Learning to Segment Publicly Accessible Green Spaces with Visual and Semantic Data

Jian Gao[1,2]
j.gao@qub.ac.uk
Niall McLaughlin[1]
n.mclaughlin@qub.ac.uk
Joanna Sara Valson[2]
j.valson@qub.ac.uk
Neil Anderson[1]
n.anderson@qub.ac.uk
Ruth Hunter[2]
ruth.hunter@qub.ac.uk

[1] School of EEECS
Queen's University Belfast
Belfast, UK

[2] Centre for Public Health
Queen's University Belfast
Belfast, UK

Abstract

The study of the health effects of Publicly Accessible Green Spaces (PAGS), such as parks and urban greenways, has received increasing attention in environmental sciences and public health research. However, the lack of relevant data and methods for PAGS mapping limits this work. To our best knowledge, most of the existing studies of PAGS mapping are manual, limited to small regions, and do not generalise geographically.

In this paper, we introduce a first-of-its-kind dataset - the Northern Ireland Publicly Accessible Green Spaces (PAGS-NI) dataset. Unlike existing datasets that typically consider only visual remote sensing data, our PAGS-NI dataset combines high-resolution, multi-band remote sensing data, geographical information data and activity data with hand-verified PAGS ground truth. Using this dataset, we develop a semantic segmentation model for automatic and scalable PAGS mapping that fuses these different data sources. Our model is able to predict PAGS on unseen places given appropriate training, which exceeds prior art. Furthermore, we show that our model trained solely on Northern Ireland can generalise to PAGS prediction for areas in the United States. Our model and dataset have the potential to advance large-scale PAGS studies in environmental science and public health research. Our dataset and code are available at https://github.com/Ellenisawake/pags-ni.

1 Introduction

Publicly accessible green spaces (PAGS) are known to promote physical and mental health while also providing increased opportunities for social inclusion and reducing inequalities [10]. The World Health Organization (WHO) classify areas such as parks, greened vacant lots, vegetated street-scapes, school yards, urban greenways, forests, and trails as

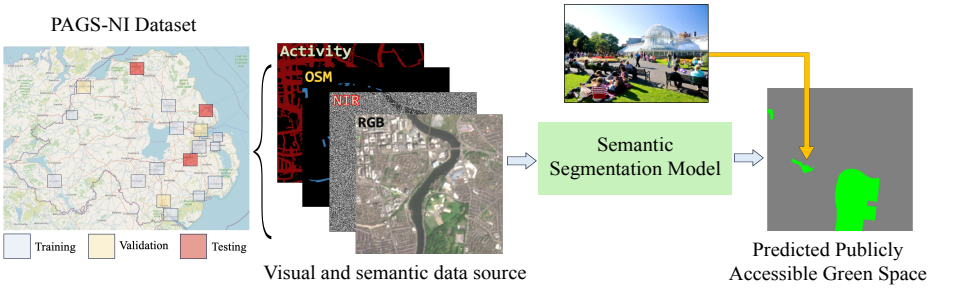

Figure 1: System Overview. We propose a new PAGS-NI dataset combining imaging, mapping and activity information and build a semantic segmentation pipeline for PAGS mapping.

PAGS [9]. With most of the world's population now living in towns and cities [10], automated mapping of urban PAGS is an important topic for environmental science and public health [25]. Yet, this is an underdeveloped area of research due to practical challenges.

To the best of our knowledge, there is currently no public datasets existing for the automated PAGS mapping task. Existing methods for green space estimation [21] tend to rely on the Normalised Difference Vegetation Index (NDVI) [14], which detects the presence of vegetation in a region using the ratio between red and near-infrared (NIR) satellite bands [32]. However, some regions with high NDVI, such as farmland, are not publicly accessible and, therefore, do not elicit the same health and well-being benefits. Additional data sources such as OpenStreetMap (OSM) [24] and data from physical activity trackers [17] are needed to distinguish PAGS from other kinds of green areas.

In this paper, we investigate the new challenging task of segmenting publicly accessible green space (PAGS), illustrated in Fig. 1. Our contributions can be summarised as threefold: 1) we introduce a new PAGS-NI dataset consisting of multiple data sources, including multi-band satellite imagery (visual), OSM geographical and activity data (semantic), as well as verified ground truth; 2) we develop a model that fuses, for the first time, visual and semantic data sources for accurately mapping PAGS; 3) finally, we demonstrate the strong generalisation capability of our model to distant geographical locations. Our proposed dataset and model can be used as a foundation to boost research on this relatively new problem and bring insights into various related topics such as environmental science and public health.

2 Related Work

Semantic Segmentation for Remote Sensing. Qin *et al* [27] present a review of land cover classification, including semantic segmentation of remote sensing images. They identify two main challenges: dataset quality and domain shift across geographical regions, which affect global-scale urban green space mapping [30]. Dedicated datasets and challenges have recently been released [3, 7]. For instance, OpenSentinelMap [13] introduces a per-pixel annotated land usage data with labels derived from Open Street Map. They show that an off-the-shelf convolutional network trained on this dataset can perform semantic segmentation at a global scale. The typical approach to performing semantic segmentation of remote sensing images is to apply a U-Net-based architecture [12]. However, the data tends to be imbalanced, with the background class outnumbering the sparse classes of interest, limiting the naive application of this approach. The foreground-aware relation network (FarSeg) [38] for geospatial segmentation with significant foreground-background imbalance. Xie *et al* [34]

proposed a multi-task method which conducts simultaneous key point detection and pixel segmentation for road extraction in satellite images.

Urban Green Space Mapping. Most existing green space studies focus on Urban Green Space mapping (UGS), which refers to any type of green space in urban areas without considering its use. Therefore, much of the work mentioned below is not directly comparable with our proposal. The earliest approaches for UGS mapping tended to rely purely on NDVI [28]. Later methods used classical machine learning to combine multiple satellite bands [15, 16, 21]. Recently, deep network approaches incorporating various contextual data sources have dominated [9, 14, 30]. Among the recent approaches, the U-Net [29] and similar encoder-decoder architectures [21] are popular choices for segmentation tasks. Zhao *et al* [37] used a U-Net model to map urban green space in 16 cities. Liu *et al* [21] compare support vector machine (SVM), fully connected networks (FCN), and U-Net with multi-spectral data input. They find DeepLabv3plus, a U-Net-style encoder-decoder model, performs best. Xu *et al* [35] fuse multi-spectral remote-sensing data captured in summer and winter. The change in vegetation between the seasons provides information unavailable at a single point in time. Like our proposed method, Ludwig *et al* [23] and Chen *et al* [2] use OSM and satellite data. However, unlike us, they use OSM data to compensate for low-resolution satellite imagery. In Chen *et al* [2], OSM data was used to derive ground truth and as an input to the model. Both above works did not consider generalization as they only tested on one city. A common issue for UGS mapping is consistent ground truth availability, especially across countries. While several public UGS datasets have become available in recent years, the definition of UGS is often inconsistent between jurisdictions, posing a challenge for consistent evaluation [18]. OpenStreetMap land usage tags have frequently been used as ground truth for UGS mapping [14]. While OSM tags have known regional variations [22], they are one of the more reliable indicators of UGS at a global level [14].

3 Dataset

In this section, we introduce our newly proposed PAGS dataset, PAGS-NI[1], which covers a total area of 2216.641 square kilometres from 18 cities/towns (details in supplementary). Greenness and accessibility are the two key characteristics of PAGS. Our dataset incorporates semantic information on accessibility in addition to visual satellite data. This enables PAGS mapping, which is not possible with existing NDVI-based UGS datasets and methods [14].

3.1 Visual Data

We use PlanetScope satellite data[2] due to the availability of high-resolution, frequently updated, multi-band images. The red, green, blue, and Near-Infrared (NIR) bands are used as similarly in relevant studies [15, 16], while we also calculate the NDVI index following the guidance from PlanetScope publisher. Our pre-processing pipeline includes four main steps: coordinate system alignment, satellite band extraction and normalisation, NDVI computation and patch vision. More details are explained in the supplementary material.

Band extraction and normalisation. We use the Planet 4-band analytical surface reflectance data, which includes the red, green, blue and Near-Infrared (NIR) channels. One prominent

[1]Dataset available at https://github.com/Ellenisawake/pags-ni.

[2]https://developers.planet.com/docs/data/planetscope/

issue in the raw satellite data is over-saturated pixels, which may be caused by specular reflection from strong surface reflectance materials such as glass or water. These over-saturated pixels suppress the rest of the pixels during normalisation, which can cause problems for our model. To remove these outliers, we perform thresholding of the raw data. For each image, we obtain the histogram of all pixel brightness values depending on the band. A threshold value that encompasses 99% of the pixel brightness values in the histogram is calculated for the band. We calculate a combined threshold for all the visible light bands (RGB) to preserve the colour balance. A separate threshold for the NIR band is obtained because the NIR value range differs significantly from the others. We clip values using the thresholds for all the bands, then normalise all pixel values to the range [0-255].

Patch division. The satellite data we use has a pixel size of around 3 metres and the resulting image for a city could be of huge size, e.g., 6550×4990. This is generally too large for most neural network models. Therefore, to prepare the data as input to the model, we crop the raw satellite images into non-overlapping patches of roughly 550×550 pixels. The 550×550 patch size allows for geometric data augmentation to be applied before cropping to 512×512 for input to the network during training and testing.

3.2 Semantic Data

The role of the semantic data source is to provide evidence of accessibility for the model. We consider two sources of accessibility data: OpenStreetMap land usage annotations and activity maps from GPS track points.

OpenStreetMap. OpenStreetMap is a crowd-sourced worldwide geographical data pool, with geometries at varying levels of granularity. The semantic information in the OSM tags, derived from local expert knowledge, provides evidence of both greenness and accessibility. However, as pointed out in [23], one of the main drawbacks of OSM data is its lack of consistency and credibility; therefore, OSM cannot be used alone to determine PAGS and must instead be combined with other credible sources of evidence.

Building on previous work using OpenStreetMap data to map UGS (different from PAGS), we select a set of OSM tags, shown in the supplementary material, as green space indicators [18]. We study two different representations of the data for input to the model. Firstly, a single binary map where all polygons containing a relevant OSM tag are filled with ones and all other areas with zeros. Secondly, individual binary maps for each tag. Our OSM maps are plotted in the same CRS as the satellite imagery, giving a one-to-one pixel correspondence between the binary OSM maps and the satellite imagery.

Activity map. The GPS track data from OSM provides an additional source of information on accessibility. Users upload their exercise GPS traces, which are used to build a global map showing where different activities, such as running, walking and cycling, occur. The map provides evidence of whether an area is accessible or not, as shown by the density of activities. However, the data is also noisy and does not directly distinguish green spaces from non-green areas. We download the GPS data in the form of discrete points, then plot the points as a binary map in the same CRS as the satellite imagery to be used as an additional input data source for PAGS mapping.

3.3 Ground Truth

Ground truth PAGS data is required to train a model to automatically map PAGS areas. Recently, Outscape Northern Ireland released a public green space map, the GreenspaceNI

Map[3], which covers the whole geographical area of Northern Ireland (NI). Our work will, therefore, mainly focus on Northern Ireland due to the availability of this data. We build a ground truth PAGS dataset containing the 18 most populous cities/towns within the GreenspaceNI Map, with details of the locations given in the supplementary. Manual checks were performed to confirm that the GreenspaceNI Map data aligns with the satellite and other data sources.

3.4 Data Splits

Our dataset consists of 857 non-overlapping patches, which are then split into geographically non-overlapping training, validation and testing sets. Specifically, we have 100 patches from three cities for validation and 97 patches from another three cities for testing, while patches from the remaining cities are used in training (details of the splits are given in the supplementary). This allows for a rigorous evaluation of model performance and allows for testing generalisation to geographical areas completely outside of the training set.

4 Method

Model. We approach PAGS mapping via the lens of semantic segmentation [36]. Our models take several visual and semantic channels as input and then learn to predict a map of PAGS areas. Applying deep neural networks to remote sensing tasks is challenging due to the lack of labelled data, leading to over-fitting. To help generalisation, we use a ResNet-50 pre-trained on BigEarthNet-MM [31] as the encoder (feature extractor) backbone for all experiments. We explore two network architectures built on this feature extractor: UNet [29], which has been shown superiority for practical real-world remote sensing tasks such as in the SpaceNet challenges [33]; and FarSeg [38], a recent state-of-the-art for geo-spatial object segmentation on remote sensing data. As there are few specialised green space mapping models to build upon, we select UNet [29] and FarSeg [38] as our two base network architectures due to their relatively low resource requirements and previously demonstrated high performance on remote sensing data. With the FarSeg model, the pre-trained ResNet-50 is used directly as the model encoder. When using the UNet model, the pre-trained ResNet-50 is used in the down blocks. UNet has about four times the number of parameters of the FarSeg model and therefore, takes up more memory space and is slower to train. However, from a results point of view, it demonstrates better generalisation capability, which we will illustrate in more detail in §5. All models accept input patches of size 512x512 pixels. To help cope with the class imbalance between PAGS and non-PAGS pixels, all our models are trained using Focal Loss [20].

Data Augmentation. During training, we apply data augmentation via horizontal flipping, random cropping and small affine transformations. During testing, we use the centre crop from each patch to preserve the original aspect ratio and fine image details.

Furthermore, we explore the use of data augmentation to address the class imbalance issue, which is a big challenge for automatic PAGS mapping. As usual in many urban settings, PAGS areas are rarer than non-PAGS areas. We find that the recently developed copy-paste augmentation (Copy-Paste) [8] can help create more PAGS samples for training. Copy-paste was proposed for instance-segmentation in natural images, where object instances are extracted and randomly pasted onto images to generate additional virtual samples. It has been

[3] https://www.out-scape.com/news/greenspaceni-map/

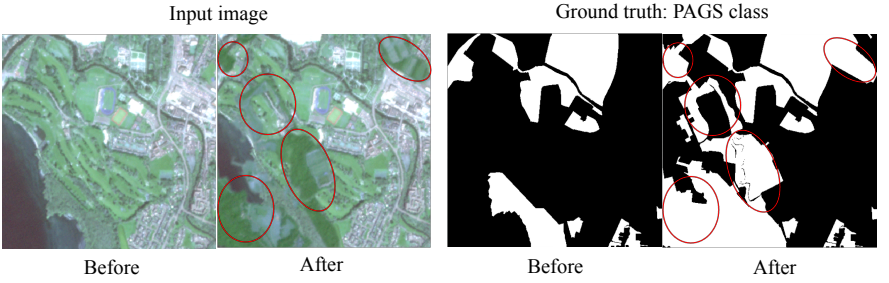

Figure 2: Copy-Paste example. Left: input image before and after Copy-Paste, right: ground truth mask for PAGS class (white for true and black for false) before and after Copy-Paste. Red lassos highlight the pasted PAGS pixels.

shown to substantially improve the performance on common daily life object segmentation benchmarks such as COCO [19], while also seen to work well on medical images[1]. Yet to our knowledge, so far there has been no application of this method to remote sensing data. Motivated by the above and the lack of positive samples in learning a good model for PAGS mapping, we explore the use of copy-paste for PAGS semantic segmentation. To do this, we randomly select PAGS areas during training and copy-paste them into non-PAGS map areas, updating all bands and the ground truth. We also add a positive sampling strategy to the original implementation to guarantee positive PAGS samples are created. This increases dataset diversity and helps alleviate the class imbalance issue. An example of this augmentation is shown in Fig. 2, where we can see that areas of PAGS significantly increased after the augmentation. Note that, an augmentation scheme like this might not guarantee contextual rationality between the background and the foreground, such as the giraffe appearing on the football field example in the original Copy-Paste paper. However, the method is still proven beneficial for pixel-level segmentation tasks because the lack of training data and labels poses a much larger challenge when training deep neural networks.

5 Experiments

In this section, we evaluate our proposed pipeline using the PAGS-NI dataset.

Data. All the experiments are conducted using our PAGS-NI dataset, except for the geographical generalisation test in §5.3. Specifically, for the training set we apply overlapping random crops instead of non-overlapping crops to increase the training sample diversity.

Evaluation. For fair evaluation under data imbalance, we mainly use F1 score (DICE) and Jaccard Index (IoU). Both metrics consider precision and recall, making them suitable for data-imbalanced settings. We also calculate two pixel-accuracy metrics: accuracy for PAGS foreground pixels (FG) and accuracy for all pixels (All). The former is more important in the PAGS mapping task as it focuses on our class of interest. For each experiment, we report the best model's performance on the test set.

Training. During training, we use a batch size of 8 unless otherwise specified, with a learning rate of 0.001 for pre-trained parameters and 0.01 for other parameters. All models are trained for 80 epochs with the model with the best F1 score on the validation set selected as the best model. We use Pytorch [26] and run all experiments on a single Nvidia RTX A5000 GPU. Generally, UNet models take about 17GB of memory and finish training in about one hour, while FarSeg models use about 13GB of memory and train in about half an hour.

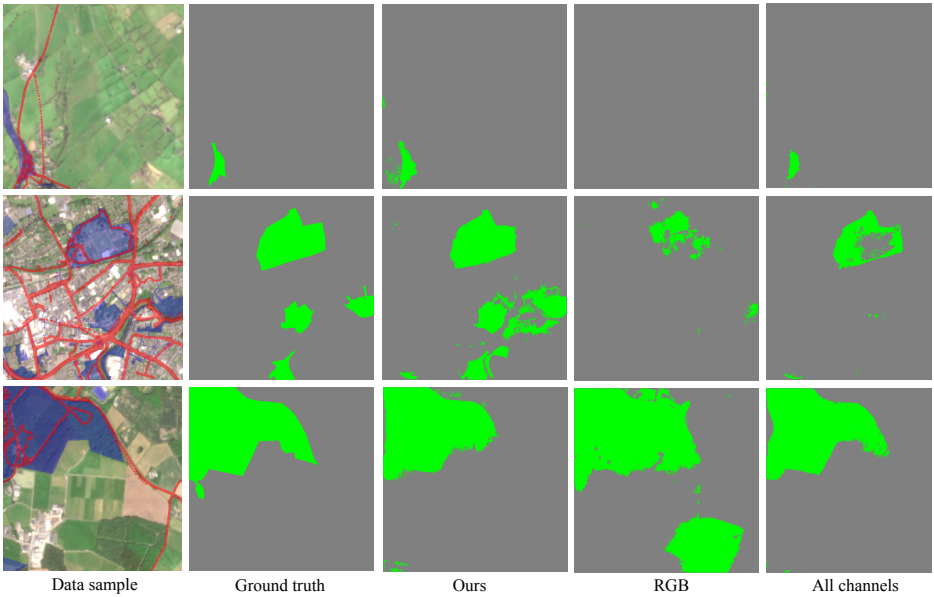

Figure 3: Qualitative results. Each row is for one example. 1^{st} column: data sample, OSM greenspace areas are plotted in blue, Activity is in red; 2^{nd} column: Ground truth, where green indicates PAGS and grey means non-PAGS; 3^{rd} to 5^{th} column: prediction from: our proposed method using NDVI+OSM+Activity channel combination, RGB and all the seven channels. *Best viewed in color.*

5.1 Channel Combination

Our dataset contains seven available input channels: five visual bands - R, G, B, NIR and NDVI - which are all data sources for greenness in PAGS mapping and two semantic bands - OpenStreetMap and Activity info which hint at the accessibility of spaces. There are more than 100 possible combinations of the seven channels, too many to try due to limited computing resources. Instead, we evaluate channel combinations guided by the literature and an understanding of the information available in each channel. Based on the input used in existing literature for UGS mapping [16, 21, 37], two channel combinations are considered as baselines: RGB, which is the most commonly adopted visible light bands in many computer vision tasks, and RGB+NDVI where NDVI has been a main criterion for green space estimation in prior arts as introduced in §2. Note that these existing methods consider only visual data sources without the semantic data source. We propose that the latter is a key element in accurate PAGS segmentation. Since accessibility is a critical part of the definition of PAGS, it is essential to compare combinations with both greenness (the five visual bands) and accessibility (the two semantic channels) against the baselines with greenness data only.

The results are shown in Table 1. We observe that the performance of baseline methods using visual data sources only are not satisfying unless combined with the semantic data source. This coincides with our motivation that the accessibility information provided by the semantic data is crucial in PAGS mapping. Overall, the best-performing channel combination is NDVI+OSM+Activity, which provides the key information (greenness and accessibility) for PAGS mapping. While we see that some other combinations, such as all-channels, also provide greenness and accessibility information, however, they also include additional

Channels							Pixel Acc %			
R	G	B	NIR	NDVI	OSM	Activity	F1	Jaccard	FG	All
Baselines										
✓	✓	✓					0.437	0.280	32.13	95.43
✓	✓	✓		✓			0.380	0.234	23.98	89.39
✓	✓	✓	✓				0.029	0.014	1.28	95.26
✓	✓	✓				✓	0.228	0.128	9.62	95.53
✓	✓	✓			✓		0.546	0.376	64.50	94.20
✓	✓	✓	✓	✓	✓	✓	0.578	0.407	63.38	94.93
✓				✓	✓		0.582	0.411	57.06	95.47
✓			✓		✓		0.612	0.441	69.32	95.20
✓			✓		✓	✓	0.627	0.457	66.35	95.64
				✓	✓		0.648	0.479	66.42	96.09
✓				✓	✓	✓	0.648	0.480	60.93	96.30
					✓	✓	0.651	0.483	64.70	96.18
				✓	✓	✓ (Ours)	**0.720**	**0.562**	68.33	93.75

Table 1: Comparison of different channel combinations.

Model	OSM Channel(s)	F1	Jaccard	Pixel Acc % FG	Pixel Acc % All
UNet	Single	**0.720**	**0.562**	68.33	93.75
	Multi	0.716	0.558	75.65	92.93
FarSeg	Single	0.694	0.531	60.29	93.83
	Multi	**0.759**	**0.612**	71.68	94.67

Table 2: Test results for different network architectures and OSM channel input formats.

Method	F1	Jaccard	Pixel Acc% FG	Pixel Acc% All
Direct OSM mapping	0.691	0.528	80.64	91.53
Our model UNet	0.720	0.562	68.33	93.75
Our model FarSeg	**0.759**	**0.612**	71.68	94.67

Table 3: Comparison of PAGS mapping performance between directly mapping the OSM green space areas and our model.

redundant information, and their results are not as good as NDVI+OSM+Activity. NDVI is specifically designed to detect green plants; therefore, the network can directly use this information without needing to rediscover the correct channel combinations from scratch. Among the two semantic data channels, OSM is seen to benefit more as the combination of RGB+OSM gives much better results than RGB+Activity. Qualitative analysis of different model outputs is shown in Fig. 3, where we observe that the proposed NDVI+OSM+Activity input is able to robustly capture most of the PAGS areas in different geographical locations. More visual results can be found in the supplementary.

5.2 Performance Analysis

OSM Channel Format. We see from results in Table 1 that OSM data is critical in accurate PAGS mapping. Next, we study the effect of presenting OSM data to our model in either single or multi-channel formats alongside the other data channels. Single channel means OSM data is in a single binary input channel as the union of all geographical areas containing the hand-selected green space indicator tags [18]. Multi-channel means that we have one channel per tag. The comparison between different channel formats are shown in Table 2. The effect of the OSM data channel format differs for UNet and FarSeg. With the UNet model, the difference between both formats is small. However, while FarSeg is more sensitive to the number of OSM channels, and it performs significantly better with multi-channel input. This indicates that how OSM data is input needs to be considered together with the network architecture used.

Comparison with Direct OSM Mapping. Previous work in the literature has explored directly using OSM green space tags as ground truth for UGS mapping [13]. In this experiment, we compare the PAGS mapping performance of a baseline system using Direct OSM

Method	F1	Jaccard	Pixel Acc %	
			FG	All
Basic aug	0.709	0.550	76.16	92.39
Basic aug + Copy-Paste	**0.720**$_{0.011\uparrow}$	**0.562**$_{0.012\uparrow}$	68.33	93.75

Table 4: Effect of adding Copy-Paste in addition to basic augmentation. Basic aug: horizontal flipping and affine transformation only.

Model		Testing Data Location	
		N.I. (Test set)	United States
FarSeg	RGB	0.501	0.022
	RGB+NDVI	0.488	0.098
	Ours	**0.759**	**0.271**
UNet	RGB	0.437	0.003
	RGB+NDVI	0.380	0.000
	Ours	**0.720**	**0.444**

Table 5: Geographical generalisation results (F1 score) from N.I. to the US. Ours: NDVI+OSM+Activity. All using Copy-Paste augmentation.

Mapping against our proposed method. The Direct OSM Mapping system returns a binary map of the union of all geographical areas containing the hand-selected green space indicator tags [18], listed in the supplementary material. The results are shown in Table 3. Although without any learning at all, using direct OSM greenspaces can help find PAGS with a reasonable level of performance, our models can accurately identify significantly more PAGS areas than the hand-input Direct OSM greenspaces, with up to 7% better F1-score for our best model. Also as mentioned in §3.2, although there is a good amount of fine-grained OSM data for Northern Ireland which is used in the test, the coverage and quality of OSM data for other countries in the world may not be as good. Our proposed method has better potential of expanding the study regions to much wider parts around the globe than simply relying on the manual-curated OSM data.

Copy-Paste Augmentation. The benefit of applying Copy-Paste to the model can be seen from Table 4, that it help improve both F1 score and the Jaccard compared to not using it.

5.3 Geographical Generalisation

We now examine the generalisation of our model to regions outside Northern Ireland. Geographical generalisation to distant locations has long been a difficult problem, even from one city to another in the same country [8]. Many existing green space mapping methods were only developed and tested for small regions, e.g., within a city [23]. Here, we conduct a challenging test of our proposed model: training purely on the training set of PAGS-NI and testing on 213 patches from three United States cities (details of which are given in supplementary material) without any modification of the model. To verify the results, we use protected green areas ground truth from the PAD-US dataset [5], which is the closest to PAGS that we could find for United States. The focus of the PAD-US dataset was originally bio-diverse areas, and it was expanded in recent years to include open public green spaces. This differs from our PAGS-NI dataset, of publicly accessible green spaces. Consistent worldwide PAGS ground truth data is very difficult, if not impossible, to obtain. Therefore, while PAD-US is not a strict match to the PAGS mapping task considered in this work, we believe it is still a valuable resource for the evaluation of PAGS generalisation.

The test data comes from three different US cities (details in the supplementary). Results are shown in Table 5. We compare our best model, NDVI+OSM+Activity, against the two baselines, RGB and RGB+NDVI, all trained on the same PAGS-NI training set. When tested within the same country (the NI test set), our proposed model performs best as shown previously. When tested on the United States cities, the performances of all models dropped significantly. The geographical distance between the two countries and the difference in the ground truth definitions may be the reason why the baseline models perform so poorly on

United States cities. However, again, our proposed model performs significantly better than the baselines at a reasonable level. For context, while not directly comparable, we note that a network designed for cross-city general object segmentation achieved a mean F1-score of 0.352 when generalising between cities within the same country [8]. The relative success of our proposed PAGS model underlines the importance of choosing appropriate inputs, including both visual and semantic data sources. Meanwhile, the difficulty of generalising to geographically distant regions remains a huge challenge for PAGS mapping methods. We leave this topic for future work.

6 Conclusion

In this paper, we study the under-explored problem of automatically mapping publicly accessible green spaces (PAGS). We introduce a novel dataset, PAGS-NI, that provides data from both visual and semantic sources, including satellite imagery, OpenStreetMap data and Activity Map data, with authorised ground truth. We propose a new semantic segmentation model that, for the first time, combines the aforementioned data sources for accurate PAGS mapping. We perform extensive experiments to assess the optimal way to utilise the multiple data sources. Our final model outperforms existing methods from the literature, especially in the challenging task of generalising to distant geographical regions. Our work has great potential to benefit various research areas including environmental science and public health.

7 Acknowledgement

This research was supported by the National Institutes of Aging (R01AG030153) and the UK Research and Innovation, Healthy Ageing Challenge, Social, Behavioural and Design Research [ES/V016075/1].

References

[1] Heeyoung Ahn and Yiyu Hong. Class-controlled copy-paste based cell segmentation for conic challenge. *bioRxiv*, pages 2022–03, 2022.

[2] Yang Chen, Qihao Weng, Luliang Tang, Qinhuo Liu, Xia Zhang, and Muhammad Bilal. Automatic mapping of urban green spaces using a geospatial neural network. *GIScience & Remote Sensing*, 58(4):624–642, 2021.

[3] Ilke Demir, Krzysztof Koperski, David Lindenbaum, Guan Pang, Jing Huang, Saikat Basu, Forest Hughes, Devis Tuia, and Ramesh Raskar. Deepglobe 2018: A challenge to parse the earth through satellite images. In *CVPRW*, 2018.

[4] WHO Europe. Urban green space interventions and health; a review of impacts and effectiveness. *WHO regional office for Europe, Copenhagen, Denmark*, 2017.

[5] U.S. Geological Survey (USGS) Gap Analysis Project (GAP). Protected areas database of the united states (pad-us) 4: U.s. geological survey data release, 2024. URL https://www.sciencebase.gov/catalog/item/65294599d34e44db0e2ed7cf.

[6] Golnaz Ghiasi, Yin Cui, Aravind Srinivas, Rui Qian, Tsung-Yi Lin, Ekin D Cubuk, Quoc V Le, and Barret Zoph. Simple copy-paste is a strong data augmentation method for instance segmentation. In *CVPR*, 2021.

[7] Ronny Hänsch, Jacob Arndt, Dalton Lunga, Matthew Gibb, Tyler Pedelose, Arnold Boedihardjo, Desiree Petrie, and Todd M Bacastow. Spacenet 8-the detection of flooded roads and buildings. In *CVPR*, 2022.

[8] Danfeng Hong, Bing Zhang, Hao Li, Yuxuan Li, Jing Yao, Chenyu Li, Martin Werner, Jocelyn Chanussot, Alexander Zipf, and Xiao Xiang Zhu. Cross-city matters: A multimodal remote sensing benchmark dataset for cross-city semantic segmentation using high-resolution domain adaptation networks. *Remote Sensing of Environment*, 299: 113856, 2023.

[9] Roberto E Huerta, Fabiola D Yépez, Diego F Lozano-García, Victor H Guerra Cobian, Adrian L Ferrino Fierro, Héctor de León Gómez, Ricardo A Cavazos Gonzalez, and Adriana Vargas-Martínez. Mapping urban green spaces at the metropolitan level using very high resolution satellite imagery and deep learning techniques for semantic segmentation. *Remote Sensing*, 13(11):2031, 2021.

[10] Ruth Fiona Hunter, Mark Nieuwenhuijsen, Carlo Fabian, Niamh Murphy, Kelly O'Hara, Erja Rappe, James Fleming Sallis, Estelle Victoria Lambert, Olga Lucia Sarmiento Duenas, Takemi Sugiyama, et al. Advancing urban green and blue space contributions to public health. *The Lancet Public Health*, 8(9):e735–e742, 2023.

[11] HABITAT III and Lucinda Hartley. The new urban agenda: Ten things you need to know. *Landscape Architecture Australia*, (153):17–20, 2017.

[12] Shunping Ji, Shiqing Wei, and Meng Lu. Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set. *IEEE Transactions on geoscience and remote sensing*, 57(1):574–586, 2018.

[13] Noah Johnson, Wayne Treible, and Daniel Crispell. Opensentinelmap: A large-scale land use dataset using openstreetmap and sentinel-2 imagery. In *CVPR*, 2022.

[14] Yang Ju, Iryna Dronova, and Xavier Delclòs-Alió. A 10 m resolution urban green space map for major latin american cities from sentinel-2 remote sensing images and openstreetmap. *Scientific Data*, 9(1):586, 2022.

[15] Nikola Kranjčić, Damir Medak, Robert Župan, and Milan Rezo. Machine learning methods for classification of the green infrastructure in city areas. *ISPRS International Journal of Geo-Information*, 8(10):463, 2019.

[16] SM Labib and Angela Harris. The potentials of sentinel-2 and landsat-8 data in green infrastructure extraction, using object based image analysis (obia) method. *European Journal of Remote Sensing*, 51(1):231–240, 2018.

[17] Andrew Ladle, Paul Galpern, and Patricia Doyle-Baker. Measuring the use of green space with urban resource selection functions: An application using smartphone gps locations. *Landscape and Urban Planning*, 179:107–115, 2018.

[18] Yiming Liao, Qi Zhou, and Xuanqiao Jing. A comparison of global and regional open datasets for urban greenspace mapping. *Urban Forestry & Urban Greening*, 62: 127132, 2021.

[19] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *ECCV*, 2014.

[20] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *ICCV*, 2017.

[21] Wenya Liu, Anzhi Yue, Weihua Shi, Jue Ji, and Ruru Deng. An automatic extraction architecture of urban green space based on deeplabv3plus semantic segmentation model. In *2019 IEEE 4th International Conference on Image, Vision and Computing (ICIVC)*, 2019.

[22] Christina Ludwig, Sascha Fendrich, and Alexander Zipf. Regional variations of context-based association rules in openstreetmap. *Transactions in GIS*, 25(2):602–621, 2021.

[23] Christina Ludwig, Robert Hecht, Sven Lautenbach, Martin Schorcht, and Alexander Zipf. Mapping public urban green spaces based on openstreetmap and sentinel-2 imagery using belief functions. *ISPRS International Journal of Geo-Information*, 10(4): 251, 2021.

[24] OpenStreetMap contributors. Planet dump retrieved from https://planet.osm.org . https://www.openstreetmap.org, 2017.

[25] World Health Organization et al. Setting global research priorities for urban health. 2022.

[26] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. In *NeurIPS*, 2019.

[27] Rongjun Qin and Tao Liu. A review of landcover classification with very-high resolution remotely sensed optical images—analysis unit, model scalability and transferability. *Remote Sensing*, 14(3):646, 2022.

[28] Isaac C Rhew, Ann Vander Stoep, Anne Kearney, Nicholas L Smith, and Matthew D Dunbar. Validation of the normalized difference vegetation index as a measure of neighborhood greenness. *Annals of epidemiology*, 21(12):946–952, 2011.

[29] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *18th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2015.

[30] Qian Shi, Mengxi Liu, Andrea Marinoni, and Xiaoping Liu. Ugs-1m: fine-grained urban green space mapping of 31 major cities in china based on the deep learning framework. *Earth System Science Data*, 15(2):555–577, 2023.

[31] Gencer Sumbul, Arne De Wall, Tristan Kreuziger, Filipe Marcelino, Hugo Costa, Pedro Benevides, Mario Caetano, Begüm Demir, and Volker Markl. Bigearthnet-mm: A large-scale, multimodal, multilabel benchmark archive for remote sensing image classification and retrieval [software and data sets]. *IEEE Geoscience and Remote Sensing Magazine*, 9(3):174–180, 2021.

[32] JD Tarpley, SR Schneider, and RL Money. Global vegetation indices from the noaa-7 meteorological satellite. *Journal of Climate and Applied Meteorology*, pages 491–494, 1984.

[33] Adam Van Etten, Dave Lindenbaum, and Todd M Bacastow. Spacenet: A remote sensing dataset and challenge series. *arXiv preprint arXiv:1807.01232*, 2018.

[34] Shenwei Xie, Wanfeng Zheng, Zhenglin Xian, Junli Yang, Chuang Zhang, and Ming Wu. Park-detect: Towards efficient multi-task satellite imagery road extraction via patch-wise keypoints detection. In *BMVC*, 2022.

[35] Zhiyu Xu, Yi Zhou, Shixin Wang, Litao Wang, Feng Li, Shicheng Wang, and Zhenqing Wang. A novel intelligent classification method for urban green space based on high-resolution remote sensing images. *Remote sensing*, 12(22):3845, 2020.

[36] Hongshan Yu, Zhengeng Yang, Lei Tan, Yaonan Wang, Wei Sun, Mingui Sun, and Yandong Tang. Methods and datasets on semantic segmentation: A review. *Neurocomputing*, 304:82–103, 2018.

[37] Jiawei Zhao. Mapping public green space using sentinel-2 imagery and convolutional neural network at a global scale. Master's thesis, 2022.

[38] Zhuo Zheng, Yanfei Zhong, Junjue Wang, and Ailong Ma. Foreground-aware relation network for geospatial object segmentation in high spatial resolution remote sensing imagery. In *CVPR*, 2020.