An Open-Source Automatic Survey of Green Roofs in London using Segmentation of Aerial Imagery

Charles H. Simpson, Oscar Brousse, Nahid Mohajeri, Michael Davies, and Clare Heaviside

UCL Institute for Environmental Design and Engineering, The Bartlett School of Environment Energy and Resources, University College London, London, United Kingdom

Correspondence: Charles Simpson (charles.simpson@ucl.ac.uk)

Abstract.

Green roofs can mitigate heat, increase biodiversity, and attenuate storm water, giving some of the benefits of natural vegetation in an urban context where ground space is scarce. To guide the design of more sustainable and climate resilient buildings and neighbourhoods, there is a need to assess the existing status of green roof coverage and explore the potential for future implementation. Therefore, accurate information on the prevalence and characteristics of existing green roofs, but this information is currently lacking. Segmentation algorithms have been used widely to identify buildings and land cover in aerial imagery. Using a machine-learning algorithm based on U-Net to segment aerial imagery, we surveyed the area and coverage of green roofs in London, producing a geospatial dataset (Simpson et al., 2022). We estimate that there was $0.19\ km^2$ of green roof in the Central Activities Zone (CAZ) of London, ($0.81\ km^2$) in Inner London, and ($1.25\ km^2$) in Greater London in the year 2019. This corresponds to 1.6% of the total building footprint area in the CAZ, and 1.0% in Inner London. There is a relatively higher concentration of green roofs in the City of London (the historic financial district), covering 3.1% of the total building footprint area. Test set accuracy was 0.996, with an f-score of 0.757. We improve on previous studies by including more negative examples in the training data, by experimenting with different data augmentation methods, and by requiring coincidence between vector building footprints and green roof patches. The survey covers 1463 km^2 of Greater London, making this the largest open automatic survey of green roofs in any city. This dataset will enable future work exploring the potential of green roofs in London and on urban climate modelling.

1 Introduction

In urban areas, green roofs (i.e. roofs deliberately covered in a growing substrate and living vegetation) can provide some of the benefits of ground-level green-space to health, well-being and the environment. Studies have examined the extent to which green roofs can directly reduce cooling energy demand and the risk of overheating in buildings (e.g. Castleton et al. (2010); Sailor et al. (2012); Sproul et al. (2014); Virk et al. (2015)), or can provide indirect benefits by decreasing the outdoor air temperature in hot weather with mixed results (e.g. Peng and Jim (2013); Virk et al. (2015); Cuthbert et al. (2022)). Green roofs have the potential to provide a range of benefits to humans, and to the wider ecological system in cities by providing habitats for wildlife (Filazzola et al., 2019; Hoeben and Posch, 2021), and can act as a carbon sink (Getter et al., 2009). Furthermore, green roofs may be able to contribute to the removal of air pollutants (Baik et al., 2012), and storm water retention (Mentens et al., 2006). Thus, green roofs are increasingly seen as an opportunity to improve health and well-being in urban environments, and as a part of climate mitigation and adaptation strategy. On the other hand, green roofs impose an structural loads and additional costs, so are not always appropriate (Losken et al., 2018); in other cases solar panels or high-albedo roofs may be more appropriate. Cities worldwide have policies that encourage the use of green roofs through quantitative planning tools (The Ecology Consultancy, 2017).

Previous technical reports commissioned by the Greater London Authority (GLA) have investigated the area of green roofs in London (Table 1). The 2019 Living Roofs and Walls report (hereafter LRW2019), surveyed existing green roofs for the years 2016 and 2017 ((European Federation of Green Roof and Green Wall Associations (EFB) and Livingroofs.org on behalf of the Greater London Authority, 2019; Livingroofs Enterprises Ltd, 2019)), although the methods are not publicly documented. The survey reports estimates for Greater London by local authority district (LAD) and for the Central Activities Zone (CAZ: a central area in London defined for planning purposes, see subsection 2.1 and Figure 1), although the survey methods are not publicly documented. The London Plan Annual Monitoring Report (AMR), another report for the GLA, green roof areas in the CAZ were estimated based on aerial imagery for years 2013, 2015, and 2017 (Greater London Authority, 2021a, Key Performance Indicator 22, page 70) ranging from $1.75 \times 10^5\ m^2$ in 2013 to over $2.9 \times 10^5\ m^2$ in 2015. Lastly, an interactive map of green roofs in the CAZ is publicly available on the GLA website (Greater London Authority, 2014); it was produced in 2013/14 and is consistent with a green roof area in the CAZ of $1.75 \times 10^5\ m^2$. Although these different estimates (Table 1) offer valuable information on recent green roof coverage in London they lack transparency about the methods used, there is a wide disagreement about the area of green roofs in the CAZ, and the full data are not publicly available for analysis.

Accurate, comprehensive, and open data documenting the location and area of green roofs can directly inform research into city-scale heat mitigation strategy and is useful for stakeholders such as urban planners, policy makers and research communities looking at urban heat mitigation and the added value of green spaces. However, there is a general lack of open data documenting the area and coverage of green roofs. In order to address this Wu and Biljecki (2021) applied a machine-learning algorithm to high-resolution satellite-imagery to identify green roofs and solar panels in a number of cities around the world, producing a ranking for which of the surveyed cities have the greatest coverage with green roofs and solar panels. London was not included in their survey.

Table 1. Previous estimates of green roof area in London. CAZ refers to a central area of London, see Subsection 2.1. Data from (European Federation of Green Roof and Green Wall Associations (EFB) and Livingroofs.org on behalf of the Greater London Authority, 2019) and (Greater London Authority, 2021a).

Survey area	Source	Survey year	Green roof area (10^5 m^2)	% of built area
CAZ	London Plan AMR 16	2013	1.75	1.5
CAZ	London Plan AMR 16	2015	2.2	1.9
CAZ	London Plan AMR 16	2017	over 2.9	2.5
CAZ	LRW2019	2016	1.5	1.3
CAZ	LRW2019	2017	2.1	1.8
Greater London	LRW2019	2016	11.0	0.43
Greater London	LRW2019	2017	15.0	0.59

In this study, we identify green roofs from aerial imagery: this is a binary segmentation problem, as a single class needs to be identified from a background. Such algorithms process an image to output a binary mask identifying areas belonging to the target class. We used a fully convolutional neural network known as U-Net to segment the imagery: this type of neural network was originally designed for biomedical image segmentation (Ronneberger et al., 2015), but have since been applied in other research fields including remote sensing: for example to map roads (Ozturk et al., 2020), parking lots (Ng and Hofmann, 2018), and green roofs (Wu and Biljecki, 2021) from imagery.

Green roofs cover only a small proportion of the planar area of London, so in aerial imagery most pixels are not part of a green roof. This means that the classification problem is imbalanced, with the negative class being many times more numerous that the positive class. This can create problems with model training if gradient descent batches often do not contain any positive examples. In Wu and Biljecki (2021), the training polygons were restricted to areas with relatively higher concentrations of green roofs and image tiles with no green roof were excluded; 1-5 km^2 of each of the 17 cities covered. Furthermore, the total number of examples for training is relatively low compared to many computer vision tasks, meaning that a computer vision model may be unable to generalise the appearance of green roofs; as such such, data augmentation is key for achieving good segmentation performance. In the original U-Net paper, elastic deformations are applied to the training images, which makes the network learn to be invariant to these deformations without the need for all possible deformations to be present in the data (Ronneberger et al., 2015), this is justified as soft tissues in medical images are often deformed in this way. In Ng and Hofmann (2018) (on which Wu and Biljecki (2021) is based), random rotations in units of 90^o and horizontal flips were applied to the images, in order enforce rotational independence to the classifier and reduce overtraining.

In this study, we build on the machine-learning based method used by Wu and Biljecki (2021) for the segmentation of green roofs from remote-sensed imagery, improving the segmentation performance by including more negative examples and experimenting with data augmentation methods. We thus provide a robust, open and documented dataset of the location and area of green roofs in London at the level of individual buildings (Simpson et al., 2022), filling a gap in publicly available data. This dataset has the greatest extent of its kind for any single city.

2 Data and Methods

2.1 Geographic Context and Data

Greater London is a region of England with an area of $1,570\ km^2$, which is divided into local authority districts (LADs), which are the 32 boroughs and the City of London. Inner London, with an area of $319\ km^2$ is defined by the Office for National Statistics; it comprises 14 LADs in the centre of London, roughly corresponding to the historic county of London (Office for National Statistics, b). The Central Activities Zone (CAZ) is the historic, governmental, and business centre of London defined by the GLA for planning purposes (Greater London Authority, 2021b, Policy SD4). The CAZ is contained within Inner London but does not align with the LAD boundaries; it intersects with 10 LADs and has an area of $33.5\ km^2$. Lower super output areas (LSOAs) are areas with 1000-3000 residents defined for the purpose of census statistics: each LSOA is within exactly one LAD, and each LAD contains multiple LSOAs. In this article we use the LSOAs defined for the 2011 census (Office for National Statistics, a).

Using the Local Climate Zone typology (Stewart and Oke, 2012; Demuzere et al., 2019) as a reference, the built form of Greater London is mostly classified as open lowrise. Inner London covers most of the area classified as open midrise and compact midrise, but also contains a large amount of open lowrise. The CAZ mainly covers the area of compact midrise in the centre and is therefore the most densely built part of London. Buildings in the CAZ, and especially the City of London, are more likely to be non-residential buildings. Figure 1 show the outlines of the LADs in Greater London and Inner London, and the outline of the CAZ.

Datasets described in this section are summarised in Table 2.

Dataset	Type	Source	Explanation
Imagery	25 cm RGB raster	Getmapping Inc.	Cloud free vertical aerial imagery mosaic.
OS VML	Vector	Ordance Survey	Large scale building outlines.
UKBuildings	Vector	Verisk Ltd.	Building footprints.
LSOA outlines	Vector	Office of National Statistics	Lower-super output areas from 2011 Census.
LAD outlines	Vector	Office of National Statistics	Local authority districts boundaries.

Table 2. Input geospatial dataset summary.

The imagery used for segmentation was colour (red, green, blue) raster images from a cloud-free mosaic of aerial imagery at $25\ cm$ horizontal resolution (from (Getmapping Plc., 2020) accessed under an academic license). The imagery collection dates for the mosaic covering Greater London are shown in Figure 2.

Two GIS datasets were used for building footprints. Ordnance Survey (OS) VectorMap Local (VML) (Ordnance Survey (GB), 2021) building footprints dated April 2019 were used in post-processing the segmentation, as inspection showed that outlines were more consistent with the aerial imagery, especially in cases of buildings with internal courtyards. UKBuild-

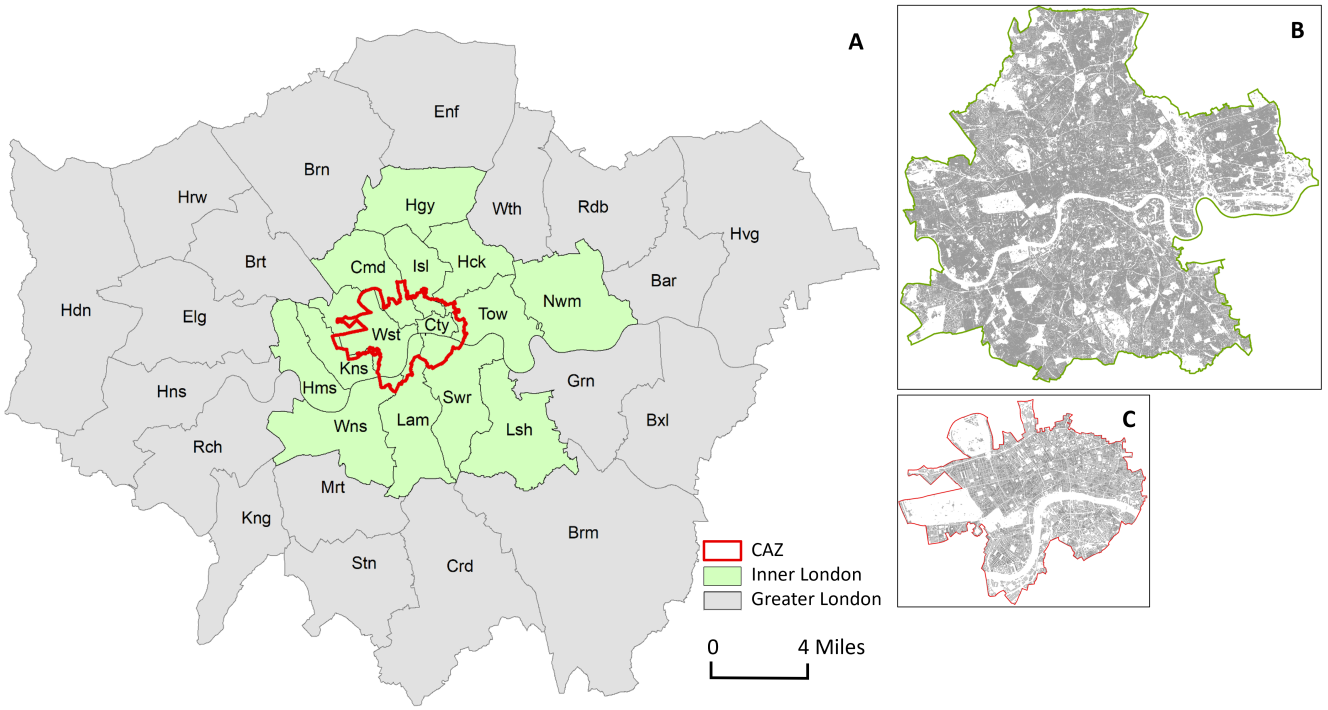

Figure 1. (A) Local authority districts in Greater London, identified by first three consonants of their names. Inner London is highlighted in green and the outline of the CAZ is shown in red. Zoomed in maps of (B) Inner London, and (C) the CAZ.

ings (Verisk Analytics, Inc., 2022) building footprints were used for building counts, as it divides buildings into individual properties.

2.2 Segmentation pipeline

Our segmentation pipeline was based on that of Wu and Biljecki (2021), which is in turn based on Ng and Hofmann (2018). The key differences are as follows:

1. we used aerial imagery rather than satellite imagery,
2. our hand-labelled areas are distributed around the city, rather than concentrated in a central area,
3. we focussed on fully surveying a single city rather than trying to cover many,
4. we experimented with additional data augmentation methods,
5. we implemented early stopping rather than training for a fixed number of epochs,
6. we performed a hyperparameter gridsearch,

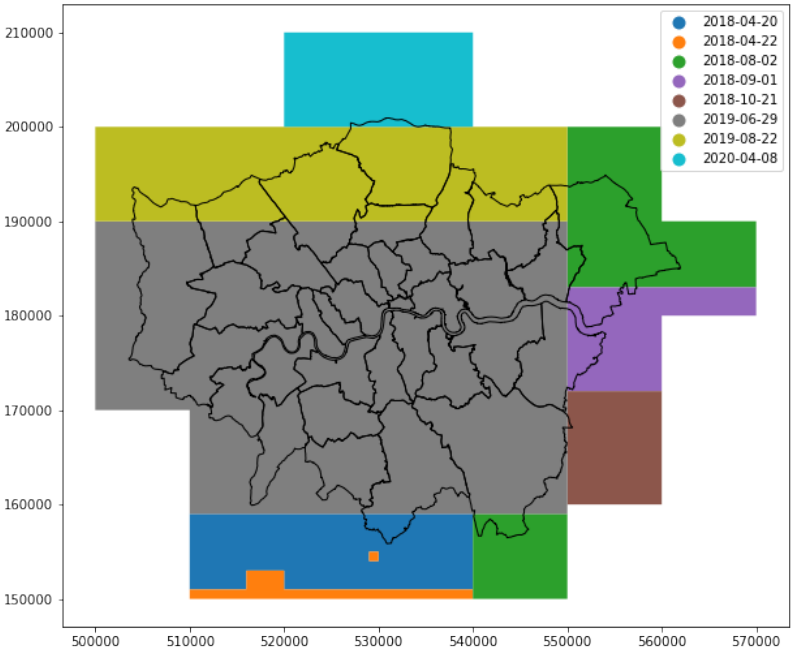

Figure 2. Collection dates for the aerial imagery mosaic covering Greater London.

7. we did not use morphological opening or closing,

8. we used building footprints provided by Ordnance Survey rather than OpenStreetMap for postprocessing,

9. we included tiles containing no positive examples.

All analysis and data management was performed using Python (Van Rossum and Drake, 2009). A general outline of the workflow is shown in Figure 3. The method is covered in more detail in the following subsections.

2.3 Imagery and labelling

To identify the locations of green roofs and estimate their covered area, we trained our U-Net with training polygons from a sample area. The encoder layers of the U-Net produce compressed abstract representations of the image at different scales, by repeatedly using convolution blocks followed by maxpool downsampling. The decoder layers apply upsampling and concate-

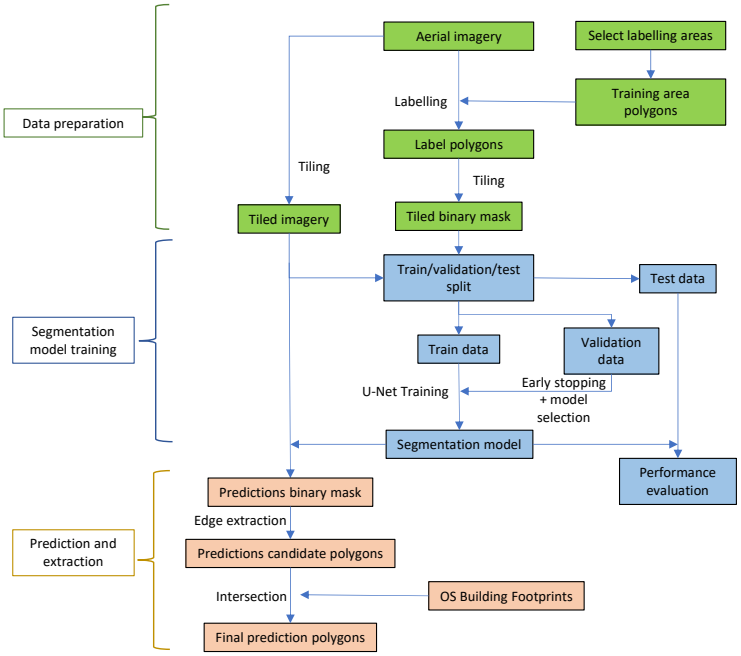

Figure 3. Workflow diagram of the overall segmentation pipeline.

nation with convolution to produce a prediction with the same dimensions as the input image, combining information from the different scales provided by each encoder layer. The relationship between the image and the classification is learned from a set of labelled examples, hereafter referred to as training polygons. To produce training data, green roofs in the imagery were labelled by hand to provide input for model training. The training label polygons and geospatial results are included in the Supplementary Material to this article for reproducibility. We selected areas for labelling based on the OS 1 km grid reference system, so each grid square is 1 km^2. Firstly, a 4 km^2 area in the CAZ was selected, known to have a relatively higher concentration of green roofs: this was to ensure that there is sufficient representation of green roofs in the data. Secondly, to increase the diversity of the data, we selected a further 21 km^2 distributed around Inner London without prior knowledge of the concentration of green roofs, aiming to represent each LAD and a variety of built forms (based on an LCZ map); these areas had much smaller amounts of green roofs. All grid references that were included are listed in Table A2 and mapped in Figure 4. Within the selected grid squares, every building in the imagery was inspected and green roofs were labelled by hand. Labelling was performed by drawing polygons using QGIS (QGIS Association, 2022); some examples of training polygons are shown in Figure 5. In total, sample areas covered 7.8% of Inner London, resulting in 4.9×10^4 m^2 of green roofs labelled inside the CAZ, and 2.2×10^4 m^2 outside the CAZ.

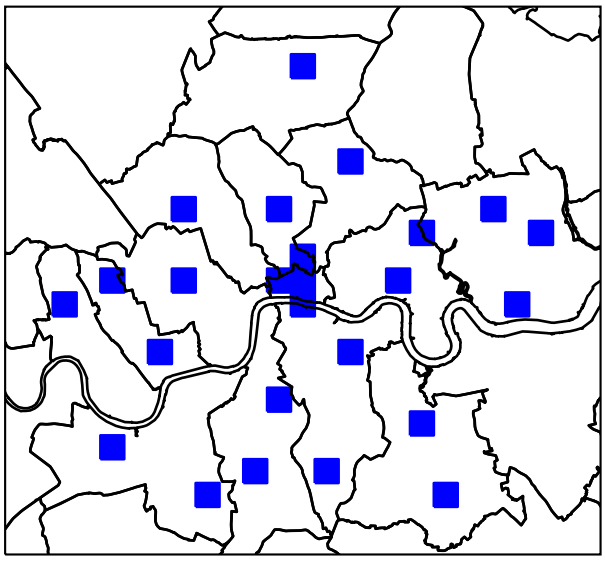

Figure 4. Map of hand labelled areas.

Once trained we applied the U-Net to a larger area (the whole of Greater London) to map existing green roofs.

2.4 Perfomance metrics

Standard metrics were calculated to assess the validity of the segmentation model. Metrics were calculated from the final vector layers, after all processing steps. The metrics are listed in Table 3. Accuracy, IoU, precision, recall, and F-score all range from 0 to 1, where 1 represents an ideal classifier. F-score is a more appropriate measure of the overall validity of a model for imbalanced classification than accuracy. As well as calculating these metrics, we examined examples of poor segmentation performance to understand the failure modes of our segmentation method.

Figure 5. Example of training polygons. The area outlined in green was manually identified as being a green roof in this aerial imagery. Imagery ©GetMapping Plc. Image location is shown in Figure A4.

Table 3. Classification performance metrics calculated in this study.

Metric	Definition
Accuracy	Proportion of area correctly classified
True positive (TP)	Area correctly classified as positive.
False positive (FP)	Area incorrectly classified as positive.
True negative (TN)	Area correctly classified as negative.
False negative (FN)	Area incorrectly classified as negative.
Intersection over union (IoU)	TP/(TP+FP+FN)
Precision	TP/(TP+FP)
Recall	TP/(TP+FN)
F-score	harmonic mean of precision and recall

2.5 Segmentation algorithm

Transfer learning refers to the practice of transferring models or parts of models between different learning tasks - in this case from a well-known image classification task to our segmentation task. Ng and Hofmann (Ng and Hofmann, 2018) used transfer learning to mitigate the small number of training examples; the U-Net encoder is replaced with a ResNet50 trained on the ImageNet dataset (He et al., 2016; Deng et al., 2009). Transfer learning can improve performance and reduce the required training resources as the model will have already learned to extract features from images that are generally informative. We initialised the U-Net encoder with a ResNet-50 model pre-trained on the ImageNet dataset (He et al., 2016).

The imagery was broken into 256×256 pixel tiles at a scale of 0.0005^o to one tile, approximately $47.5\ m$ per tile or $19\ cm$ per pixel, using OpenStreetMap's tiling conventions. Pixel values were normalized to match ImageNet during training and prediction.

We refer to areas labelled with no green roof as negative, and those labelled with green roof as positive. All tiles within the hand-labelled areas were used. In order ensure that batches would contain positive examples, we over-sampled positive tiles by repetition during training so that they were equally prevalent as the fully negative tiles. Tiles were split randomly into training (80%), validation (10%), and testing sets (10%). The random split was performed separately for positive and fully negative tiles to ensure all splits contained both classes. For the purpose of this split a tile was positive if any pixel within it was positive, and negative otherwise. The separate validation and testing datasets are required because hyperparameter tuning is performed by selecting the hyperparameters that maximise the validation dataset performance, so a testing dataset is required to properly estimate out-of-sample performance.

The algorithm was implemented in PyTorch (Paszke et al., 2019). The model was trained using the Adam optimiser (Kingma and Ba, 2014), an optimiser that dynamically adjusts learning rates for each model parameter, making training less dependent on the global learning rate and therefore reducing required training resources.

Rather than training the model for a fixed number of epochs, we implemented early stopping. Early stopping refers to stopping training when validation performance ceases to improve. This reduces the required training resources and can effective at reducing overfitting. Training was stopped if the mean validation loss in the past five epochs was greater than that of the five epochs before that.

Hyperparameter tuning experiments were performed via grid search, and the final selection made based on the validation data. Testing data were not used for training method tuning, and were only processed after the hyperparameters were finalised.

Cross-entropy, Lovasz, and Focal loss functions were tested: Lovasz is intended as a surrogate for the intersection-over-union measure (Berman et al., 2018), whereas focal loss is intended to give greater weight to hard-to-classify examples during training (Lin et al., 2017). The loss functions were weighted by the frequency of the classes to account for class imbalance. Learning rate, loss function, and data augmentation methods were tested. The hyperparameters tuned, and hyperparameter values used for the final classification, are listed in Table A1.

A key part of the U-Net methodology is data augmentation - a process wherein distortions or transformations are applied to the training data to increase robustness when training data is scarce. Augmentation can reduce overfitting, a process wherein a model memorizes certain features of the training dataset that do not generalise out-of-sample (Shorten and Khoshgoftaar, 2019). During training, augmentations were applied to the imagery tiles, and correspondingly to the label masks. We randomly flipped images in both planes, and also applied random 90° rotations, and found that this improved performance. There was no further improvement in performance from applying fully random rotations. We experimented with applying a random 90% crop to the images, with randomly manipulating the colours of the imagery, and randomly adjusting the sharpness of the image. Augmentation was applied randomly and independently each training epoch, with equal probability to positive and negative tiles.

2.6 Post-processing

Predicted segmentation masks were generated from the trained model using the same tiling method as used for training. The same prediction probability threshold of 0.5 was applied across the whole domain.

In previous work, morphological opening and closing have been used on the classification masks as a post-processing step: these are filters that remove small isolated positive areas and fill in small negative areas respectively. This can be useful for filling in small gaps resulting from the tiling of the input imagery. We tested these methods with our own models and imagery, but found that morphological opening of the classification masks increased recall but decreased precision, overall decreasing F-score; whereas, morphological closing did not have any substantial effect on F-score. Therefore, we decided not to include these post-processing steps in our final classification pipeline.

From the binary masks produced by the segmentation algorithm, we extracted green roof candidate polygons. Single pixel-wide gaps are visible in some of the candidate polygons as a result of the tiling of the input images and not using morphological closing. The intersection was then taken between the candidate polygons and the OS VML building footprints, to remove any candidate polygons that did not intersect with a building footprint. This process helped to reduce the false positive rate because the segmentation algorithm can incorrectly identify ground-level green cover as a green roof. The post-processed segmentation

results were spatially joined with the UKBuildings layer in order to identify which individual buildings have green roofs, and so calculate the number of buildings covered.

2.7 Area estimates

To estimate area of green roof in each geographic area, the polygons of green roof area identified by the segmentation are spatially overlayed with the polygons of the geographic area. A similar process is used with the building footprints to estimate building footprint area. All area calculations were applied in the OSGB36 / EPSG:27700 coordinate projection.

Area projections are scaled up by the recall of the model, based on the assumption that a fixed proportion of each green roof is missed by the model. Not doing so would lead to an underestimation of the green roof area.

3 Results

3.1 Segmentation performance

The confusion matrix and performance statistics for green roof identification are given in Tables 4 and 5 for calculations based on area. Tables 6 and 7 give the same statistics but calculated based on building counts.

Results of the hyperparameter search are shown in Table A1. The best performance was achieved by the combination of random flips, random 90° rotations, and randomly reducing the sharpness of the image by a sharpness factor 0.5. Only small differences were observed between the performance of different loss functions. The best performing augmentation had a higher f-score by 0.146 than the base case.

We found that the building-intersection step made only a small difference to the performance statistics of the validation and testing data; however, areas with no green roof are under-represented in these datasets. We found that across the whole study area, 20% of predicted green roof area was outside of building footprints and was consequently removed, showing that across the building-intersection step plays an important role in suppressing false positives.

Table 4. Confusion matrix for the green roof identification method, calculated based on area. TP, TN, FP, FN are as a proportion of total building footprint area in the hand-labelled areas.

	Land area (km^2)	Built area (km^2)	TP	TN	FP	FN
Training	21.9	6.533	0.0024	0.9969	0.0004	0.0003
Validation	2.7	0.840	0.0013	0.9973	0.0006	0.0008
Testing	2.7	0.845	0.0014	0.9967	0.0003	0.0015

Table 5. Performance metrics for the green roof identification method, calculated based on area.

	Accuracy	IoU	Precision	Recall	F-score
Training	0.9994	0.7939	0.8699	0.9008	0.8851
Validation	0.9986	0.4857	0.6847	0.6256	0.6538
Testing	0.9981	0.4353	0.8137	0.4835	0.6066

Table 6. Confusion matrix for the green roof identification method, calculated based on counts of buildings. TP, TN, FP, FN are as a proportion of total building footprint area.

	Building count	TP	TN	FP	FN
Training	50396	0.0060	0.9914	0.0021	0.0005
Validation	8279	0.0062	0.9878	0.0046	0.0014
Testing	8444	0.0066	0.9891	0.0033	0.0009

3.2 Distribution of green roofs

Table 8 gives estimates for LADs in Inner London, Table 9 for Outer London, and Table 10 for aggregated areas. *Proportion of total building footprint area* means the total green roof area divided by the total building footprint area including all buildings not only those with green roofs. *Proportion of area* means the total green roof area divided by the area of the geography (LAD, CAZ, or Inner London). *Proportion of buildings* means the count of buildings with any green roof divided by the count of all buildings. *Proportion of buildings by area* means the building footprint area of buildings that have any green roof divided by the total building footprint area. *Mean coverage* means the total area of green roof divided by the total footprint area of buildings that have any green roof.

We estimate that the CAZ contained $19.2 \times 10^4\ m^2$ of green roof on the dates of imagery collection (Summer 2019). Green roof area estimates for each LAD in Greater London, and LSOA in Inner London, are mapped in Figures 6 and 7 respectively. Most (58%) LSOAs contain no green roofs, and the maximum proportion of building footprint area covered by green roofs in any LSOA is 33.8%.

Table 7. Performance metrics for the green roof identification method, calculated based on building counts.

	Accuracy	IoU	Precision	Recall	F-score
Training	0.997	0.704	0.743	0.930	0.826
Validation	0.994	0.505	0.573	0.810	0.671
Testing	0.996	0.609	0.667	0.875	0.757

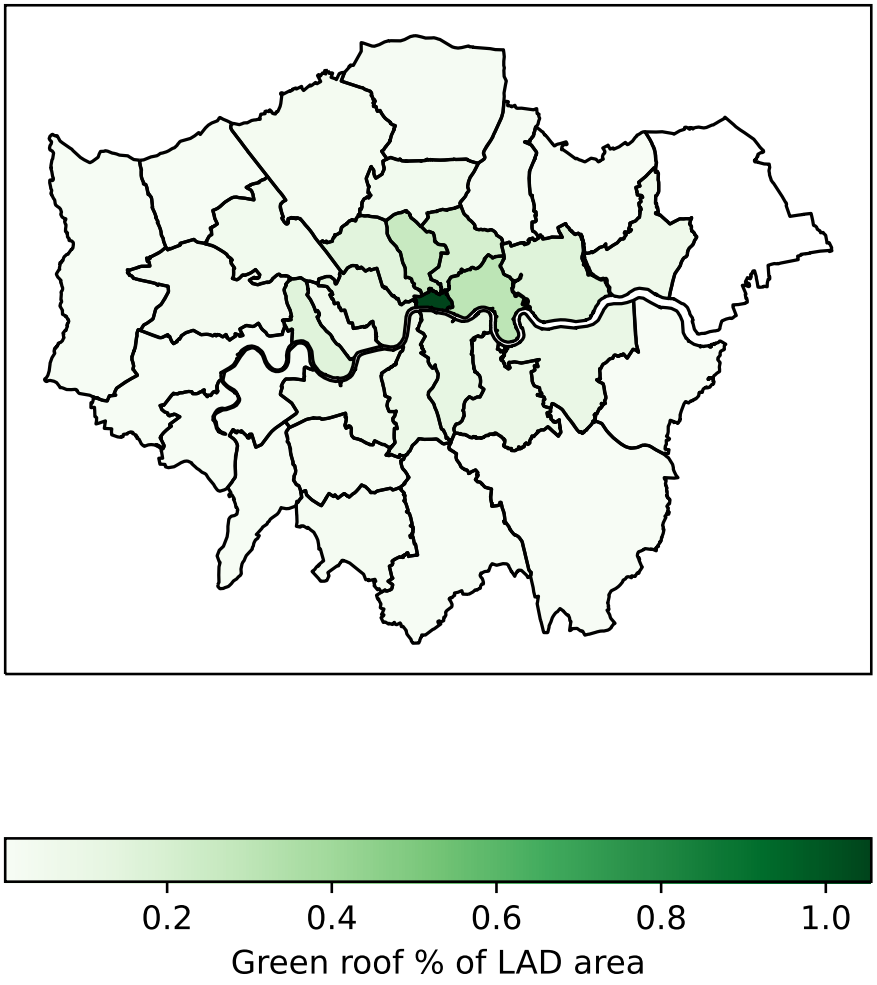

Figure 6. Area of green roof identified in LADs as a fraction of total LAD.

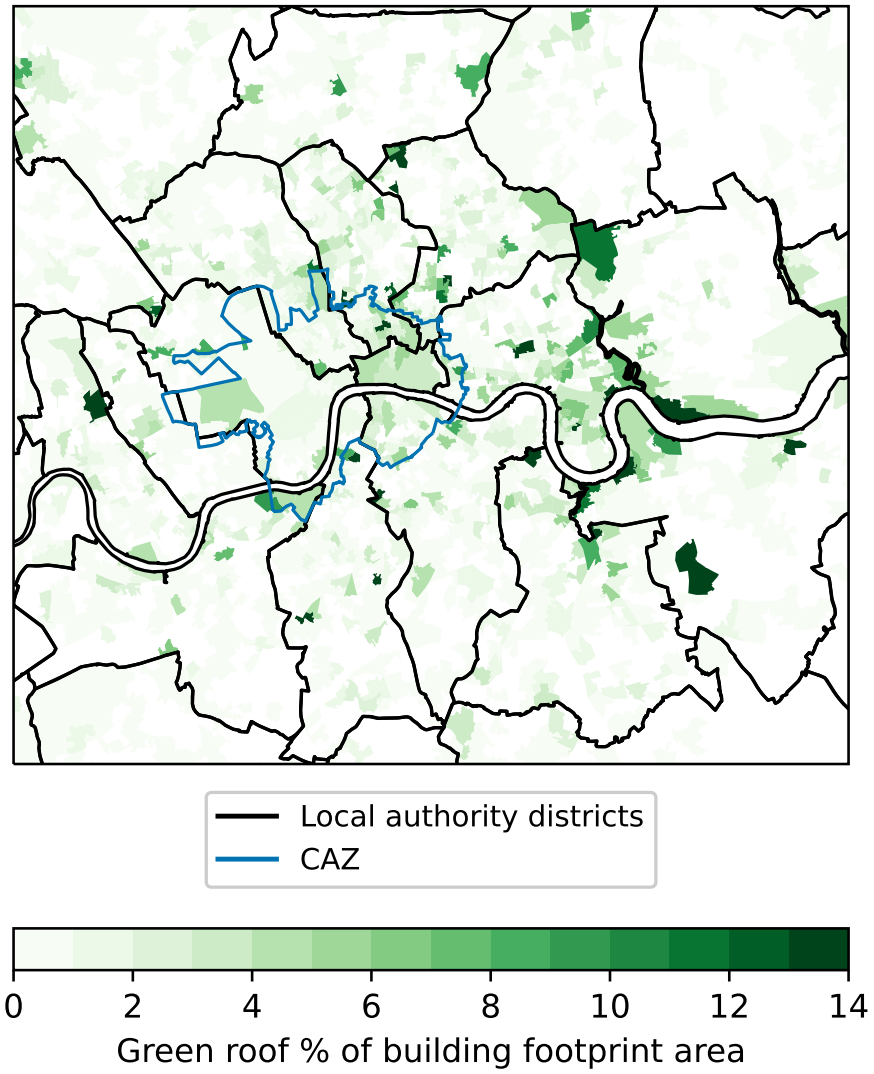

Figure 7. Area of green roof identified in LSOAs as a fraction of total building footprint area.

Table 8. Table of estimated green roof area for each LAD in Inner London.

	Green roof area (10^4 m^2)	Proportion of total building footprint area (%)	Proportion of geographic area (%)	Proportion of buildings (%)	Proportion of buildings by area (%)	Mean coverage (%)
City of London	5.0	3.2	1.7	5.8	21.0	15.2
Tower Hamlets	11.0	2.2	0.6	1.0	8.2	26.6
Islington	7.2	1.6	0.5	0.8	5.6	28.3
Hackney	7.4	1.5	0.4	0.7	5.2	29.4
Newham	10.9	1.4	0.3	0.2	4.7	29.4
Hammersmith and Fulham	4.9	1.0	0.3	0.2	5.7	18.2
Camden	5.4	1.0	0.2	0.7	5.8	16.5
Southwark	6.0	0.9	0.2	0.4	4.1	22.0
Lewisham	5.8	0.8	0.2	0.2	2.8	29.6
Westminster	4.6	0.7	0.2	0.9	5.4	12.4
Lambeth	4.1	0.6	0.2	0.3	3.4	17.3
Wandsworth	3.9	0.5	0.1	0.2	2.4	20.5
Haringey	2.9	0.4	0.1	0.1	1.2	36.9
Kensington and Chelsea	1.2	0.3	0.1	0.3	1.8	18.7

4 Discussion

4.1 Segmentation Performance

Overall performance of the segmentation model achieves a high level of accuracy (99.8%). The area precision 81% and count precision 88% are high (Tables 5 and 7) meaning that we can be confident that identified green roofs are real. Counting buildings rather than measuring areas increases the number of positives (both false positives and true positives) as an example is counted as positive if any part of the building is identified as positive: this leads to higher recall and lower precision in Table 7 compared to 5.

Given that the survey covers such a large and diverse area, and the green roof fraction is low in many areas, it is important to consider the false positive rate. Tables 4 and 6 show that we expect just 0.08% of buildings to be incorrectly identified as having a green roof, or 0.02% of the built area.

Area recall is somewhat low at 0.48, whereas the count recall is higher at 0.875 Inspection of false negatives in the results showed that many pixels classified as false positives and false negative occur at the edges of labelled green roofs. This indicates that the dataset is good at identifying whether a building has a green roof, but tends to underestimate the area.

Table 9. Table of estimated green roof area for each LAD in Outer London.

	Green roof area ($10^4\ m^2$)	Proportion of total building footprint area (%)	Proportion of geographic area (%)	Proportion of buildings (%)	Proportion of buildings by area (%)	Mean coverage (%)
Greenwich	9.3	1.2	0.2	0.2	4.1	29.4
Barking and Dagenham	4.4	0.7	0.1	0.2	1.3	51.4
Barnet	5.3	0.4	0.1	0.1	1.9	23.4
Brent	3.8	0.4	0.1	0.1	2.1	19.1
Sutton	2.1	0.3	0.0	0.1	1.5	20.0
Ealing	3.2	0.3	0.1	0.1	1.0	28.1
Enfield	2.6	0.2	0.0	0.1	1.6	14.9
Richmond upon Thames	1.6	0.2	0.0	0.1	1.0	22.7
Merton	1.4	0.2	0.0	0.0	0.8	27.4
Waltham Forest	1.4	0.2	0.0	0.0	1.0	19.9
Hillingdon	2.4	0.2	0.0	0.0	2.0	9.3
Croydon	2.3	0.2	0.0	0.1	1.2	15.3
Harrow	1.4	0.2	0.0	0.0	0.6	28.4
Hounslow	1.4	0.2	0.0	0.1	1.2	12.9
Kingston upon Thames	0.9	0.1	0.0	0.0	1.1	12.8
Bromley	1.2	0.1	0.0	0.0	0.6	14.2
Bexley	0.5	0.1	0.0	0.0	0.3	17.5
Redbridge	0.3	0.0	0.0	0.0	0.4	9.9
Havering	0.0	-	-	-	-	-

Table 10. Table of estimated green roof area for the CAZ, Inner London, and Greater London.

	Green roof area ($10^4\ m^2$)	Proportion of total building footprint area (%)	Proportion of geographic area (%)	Proportion of buildings (%)	Proportion of buildings by area (%)	Mean coverage (%)
CAZ	19.2	1.7	0.6	2.0	11.2	15.1
Inner London	80.2	1.0	0.3	0.4	4.5	22.2
Greater London	125.6	0.5	0.1	0.1	2.3	21.6

There is a substantial difference in area recall between the training and testing datasets (Table 5) which indicates over-training; suggesting that due to the high diversity in the appearance of green roofs the training dataset and model struggles to generalise the appearance of green roofs to unseen areas. However, area precision is consistently high. Together, this suggests that the training data contained a good quantity of negative examples, but would benefit from greater diversity in the positive examples. This could be improved by labelling more data thus increasing the size of the training dataset. In our experimentation with augmentation methods, we found that augmentation substantially decreased the difference in F-score between training and validation datasets. It may be possible to reduce the performance difference using additional methods of data augmentation or model regularization.

The IoU score for counts of buildings (0.43, Table 7) of our segmentation model is higher than that reported in Wu and Biljecki (2021) (0.396, see their sec 3.2.2). This could be because we used higher quality imagery and focused on a single city, although because the two studies are in different cities it is not a like-for-like comparison. Wu and Biljecki (2021) covered a total of 2217 km^2 across 12 cities, with the largest being 302 km^2 in Las Vegas, Nevada; our survey covered 1463 km^2, making ours the largest survey of green roofs in a single city.

This method can be applied to other cities, and we have explored how the segmentation methods can be improved. We found that including a large number of negative examples and over-sampling positive examples was very effective at suppressing the false positive rate in unseen areas, and would recommend this approach in general. Furthermore, we found small improvements in performance from augmenting the training dataset with random adjustments in sharpness, but this will not necessarily be effective with different imagery.

4.2 Limitations

As is clear from this study, automatic methods are scalable, allowing large areas to be surveyed and monitored; however, they have limitations. Green roofs can only be identified by this method if they are visible in the imagery, and small areas of vegetation (that is, not visible at 25 cm pixel size) are necessarily left out. Hand labelling also has limitations; there are cases where it is difficult to determine visually from the imagery whether a building has a green roof, or where the edge of the green roof is.

The majority of the Greater London region is covered by a single acquisition date (2019-06-29, see Figure 2), and this also contains the majority of the data included in training. Summer collection of optical aerial imagery is preferred because a high solar angle means better light conditions. We found that there was a higher rate of false positives east of OSGB37 Easting 5.5×10^5 m, for which imagery was acquired on 2018-08-02 and 2018-09-01 (see Figure 2). Although part of the north of the region is covered by another acquisition date (2019-08-22), some data from this acquisition date was included in training, as can be seen by comparing Figures 2 and 4, and generally classification performance is not visibly different in this area. Performance was good for acquisition dates which were included in training. As a result, data in the 2018-08-02 and 2018-09-01 area were excluded from the dataset, affecting parts of Bromley, Havering, Bexley, and Barking and Dagenham. The excluded area contains a large amount of agricultural land, woodland, surburban streets, and industrial areas which typically do not have green roofs, so few if any green roofs are missed.

Generalisation to different imagery sets is achievable provided that both positive and negative examples from each imagery dataset are included. The most northern training area, TQ3290 was included specifically to improve performance in the northern area covered by the collection date 2019-08-22. This is not viable for the eastern areas covered by 2018-08-02 and 2018-09-01 because we could not find any positive examples in this area with which to train. This means that updating this dataset in future using new imagery may require further training of the model on the new imagery.

More broadly, generalisation to completely different imagery sets (for example satellite imagery) would be best achieved by including examples from those sets during training. The trained model would not be expected to perform well on a completely unseen source of imagery without further training, as a diversity of imagery sources was not present during training. While relatively high-resolution satellite-imagery is available covering most cities in the world, these are generally not as high quality as the aerial imagery available in London; therefore, the same method applied to other cities may yield worse performance.

While performance was generally good as measured by the performance metrics, we collected some examples of poor classification performance, which are shown in Figure 8. Many false positives were observed where part of a green roof was correctly identified but another part nearby was misclassified, exemplified by Figure 8 A and B. It could be that relatively small variations in colour lead to the misclassification, but we found that random augmentations in brightness and hue did not improve performance. Areas in shadow in the imagery are generally poorly classified (for example in Figure 8 C). This could be because the shapes and colours are simply less distinct in shadow, but there are also few examples of this to learn from in the training data. Multi-spectral imagery could help deal with shadows and variations in vegetation colour. However, multi-spectral aerial imagery is collected more rarely and is less available; satellite multi-spectral imagery is available but resolution is poorer. Therefore, visible-spectrum aerial imagery has some practical advantages over multi-spectral imagery. Combining layers of multi-spectral imagery at lower resolution with aerial-imagery is technically challenging, but could be effective for this task.

We have not attempted to separate different types of green roof (e.g. intensive, extensive, roof gardens). While types of plant may be differentiated to some extent in aerial imagery, important features like depth of substrate cannot. Some green roofs may be in poor condition from lack of water, and there may be cases of fake turf or other imitation vegetation being detected as green roofs: both of these could be better identified using multi-spectral imagery.

Performance of the building-intersection step is reliant on the alignment of the buildings footprints with the imagery. The OS building footprints are very accurate, especially for identifying courtyards within building footprints. We found that alignment between other imagery sets, and with other building footprint sources, was not as reliable. However, OS maps are only available in the Great Britain, as opposed to OpenStreetMap which has a more global coverage.

4.3 Comparison to other estimates

Our estimate of green roof area in the CAZ in 2019 ($1.92 \times 10^5\ m^2$) is higher than the LRW2019 estimates and the AMR estimate for 2013, but lower than the AMR estimates for 2015 and 2017. For Greater London, the identified area is between the 2016 and 2017 estimated areas from LRW2019. In Figure A1, we compare our results the estimates for each LAD in 2017 from LRW2019 (Livingroofs Enterprises Ltd, 2019; European Federation of Green Roof and Green Wall Associations (EFB) and Livingroofs.org on behalf of the Greater London Authority, 2019): the results are strongly correlated, but some LADs have

(a)

(b)

(c)

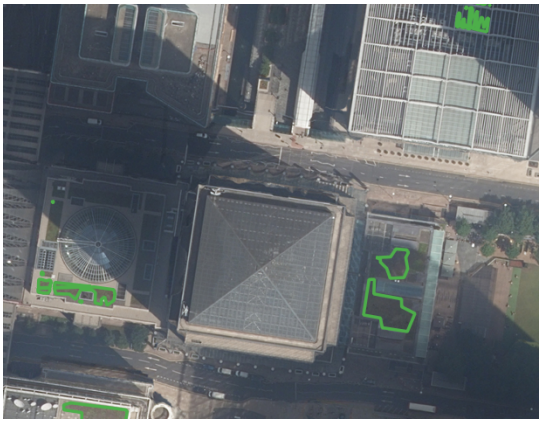

Figure 8. The green outline shows the area identified as green roof. (a) Barking Academy, (b) Royal Docks, (c) Canary Wharf. Each example shows some green roofs that are well classified and some that are missed. Typically, those that are missed are a darker red or brown colour, or are shadowed in the imagery. Imagery ©GetMapping Plc. Image location is shown in Figure A4.

quite different results. Our estimate is substantially greater for Newham, but is generally lower for the LADs with the greatest green roof area (e.g. Tower Hamlets, Greenwich). A slightly smaller area of green roofs is identified by our study in Tower Hamlets.

Examining the GLA's geospatial data (which is only public for the CAZ) (Greater London Authority, 2014) and infographics (Livingroofs Enterprises Ltd, 2019), we see multiple instances of ground-level parks being incorrectly identified as green roofs (e.g. Finsbury Square in Islington, Figure A2). Making use of the building footprint data enables us to avoid such misclassifications. There is also disagreement for the Barbican Centre (Figure A3), of which the full area is counted as a roof by the GLA results: this is a difficult edge-case, as the OS building footprints do not include the full area of the complex as a building. Over the CAZ, we find that 4% of the the area of the Greater London Authority (2014) dataset does not intersect with OS building footprints. It also appears that in the GLA's geospatial data, an area slightly larger than the vegetation is usually selected, which may be due to the resolution of the input data. This demonstrates the utility of ensuring the coincidence of identified green roof patches with building footprints.

Wu and Biljecki (2021) report that proportion of buildings by area which have a green roof is 41.6% in Zurich, 24.8% in Berlin, and 17.2% in New York (London was not included in their survey). Comparing this with the results in Tables 8, 9, 10, we see that the City of London ranks between Berlin and New York at 21.0%. This method of ranking is sensitive to the geographic area included in the calculation if the concentration of green roofs varies between districts within a city. Furthermore, given our interest in rooftop vegetation as a climate adaptation strategy, the actual amount of vegetation seems more relevant than the total area of the building.

4.4 Distribution of green roof areas

As shown in Table 8, although larger total areas of green roof are present in some LADs, the City of London is unique within Greater London for its' relatively higher concentration of green roofs. High concentrations of green roofs are also seen in the former dockland areas in Tower Hamlets, Newham and Greenwich, as well as in Stratford around the Olympic Park and the Kidbrooke development in Greenwich.

Distribution of green roofed buildings within LADs is heterogeneous (see Fig 7). When LSOAs stand out as having relatively high green roof coverage, it is often due to a single large building or a cluster of buildings with green roofs.

Despite having the highest green roof coverage out of the LADs, only 3.2% of the building footprint of the City of London is covered by green roofs. The City of London has very low amounts of green cover generally, so it is consistent with policy (e.g. (Greater London Authority, 2021b, Policy G5)) that green roofs would be adopted there. However, the LRW 2008 report (Design for London et al., 2008) found that 32% of roof area in the City of London could be suitable for retrofitting with green roofs, so the current status is a long way from that proposed. As the dataset identifies individual builtings, in future work we will explore what kinds of buildings, and what areas, have adopted green roofs.

Given that the area of vegetation in the City of London is overall quite low, it is possible that existing green roof coverage is making a difference to the thermal environment: a possibility that we will explore in a urban climate modelling study enabled by this data.

5 Conclusions

In this study, we produced a survey of green roofs in London using automatic segmentation of aerial imagery. The resulting geospatial dataset is made available for further research. We identified areas which have a high prevalence of green roofs; especially the City of London, the former docklands in Tower Hamlets and Newham, and the Olympic Park in Stratford. We highlighted some of the difficulties of producing such a dataset: especially that low prevalence of green roofs means that the classification problem is highly imbalanced, which can create problems for machine-learning algorithms. Furthermore, we demonstrate the importance of excluding ground-level vegetation from surveys of green roofs by ensuring areas classified as green roofs are coincident with building footprints.

In future work, we will use this geospatial dataset to further explore the characteristics and uses of buildings and neighbourhoods which have green roofs as well as those with potential for more green infrastructure, and to quantify the thermal effects of green roofs on London's micro-climate.

6 Code and data availability

Code and data generated by this project are available for download DOI:10.5281/zenodo.6861929 (Simpson et al., 2022). The geospatial data is stored in GeoJSON format, and can be read with GIS applications such as QGIS, ArcMap, or Fiona.

Aerial imagery was used under license from GetMapping Plc. Ordnance Survey data was used under license. These licensed data are available under educational license https://digimap.edina.ac.uk.

Author contributions. All authors participated in the conceptualisation of the paper. CS did the investigation and wrote the initial draft, with all others participating in review and editing.

Competing interests. The authors declare no competing interests.

Acknowledgements. CS, OB, and CH are supported by the Wellcome HEROIC Project (216035/Z/19/Z). CH is also supported by a NERC fellowship (NE/R01440X/1). The authors acknowledge the use of the UCL Myriad High Performance Computing Facility (Myriad@UCL), and associated support services, in the completion of this work. Thanks to Zaid Chalabi, Lora Fleming, James Grellier, Jon Taylor and Tim Taylor who all provided feedback on an early draft.

Table A1. Table listing the hyperparameters that were tuned, which values were tested, and the final value used for classification.

Parameter	Tested values	Final value
Loss function	Cross entropy, Lovasz, Focal	Focal
Learning rate	5.e-3, 5.e-4, 5.e-5, 5.e-6	5.e-5
Random augmentations	None; flips and 90° rotations; crops and flips and 90 ° rotations; flips and fully random rotations; 90% crops and flips and 90° rotations; flips and 90° rotations and sharpeness; flips and 90° rotations and sharpness	flips and 90° rotations and reduce sharpness
Max epochs	100	100
Pretrained model frozen	True, False	False

Table A2. Grid references of hand-labelled areas. Grid references are in the OSGB 1936 system.

TQ2280	TQ2474	TQ2481	TQ2678	TQ2781	TQ2784
TQ2872	TQ3073	TQ3176	TQ3181	TQ3184	TQ3280
TQ3281	TQ3282	TQ3290	TQ3373	TQ3478	TQ3486
TQ3681	TQ3775	TQ3783	TQ3872	TQ4084	TQ4180
TQ4283					

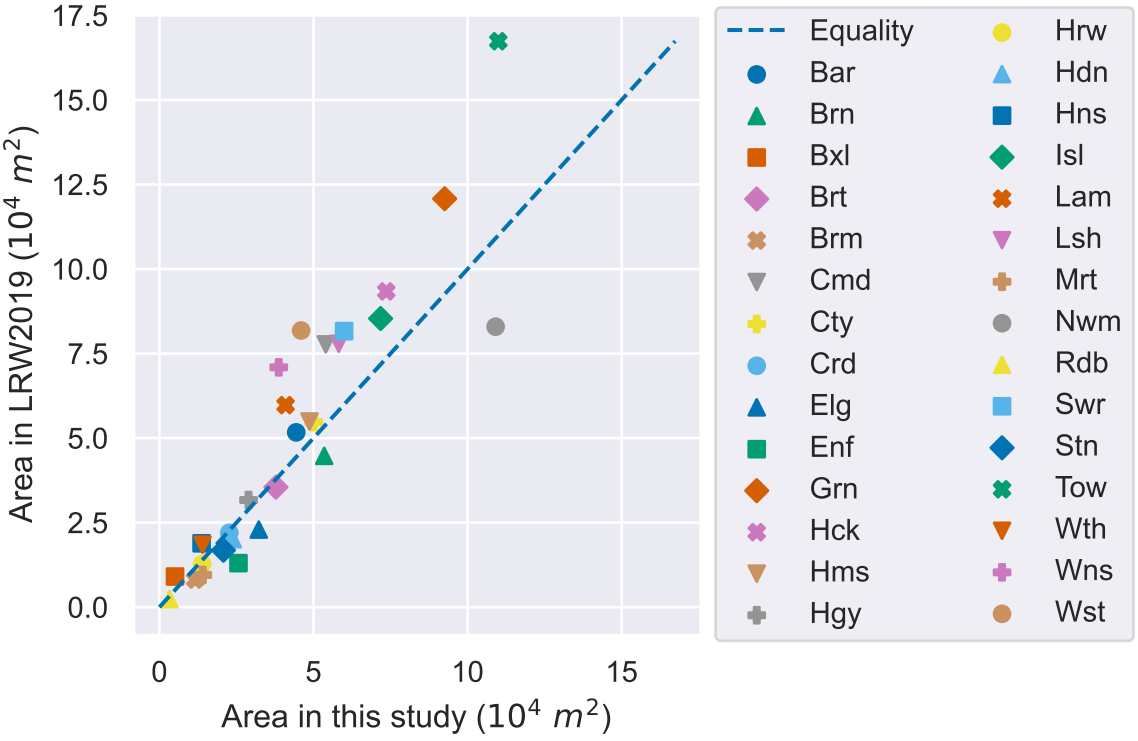

Figure A1. Scatter plot showing estimated green roof area in LADs of Inner London, comparing the estimates from Livingroofs Enterprises Ltd (2019) to our estimates.

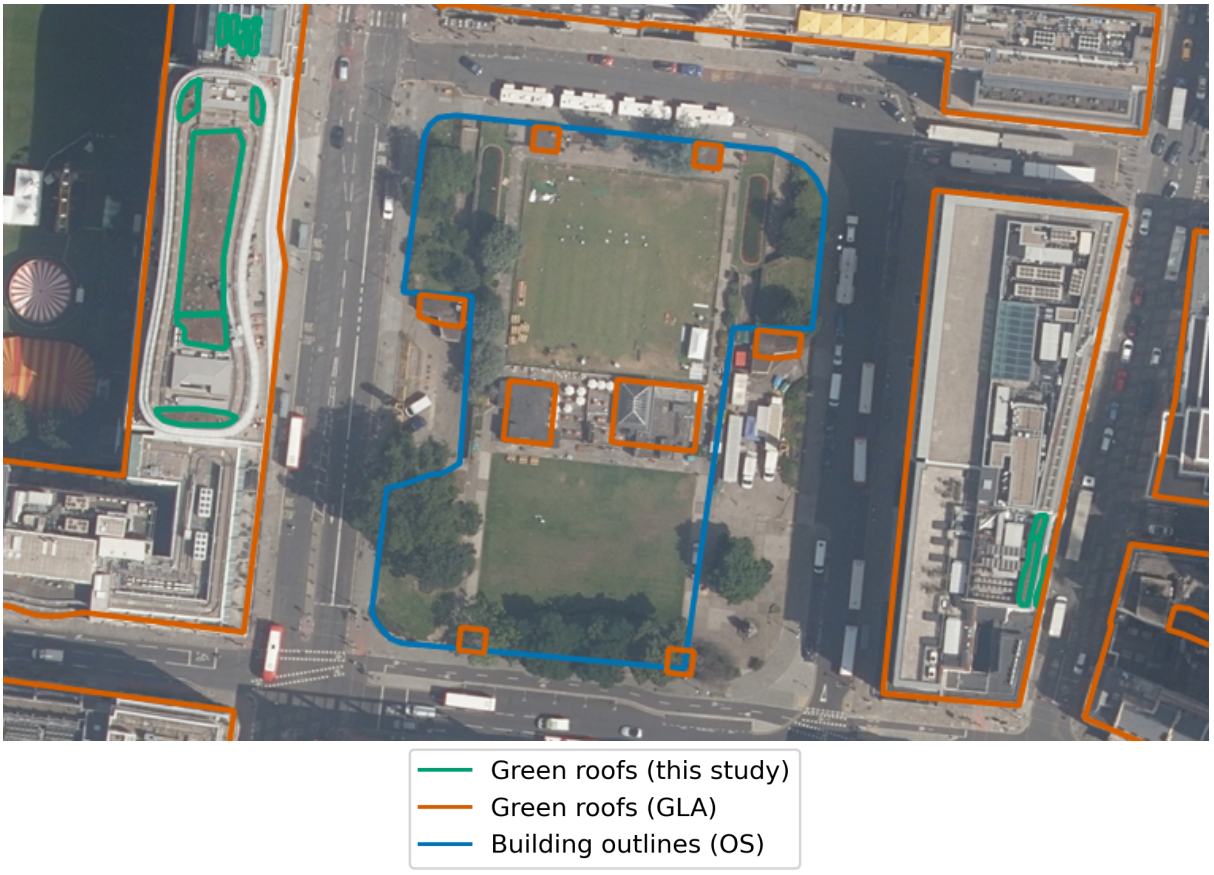

Figure A2. Example of ground-level green space misclassified as green roof in GLA dataset (in blue) (Greater London Authority, 2014). Building outlines according to OS VML are shown in orange, our results are shown in green. Image location is shown in Figure A4. Imagery ©GetMapping Plc. Building polygons are OS data ©Crown copyright and database rights 2022.

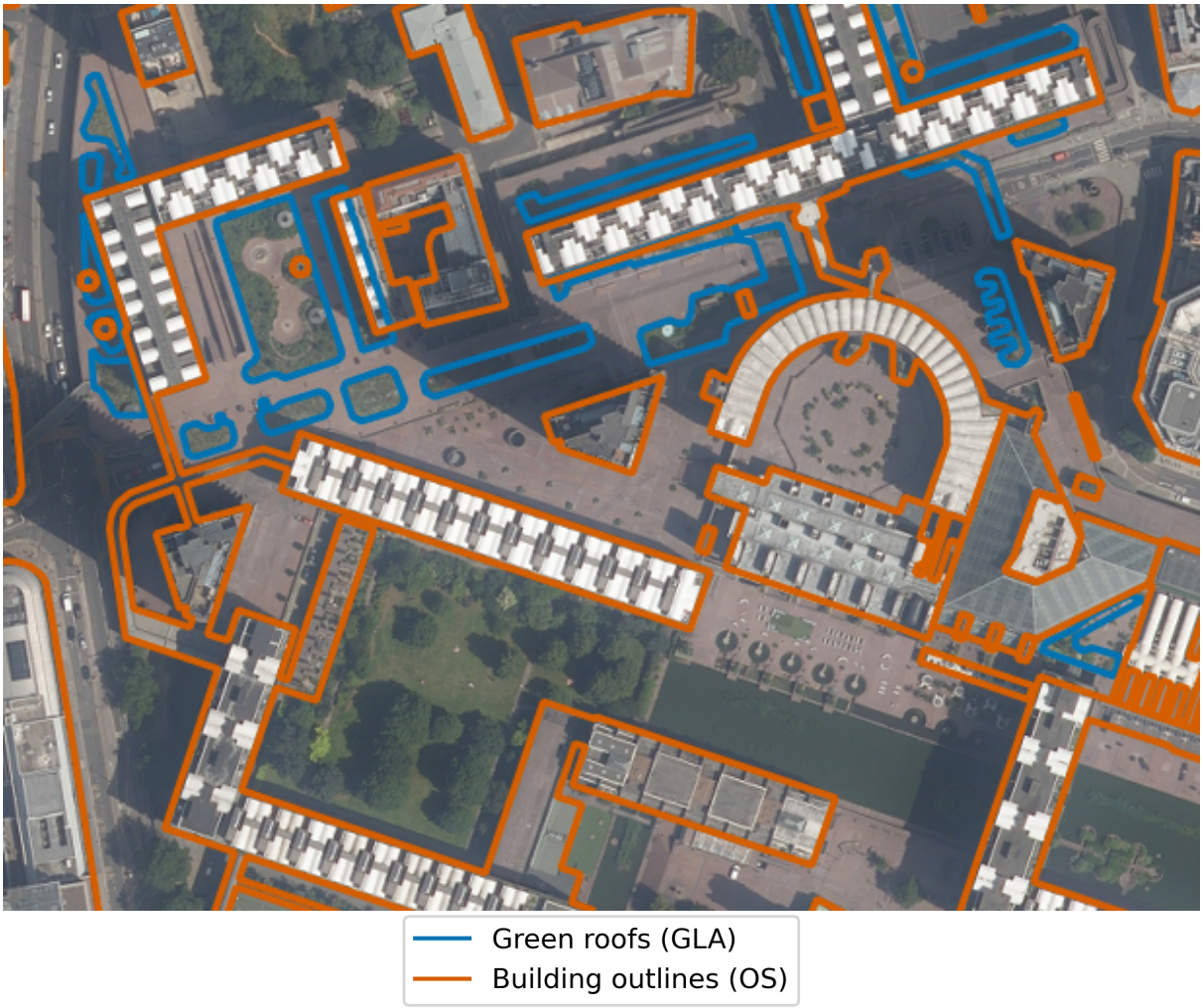

Figure A3. Example of disagreement between our result and GLA dataset (Greater London Authority, 2014) (in blue) due to building outlines. Building outlines according to OS VML are shown in orange. The green areas of the Barbican Centre are excluded in our analysis, as the OS VML does not identify them as within a building footprint. Image location is shown in Figure A4. Imagery ©GetMapping Plc. Building polygons are OS data ©Crown copyright and database rights 2022.

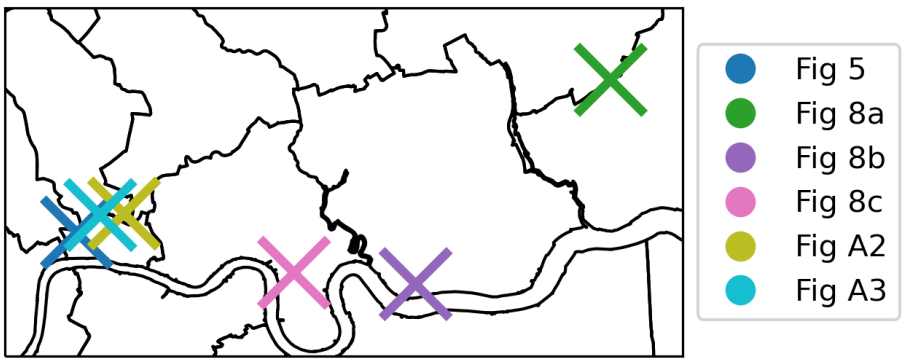

Figure A4. Locations of images in this paper.

References

Baik, J.-J., Kwak, K.-H., Park, S.-B., and Ryu, Y.-H.: Effects of Building Roof Greening on Air Quality in Street Canyons, Atmospheric Environment, 61, 48–55, 2012.

Berman, M., Triki, A. R., and Blaschko, M. B.: The lovász-softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4413–4421, 2018.

Castleton, H. F., Stovin, V., Beck, S. B., and Davison, J. B.: Green Roofs; Building Energy Savings and the Potential for Retrofit, Energy and buildings, 42, 1582–1591, 2010.

Cuthbert, M. O., Rau, G., Ekström, M., O'Carroll, D., and Bates, A.: Global Climate-Driven Trade-Offs between the Water Retention and Cooling Benefits of Urban Greening, Nature Communications, 13, 1–10, 2022.

Demuzere, M., Bechtel, B., Middel, A., and Mills, G.: Mapping Europe into local climate zones, PloS one, 14, e0214 474, 2019.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L.: Imagenet: A large-scale hierarchical image database, in: 2009 IEEE conference on computer vision and pattern recognition, pp. 248–255, Ieee, 2009.

Design for London, Greater London Authority, and London Climate Change Partnership: Living Roofs and Walls, Greater London Authority, 2008.

European Federation of Green Roof and Green Wall Associations (EFB) and Livingroofs.org on behalf of the Greater London Authority: Living Roofs and Walls: From Policy to Practice, https://livingroofs.org/wp-content/uploads/2019/05/LONDON-LIVING-ROOFS-WALLS-REPORT_MAY-2019.pdf, accessed: 2022-02-11, 2019.

Filazzola, A., Shrestha, N., and MacIvor, J. S.: The contribution of constructed green infrastructure to urban biodiversity: A synthesis and meta-analysis, Journal of Applied Ecology, 56, 2131–2143, 2019.

Getmapping Plc.: High Resolution (25cm) Vertical Aerial Imagery, EDINA Aerial Digimap Service, https://digimap.edina.ac.uk, updated: 2020-09-27, Downloaded: 2021-10-14, 2020.

Getter, K. L., Rowe, D. B., Robertson, G. P., Cregg, B. M., and Andresen, J. A.: Carbon Sequestration Potential of Extensive Green Roofs, Environmental science & technology, 43, 7564–7570, 2009.

Greater London Authority: Green roof map, https://www.london.gov.uk/what-we-do/environment/parks-green-spaces-and-biodiversity/green-roof-map, accessed: 2021-10-28, 2014.

Greater London Authority: London Plan Annual Monitoring Report 16 2018/19, Greater London Authority, Accessed: 2022-02-18, 2021a.

Greater London Authority: The London Plan, Greater London Authority, 2021b.

He, K., Zhang, X., Ren, S., and Sun, J.: Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778, 2016.

Hoeben, A. D. and Posch, A.: Green Roof Ecosystem Services in Various Urban Development Types: A Case Study in Graz, Austria, Urban Forestry & Urban Greening, 62, 127 167, 2021.

Kingma, D. P. and Ba, J.: Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980, 2014.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P.: Focal loss for dense object detection, in: Proceedings of the IEEE international conference on computer vision, pp. 2980–2988, 2017.

Livingroofs Enterprises Ltd: London borough green roof infographics and maps, https://livingroofs.org/borough-green-roof-infographics-maps-london-green-roof-report/, accessed: 2022-02-11, 2019.

Losken, G., Ansel, W., Backhaus, T., Bartel, Y.-C., Bornholdt, H., Bott, P., Henze, M., Hokema, J., Kohler, M., Krupka, B. W., Mann, G., Munster, M., Neisser, H., Roth-Kleyer, S., Ruttensperger, S., Schenk, D., Sprenger, D., Upmeier, M., and Westerholt, D.: Guidelines for the planning, construction and maintenance of green roofs, Landscape Development and Landscaping Research Society e.V., Bonn, 6th edn., 2018.

Mentens, J., Raes, D., and Hermy, M.: Green roofs as a tool for solving the rainwater runoff problem in the urbanized 21st century?, Landscape and urban planning, 77, 217–226, 2006.

Ng, V. and Hofmann, D.: Scalable feature extraction with aerial and satellite imagery, in: Proceedings of the 17th Python in Science Conference (SCIPY 2018), Austin, TX, USA, pp. 9–15, 2018.

Office for National Statistics: 2011 Census geography products for England and Wales, https://webarchive.nationalarchives.gov.uk/ukgwa/20160105225829/http://www.ons.gov.uk/ons/guide-method/geography/products/census/index.html, accessed 2022-03-22, a.

Office for National Statistics: Definitions of terms and phrases used in products and statistical outputs from the 2001 Census, https://www.ons.gov.uk/census/2001censusandearlier/glossary, accessed 2022-03-31, b.

Ordnance Survey (GB): OS VectorMap® Local, https://digimap.edina.ac.uk, updated: 2 July 2021, 2021.

Ozturk, O., SARITÜRK, B., and Seker, D. Z.: Comparison of Fully Convolutional Networks (FCN) and U-Net for Road Segmentation from High Resolution Imageries, International Journal of Environment and Geoinformatics, 7, 272–279, 2020.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., and Chintala, S.: PyTorch: An Imperative Style, High-Performance Deep Learning Library, in: Advances in Neural Information Processing Systems 32, edited by Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., and Garnett, R., pp. 8024–8035, Curran Associates, Inc., http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf, 2019.

Peng, L. L. and Jim, C. Y.: Green-Roof Effects on Neighborhood Microclimate and Human Thermal Sensation, Energies, 6, 598–618, 2013.

QGIS Association: QGIS Geographic Information System, https://QGIS.org, 2022.

Ronneberger, O., Fischer, P., and Brox, T.: U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical image computing and computer-assisted intervention, pp. 234–241, Springer, 2015.

Sailor, D. J., Elley, T. B., and Gibson, M.: Exploring the Building Energy Impacts of Green Roof Design Decisions–a Modeling Study of Buildings in Four Distinct Climates, Journal of Building Physics, 35, 372–391, 2012.

Shorten, C. and Khoshgoftaar, T. M.: A survey on image data augmentation for deep learning, Journal of big data, 6, 1–48, 2019.

Simpson, C., Brousse, O., Mohajeri, N., Davies, M., and Heaviside, C.: An Open-Source Automatic Survey of Green Roofs in London using Segmentation of Aerial Imagery: Dataset, https://doi.org/10.5281/zenodo.6861929, 2022.

Sproul, J., Wan, M. P., Mandel, B. H., and Rosenfeld, A. H.: Economic Comparison of White, Green, and Black Flat Roofs in the United States, Energy and Buildings, 71, 20–27, 2014.

Stewart, I. D. and Oke, T. R.: Local climate zones for urban temperature studies, Bulletin of the American Meteorological Society, 93, 1879–1900, 2012.

The Ecology Consultancy: Urban Greening Factor for London, https://www.london.gov.uk/sites/default/files/urban_greening_factor_for_london_final_report.pdf, Accessed: 2022-02-18, 2017.

Van Rossum, G. and Drake, F. L.: Python 3 Reference Manual, CreateSpace, Scotts Valley, CA, 2009.

Verisk Analytics, Inc.: UKBuildings, https://www.verisk.com/en-gb/3d-visual-intelligence/products/ukbuildings/, accessed: 2021-12-20, 2022.

Virk, G., Jansz, A., Mavrogianni, A., Mylona, A., Stocker, J., and Davies, M.: Microclimatic effects of green and cool roofs in London and their impacts on energy use for a typical office building, Energy and Buildings, 88, 214–228, 2015.

Wu, A. N. and Biljecki, F.: Roofpedia: Automatic mapping of green and solar roofs for an open roofscape registry and evaluation of urban sustainability, Landscape and Urban Planning, 214, 104 167, 2021.