which appears to be the case with respect to the density of positive (vegetation) examples in the L*a*b* colour space.

Given as input the two a*b* features, a random forest model has been trained to classify pixels into vegetation and non-vegetation classes. A random forest has been selected primarily due to the minimum number of hyper-parameters, of which the number of estimators and estimator maximum depth have been selected using Bayesian parameter optimisation with the MCC objective function yielding a minimal model with just 11 estimators restricted to a depth of 14.

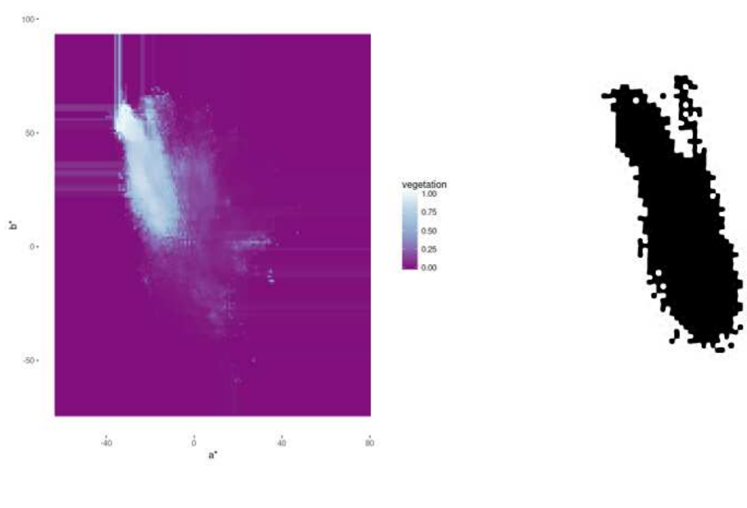

Figure 8: Generation of an elliptic L*a*b* bitmap mask by thresholding random-forest class probability

Having trained the model, it is then possible to enumerate all possible `(a*,` b*) combinations to produce a matrix containing the model's confidence a pixel would belong to the vegetation class with respect to its a* b* feature. Figure 8 shows a visualisation of the class confidence matrix for all a* b* combinations and a bitmap decision mask which has been derived by applying a class confidence threshold. Note that the model has been fit to an elliptic (as opposed to rectangular) region of the colour-space. Figure 9 illustrates the same confidence matrix, emphasising confidence around the greener regions of the colour space.

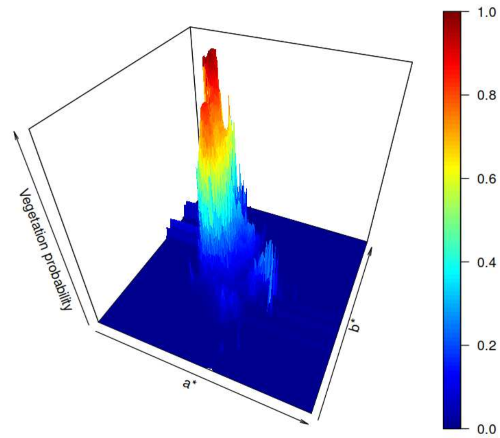

Figure 9: Random forest: confidence a pixel belongs to the vegetation class given it's a* b* value

To extract an elliptic-shaped binary mask, the class confidence of the model has been clipped above a specific threshold of 0.32, which has been found by enumerating all possible (discretised) threshold values within [0, 1] whilst looking to maximise the mask's MCC score over the ground-truth image data. The maximum MCC score located with this method was 0.26 which approximates the point at which precision equals recall. A higher threshold value of 0.32 has been selected in favour of recall and a higher R^2 with respect to predicted compared with actual percentage vegetation over all images in the ground-truth data.

Both the random forest a* b* mask and linear threshold method described so far are superior to thresholding pixels using the a* channel alone. However, despite our best efforts, we have not been able to significantly improve on the a* b* threshold method beyond a slight increase in R^2 with respect to predicted compared with ground-truth percentage vegetation. Having experimented and optimised a number of different approaches, at this point we consider the predictive power of the L*a*b* *feature* in isolation to be exhausted.

So far, three different models have been developed by thresholding the L*a*b* colour space. First by the a* channel, second both a* and b* channels and finally a non-linear a*b* mask-based model generated with a random forest.

The use of a random forest to generate a mask is computationally efficient and decoupled from the machine learning method. The bitmap mask created from the thresholding method described previously can later be loaded as a boolean matrix from which future instances can be classified. Thus, the random forest model may be discarded. Outside of the scope of this project, the bitmap mask approach may be of use when an approximate, high-performance classification scheme is required.

Whilst computationally efficient, the method is restricted to the very limited information provided by the L*a*b* colour-space features. Specifically, the method may be better suited to more controlled conditions such as in the plant phenotyping domain in which the visual complexity of an urban environment would be absent.

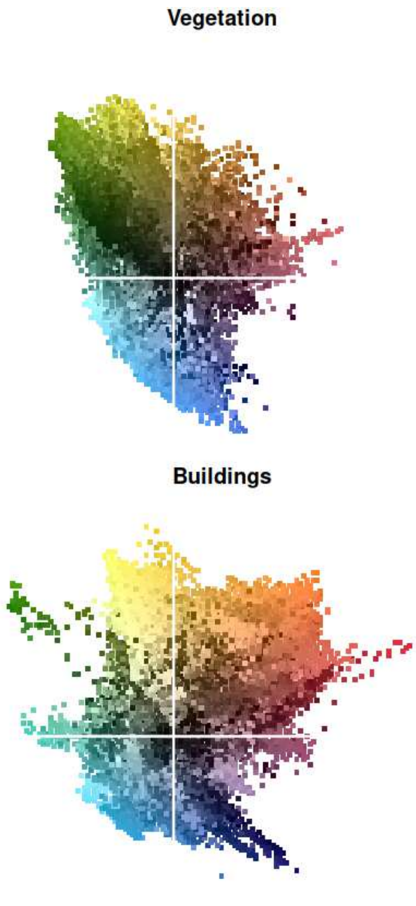

Figure 10: L*a*b* colour space for four classes in the Mapillary dataset

Figure 10 illustrates the main drawback of the method with respect to scene complexity. Each plot represents the colour-space distribution for four specific classes in the Mapillary dataset, namely: vegetation, cars, buildings and sky. Each of the images have been divided up into four quadrants, delimited by `a* = 0` (vertical) and `b* = 0` (horizontal). Although each class exhibits a somewhat different colour-space distribution (lots of red cars, blue sky), there is a high degree of overlap with respect to the `a* b*` co-ordinates.

In other words, it is not possible to differentiate between green cars and green trees by using the L*a*b* features in isolation: the models developed using only these features, whilst exhibiting a strong correlation with respect to predicted compared with actual vegetation exhibit high false positive rates.

Third approach: Deep image segmentation

Taking a leap forward, the current state-of-the-art in semantic image segmentation has been due to recent advancements in deep learning and convolutional neural networks. Of particular relevance to this work, recent research in the specialised domain of street-level image segmentation has resulted in the development of a number of sophisticated models including SegNet [05], PSPNet [06] and DeepLabV3 [07].

We have opted to use PSPNet [06] - Pyramid Scene Parsing Network - a current state-of-the-art image segmentation network, to segment street-level images into a number of different classes, including cars, buildings, sky, people, vegetation and so on. Specifically, our project has made use of a Chainer [08, 25] implementation [20] of PSPNet using the pre-trained Cityscapes [01] and ADE20K [02] weights from the author's original Caffe implementation [21], which came first place in the ImageNet scene parsing challenge 2016 [09]. We have chosen to use PSPNet due to its high performance in existing street-level image segmentation tasks.

Using the Mapillary Vistas dataset [12], The performance of the PSPNet models pre-trained on both Cityscapes and ADE20K have been evaluated. Both ADE20K and Cityscapes models contain a vegetation class (amongst many others). As such, only the model's ability to identify vegetation has been evaluated here.

For each image, we evaluate the performance of the PSPNet using a number of metrics that assess the model's ability to classify pixels as vegetation or non-vegetation. For comparative purposes, we also include the performance of our early L*a*b*-based prototypes.

Table 1: Progressive improvement of our three models and later evaluation of the PSPNet pre-trained models. BACC = Balanced accuracy, Pre/Rec = Precision or recall, F1 = Sørensen–Dice coefficient, MCC = Matthews correlation coefficient, τ = Kendall's tau.

Model	BACC	Pre	Rec	F1	MCC	R^2	τ
PSPNet (city)	**90%**	0.66	**0.87**	0.75	0.72	**0.83**	**0.77**
PSPNet (ade20k)	85%	**0.82**	0.73	**0.77**	**0.74**	**0.83**	0.76
Random forest	62%	0.48	0.29	0.36	0.31	0.25	0.32
Lab (a* b*)	62%	0.47	0.28	0.35	0.29	0.20	0.28
Lab (a*)	55%	0.33	0.14	0.19	0.15	0.04	0.15

We found that (as expected) the vegetation classification performance for the PSPNet model pre-trained on the Cityscapes and ADE20K data is vastly superior to the three L*a*b* threshold methods.

It should be noted that the L*a*b* threshold methods have been optimised and evaluated against approximately 10,000 images in the data. That is, the (L*a*b*) models have been fit perfectly to the data, and likely have been overfitted.

On the other hand, the PSPNet models have been trained on *completely different* datasets (Cityscapes and Ade20k respectively). Subsequently, the results of the PSPNet presented here are indicative of the expected performance on the Google Street View imagery used in our production deployment.

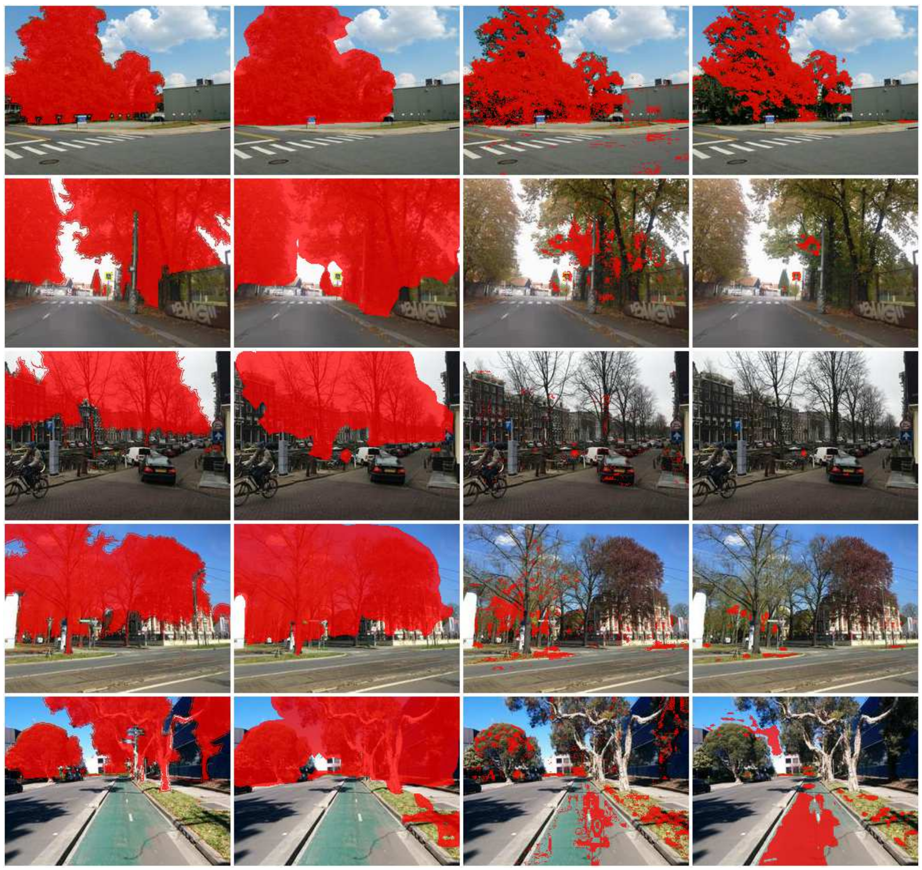

Figure 11: Comparison of vegetation segmentation methods, from left to right: ground-truth (Mapillary data), PSPNet, Random Forest, a* threshold. Images copyright Mapillary.

To further illustrate these results, Figure 11 depicts the segmentation labels produced by the different methods. The first row consists of an easily identifiable group of trees, which have been labelled nearly perfectly by the PSPNet model and with some success by the random forest and a* threshold methods. The second and third rows illustrate one of the main issues with the early L*a*b* methods: Green vegetation will be less prevalent in autumn and winter months.

The forth and fifth rows highlight another limitation of the L*a*b* method: not all tree species are green and not all green objects are trees. The PSPNet model is robust to all of these situations.

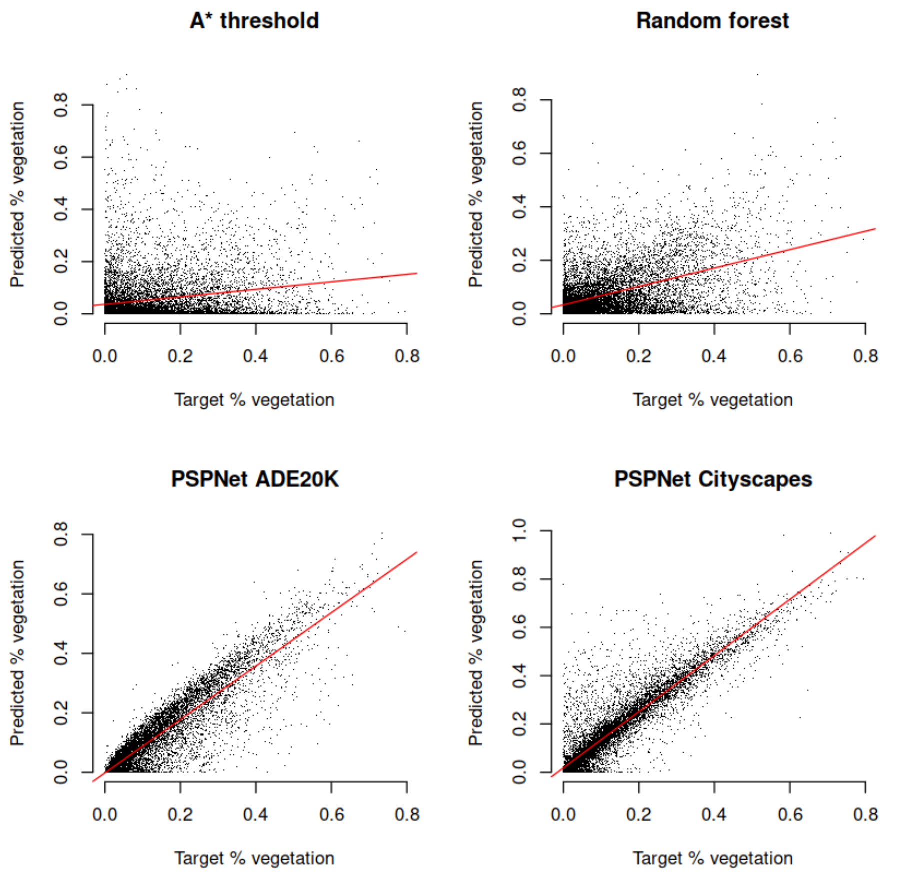

Figure 12: Actual compared with predicted percentage vegetation: our random forest model achieves $R^2 = 0.25$ vs $R^2 = 0.83$ compared with both PSPNet models

The objective of the method described here is to predict the percentage vegetation present in an image. Figure 12 illustrates the relationship between the *actual* percentage vegetation compared with *predicted* percentage vegetation over approximately 10,000 images in the dataset. The difference in performance, in terms of R^2 is quite remarkable given that the PSPNet models have been trained to classify multiple object classes besides vegetation. So, given the impressive PSPNet results, our early work focusing exclusively on L*a*b* thresholds is rendered obsolete.

In light of this, the remainder of this article and our final prototype system have made use of the PSPNet pre-trained on the Cityscapes dataset, which has shown to produce a class-balanced accuracy of `90%` and `0.77` Kendall's tau with respect to actual compared with predicted percentage vegetation.

To summarise, we have so far described a vegetation classification model that maps an arbitrary image to a single percentage vegetation index. The percentage vegetation is defined as the ratio of pixels in the scene classified as vegetation. Our early attempts at this problem relied on two features. Namely, the a* and b* components of the L*a*b* colour space from which we created 3 models of increasing performance based on the a* (green or red) feature, a* *and* b* (yellow or blue) features, and finally, a random forest model, again using the a* b* features.

Having reached the limits of the various models' performance by means of hyper-parameter optimisation, we turned our attention to more sophisticated detection methods by considering the task at hand as a generic image segmentation problem. Having explored various options, we decided to trial an implementation of PSPNet using highly-relevant pre-trained weights from a *separate* street-level segmentation task. We found that the PSPNet model completely overshadowed our L*a*b* space prototype in terms of classifier performance and as such, have selected the PSPNet as the core component in our overall prototype. In the sections that follow, we make use of percentage vegetation derived by the (pre-trained) PSPNet model.

Evaluation 2: Comparison with Cardiff LSOA percentage canopy cover from Natural Resources Wales

Having evaluated the model's effectiveness at identifying vegetation in street-level images, we now attempt to compare the overall methodology with an existing approach.

In 2006, 2011 and 2013, Natural Resources Wales (herein referred to as NRW) conducted the world's first nationwide urban tree mapping survey, covering 220 urban areas in Wales [26]. The survey has been conducted in three phases using aerial photography captured from a survey plane, from which individual trees have been identified using a "desk-based analysis". The study has focused on reporting tree crown cover for three tree sizes, categorised as small (3-6 metres), medium (6-12 metres) and large (12 metres or more), in a variety of contexts including, but not limited to, green open space, transport corridors, commercial areas and woodland. Woodland data from the existing National Forest Inventory

[14] has been used to augment the dataset, which we have obtained from the NRW and Welsh Government Lle geo-portal [15].

In addition to a nationwide report, an auxiliary study was published [27] detailing Cardiff's urban canopy cover at Lower layer Super Output Area (LSOA) [28] level from which the report also explores the relationship between green space and the Welsh Index of Multiple Deprivation (WIMD) [37].

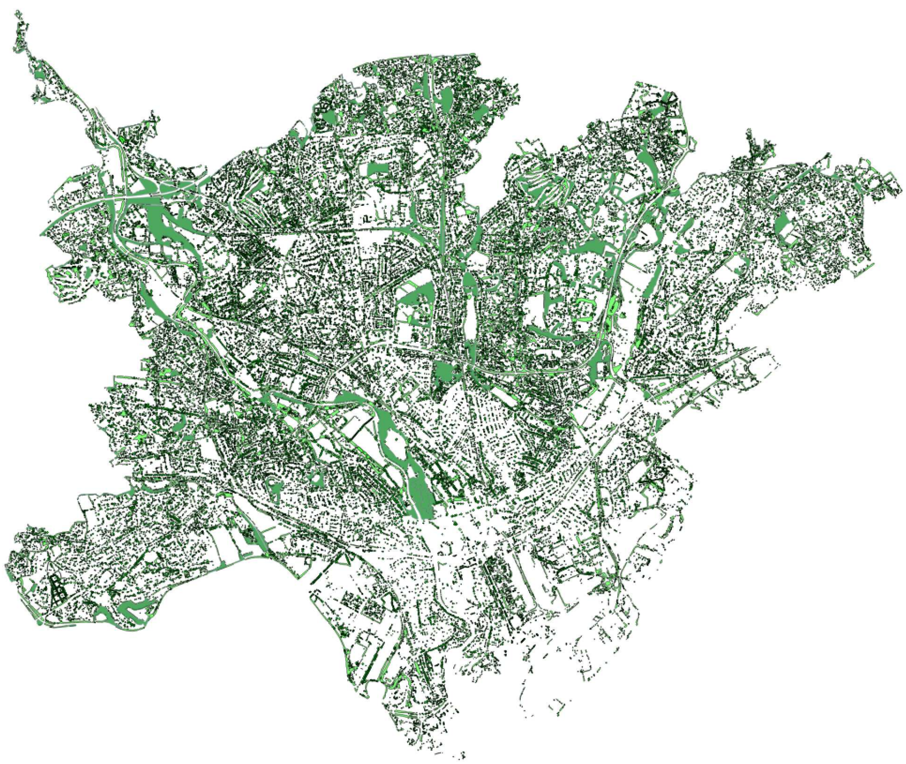

Figure 13: Cardiff's urban woodland and trees - visualisation of the Natural Resources Wales dataset

In preparation for this evaluation, the data obtained from the 2013 NRW study have been pre-processed using QGIS [22]. Specifically, groups and locations of trees from the Cardiff urban area have been extracted and merged with data from the National Forest Inventory.

Figure 14: Individual trees, groups of trees and areas of woodland from the NRW dataset. Satellite imagery copyright Google

The dataset consists of polygons representing small groups of *amenity* trees and larger areas of urban woodland from the NFI. In the original published data, individual trees have been represented as points. For the purpose of this study, these points have been expanded to circular polygons, where the circumference of each circle corresponds to the average canopy size (4.5 metres, 9 metres, 1 metres respectively) as defined in the Town Tree Cover in the City of Cardiff report from NRW.

Figure 15: In addition to locations of individual trees, the NRW dataset includes crown diameter

The resulting NRW dataset represents the current (and only, to our knowledge) detailed inventory of urban trees in Cardiff. This geospatial data, along with the tabular data provided in the original report, have been used here for comparative purposes. However it should be noted that while the data obtained from NRW is extensive (as illustrated in Figures 13, 14 and 15) and detailed, it contains a number of false negatives and positives. This is likely due to the tree labelling process and aerial image quality, which the report indicates was not sufficient to detect tree canopy less than 3 metres in circumference, which is more likely the case for trees growing along transport corridors (TRN in the report). Furthermore, the report indicates that a ground-truth study to assess the accuracy of the small, medium and large tree classification was not conducted. Nonetheless, the data are extensive and represent the current state-of-the-art.

The Town Tree Cover in the City and County of Cardiff study [27] from NRW, reports the percentage of tree crown cover for each LSOA in Cardiff in 2006, 2011 and 2013 respectively. The data from the 2013 study have been used here to illustrate the relationship between our street-view vegetation index and LSOA canopy cover.

To conduct this study, 220,068 street-view images have been sampled from the left-hand and right-hand side of the road at 10-metre intervals along the entire Cardiff road network. We then extract the percentage vegetation from each image using the methodology described previously to create a high-resolution urban vegetation dataset. Using this street-level dataset, we then derive an LSOA percentage vegetation index using the mean percentage vegetation present for all images in each LSOA polygon.

It is important to note that the NRW study reports a percentage *canopy* cover for an entire LSOA. Specifically, the ratio of combined tree crown area compared with non-tree covered area for each LSOA polygon. In contrast, the LSOA vegetation percentage produced by the methodology reported in this study describes the observed vegetation at street level, as opposed to an aerial view. As such, whilst the variables under study are different, we expect to observe a relationship between the two variables: we assume that both variables (aerial canopy cover and street-level vegetation) are both a proxy indicator for biomass.

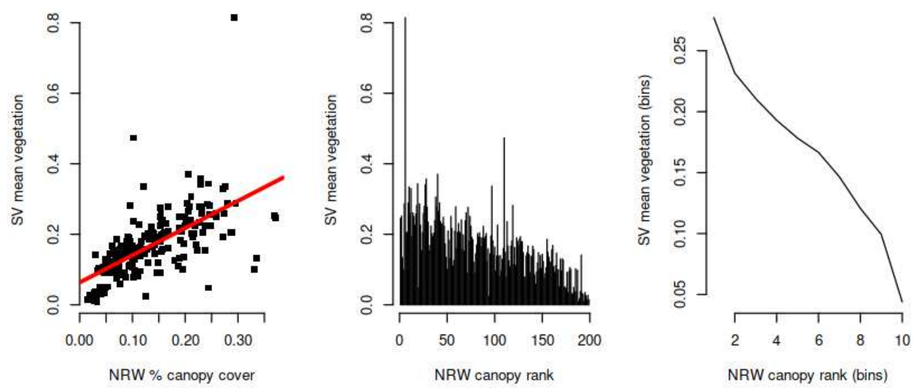

Figure 16: NRW percentage Lower layer Super Output Area (LSOA) canopy cover compared with Mean percentage LSOA street-view vegetation. Rˆ2 = 0.41.

At aggregated LSOA level, our street-level vegetation index is comparable with the results presented in the NRW study. For each LSOA, there exists a linear relationship between the percentage canopy cover reported in the NRW study and our mean percentage street-level vegetation.

Figure 16 illustrates this relationship. From left to right, the first plot shows percentage canopy and percentage street-level vegetation for each LSOA in Cardiff. A clear relationship exists (with $R^2 = 0.41$): higher reported levels of canopy cover in the NRW report coincide with higher levels of vegetation found at street level with our approach. The second plot further demonstrates this relationship by showing our percentage LSOA vegetation (y-axis) relative to the order of percentage LSOA canopy (x-axis). This demonstrates that as percentage canopy decreases, percentage vegetation at street level also decreases. The third graph emphasises this observation by aggregating, and thus smoothing, the percentage LSOA canopy and percentage vegetation into bins, each containing mean vegetation for 10 LSOAs.

Figure 17: Natural Resources Wales Lower layer Super Output Area percentage canopy cover (left) compared with street-view LSOA average percentage vegetation (right)

The same results can be visualised in the form of an LSOA hexagon [23] heatmap. Figure 17 shows the density of Cardiff LSOA canopy cover from the NRW data on the left, compared with the density of street-level vegetation from our methodology on the right. Both approaches result in similar *ordering* of LSOA tree density.

Aggregating further from LSOA to ward level, both approaches yield similar results. Table 2 lists Cardiff wards, ordered by mean street-level vegetation.

Table 2: List of all Cardiff wards, ordered by our mean street-level vegetation (SV mean) compared with percentage canopy cover from the NRW data (NRW mean). The third row, `Diff`, simply shows the difference between the two approaches, highlighting similar and outlying wards.

	SV mean	NRW mean	Diff
Pentyrch	0.47	0.10	0.37
Lisvane	0.32	0.26	0.06
Pontprennau Old St Mellons	0.30	0.19	0.12
Cyncoed	0.27	0.24	0.03
Whitchurch and Tongwynlais	0.24	0.16	0.08
Radyr	0.24	0.20	0.05

	SV mean	NRW mean	Diff
Llandaff	0.23	0.20	0.03
Creigiau St Fagans	0.22	0.15	0.07
Pentwyn	0.22	0.21	0.01
Penylan	0.21	0.17	0.04
Rhiwbina	0.21	0.15	0.06
Llandaff North	0.20	0.12	0.08
Llanishen	0.20	0.19	0.01
Fairwater	0.19	0.14	0.05
Trowbridge	0.19	0.12	0.06
Rumney	0.17	0.10	0.07
Heath	0.16	0.12	0.05
Ely	0.15	0.10	0.05
Llanrumney	0.13	0.12	0.01
Caerau	0.12	0.14	-0.03
Butetown	0.11	0.05	0.06
Gabalfa	0.11	0.11	0.00
Riverside	0.10	0.10	-0.01
Canton	0.09	0.10	-0.01
Plasnewydd	0.09	0.08	0.02
Grangetown	0.09	0.06	0.03
Splott	0.09	0.08	0.00
Cathays	0.08	0.11	-0.03
Adamsdown	0.07	0.05	0.02

Whilst it has been shown that both approaches yield correlated results in terms of LSOA vegetation ranking, it should be noted that the two approaches are measuring different, although similar, quantities and that our street-level methodology is designed to capture *only* vegetation visible from the *field of view of a pedestrian*. As such, there are a number of outliers in the results presented in Table 2.

Interestingly, our approach has produced two significant outliers where our reported percentage vegetation is much greater than the reported percentage canopy cover in the NRW study. Specifically, two LSOAs, (codes `W01001893` and `W01001820`), belonging to Tongwnlais and Pentyrch wards respectively, have been over-reported in comparison with the NRW study. This can be explained by the presence of long stretches of road within the two wards that pass through high-density road-side woodland. Images obtained along these roads consist of predominantly more than 99% detected vegetation.

On the other hand, there are some instances where our method under-reports the amount of LSOA vegetation compared with the NRW study. The under-reporting in one LSOA (`W01001922`), belonging to the Cathays ward, can be explained by the prevalence of terraced housing: at street-level there is little vegetation whereas the area contains numerous trees in back-gardens.

In summary, the mean LSOA percentage vegetation derived from our street-level methodology has been compared with the LSOA % tree crown cover from the most recent NRW survey which is a high-quality, thorough inventory of the study area. We have sought to demonstrate that while the two methods differ in their approach, they are strongly related. It should be noted that the NRW study *does not* represent a *ground truth* dataset. We use it here specifically to demonstrate the validity of our method *at LSOA level* and to highlight the main characteristic of our approach as a method for sampling vegetation density along transit corridors *from the point of view of a pedestrian*.

Evaluation 3: Comparison with street-level Cardiff tree inventory data from NRW

The final evaluation in this article explores the relationship between predicted vegetation at street level and canopy cover reported in the Natural Resources Wales (NRW) study. In the previous section using the NRW data, we compared Lower layer Super Output Areas (LSOAs) ranked by the predicted vegetation from both methods. Since our method has been designed to quantify the amount of vegetation present at much higher resolution than LSOA level, at every 10 metres along a city's road network, our evaluation can be extended to include a finer-grained comparison.

The ranked LSOA evaluation used tabular data reported in the NRW study. As mentioned previously, the data used for the NRW study have been made publicly available [15] in the form of GeoJSON data, describing the location of individual trees. Using these data, our objective is to reconstruct a dataset of the same dimensions as our street-level derived data. Specifically, when expressed in simplified tabular form, our data contains 110,034 rows of sample points that represent 10-metre intervals over the entire Cardiff road network:

```
latitude, longitude, percent_vegetation
```

Here, the `percent_vegetation` produced at each sample point is the average of the percent vegetation detected on the left-hand and right-hand side of the road.

We wish to demonstrate that for each sample point, our "detected vegetation" coincides with the vegetation present in the NRW data for that specific point. As such, Using the NRW data, the objective is to construct a new dataset for comparative purposes:

```
latitude, longitude, sv_vegetation, nrw_vegetation
```

Where `sv_vegetation` equals the percentage vegetation detected by our method at the `(latitude, longitude)` sample point, and `nrw_vegetation` equals the (estimated) percentage vegetation at the sample point in the NRW data.

A rigorous approach to forming this dataset would involve a view-shed [38] analysis in which for every 10-metre point along the road a view-port is constructed to match the field of view [39] of the associated street-view image. The field of view would exclude areas behind buildings and other objects that would not be visible from the street-level view-port. The intersection of visible tree canopy in the remaining field of view would then be used to derive an percentage visible vegetation.

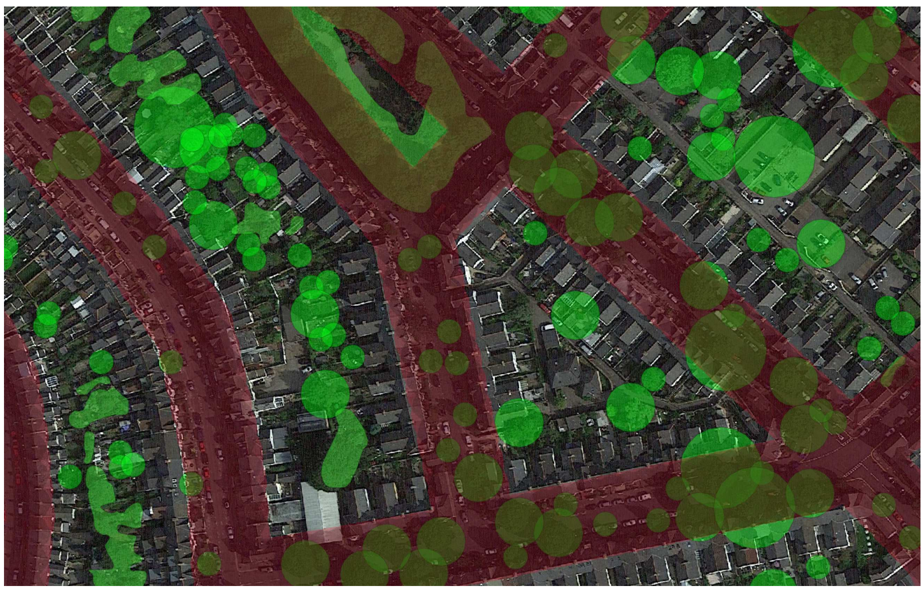

Figure 18: Projected buffer zone (red) designed to include road area and building facades. Intersecting trees (green) are counted as roadside trees

We have not implemented the view-shed approach for this study, but have instead opted for a simplified method. Firstly, we a construct a fixed width buffer around the Cardiff road network, which from observation, tends to extend to the roof tops of buildings along each road. The objective of the buffer is to *exclude* non road-side vegetation such as trees in back gardens and private spaces.

As illustrated in Figure 18, the buffer is composed of a sequence of overlapping fixed radius circles around the 10m sample points along the road. Then, the percentage canopy present in a particular circular buffer zone is defined as the ratio of intersecting tree canopy. Repeating this process for each circular buffer zone, yields the dataset defined above.

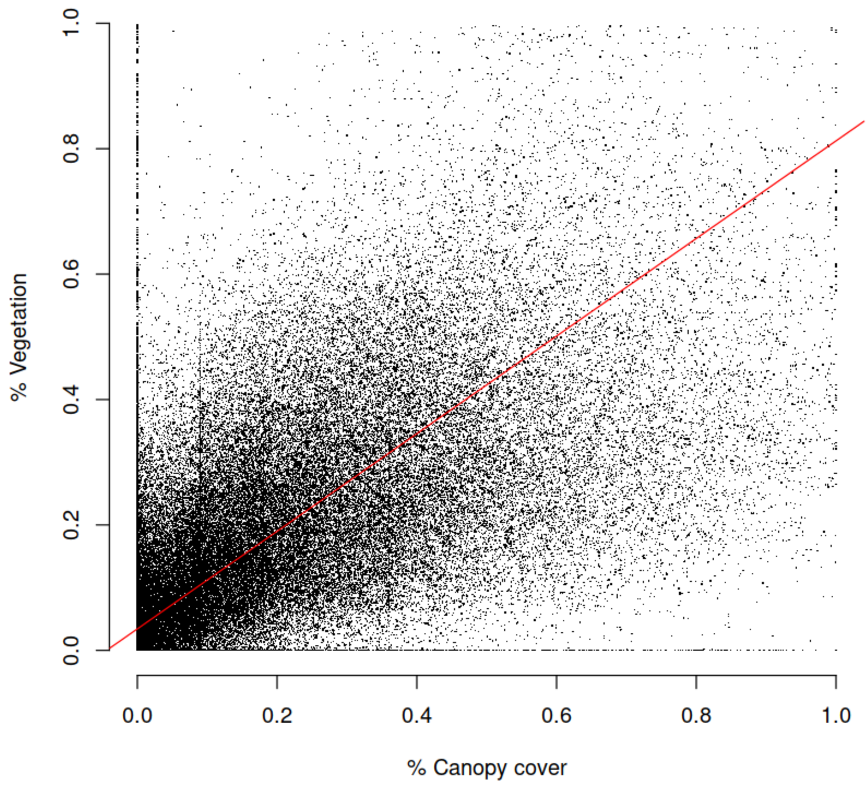

Figure 19: Relationship between street-level vegetation and canopy cover

We expect to demonstrate a relationship between the percentage vegetation derived from our method and the percentage canopy derived using this approximate field of view. While there does indeed exist a linear relationship between the two variables as visible in Figure 19 ($R^2 = 0.39$), there is a high residual standard error (~0.16). This is due to a flaw in the evaluation method, where a tree crown inside a circular sampling buffer will yield the same intersecting area regardless of its position within the buffer. The buffer zones used here could be described as panoramic, whereas the vegetation captured by our model is constrained by the left and right view-port at each sample point.

Summary

In this article, we have considered three different approaches to demonstrate the effectiveness of our approach at identifying street-level vegetation. In the first evaluation, we described, in detail, three different classification techniques of increasing complexity along with comparative performance metrics. The first evaluation was designed to demonstrate the performance of the approach specifically in terms of its ability to identify vegetation in arbitrary images, in a non-geospatial context.

In the second evaluation, we have shown that our approach yields similar results as an existing study when used to rank Lower layer Super Output Areas (LSOAs) by vegetation. Besides demonstrating the validity of our prototype from a geospatial perspective, more generically, the result demonstrates a relationship between the observed density of trees at street-level and overall “greenness” of an area.

We had hoped to demonstrate a higher resolution relationship between the Natural Resources Wales (NRW) study and our own method by attempting to reconstruct a vegetation index by estimating visible tree density at specific points from the NRW data. Whilst we found a relationship, we note that the third evaluation methodology is flawed and would require a detailed view-shed analysis to produce more meaningful results.

Our initial attempts at the problem were based on the assumption that the presence of green in a street-level scene would be a crude, although approximate indication of vegetation. A green pixel thresholding method based on the L*a*b* colour space was then developed as to provide a baseline or minimal viable prototype.

Whilst this technique can work well in controlled environments such as in the plant-phenotyping domain, the reliability of the thresholding technique breaks down in complex urban scenes. We attempted to refine the threshold model by parameter optimisation and later by introducing a non-linear threshold method based on a binary threshold mask derived from a random forest model.

Despite extracting the maximum possible performance from our thresholding method by means of hyperparameter optimisation, our best model came nowhere close to the performance of the (pre-trained) PSPNet model. We reported the results of our early model and initial attempts to improve it as to illustrate the progression of this project.

Although not covered here, a significant component of this work focused data engineering, in which we developed an end-to-end distributed image processing pipeline, API and geospatial backend. During that phase we were faced with a number of technical challenges relating to the scalability of the approach (we would require 80 million images to sample the entire UK road network) and as such, the intention of the L*a*b*-based threshold method served its purpose well as a minimum viable product. Details of our image processing pipeline and associated code have been published on our Github page [42].

Future work

The performance of the PSPNet in terms of its ability to identify vegetation is somewhat remarkable given the fact that the model used here had previously been trained on a *completely different* dataset. Furthermore, we only consider the use of the model as a binary vegetation classifier: the pre-trained model can segment a scene into a number of classes.

The pre-trained model used in this evaluation represents a high-quality benchmark for future work to improve on. Given the already high level of vegetation segmentation accuracy achieved by the model, we hope to focus later iterations of the work on tree *species* identification.

One of the most exciting outcomes of this project has been the creation of a high-resolution dataset describing the observable vegetation density at 10-metre intervals across an entire city. Given the availability and quality of the Natural Resources Wales (NRW) study, we have used the city of Cardiff for comparative purposes. Our generated (Cardiff) dataset includes approximately 220,000 sample points. In addition, since the start of the project we have also sampled Manchester (approximately 330,000 points), Newport and Walsall. Furthermore, we have partially sampled another 108 cities and having deployed our prototype image processing pipeline, our dataset is improving on a daily basis.

In addition to an urban-vegetation dataset, we have used the additional classes predicted by the PSPNet model to build up a database of non-vegetation related classes for each 10-metre point in a city. For example, we are now able to describe, in detail, the visual components of a city in high resolution, including the percentage observed building density, number of cars, bicycles, people, signage, street furniture and various other objects used to describe an urban scene. This is a highly interesting geospatial dataset from which we aim to produce a textual representation of towns and cities. We plan to extend this approach to form a topological description of a city, combining the quantitative information (for example, percentage vegetation) detected at specific locations with abstract descriptions derived from image captioning techniques.

Beyond the production of this dataset, our work is intended to be of use in the urban-analytics domain. The NRW report used for comparative purposes in this article, describes a relationship between urban green-space and levels of deprivation including health, income and presence of crime. There are numerous studies linking green space to various social, environmental and economic indicators. Exploring the relationship between green-space (and other features), *from the point of view of a pedestrian* and other factors such as indicators of well-being offer an exciting direction for future research.

References and additional resources

Papers

1. **The Cityscapes Dataset for Semantic Urban Scene Understanding** M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele. Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016.

2. **Scene Parsing through ADE20K Dataset** B. Zhou, H. Zhao, X. Puig, S. Fidler, A. Barriuso and A. Torralba. Computer Vision and Pattern Recognition (CVPR). 2017.

3. **Leaf segmentation in plant phenotyping: a collation study** Scharr, Hanno and Minervini, Massimo and French, Andrew P. and Klukas, Christian and Kramer, David M. and Liu, Xiaoming and Luengo, Imanol and Pape, Jean-Michel and Polder, Gerrit and Vukadinovic, Danijela and Yin, Xi and Tsaftaris, Sotirios A. Machine Vision and Applications, 27 (4). pp. 585-606. ISSN 1432-1769. 2016.

4. **Comparison of the predicted and observed secondary structure of T4 phage lysozyme** Matthews, B. W. Biochimica et Biophysica Acta (BBA) - Protein Structure. 405 (2): 442–451. 1975.

5. **SegNet: A Deep Convolutional Encoder-Decoder Architecture for Robust Semantic Pixel-Wise Labelling.** Vijay Badrinarayanan, Ankur Handa and Roberto Cipolla. PAMI. 2017.

6. **Pyramid Scene Parsing Network** Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, Jiaya Jia. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017.

7. **Rethinking Atrous Convolution for Semantic Image Segmentation** Liang-Chieh Chen, George Papandreou, Florian Schroff, Hartwig Adam. Google technical report, 2017.

8. **Chainer: a Next-Generation Open Source Framework for Deep Learning** Tokui, S., Oono, K., Hido, S. and Clayton, J. Proceedings of Workshop on Machine Learning Systems(LearningSys) in The Twenty-ninth Annual Conference on Neural Information Processing Systems (NIPS), 2015.

9. **Semantic understanding of scenes through the ADE20K dataset** B. Zhou, H. Zhao, X. Puig, S. Fidler, A. Barriuso, and A. Torralba. MIT technical report, 2016.

10. **Assessing street-level urban greenery using Google Street View and a modified green view index** X Li, C Zhang, W Li, R Ricard, Q Meng, W Zhang Urban Forestry & Urban Greening 14 (3), 675-685, 2015.

11. **Green streets - Quantifying and mapping urban trees with street-level imagery and computer vision** I Seiferling, N Naikc, C Ratti, R Proulx Landscape and Urban Planning 165: 93–101. 2017.

Datasets

12. **Mapillary Vistas Dataset** Diverse training data for segmentation algorithms. Mapillary Research 2017, 2018.
13. **Appoloscape dataset** Advanced open tools and dastasets for autonomous driving. Baidu. 2018.
14. **National Forest Inventory** Forest Research.
15. **Urban Tree Cover 2013 Polygons, points and urban extents** Welsh Government and Natural Resources Wales, Lle Geo-Portal.
16. **Google street view API**
17. **Open street map**
18. **Bluesky - National Tree Map** Bluesky International Limited.

Tools

19. **A Python implementation of global optimization with gaussian processes** Fernando Nogueira & contributors.
20. **PSPNet - chainer implementation** Shunta Saito. 2017.
21. **PSPNet - original Caffe implementation** Hengshuang Zhao. 2017.
22. **QGIS - Open Source Geographic Information System**
23. **geogrid - Turning geospatial polygons into regular or hexagonal grids** R package, Joseph Bailey, 2018.
24. **Open trip planner**
25. **Chainer - deep learning framework**

Reports

26. **Tree Cover in Wales' Towns and Cities** Natural Resources Wales. 2016.
27. **Town Tree Cover in the City and County of Cardiff** Natural Resources Wales. 2016.

28. **An overview of the various geographies used in the production of statistics collected via the UK census** ONS Geography.

29. **The case for trees in development and the urban environment** Forestry Commission England. 2010.

30. **Major Towns and Cities** ONS, Methodological Note and User Guidance. 2016.

31. **Semantic Image Segmentation with DeepLab in TensorFlow** Google research blog. 2018.

32. **Valuing London's urban forest** Treeconomics London. Results of the London i-Tree Eco Project. 2015.

33. **Road Lengths in Great Britain 2016** Department for Transport. 2017.

34. **Street tree valuation systems** Forest Research, 2011.

35. **Developing Estimates for the Valuation of Air Pollution Removal in Ecosystem Accounts** Jones, L., Vieno, M., Morton, D., Cryle, P., Holland, M., Carnell, E., Nemitz, E., Hall, J., Beck, R., Reis, S., Pritchard, N., Hayes, F., Mills, G., Koshy, A., Dickie I. Final report for Office of National Statistics, July 2017.

Links and additional resources

36. **Lab colour space**

37. **Welsh Index of Multiple Deprivation** Welsh Government.

38. **Viewshed analysis**

39. **Field of view**

40. **ONS Natural Capital**

41. **MIT Treepedia** Exploring the Green Canopy in cities around the world. MIT Senseable City Lab.

42. **Data Science Campus - Github projects**

43. **UK air pollution removal: how much pollution does vegetation remove in your area?** July 2018.

Code and data

- Distributed Google Street View image processing pipeline. (Blog post).
- OpenStreetMap road network sampling.