Urban Footpath Image Dataset to Assess Pedestrian Mobility

Venkatesh G M
Insight SFI Research Centre for Data Analytics
Dublin City University
Dublin, Ireland
venkatesh.gurummunirathnam@insight-centre.org

Bianca Pereira
Insight SFI Research Centre for Data Analytics
National University of Ireland
Galway, Ireland
bianca.pereira@insight-centre.org

Suzanne Little
Insight SFI Research Centre for Data Analytics
Dublin City University
Dublin, Ireland
suzanne.little@dcu.ie

Figure 1: Urban footpath images captured via mapillary mobile application

ABSTRACT

This paper presents an urban footpath image dataset captured through crowdsourcing using the mapillary service (mobile application) and demonstrating its use for data analytics applications by employing object detection and image segmentation. The study was motivated by the unique, individual mobility challenges that many people face in navigating public footpaths, in particular those who use mobility aids such as long cane, guide digs, crutches, wheelchairs, etc., when faced with changes in pavement surface (tactile pavements) or obstacles such as bollards and other street furniture. Existing image datasets are generally captured from an instrumented vehicle and do not provide sufficient or adequate images of the footpaths from the pedestrian perspective. A citizen science project (Crowd4Access[1]) worked with user groups and volunteers to gather a sample image dataset resulting in a set of 39,642 images collected in a range of different conditions. Preliminary studies to detect tactile pavements and perform semantic segmentation using state-of-the-art computer vision models demonstrate the utility of this dataset to enable better understanding of urban mobility issues.

CCS CONCEPTS

• **Computing methodologies** → **Object detection**; **Image segmentation**.

KEYWORDS

urban elements; convolution neural network; object detection; semantic segmentation; street-view analytics

ACM Reference Format:
Venkatesh G M, Bianca Pereira, and Suzanne Little. 2021. Urban Footpath Image Dataset to Assess Pedestrian Mobility. In *Proceedings of the 1st International Workshop on Multimedia Computing for Urban Data (UrbanMM '21), October 20–24, 2021, Virtual Event, China.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3475721.3484313

1 INTRODUCTION

Mobility is the ease with which one can move freely in an environment without any constraints. Footpaths are built to enable the mobility of pedestrians in urban areas. However, a number of pedestrians still experience reduced mobility when using public footpaths, in particular those who use mobility aids (e.g. long cane, guide dogs, crutches, and wheelchairs) who often require footpaths to have particular characteristics. According to a report

[1]https://crowd4access.insight-centre.org/

UrbanMM '21, October 20–24, 2021, Virtual Event, China

ACM ISBN 978-1-4503-8669-2/21/10…$15.00
https://doi.org/10.1145/3475721.3484313

published by World Heath Organisation (WHO), "About 15% of the world's population lives with some form of disability, of whom 2-4% experience significant difficulties in functioning." [3].

Assessing and understanding the accessibility of footpaths and other urban pathways requires continuous monitoring. Issues such as surface consistency, permanent and temporary obstacles or obstructions and pavement width are just some of the potential constraints to free mobility. To analyse mobility in urban areas, street-view image datasets, captured from instrumented vehicles, have been used with computer vision and routing applications to detect the characteristics of specific routes such as the state of pavement, presence of street crossings, traffic lights and other traffic signals. However, current urban mapping datasets present a number of challenges when used for the purpose of assessing the mobility of pedestrians on footpaths.

Firstly, these datasets are collected from a *driver's* perspective rather than a pedestrian's and, while high volumes of images can be generated, they lack the variety of routes and environments (time of day, weather, individual mobility requirements, etc.) that are necessary to better understand and assess urban mobility. Secondly, this leads to challenges in automatic image content analysis (segmentation, object identification, etc.) when the images are from the perspective of a camera on the road. Such images are often unable to capture the footpath due to parked vehicles and the images have a poor angle of view to adequately assess pedestrian mobility issues such as pavement surface and width.

To address these issues, Crowd4Access, a citizen science project, was initiated to gain understanding of urban mobility needs, provide tools and a methodology for systematic crowdsourced data capture of footpaths in urban environments and produce a preliminary dataset to evaluate the utility of the project.

In this paper we introduce a footpath-view image dataset built from a pedestrian's perspective[2]. First we outline the type of data needed to assess pedestrian mobility (Section 2), then we present our data acquisition process via crowdsourcing (Section 3) and conclude with an investigation of the dataset's usability and effectiveness through the development of an application for automatic detection of tactile paving (Section 4).

2 PEDESTRIAN MOBILITY

The Crowd4Access project aims to assess the accessibility of footpaths from a diversity perspective. We begin from the assumption that people use the footpath differently and, therefore, a given characteristic of a footpath may affect diverse pedestrians in different ways. For instance, the placement of a tactile paving (Figure 2, top row, right image) supports the mobility of pedestrians using a long cane by indicating the direction of movement or the presence of hazard. However, the same element hinders the mobility of users of crutches who may have difficulties in keeping their balance while on the tactile paving, and users of wheelchairs and adults pushing a pram who would suffer moderate shaking from passing on top of the irregular surface of the tactile pavement.

With diversity in mind, an image dataset used for the assessment of pedestrian mobility needs to take into consideration what urban elements either support or hinder the mobility of various footpath users, with or without disabilities. Therefore, rather than containing only images about specific urban elements (e.g. tactile pavements, lowered kerbs, street crossings with traffic lights), such dataset should provide imagery of complete footpaths under different conditions.

There are limited footpath image datasets that are capable of capturing these challenges. Open source datasets related to the automotive domain, such as CityScape [4], mapillary vistas dataset [9]), Bdd100k [16] etc., are captured using an instrumented vehicle from the driver perspective focusing on the road elements with limited footpath attributes visible. Though the datasets consist of annotated general instances of footpath or sidewalks, the data cannot be directly used for determining the mobility support of the footpath. Other research has focused primarily on indoor navigation [2, 5].

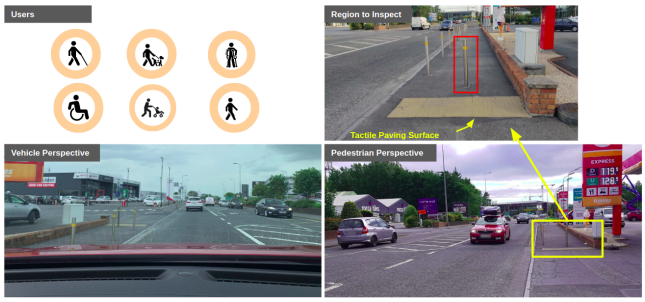


Figure 2: Illustration vehicle vs pedestrian of a footpath

Figure 2 illustrates the viewpoint of a footpath from both a vehicle and pedestrian's perspective. The inference drawn using the vehicle perspective contains minimal information about the footpath attributes and the resulting analysis on footpath mobility will fail to detect the tactile pavement surface and the placement of a bollard. The pedestrian view (bottom row, first image) clearly shows these attributes and enables better understanding of potential mobility impacts.

The inadequacy of footpath related information in available public datasets motivated us to create an open source dataset of urban street view images focusing on the footpath and aiming to better capture the perspective of prospective users.

3 CROWDSOURCING URBAN MAPPING

Interest in citizen science has been increasing [11]. Recently, Haklay et al. [6] discuss the challenges of defining citizen science but for the context of this project we highlight two of the main criteria – the direct involvement of end users in project development and the engagement of volunteers in the data gathering. This is especially critical in addressing mobility challenges where end user feedback generated from the online focus groups was critical in understanding the needs for the multimedia content analytics components. In addition the participation of the general public in the data gathering is a valuable opportunity for education both about mobility issues and the latest computer vision technologies.

To properly crowdsource the data and to better understand the mobility implications, we had direct or indirect engagement with multiple stakeholders categorised into individuals with reduced mobility (people using long cane, wheelchair, guide dogs), researchers,

[2]available under an open license at https://crowd4access.insight-centre.org/

local authorities and organisations that advocate on footpath accessibility. While the project was designed prior to 2020, the work started during COVID-19 pandemic which imposed certain restrictions on the execution of the project and the requirements gathering sessions for the identified mobility stakeholders. Direct online meetings were held with 9-12 individuals and all the information and feedback provided by the stakeholders via experience sharing workshops, mapping workshops and information session were recorded for further analysis and to define the problem statement for the pilot study.

This section describes the process of data collection using the Mapillary mobile phone application[3] and the resulting dataset preparation including annotation of tactile pavements and other key urban elements.

3.1 Image Acquisition and Retrieval

Volunteers were trained at specific online workshops in the process of data gathering using mapillary, provided with documentation and walked through the process of creating an account and uploaded their photographs. 20-25 people drawn from students, researchers, administrative staff, family members and community groups completed the mapillary training. For the dataset, 15 individuals from different parts of Ireland collected data between July 2020 and January 2021, mostly limited to a 2-5km region due to the COVID-19 lockdown restrictions. The training workshops are ongoing and data continues to be collected (as of July 2021).

All the images were captured using smart mobile phones (Android or iOS) using the mapillary phone application. Mapillary services were chosen as it is open source allowing the data (processed photos and metadata) to be used for further development and the privacy aspect of the data is ensured by automatically detecting and blurring all faces and license plates. Photos and metadata about their capture hosted in the mapillary server are made available for use within few hours of upload.

Once the uploaded image of the footpath is available, we can retrieve the data via a Python API by specifying the clientID, usernames of the volunteers who captured the data and a bounding box of the latitude and longitude values of the region. If the bounding box position is not known then all the images related to an organisation key will be extracted. A representative query url command: 'https://a.mapillary.com/v3/images?client_id=''&bbox=''&usernames=''&per_page=500&start_time=''&end_time=''&sort_by=key

The headers parameter uses image token value to "Authorization":"Bearer token" and the data is retrived using requests.get(url,headers=header). the data is saved in json file format using 'response.json()' command.

3.2 Data Annotation and Description

Data annotation is generally a time consuming task that requires huge human resources and effort to manually label the data required to train computer vision deep learning models. Based on the requirements gathered during the experience sharing workshops to assess urban elements and user concerns, we identified some of the urban elements with the greatest potential impact:

[3]https://www.mapillary.com/mobile-apps, accessed July 2021

Figure 3: Illustration of some of the challenging scenes where detection of tactile pavement surface

- footpath pavement type (e.g., tactile pavement surface, damaged surface),
- kerbs (e.g., raised or lowered, present or absent),
- bollards,
- street furniture (e.g. benches, "sandwich" advertising boards, enclosed areas for outdoor tables),
- traffic lights, and
- crosswalk lane marking.

Due to the time required to fully annotate data at the pixel level, we decided to focus the pilot study on tactile pavement surface detection to assess the suitability and performance of the gathered dataset for judging the footpath mobility for visually impaired pedestrian users.

During the data acquisition phase 39,642 images were captured between July 2020 and January 2021 under different weather conditions (e.g., sunny, partially cloudy, cloudy, drizzle) and times of the day (e.g., morning, noon, evening, dusk). A total of 2,119 images with two different resolutions – 1024x576, 1024x768 – were selected by visually inspecting each image for the presence of tactile pavement.

The selected images were manually annotated using third party software, LabelImg [13], with the label annotations saved to corresponding TXT file format as (Class_name, bbox_left, bbox_top, bbox_right, bbox_bottom) and we had 2,769 instances of tactile pavement surface. The dataset was divided into trainval (85%) and test (15%) set and again the trainval set was further divided into train (70%) and validation (30%) sets consisting of 1,261 and 540 samples respectively.

To improve the generalization performance of the model and alleviate the bias we employed image augmentation techniques along with random rotation and Gaussian blur to enhance the dataset. The test set with 318 samples was only used for evaluating the final performance of the YOLO-V5 model.

Some of the challenging scenarios observed during annotation of the data are presented in Figure 3. Based on the visual inspection of labelled data, problems like viewpoint variation, occlusion due to people using footpath, illumination conditions, cluttered or similar textured background were evident and present in the proposed dataset.

4 PILOT STUDY: TACTILE PAVEMENT DETECTION AND SEMANTIC SEGMENTATION

The focus of the study was not to come up with new algorithm or approach to improve tactile pavement detection specifically but to investigate the effectiveness of the data acquired and how the data can further enhanced for researchers who are interested in this task. Detection of tactile pavement surfaces still remains an open research problem due to the limited availability of public data. In [5], automatic detection of warning and directional tactile paving surfaces in indoor environment is done by making use of Grey Level CoOccurrence Matrix (GLCM) feature and Support Vector Machine (SVM) to detect and classify the pavement surface. A deep learning based approach was proposed in [2] by combining YOLO-V3 [10] with the DenseNet model to create the YOLOV3Dense model and trained the proposed models on the Marmara Tactile Paving Surface (MDPY) dataset consisting 4580 image samples. However, the dataset used by the authors is not made available for further exploration.

In this section, we evaluate the feasibility of applying computer vision algorithms to detect tactile pavement surfaces to aid users with long canes or other visual impairments. Semantic segmentation was included in the pilot study to support better localisation of the search region and to minimise the computation cost during visual inspection of urban elements and the footpath.

4.1 Tactile Surface detection using YOLO-V5

You only look once (YOLO) is the one of the most popular single stage, state-of-the-art, real-time object detectors widely used among researchers for the task of object recognition. For the detection of tactile pavement surfaces we used the latest architecture from the yolo family: YOLO-V5 (Version 5) [1, 12, 15] shown in Figure 4 to predict class probabilities and bounding box offsets simultaneously from feature maps extracted on an image scene.

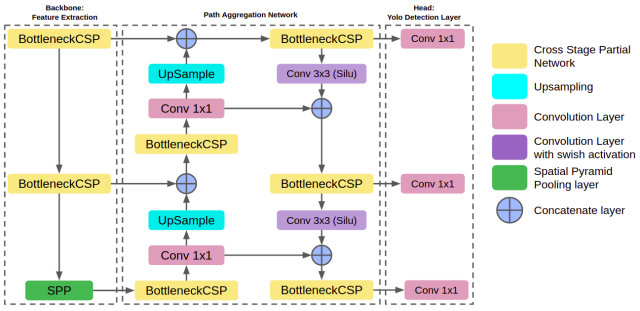


Figure 4: YOLO-V5 network architecture [15] for detecting tactile surface.

The first stage in the YOLO-V5 architecture uses cross stage partial network (CSPNet) [14] as its backbone to extract the features from the input image. With the use of CSPNet architecture the number of model parameters and FLOPS (floating-point operation per second) are reduced as gradient change information in large-scale backbones is integrated into the feature map. This not only ensures improvement in the inference speed and accuracy, but also reduces the overall model size.

After the feature extraction stage, a path aggregation network (PANet) [8] is applied to boost information flow. PANet adopts a modified feature pyramid network (FPN) structure with enhanced bottom-up path augmentation employed to shorten the information path between lower layers and topmost feature. Adaptive feature pooling is used to aggregate the feature grid and all feature levels to make useful information in each feature level propagate directly to subsequent sub-network. Since lower layer features are effectively utilized and propagated between the layers, the accuracy of localising smaller objects is improved and this also helps models to perform well on unseen data.

The last stage is the detection stage which uses the yolo layer with three different sizes (20×20, 40×40, 80×80) of feature maps to achieve multi-scale prediction, enabling the model to handle small, medium, and big objects effectively. This is particularly useful in analysing urban footpath elements as significant differences in object size and viewpoint creates scale variation in the images.

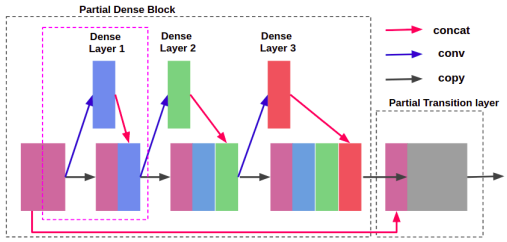


Figure 5: Bottleneck architecture of Cross stage partial dense net(CSPNet) [14]

The code repository of YoloV5 was accessed from the github repository of Ultralytics [1] and is pre-trained on the MS-COCO [7] dataset. The size of the anchors are estimated using the k-mean clustering method and each type of the down-sampling scale is set with three kinds of anchors depicted in Table 1 and a total of nine sizes of anchors are clustered.

The model is trained on an Intel i7 processor with 24GB RAM and a single GPU (GeForce GTX TITANX, 12GB RAM) for 300 epochs. All the input images are resized to 640x640 with batchsize of 8 and the gradients are updated using Adam optimiser with momentum 0.9. Initial learning rate is set to 0.001 and the learning rate is adjusted using the reduce on plateau schedule strategy, which decreases the learning rate by the specified decay percentage when the given metric is stagnating longer than allowed (Patience).

Figure 9 shows the loss (box loss and objectness loss) and performance metric (mAP) curves obtained during training of YOLO-V5 model.

4.2 Semantic Segmentation using ICNet

Image cascade network (ICNet) [17] makes use of the processing efficiency of low resolution images and high inference quality of high-resolution images by cascading the feature maps from various resolution. The network architecture employs cascade feature fusion (CFF) technique to merge extracted feature maps from different

Figure 6: Snapshots of the tactile detection result on test data using YOLO-V5 [15] for detecting tactile surface.

Figure 7: Snapshots of the prediction results of the segmentation on test data using ICNet for detecting tactile surface. Input images are presented in the top row and the corresponding segmented results are presented in the bottom row.

Table 1: Anchors generated using k-mean clustering, RF represents receptive field

small RF	medium RF	large RF
77x16	217x 36	116x 94
123x24	188x 61	156x 95
132x41	311x 64	373x155

resolution input. The network architecture is shown in Figure 8 and it can be seen that the input image is fed into the three different branches where feature maps are generated at 1/4, 1/2 and 1 resolutions respectively. Pyramid scene parsing network (PSPNet) [18] is employed as backbone network at each branch to extract the feature maps and the spatial resolution of the image passed to each of the branch are reduced by 1/8, due to which the predicted output to have resolution 1/32, 1/16 and 1/8 of the original image.

The ICNet was pre-trained on the Cityscape dataset for network initialisation and the model then trained using the mapillary vistas dataset [9] on an Intel i7 processor with 24GB RAM and a single GPU (GeForce GTX TITANX, 12GB RAM) for 100,000 iterations. The Mapillary dataset consists of 80 classes that are reduced to 21 by grouping multiple classes and newly mapped classes along with the class details are depicted in 3. All input images are resized to 1024x2048 with a batchsize of 4 and the gradients are updated using Adam optimiser with momentum 0.9. Initial learning rate is set to 0.0001 and a polynomial learning rate policy is adopted with power 0.9, together with the maximum iteration number set to 100K. Data augmentation contains random mirror and rand resizing between 0.5 and 2.

Table 2: Performance of YOLO-V5 model on the proposed dataset. In the table, N = number of Images, nl = number of labels, Prcn = Precision, Rcll = Recall, mAP = mean Average Precision

	N	nL	Prcn (%)	Rcll (%)	mAP @0.5	mAP @(0.5-0.95)
val	540	690	91.4	84.3	83.7	56.7
test	318	425	91.7	86.4	84.6	55.9

The trained model is used to predict full semantic segmentation for these class on the proposed dataset set and the results are visually inspected for qualitative analysis as full ground truth labels are not available. Figure 10 represents loss and performance metric (Frequency-weighted IoU, mean Accuracy and Overall Accuracy) curves obtained during training of ICNet model.

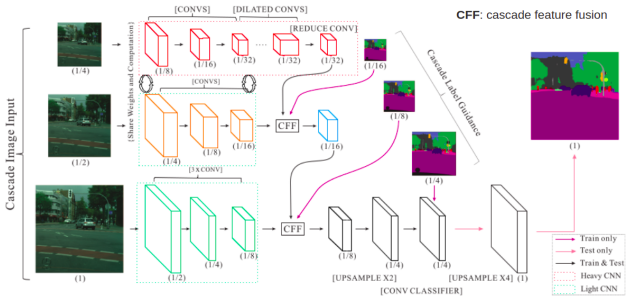


Figure 8: Network architecture of ICNet. The figure depicted here is taken from the original article [17].

5 RESULTS AND DISCUSSION

In this section, we report the performance of the YOLO-V5 and ICNet deep learning models on the proposed dataset. A snapshot of detected tactile pavement surface images are presented in Figure 6 and the performance of YOLO-V5 model on the proposed validation and test data are depicted in Table 2. The performance in mean average precision (mAP) of the model is ≈ 86.4 % at 0.5 intersection over union (IOU) threshold value and the value is ≈ 55.9 % when the threshold value is varied between 0.5 to 0.95 on test dataset. Based on the obtained results we can conclude that the data can be used for further exploration with multiple urban elements.

Table 3 depicts the performance of the ICNet on mapillary vistas training dataset with 21 classes and the predicted segmented results are presented in Figure 7. Qualitative analysis of ICNet model is not done on proposed data but from the visual inspection we can localise the search region to perform in depth analysis to extract useful inference. One of the use cases derived based on this inference is presented in Figure 11a where the footpath attributes need to be updated based on the visual inspection of the region. This can be simplified by employing the segmentation and detection as plug-in modules that filter and highlight the segments of footpath region in the scene and the detection module helps in the localising the position of the tactile pavement surface to map the region in OpenStreetMap. The feasibility analysis of the plug-in modules to aid the mappers while validating the mapped information is planned for the next phase of work focusing on enhancing the proposed dataset with more urban scenarios and investing time required to perform pixel level annotation for semantic segmentation.

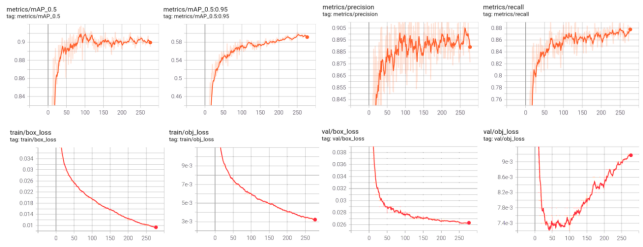


Figure 9: snapshot of the training and validation graph obtained during training of YOLO-V5. Top row: from left to right represents the mean Average Precision (mAP) with IOU=0.5, mAP with IoU between range 0.5-0.95, precison and recall curve respectively. Bottom row: from left to right represents box loss and objectness loss at training and validation stage. legend x-axis represents number of epochs and y-axis represents the value

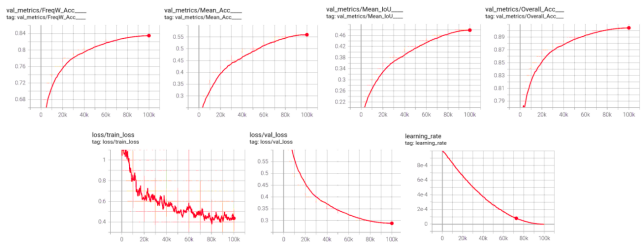


Figure 10: snapshot of the training and validation graph obtained during training of ICNet. Top row: from left to right represents validation metric graph of Frequency-weighted IoU, Mean Accuracy, Mean Iou and Overall Accuracy respectively. Bottom row: from left to right represents graph of training loss, validation loss and learning rate. legend x-axis represents number of iterations and y-axis represents the value

6 CONCLUSION

In this paper, we describe the aims and data acquisition process from the Crowd4Access project, propose an urban street-view image dataset to assess pedestrian mobility on footpaths and make the resulting data available to researchers to further explore this topic.

The data acquisition phase was a very interesting and challenging task as it required engaging and understanding the needs and challenges that of those who use mobility aids (e.g. long cane, guide dogs, crutches, and wheelchairs) and who would therefore require footpaths to have particular characteristics. The crowdsourced and citizen-led efforts to take photographs of urban footpaths and to use

Table 3: Predicted mIoU in % on mapillary dataset [9] with image resolution 1024 × 2048. Downsampling ratio (DR) during testing is set to 3 (e.g, DR=4 represents testing at resolution 256 × 512). In the table, m_id, n_id, LM, CW, TS and NA represents original class id in mapillary dataset, new class id assigned after grouping, Lane Marking, CrossWalk, Traffic Sign and Not Applicable respectively

class	m_id	n_id	mIoU(%)	class	m_id	n_id	mIoU(%)	class	m_id	n_id	mIoU(%)
unlabelled	65	0	NA	Motorcyclist	21	4		Other Vehicle	59	8	
Bird	0	1	74.16	Other Rider	22	4		Trailer	60	8	
Ground Animal	1	1		CW-plain	8	5	48.31	Truck	61	8	
Barrier	5	1		LM-CW	23	5		Wheeled Slow	62	8	
Wall	6	1		LM-general	24	5		Car Mount	63	8	
Parking	10	1		Sky	27	6	97.25	Ego Vehicle	64	8	
Rail Track	12	1		Banner	32	7	17.86	Curb	2	9	42.03
Bridge	16	1		Billboard	35	7		Curb Cut	9	9	
Building	17	1		Catch Basin	36	7		Pedestrain Area	11	10	56.50
Tunnel	18	1		CCTV Camera	37	7		Side Walk	15	10	
Mountain	25	1		Junction Box	39	7		Bench	33	11	NA
Sand	26	1		Mailbox	40	7		Bike Rack	34	12	NA
Snow	28	1		Manhole	41	7		Pole	45	13	30.64
Terrain	29	1		Phone Booth	42	7		Traffic Light	48	14	38.42
Fire Hydrant	38	1		Street Light	44	7		TS-(Frame)	46	15	45.57
Fence	3	2	39.78	Bicycle	52	8	81.12	TS-(Back)	49	15	
Guard Rail	4	2		Boat	53	8		TS-(Front)	50	15	
Bike Lane	7	3	82.61	Bus	54	8		Trash Can	51	16	11.26
Road	13	3		Car	55	8		Utility Pole	47	17	30.79
Service Lane	14	3		Caravan	56	8		Pothole	43	18	NA
Person	19	4	45.57	Motorcycle	57	8		Vegetation	30	19	84.68
Bicyclist	20	4		On Rails	59	8		Water	31	20	14.85

the mapillary framework to upload, process and retrieve the images demonstrated how crowdsourcing of quality images in a variety of environments (location, lighting, weather) enables images to be collected in a structured manner by non-experts and to produce a dataset suitable for use in computer vision models.

From the initial pilot study it was evident from the performance on tactile pavement surface detection using YOLO-V5 and the predicted results of semantic segmentation task using ICNet that the proposed dataset can be used for computer vision applications. The outcome of the pilot study indicated the need for more footpath images to improve the performance of the deep learning model and to handle the challenging use cases. Furthermore, the study also motivated us to enhance the dataset by annotating other urban elements for the detection task and demonstrated why it would make sense to invest time in pixel-level annotation of scenes to generate rich contextual information with pedestrian perspective images for better understanding of the footpath and its attributes in the context of mobility.

Finally the experience of conducting a citizen science project on the important topic of urban mobility has enabled us to bring together disparate groups of researchers (social policy, usability, computer vision, data analytics) and to talk directly with effected end users about their experiences and needs.

ACKNOWLEDGMENTS

This publication has emanated from research supported by Science Foundation Ireland (SFI) under Grant Numbers SFI/12/RC/2289_P2 and 16/SP/3804, co-funded by the European Regional Development Fund.

REFERENCES

[1] [n. d.]. ultralytics/yolov5: v4.0 - nn.SiLU() activations, Weights & Biases logging, PyTorch Hub integration | Zenodo. https://zenodo.org/record/4418161#.YQXtK6IzZuQ. (Accessed on 07/25/2021).

[2] Abdulsamet Aktaş, Buket Doğan, and Önder Demir. 2020. Tactile paving surface detection with deep learning methods. *Journal of the Faculty of Engineering and Architecture of Gazi University* 35, 3 (2020), 1685–1700.

[3] Jerome Bickenbach. 2011. The world report on disability. *Disability & Society* 26, 5 (2011), 655–658.

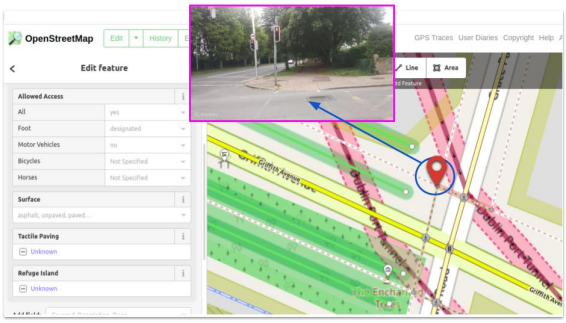
[4] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Scharwächter, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. 2015. The cityscapes dataset. In *CVPR Workshop on the Future of Datasets in Vision*, Vol. 2.

[5] Daniel Centeno Einloft, Marcelo Cabral Ghilardi, and Isabel Harb Manssour. 2016. Automatic detection of tactile paving surfaces in indoor environments. In *Workshop of Undergraduate Works (WUW) in the 29th Conference on Graphics, Patterns and Images (SIBGRAPI'16), 2016, Brasil.*

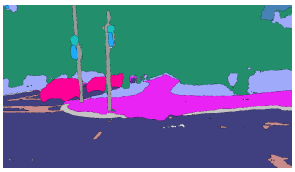
[6] Mordechai Muki Haklay, Daniel Dörler, Florian Heigl, Marina Manzoni, Susanne Hecker, and Katrin Vohland. 2021. What is citizen science? The challenges of definition. *The Science of Citizen Science* (2021), 13.

[7] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. 2014. Microsoft coco: Common objects in context. In *European conference on computer vision*. Springer, 740–755.

[8] Shu Liu, Lu Qi, Haifang Qin, Jianping Shi, and Jiaya Jia. 2018. Path aggregation network for instance segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition.* ", ", 8759–8768.



(a) Usecase scenario of localising tactile surface in OSM



(b) Segmentation applied on retried image

(c) Detection and localisation of Tactile pavement surface

Figure 11: OSM Uscase: (a) Represents the scenario of missing attribute in the footpath/sidewalk in OSM, (b) is the retrieved mapillary image at the pinned location is sBvJZr0JbT6Bh-bynZKc7Q and (c) Segmented image regions obtained using ICNet

[9] Gerhard Neuhold, Tobias Ollmann, Samuel Rota Bulo, and Peter Kontschieder. 2017. The mapillary vistas dataset for semantic understanding of street scenes. In *Proceedings of the IEEE international conference on computer vision.* ", ", 4990–4999.

[10] Joseph Redmon and Ali Farhadi. 2018. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767* (2018).

[11] Joseph Roche, Aoibhinn Ní Shúilleabháin, Peter Mooney, Gillian Barber, Laura Bell, and Cliodhna Ryan. 2021. Citizen Science in Ireland. *Frontiers in Communication* 6 (2021), 19. https://doi.org/10.3389/fcomm.2021.629065

[12] Jacob Solawetz. 2021. How to Train YOLOv5 On a Custom Dataset. https://blog.roboflow.com/how-to-train-yolov5-on-a-custom-dataset/

[13] Tzutalin. 2021. Tzutalin. LabelImg. Git code (2015). https://github.com/tzutalin/labelImg. https://github.com/tzutalin/labelImg (Accessed on 07/29/2021).

[14] Chien-Yao Wang, Hong-Yuan Mark Liao, Yueh-Hua Wu, Ping-Yang Chen, Jun-Wei Hsieh, and I-Hau Yeh. 2020. CSPNet: A new backbone that can enhance learning capability of CNN. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops.* ", ", 390–391.

[15] Renjie Xu, Haifeng Lin, Kangjie Lu, Lin Cao, and Yunfei Liu. 2021. A Forest Fire Detection System Based on Ensemble Learning. *Forests* 12, 2 (2021), 217.

[16] Fisher Yu, Haofeng Chen, Xin Wang, Wenqi Xian, Yingying Chen, Fangchen Liu, Vashisht Madhavan, and Trevor Darrell. 2020. Bdd100k: A diverse driving dataset for heterogeneous multitask learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.* 2636–2645.

[17] Hengshuang Zhao, Xiaojuan Qi, Xiaoyong Shen, Jianping Shi, and Jiaya Jia. 2018. Icnet for real-time semantic segmentation on high-resolution images. In *Proceedings of the European conference on computer vision (ECCV).* ", ", 405–420.

[18] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. 2017. Pyramid scene parsing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition.* 2881–2890.