International Journal of Applied Earth Observations and Geoinformation 108 (2022) 102748

Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

journal homepage: www.elsevier.com/locate/jag

Power to the people: Applying citizen science and computer vision to home mapping for rural energy access

Alycia Leonard *, Scot Wheeler, Malcolm McCulloch

Department of Engineering Science, University of Oxford, Oxford, United Kingdom

ARTICLE INFO

Keywords:
Citizen science
Computer vision
Electricity access
Rural mapping
Satellite imagery
Remote sensing

ABSTRACT

To implement effective rural electricity access systems, it is fundamental to identify where potential consumers live. Here, we test the suitability of citizen science paired with satellite imagery and computer vision to map remote off-grid homes for electrical system design. A citizen science project called "Power to the People" was completed on the Zooniverse platform to collect home annotations in Uganda, Kenya, and Sierra Leone. Thousands of citizen scientists created a novel dataset of 578,010 home annotations with an average mapping speed of 7 km^2/day. These data were post-processed with clustering to determine high-consensus home annotations. The raw annotations achieved a recall of 93% and precision of 49%; clustering the annotations increased precision to 69%. These were used to train a Faster R-CNN object detection model, producing detections useful as a first pass for home-level mapping with a feasible mapping rate of 42,938 km^2/day. Detections achieved a precision of 67% and recall of 36%. This research shows citizen science and computer vision to be a promising pipeline for accelerated rural home-level mapping to enable energy system design.

1. Introduction

This paper addresses the question: *Can we accurately map remote off-grid populations from satellite imagery at scale for energy access system design using citizen science and computer vision?*

To implement effective electrical systems, it is fundamental to identify where potential consumers live. Without location data, it is impossible to specify a topology to connect all consumers at lowest cost. Additionally, location data is needed to determine electrical demand, which derives from a complex mix of continuously evolving temporally- and spatially-dependent social practices (Shove and Walker, 2014; Hui and Walker, 2018). The effects of spatial demand variance are particularly important in small-scale electrical systems (e.g. mini-grids, solar home systems), which may be the most affordable option in many remaining rural off-grid contexts (Energy Sector Management Assistance Program (ESMAP), 2019), and whose designs are more tightly bound to the spatial distribution of demands than larger grids.

As households constitute the majority of potential consumers in rural off-grid areas, home-level location data are necessary for system design. Commonly-available aggregated population density data does not always provide all the information required. Consider, for instance, designing an electrical system for a clustered community far from roads, a community linearly distributed along a major highway, and homes dotted on plots of agricultural land. While each could have the same population density per km^2, different topologies and technologies may be required based on the configuration of homes.

Accurate and complete home-level location data for off-grid populations are unfortunately difficult to find. Typical population raster dataset resolutions range from 30" (i.e. 1 km at equator) in the Gridded Population of the World (GPW) dataset (Center for International Earth Science Information Network (CIESIN), 2018), to 1" (i.e. 30 m at equator) in the Facebook High Resolution Settlement Layer (HRSL) (Facebook Connectivity Lab, 2016). These resolutions can misrepresent small communities, or those whose homes are unevenly distributed. Census data suffer the same limitations, and can also be hindered by the irregular shapes and sizes of enumeration areas or political districts used to aggregate data. Those datasets that do capture home-level data tend to suffer from accuracy, completeness, or recency issues. For instance, the OpenStreetMap (OSM), a community-driven global open-source effort to map homes and geographic features (OpenStreetMap Contributors, 2021), has no regulated update frequency or temporal data uniformity. It also has gaps which unhelpfully tend to correspond with rural off-grid regions (see Fig. 1 for an example). Furthermore, "non-Western" housing styles such as small thatched-roof homesteads far from paved

* Corresponding author at: Department of Engineering Science, University of Oxford, Parks Road, Oxford OX1 3PJ, UK.
E-mail addresses: alycia.leonard@eng.ox.ac.uk (A. Leonard), malcolm.mcculloch@eng.ox.ac.uk (M. McCulloch).

https://doi.org/10.1016/j.jag.2022.102748
Received 10 November 2021; Received in revised form 8 March 2022; Accepted 13 March 2022
Available online 13 April 2022

roads tend to be underrepresented. While OSM has been tackling this issue in specific areas for humanitarian purposes (Humanitarian OpenStreetMap Team, 2021), the broader issue persists.

Citizen science represents one possible avenue to fill the home-level mapping data gap. This method integrates enthusiastic non-specialists (i.e. citizens) into the scientific process to collect or process data, collaborate on research design, or co-create research projects. It can enable substantial data collection or processing at low cost and bring diverse perspectives to the research, improving inclusivity. While citizen science has traditionally been most successful in ecology and astronomy (Dickinson et al., 2010), it is increasingly being applied in other natural and social sciences. It has been used to collect various types of geolocated data: for instance, projects like iNaturalist (Horn et al., 2018; Wilson et al., 2020; Aristeidou et al., 2021) engage citizens to capture geolocated information using web-based interfaces and a smartphone camera or GPS, while others like Missing Maps (Scholz et al., 2018; Givoni, 2016; Herfort et al., 2016) engage citizens to identify specific features (e.g. roads, buildings, or land use) in existing maps or geographic datasets. Furthermore, it can be used to gather or classify large datasets for artificial intelligence (AI) research. Large annotated image datasets generated through citizen science – such as the large-scale Penguin Watch (Jones et al., 2018) and Snapshot Serengeti (Swanson et al., 2015) projects – can be used in AI-based computer vision (CV) to classify images or detect features. Where image features are geolocated, there is exciting potential to train AI algorithms to map features such as buildings at scale.

While the use of CV for automated building detection is not new, its combination with robust and representative citizen-science-generated training data is a powerful and novel contribution. Object-detection-based home mapping research in the literature includes the "You Only Look Twice" algorithm proposed by Van Etten, 2018, which detects cars and homes amongst other features on Planet and DigitalGlobe satellite images alongside aerial images with GSD as high as 0.15 m, achieving an F1 score of 0.6 to 0.9 depending on the category detected. In this case, the SpaceNet (Van Etten et al., 2018) and Cars Overhead with Context (Mundhenk et al., 2016) datasets were used alongside researcher-generated annotations. Others have used the Xview dataset (Lam et al., 2018) and super-resolution techniques or the UC Merced land use dataset[1] and a two-stage classification and object detection pipeline (Shermeyer and Van Etten, 2019) to achieve similar results. Semantic segmentation has also been proven as a feasible method for building detection, for instance being used to generate three-channel (e.g. road, building, or background) labelled images (Saito et al., 2016) using the 1 m GSD Mnih dataset (Mnih, 2013), or for pixel-wise building prediction from 0.3 m GSD imagery based on GIS data for home locations in Washington DC (Yuan, 2018). Note, however, that these applications are largely based in urban contexts where plentiful training data is more likely to be available. Datasets such as the Mnih dataset or the Inria dataset covering urban contexts in the United States and Austria (Maggiori et al., 2017) are frequently applied in training and testing (e.g. Pan et al., 2016). Other similar works have leveraged 0.075 m aerial data openly available in Christchurch New Zealand and 0.8 m urban imagery in China (Yi et al., 2019; Qin et al., 2019). While the algorithmic improvements represented in such studies are commendable and valuable, without availability of high-quality training data suitable to the rural context, they cannot be applied to map buildings in off-grid LMIC regions. This is where the power of citizen science can create impact in automated building detection.

In this paper, we argue that a combined citizen science and CV pipeline can effectively map rural off-grid homes at scale for energy system design. To test this, a project called "Power to the People"[2] (PTTP) engaged citizen scientists to locate homes in rural Kenya, Uganda, and Sierra Leone, three countries with ongoing rural electrification efforts, home location data gaps, and diverse rural home styles. A sample of rural areas in each country was mapped by citizen scientists using very high resolution satellite imagery. Thousands of citizen scientists classified hundreds of thousands of images, which were used to experimentally train an object detection algorithm to map homes at scale.

To benchmark this work, we draw on the rich literature of land administration in rural areas of LMICs, given the dearth of automated

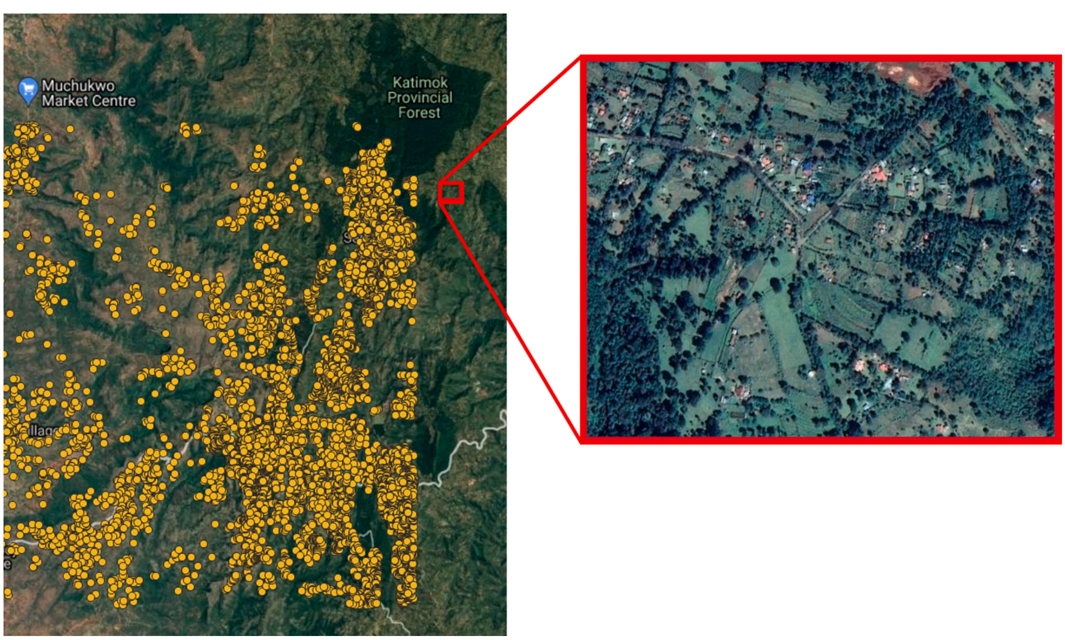

Fig. 1. Sample of OpenStreetMap (OSM) building locations in Baringo County, Kenya (0°34'01"N 35°47'37"E). Note the sharp edge to the right of which there are no buildings identified. However, as shown in the expanded area, there are buildings present. Such gaps are common in rural areas. Data from Geofabrik (2018) overlaid on Google Earth imagery.

[1] http://weegee.vision.ucmerced.edu/datasets/landuse.html
[2] www.zooniverse.org/projects/alycialeonard/power-to-the-people

building detection research focused in these areas. Success criteria to validate the presented argument were:

- Citizen scientists must produce home annotations on very high resolution satellite imagery with accuracy ⩾62% and speed ⩾0.25 km^2/day [3].
- A CV algorithm must detect homes at a similar level of accuracy, a lower cost, and a faster rate than the citizen science approach.
- The entire pipeline should map off-grid areas at a cost < $465[4] per km^2.

The remainder of this paper proceeds as follows. Section 2 outlines the data and materials used and Section 3 discusses the methodology, including the citizen science workflow, the CV algorithm, and the metrics to evaluate the work. Section 4 details the results of the citizen science project and the CV training and application. Section 5 elaborates upon insights and implications and Section 6 offers conclusions.

2. Data and materials

The data and materials used in this research included the satellite imagery to be annotated, the online citizen science platform, and CV tools.

2.1. Satellite imagery

Satellite imagery was used as input data for citizen science annotation. This allowed for high temporal specificity (i.e. the resulting home maps/annotations represent the point in time at which imagery was captured). The near-worldwide coverage of satellite imagery is useful in off-grid rural regions which can be data-deserts.

For PTTP satellite images, a ground sample distance (GSD) or resolution of ⩽1 m and minimal cloud cover were required to enable home-level annotation by citizen scientists. Panchromatic and multispectral bands were used to produce pansharpened images. Recent images (i.e. captured from 2015 to 2019) were selected to ensure that recent housing styles were captured in the resulting training set. Based on these requirements, satellite images for this project were captured by the Superview-1 constellation, the Disaster Monitoring Constellation 3 (DMC-3), and the Korea Multi-Purpose Satellite (KOMPSAT) 3A. Details on each satellite are presented in Table 1. Both archive and new-tasked imagery were used.

Satellite imagery samples in Kenya, Sierra Leone, and Uganda were chosen as these countries have ongoing rural electrification efforts at different stages, home location data gaps in OSM, and diverse rural home styles (e.g. agricultural, clustered community, refugee, etc.). Image samples were selected which represented rural under-mapped and under-electrified areas. Areas were also chosen with home styles which were under-represented in available training sets for home detection: these include thatched roofs; homesteads constituted by multiple structures; homes far from roads or surrounded by vegetation; and homes in refugee settings.

2.2. Citizen science platform

Online citizen science was applied here to leverage enthusiasm amongst the existing citizen science community while avoiding exacerbating disproportionate time poverty experienced by those in study regions (Charmes, 2006; Kes and Swaminathan, 2006). This approach required an attractive web-based citizen science interface with phone, tablet, and computer compatibility. A platform with a large and diverse existing volunteer base and a track record of successful projects was desired, and a dedicated project landing page was required to be shared in project promotion. The interface needed to be free and allow for rapid, easy project construction, as no budget was available for extended web development. Built-in image annotation tools were required, as were the abilities to host large imagery datasets and to export pixel-level annotations. The Zooniverse online citizen science platform was selected to meet these requirements. This is the largest and most popular platform for online citizen science, boasting a contributor base of over 2 million contributors (The Zooniverse Team, 2021).

2.3. Computer vision toolkit

Computer vision (CV) was used to scale the citizen science results given its ability to automate tasks which humans can accomplish by eye. A free and open source CV toolkit was used to maximise the usefulness of the research in resource-constrained contexts (i.e. rapidly electrifying low- and middle-income countries). Straightforward implementations of state-of-the-art CV algorithms compatible with graphics processing units (GPU) were required. To meet these needs, Tensorflow was selected. Tensorflow offers a user-friendly, GPU-compatible computer vision application programming interface (API) for many cutting edge algorithms. This API has minimal barrier to entry for use aside from basic Python literacy, and so is well-suited for practical application in home mapping for grid design. All training and detection here were completed on two Nvidia GeForce RTX 2080 GPUs.

3. Methods

The general methodology for this work is illustrated in Fig. 2. Satellite imagery is pre-processed and annotated by citizen scientists. Annotation data is then clustered and used to train an object detection algorithm to map at scale. Citizen science and object detection outputs are evaluated.

3.1. Satellite imagery pre-processing

Satellite imagery was pre-processed to facilitate online annotation through pansharpening, rendering, tiling, upsampling, and metadata recording. The original 16-bit GeoTIFF satellite images were first pansharpened (i.e. the lower-quality multispectral and higher-quality panchromatic images were merged to create a higher-quality colour image). Image histograms were also manually colour-corrected for a natural colour appearance. Images were then rendered in 8-bit color and split into small square[5] tiles (i.e. 256 × 256 for panchromatic and pansharpened images, and 64 × 64 for multispectral images). Tiles with any heavy cloud cover were discarded. Georeferencing information for each tile (i.e. maximum and minimum coordinates) was saved. Panchromatic and pansharpened tiles were upsampled by 200% while multispectral tiles were upsampled by 800% to obtain 512 × 512 tiles. These were converted to PNG format and uploaded to the Zooniverse platform. Each "subject" classified by citizen scientists included a pansharpened, panchromatic, and multispectral tile of the same area. An example of the three imagery types is shown in Fig. 4.

3.2. Citizen science annotation workflow

The citizen science workflow was designed based on the goals of the research: the main goal being to map home locations for electrical

[3] Accuracy and speed from (Asiama et al., 2017). Assumes one project employee.

[4] Lowest cost achieved in the case studies referenced in Enemark et al. (2014). Assumes each parcel is 1.29 hectare, the average parcel size from Asiama et al. (2017).

[5] Some samples on image edges were not square. These were discarded if too thin to be practical for annotation, or if there were likely to be no homes based on visual inspection.

Table 1
Satellite constellations which acquired imagery for the PTTP citizen science project.

Name	Operator	Launch date(s)	Number of satellites	Panchromatic GSD (m)	Multispectral GSD (m)
Superview-1	Beijing Space View Tech Co Ltd	2016–12-28 2018–01-09	4	0.5	2
DMC-3	DMC International Imaging Ltd	2015–07-10	3	1	4
KOMPSAT-3A	Korea Aerospace Research Institute	2015–03-25	1	0.55	2.2

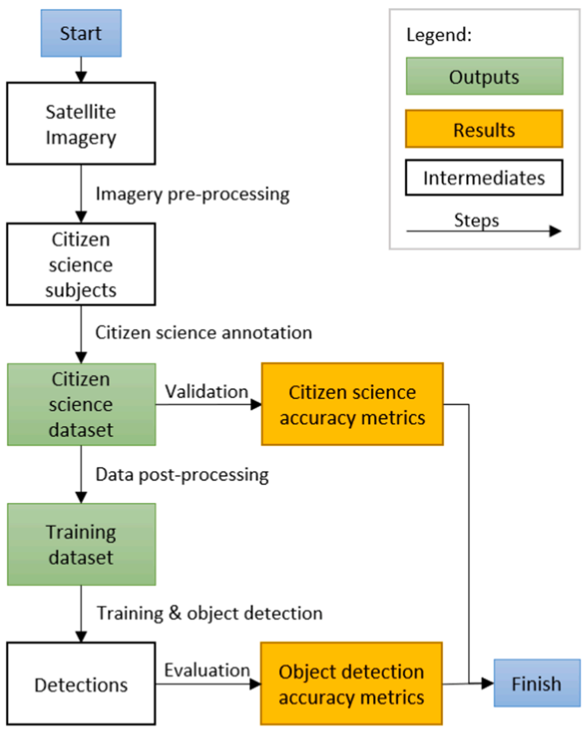

Fig. 2. Overview of the research methodology, using an activity-on-arrow representation.

system planning, and a secondary goal being to collect information about home size and roof material (e.g. thatch or metal) to serve as proxies for demand or willingness to pay, which are key factors in electrical system design. The material and size of a roof also dictate whether photovoltaic (PV) panels can be mounted, and if so, the capacity which can be installed. These are important parameters to electrical design, particularly in rural developing regions where rooftop PV might be the most economic option.

Given these goals, the annotation workflow proceeded as shown in Fig. 3. The citizen scientist was presented with a subject, and asked whether they could see any homes. The word “home” was used instead of building as we were uncertain whether citizen scientists would interpret all of the rural home structures as “buildings” (i.e. given their different construction styles including semi-permanent refugee homes and thatched-roof buildings). However, we did believe they would recognize them as homes. If they did not see any homes, they were presented with a new subject. If they did, they were asked to annotate them. For each annotation, they were asked to identify the roof color, under the assumption that thatch would appear brown and metal would appear light or white. Options were “white or light-coloured”, “brown”, or “other”, as shown in Fig. 5. They were also asked to identify roof shape from the options “square or rectangular”, “circular or rounded”, or “other”. When finished, they were asked whether they had annotated all visible homes, or whether there were too many to annotate them all[6]. They could then confirm their annotations and move on.

Key to this workflow is the specific annotation tool selected. The Zooniverse platform offers many tools (e.g. bounding boxes, circular and point annotations, free-form polygons). Here, the type of annotation dictated the types of CV which could subsequently be used for AI-driven mapping. Two CV algorithm families are most appropriate for home mapping: segmentation (i.e. particular features or classes of objects are identified at pixel-level as arbitrary shapes) and object detection (i.e. particular classes of objects are identified within bounding boxes). Segmentation requires training annotations which follows feature contours, while object detection requires bounding boxes. Annotating accurate contours is more challenging than accurate bounding boxes using Zooniverse tools. We suspected that this might cause citizen scientist frustration, harming data quality and slowing the project. So, a rotatable bounding box annotation tool was chosen to prioritize simplicity and user-friendliness, and a tandem decision was made to use object detection.

Each subject was classified (i.e. viewed and annotated) by multiple citizen scientists. If the first five citizen scientists to classify a subject all indicated that there were no visible homes, that subject was “retired” (i.e. removed from the workflow). If any of the first five citizens scientists did see any homes, the subject was classified by ten citizen scientists before retirement.

3.3. Annotation post-processing

Annotation data was exported from Zooniverse as a CSV containing nested JSON. Annotations made when the project was not “live” (i.e. during development, beta testing, or after termination) were discarded. They were then post-processed to identify where contributors agreed that there is a home using clustering. While there are many clustering algorithms to choose from (based on density, centroid, hierarchy, and other definitions of similarity), here, we wanted to cluster based on the density of annotations in two-dimensional space. An algorithm that allows outliers and which can adapt to variable annotation accuracy and quantity was desired. To this end, Hierarchical Density-Based Clustering of Applications with Noise (HDBSCAN*) was selected. HDBSCAN* builds on the popular DBSCAN algorithm by producing a hierarchical set of DBSCAN clusters as the ϵ parameter is varied. It includes a method to optimally “cut” through the hierarchical tree where the resulting clusters are most stable and persistent (Campello et al., 2013). As the results of HDBSCAN* depend highly on the minimum cluster size (m_{clSize}) hyperparameter selected, a range of values were investigated to determine where performance was best based on the metrics in Section 3.4. Annotation pixel coordinates were mapped to geographic coordinates in the satellite image’s original coordinate reference system based on the minimum and maximum x and y coordinates of the tile.

3.4. Citizen science data validation

A set of “gold standard” annotations was generated by the research team on a random sample of 188 images which contained HDBSCAN*

[6] This question was added based on urgent feedback that some images contained too many homes to label. These images were typically from peri-urban regions or crowded refugee camps.

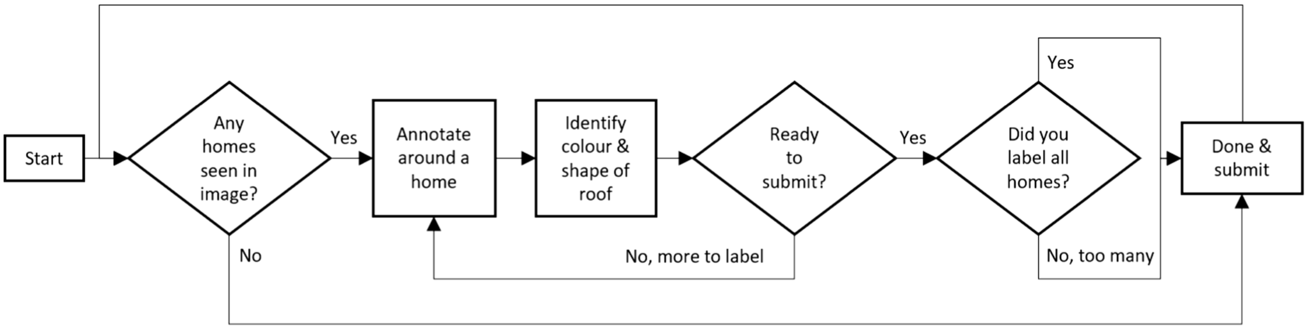

Fig. 3. Satellite imagery annotation workflow in the PTTP citizen science project.

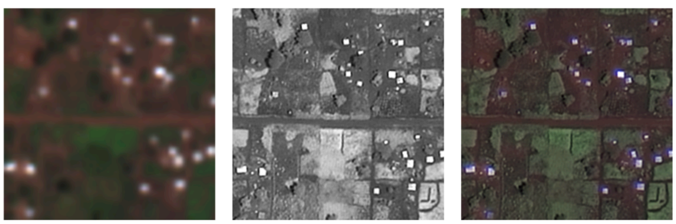

Fig. 4. Examples of the three imagery types used in annotation in the PTTP citizen science project: multispectral visualised in natural colour (left), panchromatic (center), and pansharpened (right).

clusters for $m_{clSize} = 5$. This sample was used to calculate precision (P), recall (R), and F_1 score (i.e. the harmonic mean of P and R) for citizen science annotations, where these were treated as predictions and gold standard data were treated as ground truths. P, R, and F_1 are defined as:

$$P = \frac{T_P}{T_P + F_P} \tag{1}$$

$$R = \frac{T_P}{T_P + F_N} \tag{2}$$

$$F_1 = 2\frac{P \times R}{P + R} \tag{3}$$

where T_P represents true positives (i.e. positive predictions which agree with ground truth), F_P represents false positives (i.e. positive predictions which disagree with ground truth), and F_N represents false negatives (i.e. negative predictions which disagree with ground truth).

To use these metrics, a threshold value must be defined to differentiate between T_P and F_P, and to identify F_N (i.e. ground truths which had been missed). This threshold, defined as the intersection over union (IoU), can be intuitively understood as a certain amount of overlap between a prediction and a ground truth to classify that prediction as correct. IoU can be calculated for a prediction A and ground truth B as:

$$IoU = \frac{A \cap B}{A \cup B} \tag{4}$$

An IoU threshold of 0.5 was used here. This means that $IoU \geqslant 0.5$ between a prediction and any ground truth was counted as T_P, $IoU < 0.5$ between a prediction and any ground truth was counted as a F_P, and $IoU < 0.5$ between a ground truth and any prediction was counted as a F_N.

3.5. Object detection algorithm

The object detection algorithm used here had to meet speed and accessibility requirements pertinent to mapping for rural electrification. While algorithm speed had to be suitable for the electrical design process, there was no need for real-time results (e.g. training could take days and detection minutes to hours). As discussed in Section 2.3, a user-friendly and open-source method accessible in low-resource contexts was required.

We chose to use an R-CNN algorithm for the user-friendliness of

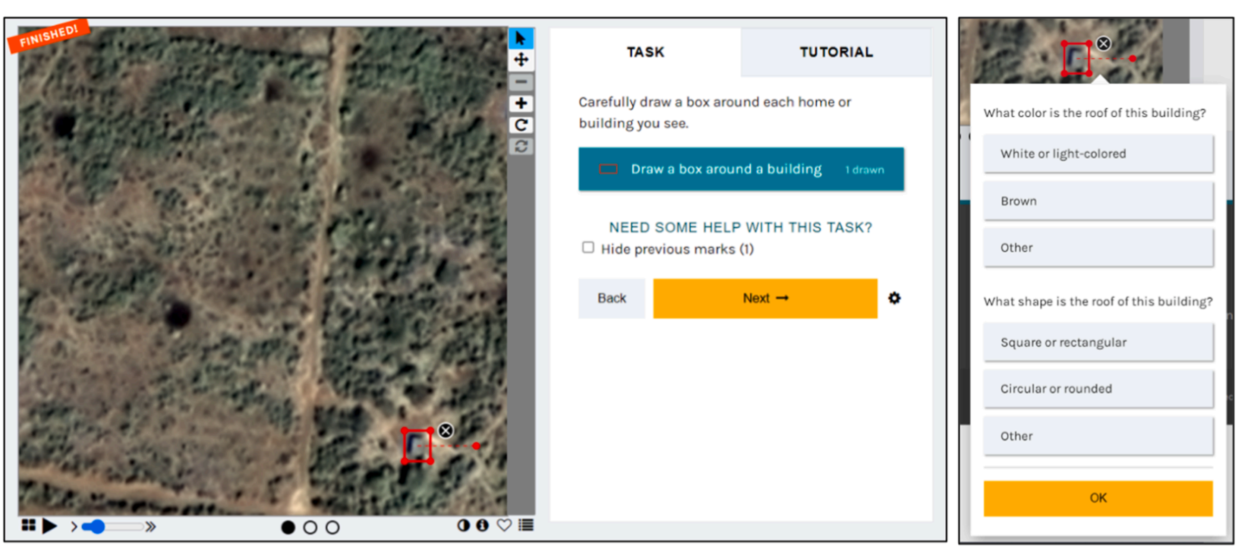

Fig. 5. Annotation interface (left) and follow-up questions (right) on the PTTP citizen science project. The red rectangle is a home annotation, which can be rotated by dragging the red dot to its right. On-screen controls are available to cycle through three image types, invert colour, zoom in or out, shift the image, hide previous annotations, and rotate the image. Please disregard the "finished" banner.

open-source implementations compared to alternatives like YOLO. We made a judicious choice of the Faster R-CNN variant (Ren et al., 2015), a high performing R-CNN which uses a region proposal network (RPN) to predict where there may be an object in an image and then predicts object class for these regions using a single pass through a CNN. Its RPN shares convolutional features with the detection network, minimising the region proposal bottleneck found in previous R-CNN versions (Ren et al., 2015).

Specifically, a Faster R-CNN implementation from the Tensorflow 1 model zoo (Shi and Rathod, 2021) with an Inception v2 RPN was used. This implementation was pretrained with the Common Objects in Context (COCO) dataset; we made use of transfer learning to leverage low-level features (e.g. corners, edges, etc.) learned from generalized object detection in COCO. It has been timed at 58 ms per 600×600 image using an NVIDIA GeForce GTX TITAN X GPU[7] (Shi and Rathod, 2021). As each tile of the highest-resolution imagery used in this work represents approximately 0.0164 km^2, approximately 24,400 km^2/day could be mapped per day with the specified GPU using this implementation according to the benchmarked speed. For context, this translates to mapping the entirety of London in about an hour and a half, or all of Kenya in about 24 days. Note that since this speed was benchmarked on a single Nvidia GeForce GTX Titan X GPU, a somewhat dated model first released in 2015, it is not expected to exactly match the speeds achieved in the following experiments, which used two more up-to-date Nvidia GeForce RTX 2080 GPUs.

This model was trained with the citizen science annotations. Minimum and maximum x and y values from annotation corners were used as training labels. Training parameters were adapted from sample parameters published by Tensorflow for our selected Faster R-CNN implementation[8] initially configured for transfer learning with the Oxford-IIIT Pets Dataset[9]. We added the following data augmentation options to the random horizontal flipping specified: random brightness, contrast, hue, and saturation adjustments; random colour distortions; and random image cropping. We tested learning rates from 10^{-2} to 10^{-9} for 10^5 steps to choose a rate for our work. The rate which produced the best results was 10^{-3}, as shown in Fig. 6, so this rate was used going forward.

3.6. Object detection evaluation

The performance of the object detection algorithm is first evaluated using the average precision (*AP*) and average recall (*AR*) metrics

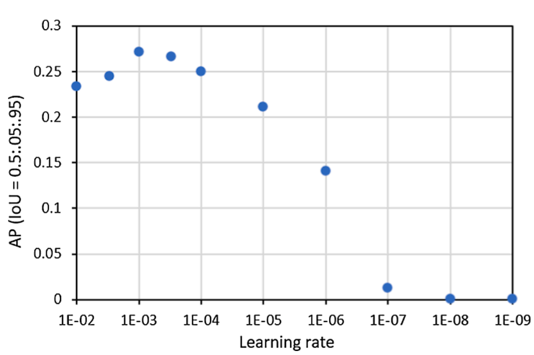

Fig. 6. Experimental results to select learning rate for object detection. Best performance is at 10^{-3} (i.e. $1E-3$).

employed in the COCO competition (Common Objects in Context (COCO), 2021). *AP* describes how many predictions were T_P over various R values (i.e. as confidence varies). It is defined as the area under the interpolated *P-R* curve for any object class. *AR* describes how many T_P were "recalled" (i.e. predicted) over various *IoU* values. It is defined as the mean R over *IoU* values between 0.5 and 1.0, equal to twice the area under the *R-IoU* curve.

The specific metrics used here are listed in Table 2. They evaluate performance over a range of *IoU* values (i.e. 0.5, 0.75, and 0.5 to 0.95 in 0.05 steps), object sizes (i.e. "small" area $< 32 \times 32$, "medium" $32 \times 32 <$ area $< 96 \times 96$, "large" area $> 96 \times 96$) and detections per image (i.e. 1, 10, 100). Ten-fold cross-validation was used – that is, training images were divided into ten random partitions and the model was trained ten times, during each of which a different partition was set aside for testing. The *AP* and *AR* values listed in Section 4 provide results for each iteration as well as their average across the entire cross-validation.

AP and *AR* evaluate model performance against the training data itself (i.e. not against gold standard data). To directly compare object detection to gold standard data, one additional detector was trained while intentionally withholding the 188 images for which gold standard labels were produced. P, R, and F_1 were then calculated using detections on these images as predictions, as described in Section 3.4. These results were then compared directly to the citizen science results and to the benchmark.

4. Results

4.1. Citizen science project execution

PTTP launched on 2nd March and ran until 28th August 2020. A total of 519,420 classifications including 578,010 home annotations were made on 74,802 subjects. These data are made available by Leonard et al. (2022). Examples of raw citizen science annotations in various contexts are shown in Fig. 7. Cumulative classifications over the project duration are shown in Fig. 8. Engagement spiked as COVID-19 lockdowns increased (i.e. mid-March through early April) and on Easter weekend (i.e. 12th April). The project was promoted by various institutions and at virtual events[10], which encouraged engagement. Further study on the composition of the contributor community and their motivations for engaging is undertaken in Leonard et al. (2022).

Throughout the project, 71,251 SuperView subjects, 2,915 KOMPSAT 3A subjects, and 636 DMC-3 subjects were mapped, totalling approximately 1,267 km^2 of imagery. So, over the duration of the project (179 days), the average retirement speed was 7.1 km^2/day. While the Zooniverse platform does not log time spent by volunteers on-task, they estimate that for similar tasks on their platform, between 20–75 classifications can be completed per hour (The Zooniverse, 2020). Using an estimate of two minutes per classification, an estimated 17,314 hours of volunteer time were contributed throughout the project, or 13.7 volunteer hours per km^2 mapped.

4.2. Citizen science data validation

P, R, and F_1 were calculated using annotations as predictions and gold standard data as ground truth. For the raw annotations, results were $P = 0.492$, $R = 0.933$, and $F_1 = 0.644$. The high R value may indicate that citizen scientists found most ground truths, but is also influenced by the fact that multiple overlapping raw annotations each count as a distinct T_P. The lower P value indicates a substantial amount of F_P. This validates our assumption that clustering may be needed to

[7] Our tiles are actually 512×512, which may improve application time, but we assume the same speed for simplicity.
[8] https://github.com/tensorflow/models/tree/master/research/object_detection/samples/configs
[9] https://www.robots.ox.ac.uk/ vgg/data/pets/

[10] https://www.northwestern.edu/sustainability/news/2020/citizen-science-project-spotlights.html, https://www.societyforscience.org/blog/regeneron-isef-attendees-contribute-900-volunteer-hours/.

Table 2
Average precision (*AP*) and average recall (*AR*) metrics calculated through 10-fold cross-validation of the Faster R-CNN object detecion model, and the average of these.

	Cross-validation iteration										
Metric	0	1	2	3	4	5	6	7	8	9	Average
$AP^{IoU=0.5:.05:.95}$	0.250	0.255	0.277	0.273	0.261	0.250	0.284	0.272	0.260	0.260	**0.264**
$AP^{IoU=0.5}$	0.556	0.546	0.572	0.570	0.553	0.558	0.583	0.570	0.560	0.558	**0.563**
$AP^{IoU=0.75}$	0.193	0.202	0.240	0.230	0.214	0.191	0.243	0.227	0.204	0.204	**0.215**
AP^{small}	0.175	0.182	0.213	0.210	0.196	0.204	0.213	0.204	0.198	0.213	**0.201**
AP^{medium}	0.344	0.329	0.346	0.345	0.346	0.311	0.357	0.346	0.331	0.317	**0.337**
AP^{large}	0.341	0.301	0.369	0.370	0.350	0.395	0.412	0.307	0.348	0.405	**0.360**
$AR^{max=1}$	0.078	0.084	0.083	0.080	0.081	0.078	0.089	0.082	0.081	0.087	**0.082**
$AR^{max=10}$	0.313	0.326	0.326	0.328	0.320	0.310	0.344	0.332	0.316	0.333	**0.325**
$AR^{max=100}$	0.431	0.435	0.443	0.452	0.436	0.427	0.456	0.448	0.435	0.443	**0.441**
AR^{small}	0.371	0.373	0.387	0.407	0.377	0.372	0.395	0.396	0.377	0.386	**0.384**
AR^{medium}	0.510	0.504	0.509	0.511	0.520	0.497	0.523	0.513	0.506	0.512	**0.511**
AR^{large}	0.501	0.505	0.532	0.494	0.492	0.573	0.546	0.455	0.529	0.564	**0.519**

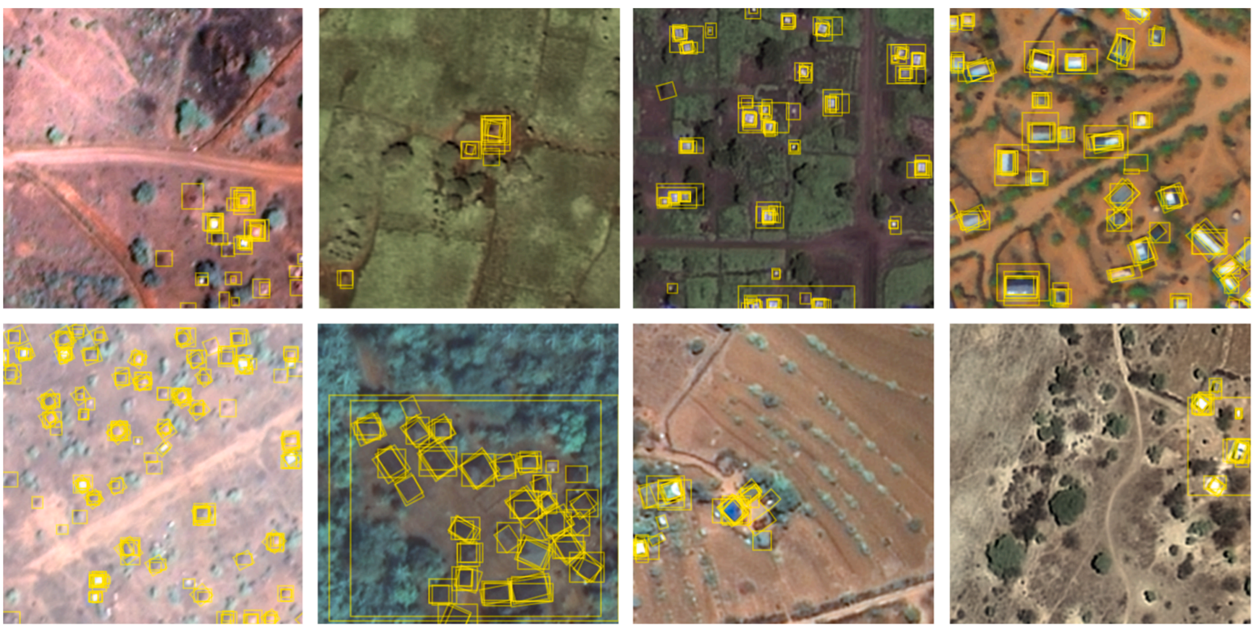

Fig. 7. Example raw citizen science annotations from PTTP (yellow boxes) in various contexts. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

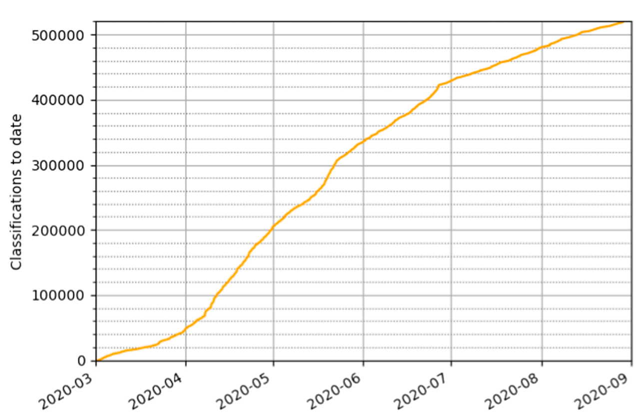

Fig. 8. Cumulative image classifications on the PTTP citizen science project.

help find consensus and eliminate noise. P, R, and F_1 were also calculated for clusters produced by HDBSCAN* with different m_{clSize} values. Results are shown in Fig. 9. The highest P (0.689) was seen for $m_{clSize} = 4$ while the highest R (0.487) and F_1 (0.568) were seen at $m_{clSize} = 2$. Clustering improved P, but R for clusters was lower than R for raw annotations.

Next, raw citizen science annotations, clusters for HDBSCAN* with $m_{clSize} = 2$, and gold standard annotations were visualised over pan-sharpened image tiles for inspection. Various interesting observations arose from comparing these, each of which is visualized in Fig. 10. The following observations appeared to have negative impacts on data quality:

- **Clustering issues**: HDBSCAN* occasionally "split the difference" between (groups of) annotations which are close together yet too few to form distinct clusters.
- **Crowding**: Annotation quality is highly variable on crowded images. Homes are often missed in these cases. Crowded images with rectangular high-contrast homes generally had higher-quality annotations than those with irregular lower-contrast homes.
- **Rotation**: Some citizen scientists elected not to rotate their annotations to align with roof edges. This made little difference for homes with circular roofs or homes already aligned with the initial rotation of the annotation tool. However, for rectangular homes on an angle,

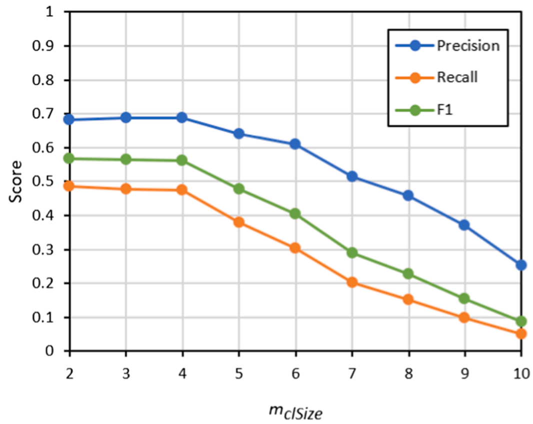

Fig. 9. Precision, recall, and F1 results for home annotation clusters generated using HDBSCAN* with m_{clSize} between 2 and 10.

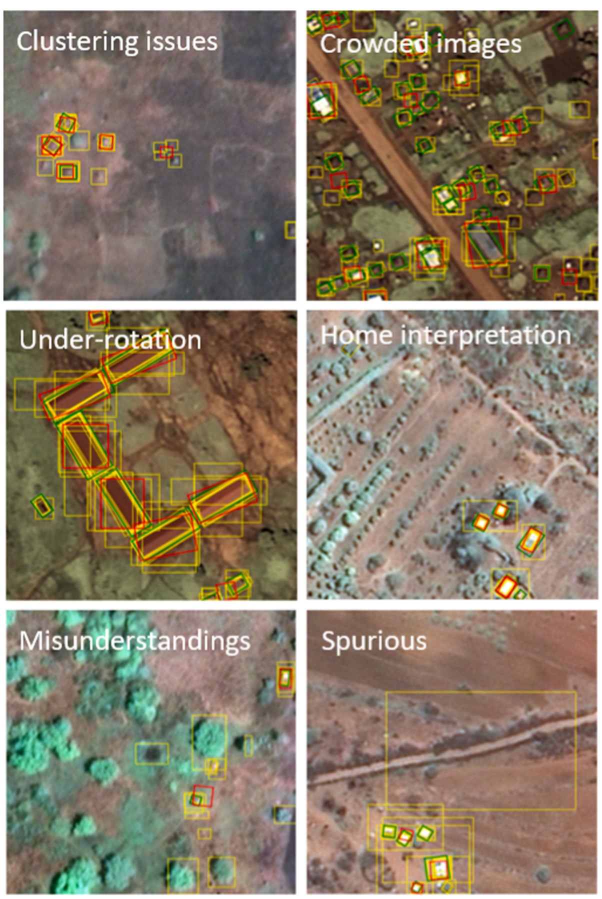

Fig. 10. Trends observed in the PTTP citizen science annotations and clusters include: clustering issues, difficulties with crowded images and under-rotation, different home interpretations, misunderstandings, and spurious annotations. Citizen science annotations are shown in yellow, clusters in red, and gold standard data in green. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

this caused annotations and subsequently clusters to be under-rotated.

The following observations did not appear to dramatically affect data quality but were still interesting to note:

- **Home interpretation**: There were differences in interpretation between citizen scientists on what constitutes a home. In images with groups of small buildings, some annotated each structure as a home, while some interpreted the whole group as one home. HDBSCAN* placed clusters at the locations with the highest density of annotations and disregarded other interpretations.
- **Misunderstandings**: There were various images in which citizen scientists seem to have misunderstood non-home objects to be homes. This is somewhat expected, as annotating diverse home styles and settlement types is challenging. In most instances, misunderstandings were the minority and HDBSCAN* eliminated these from the clusters.
- **Spurious annotations**: Some images contained annotations which appeared to be completely incorrect or accidental. These differ from misunderstandings in that there was no obvious feature which could be misunderstood as a home. Again, HDBSCAN* typically eliminated these from clusters.

4.3. Object detection training and evaluation

Faster R-CNN was trained using satellite imagery tiles and citizen science annotation clusters ($m_{clSize} = 2$) to detect homes, using the methods described in Section 3.5. After training, model graphs were exported and their performance evaluated using the metrics outlined in Section 3.6. The average of these results across the ten-fold cross validation are provided in Table 2 and Fig. 11. Sample detections were also superimposed on satellite imagery tiles for a qualitative understanding of performance. Examples from various contexts are visualised in Fig. 12. As both *AP* and *AR* values range between 0 and 1 with 1 being optimal, it is evident that the performance of the object detection model could be improved. Nevertheless, these results show promise. *AP* values for state-of-art models can range from 0.5 to 0.8 depending on the training data and task at hand. The *AP* results for $IoU = 0.5$ (i.e. the more standard and less strict threshold) fall within this range. The results also provide some indication of detector performance in different types of images with different sizes and quantities of homes. Medium and large objects (i.e. with area $\geqslant 32 \times 32$) are detected more accurately than small objects. This indicates that larger homes are more easily detected than smaller homes. It can be theorized that these homes were better annotated in the training data by citizen scientists, generating higher performance. Additionally, performance improves as more homes are allowed to be detected per image. This makes sense, as some images (e.g. those representing village centers or refugee camps) do contain many homes, and if only one or ten detections are allowed, many homes will be missed.

It is important to remember that these metrics are evaluating detections against annotation clusters, not against gold standard. Evaluating a detector trained on imperfect data against different imperfect data does not indicate much about the ability of the detector to do the task well; rather, it speaks to how well the detector is able to replicate the imperfect data.

For a more intuitive understanding of performance, a version of the model was also trained where the 188 images with gold standard annotations were purposefully withheld. The model was then used for detection on these images, and P, R, and F_1 was calculated using the gold standard annotations as ground truth using the same process described in Section 3.6. The results were $P = 0.672$, $R = 0.361$, and $F_1 = 0.470$.

There were also various interesting observations from the detection visualizations, which are illustrated in Fig. 13. First, the model replicated issues encountered by citizen scientists. Just as citizen scientists struggled most with crowded images, so did the trained detector. Furthermore, the detector also struggled with lower resolution and lower contrast images, and images blurred due to atmospheric effects. However, the detector did not only struggle in the same ways as the citizen scientists. It appeared to have learned features which, when

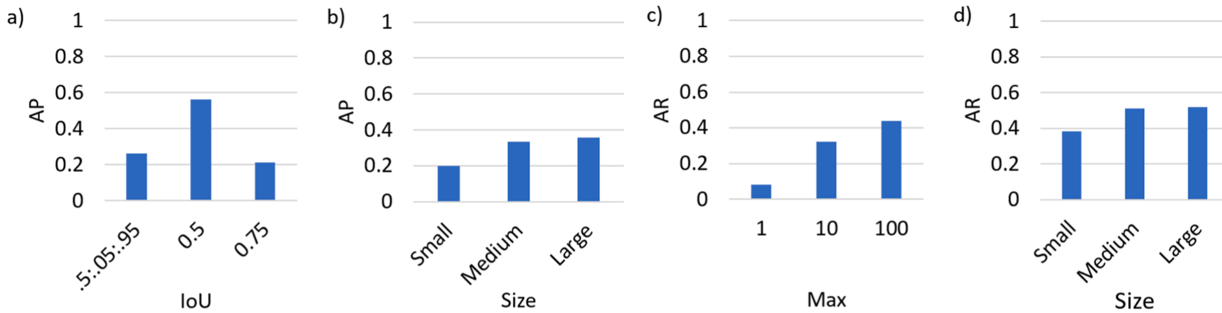

Fig. 11. Cross-validation results for the object-detection-based mapping: (a) *AP* for different *IoU* thresholds; (b) *AP* for different detection sizes; (c) *AR* for different maximum numbers of detections; and (d) *AR* for different detection sizes.

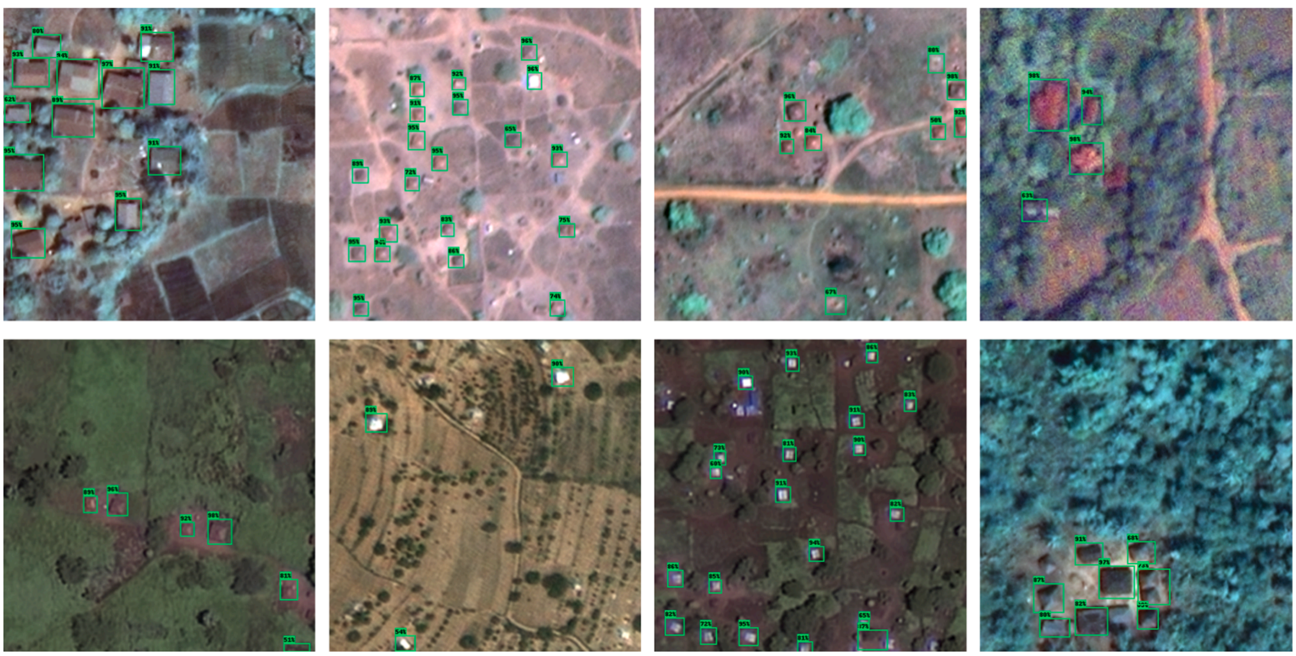

Fig. 12. Detection examples in various satellite imagery contexts with a probability of 50% or higher, visualized as green boxes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

cross-applied in different contexts, led to false detections. For instance, the detector seemed to have learned that a high contrast bounded edge is likely to indicate a roof. However, dark fences or hedges can create the same type of high-contrast feature. While it is easy for a human to tell that this is not a roof, the detector mistakenly predicts it to be one. The detector also performed poorly in certain agricultural contexts, but not all such contexts, as evident in Fig. 12, where various examples showing good performance are in agricultural contexts. On the whole, while various homes are missed, and some non-home features are predicted to be homes, the detector does manage to predict homes in each context. The model learned imperfectly from imperfect input data.

The model was timed during both training and application. On two Nvidia GeForce RTX 2080 Ti GPUs, training proceeded at approximately 0.086 s per step, and detections took an average of 0.033 s per detection. To put this in more intuitive terms, this means that 100,000 steps of training would take approximately 2.4 h, and approximately 30 images could be put through the detector each second for detection. This is an improvement over the benchmark speed provided for this algorithm implementation, likely due to the use of more advanced GPUs and slightly smaller image files.

Generally, the results of this work show that the citizen science project was completed successfully and produced data of a quality suitable to train a CV model, albeit with some issues (e.g. crowded images, under-rotated annotations) which require further work.

5. Discussion

The results of the citizen science and CV approaches detailed above are summarized and compared to the benchmark in Table 3. These results show citizen science paired with object detection to be a promising method for scalable mapping of rural off-grid areas. Some key insights and challenges faced in this work are detailed in the following sections, with a focus how this work can support better energy access system design in future work.

5.1. Citizen science performance and design implications

Citizen scientists struggled to annotate images with many homes. As one of the target off-grid settlement types in this work was the often-crowded refugee camp context, the struggle with crowded images is problematic. Poor mapping of such under-resourced crowded areas could engender poor electrical system design, perpetuating existing inequities. It appears that it was not the relative size of homes that contributors found to be challenging, as non-crowded images had homes of similar sizes, but the quantity of homes per tile. These issues were subsequently replicated by the trained detector.

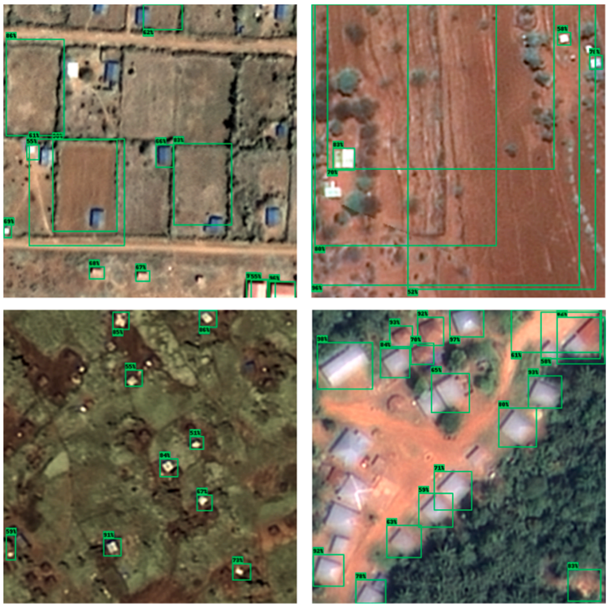

Fig. 13. Issues observed in object detection. The trained detector struggled with crowded (bottom left) and blurred images (bottom right), misdetected some fenced homesteads (top right), and performed poorly in particular agricultural settings (top left).

Table 3
Summary of results compared to the benchmark. It is assumed that the benchmark case would achieve 100% precision (i.e. no false positives). Other benchmark figures are provided in Section 1. Costs for citizen science and object detection are calculated in Section 5.4.

Metric	Benchmark	Citizen science (raw)	Citizen science (clusters)	Object detection
Precision (%)	100%	49%	69%	67%
Recall (%)	62%	93%	49%	36%
Speed (km^2/day)	0.25 km^2/day	7 km^2/day	7 km^2/day	42,938 km^2/day
Cost ($/$km^2$)	$465	$20.84	$20.84	$11.40

Several strategies could be used to mitigate this in future projects. Crowded images could be presented at a higher zoom level to reduce the quantity of homes per tile. Gridded population density datasets (e.g. HRSL, GPW) could be used to identify higher-density areas needing higher zoom. Depending on the distribution of homes in each grid cell (i.e. settlement style), this could result in some empty or sparse areas being presented at higher zoom than needed, but this would be minimally harmful aside from slowing the annotation process. Crowded images could also be excluded altogether based on a density threshold. It would be interesting to see whether CV algorithms trained only on annotations in non-crowded images could still accurately map crowded images. Another option could be a multi-stage citizen science pipeline in which citizens themselves identify crowded images requiring increased zoom in a preliminary classification workflow. These images could then be reprocessed at a higher zoom level before annotation. Alternatively, citizens could be integrated in an iterative verification feedback loop. An object detection algorithm could be periodically retrained throughout the citizen science project, and its detections could be fed back into a citizen science workflow for verification. The verifications could be used to refine training data on challenging or crowded images over time. These potential mitigation strategies can be applied not only to home mapping, but to other citizen science projects which involve annotating crowded images.

Additionally, not all citizen scientists rotated their annotations to align with roof edges. This reduces the accuracy of the annotations, and leads to inconsistent information about roof size (i.e. home footprint), reducing its viability as a proxy for electrical demand. Contributors may have found rotating annotations to be difficult or time consuming, encountered issues with the rotation interface, or have been influenced by experience with previous projects with non-rotating tools. Further follow-up would be required to determine the reason. Regardless, this validates our initial assumption that more difficult annotation tools (i.e. with more steps) can lead to more variable data quality. Perhaps a simple box (i.e. no rotation) or circular annotation tool may create more consistent data in future projects.

5.2. Clustering performance

There is room for improvement in annotation clustering methods for home mapping. While theoretically well-suited to the task, HDBSCAN* nevertheless had some issues (e.g. the tendency to "split the difference" discussed in Section 4.2). Future work may wish to explore different clustering methods or variations on HBDSCAN* tailored to home mapping. Additionally, it could be explored whether clustering parameters tuned to each context (e.g. region, settlement type, or tile) would produce more accurate clusters. This is an area for future work. Low recall is a known challenge in home mapping: OSM data also has low recall (Bonafilia et al., 2019).

5.3. Object detection performance

It was observed that object detection algorithms can only work as well as the data you train them with. To train a high-performing object detection algorithm, meticulous training data are needed, with precision and recall values approaching 1. Despite this, results indicate that even this imperfect algorithm could make a useful first pass home-level mapping tool. Small improvements in training data accuracy, achieved using various citizen science process improvements detailed in Section 5.1, could improve object detection performance. Specific data processing methods could also improve the quality of data used in object detection training. For instance, annotations could be clipped closely to minimize the number of non-labelled homes the algorithm was exposed to as "false" samples. Additionally, images could be pre-processed for high contrast to help improve annotation accuracy. In this work, images were pre-processed to achieve a "natural" colour profile. Pre-processing for high contrast instead of natural colour could help to improve the accuracy of annotation in blurry or overexposed images. This may create a trade-off between the accuracy of locations and roof colour information, which would be acceptable here as location was the priority.

5.4. Method scalability

To validate our argument, the methods used here must be shown to be scalable, measured as the speed (km^2/day) and cost ($/$km^2$) of mapping.

PTTP achieved an average mapping speed of 7 km^2/day. This improves on the benchmark by a factor of 28. As the Zooniverse platform is free to use, the main material cost of this approach is satellite imagery. The highest resolution imagery used in this work is commmercially available at $10/$km^2$ (GeospatialWorld.net, 2020). Other costs include researcher or practitioner time and computing facilities, though it is likely that practitioners and researchers already have access to the facilities required to utilize the Zooniverse platform and prepare and preprocess data. These costs can be illustrated using the example of replicating this project[11]. Assuming a cost of $10/$km^2$ for satellite imagery,

[11] Costs provided in this example are representative, not exact.

the total imagery cost would be $12,670. If one researcher worked half-time on this project over the project duration (26 weeks) at a pay rate of $28/hour (Salary.com, 2022), assuming a 37.5 h work week, researcher labour would cost $13,650. Therefore, PTTP could be replicated with a total cost of $26,410, or $20.84/km^2.

Once a detector is trained using these data, each additional km^2 mapped tends towards a cost of $10 (i.e. the cost of imagery). For instance, if another 10,000 km^2 of imagery were mapped using the trained detector produced using PTTP annotations, the cost per km^2 mapped would drop to $11.40, assuming a $2,000 investment cost in a GPU is required to run the detector[12]. The only other cost besides the imagery and the GPU is researcher time – which, given the speed of the detector trained here, is negligible in this instance (i.e. 10,000 additional km^2 could be mapped in less than one day).

The Faster R-CNN implementation used here was found to map at 42,938 km^2/day using two Nvidia GeForce RTX 2080 Ti GPUs. This is a speed improvement of multiple orders of magnitude over the benchmark and PTTP. To illustrate, at this speed, the 78 million km^2 of rural land area in low- and middle-income countries globally could be mapped at this rate at home-level in 1,817 days, or 5 years. By increasing the quantity of GPUs used and splitting the detection task amongst them, this could be reduced to a year or less. Of course, this would require adequate representative training data for all regions, which could be acquired through a series of low-cost PTTP-style citizen science projects.

5.5. Practical implications

Many popular electrical system design models take home location data as an input to produce grid designs including topology and costing. For instance, home locations are required by the Reference Electrification Model (REM) (Ciller et al., 2019), an innovative models at the intersection of large-area energy access planning and local topology design. CV is used to generate this data for REM where it is missing from available sources. To train their CV algorithm, images hand-annotated by research assistants had to be used, as even training data generated by paid “crowdworkers” from Amazon Mechanical Turk was found to be of an inadequate quality (Lee, 2018). Meanwhile, this work demonstrates that citizen science can produce data which can be directly used as input to models like REM, or used to train CV algorithms to generate this data, quickly and at far less expense. Higher accuracy of citizen scientists over alternatives like Mechanical Turk may be attributed to their familiarity with the annotation interface or the task, interest in the research question, enthusiasm, or investment in good research outcomes. Further study on this is needed.

6. Conclusions

This work has found citizen science and computer vision to be a feasible method to map remote off-grid populations at scale for energy access system design. Through the execution of a large-scale online citizen science project called “Power to the People” (PTTP), it was found that citizen scientists are able to map rural off-grid areas in Kenya, Sierra Leone, and Uganda at home-level using satellite imagery. A novel dataset of 578,010 home annotations was generated at a mapping rate of 7 km^2/day. Raw citizen scientist annotations achieved a precision of 49% and recall of 93%; precision could be increased to 69% through data clustering and post-processing. While we find that citizen scientists struggle with crowded and low contrast images, we suggest ways to pre-process images in the future for higher accuracy and completeness. Post-processed citizen science data were used to train a Faster R-CNN object detection model which produced detections useful as a first pass for home-level mapping. Detections achieved a precision of 67% and recall of 36%. This performance would be improved through various adjustments to the citizen science method that we recommend to improve training data quality. Faster R-CNN could be applied at 42,938 km^2/day, an increase of mapping speed of several orders of magnitude beyond the benchmark or citizen science alone. The cost of mapping is $20.84/km^2 for citizen science alone and reduces drastically as additional area is mapped – for instance, scaling mapping by an additional 10,000 km^2 using a trained detector lowers costs to $11.40/km^2. This research shows citizen science and computer vision to be a promising pipeline for rural mapping for energy access system design.

CRediT authorship contribution statement

Alycia Leonard: Conceptualization, Methodology, Software, Validation, Formal analysis, Data curation, Writing - Original draft, Writing - Review and editing, Visualization, Funding acquisition, Project administration. **Scot Wheeler**: Conceptualization, Writing - Review and editing, Visualization. **Malcolm McCulloch**: Conceptualization, Methodology, Writing - Review and editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research would be impossible without the dedicated work of the Zooniverse citizen science community. The project was supported by a United Kingdom Research and Innovation grant (BB/T01833X/1). Satellite imagery was provided by Earth-i through the Satellite Applications Catapult. Many thanks to the Zooniverse team for their support of this project and to volunteer project moderator Ravi Kohli. Additionally, this material has been produced in part under the Climate Compatible Growth programme, which is funded by UK aid from the UK government. However the views expressed herein do not necessarily reflect the UK government’s official policies.

References

Aristeidou, M., Herodotou, C., Ballard, H.L., Young, A.N., Miller, A.E., Higgins, L., Johnson, R.F., 2021. Exploring the participation of young citizen scientists in scientific research: The case of iNaturalist. PLoS ONE 16 (1 January), pp. 1–13. doi: 10.1371/journal.pone.0245682.

Asiama, K., Bennett, R., Zevenbergen, J., 2017. Participatory land administration on customary lands: A practical VGI experiment in Nanton, Ghana. ISPRS Int. J. Geo-inform. 6 (7), 186. https://doi.org/10.3390/ijgi6070186.

Bonafilia, D., Gill, J., Kirasanov, D., Sundam, J., 2019. Mapping the world to help aid workers, with weakly, semi-supervised learning, URL: https://ai.facebook.com/blog/mapping-the-world-to-help-aid-workers-with-weakly-semi-supervised-learning.

Campello, R., Moulavi, D., Sander, J., 2013. Density-based clustering based on hierarchical density estimates. In: Pei, J., Tseng, V., Cao, L., Motoda, H., Xu, G. (Eds.), Advances in Knowledge Discovery and Data Mining. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 160–172 doi:10.1007/978-3-642-37456-2_14.

Center for International Earth Science Information Network (CIESIN), 2018. Columbia University, Gridded Population of the World, Version 4 (GPWv4): Population Density, Revision 11. doi:10.7927/H49C6VHW.

Charmes, J., 2006. A Review of Empirical Evidence on Time Use in Africa from UN-Sponsored Surveys. In: Blackden, C.M., Wodon, Q. (Eds.), Gender, time use, and poverty in Sub-Saharan Africa, no. World Bank Working Paper No. 73, The International Bank for Reconstruction and Development/The World Bank, Washington, D.C., 2006, Ch. 3. doi:10.1596/978-0-8213-6561-8.

Ciller, P., Ellman, D., Vergara, C., González-García, A., Lee, S.J., Drouin, C., Brusnahan, M., Borofsky, Y., Mateo, C., Amatya, R., et al., 2019. Optimal electrification planning incorporating on-and off-grid technologies: the reference electrification model (rem). Proc. IEEE 107 (9), 1872–1905. https://doi.org/10.1109/JPROC.2019.2922543.

Common Objects in Context (COCO), 2021. COCO: Detection Evaluation, URL: https://cocodataset.org/#detection-eval.

[12] NVIDIA graphics cards compatible with Tensorflow are commercially available in the $300-$2,000 range – see https://shop.nvidia.com/en-gb/geforce/store. Alternatively, practitioners and researchers could make use of affordable cloud computing (e.g. Google Colab, Microsoft Azure).

Dickinson, J.L., Zuckerberg, B., Bonter, D.N., 2010. Citizen science as an ecological research tool: Challenges and benefits. Annu. Rev. Ecol. Evol. Syst. 41, 149–172. https://doi.org/10.1146/annurev-ecolsys-102209-144636.

Enemark, S., Bell, K.C., Lemmen, C., McLaren, R., 2014. Fit-for-purpose land administration, International Federation of Surveyors and World Bank, 2014. URL: https://fig.net/resources/publications/figpub/pub60/figpub60.asp.

Energy Sector Management Assistance Program (ESMAP), 2019. Mini Grids for Half a Billion People: Market Outlook and Handbook for Decision Makers - Executive Summary, Tech. rep., World Bank, Washington DC. URL: http://hdl.handle.net/10986/31926.

Facebook Connectivity Lab and CIESIN, 2016. Columbia University, High Resolution Settlement Layer (HRSL). URL: https://www.ciesin.columbia.edu/data/hrsl/.

Geofabrik GmbH, 2018. Download OpenStreetMap data for this region: Kenya, URL: http://download.geofabrik.de/africa/kenya.html.

GeospatialWorld.net News Desk, 2020. Price breakthrough in high resolution satellite imagery via Soar platform, URL: https://www.geospatialworld.net/news/price-break-through-for-new-tasked-high-resolution-satellite-imagery-via-the-soar-platform/.

Givoni, M., 2016. Between micro mappers and missing maps: Digital humanitarianism and the politics of material participation in disaster response. Environ. Plan. D: Soc. Space 34 (6), 1025–1043. https://doi.org/10.1177/0263775816652899.

Herfort, B., Eckle, M., De Albuquerque, J.P., 2016. Being specific about geographic information crowdsourcing: A typology and analysis of the Missing Maps project in South Kivu. iN: Proceedings of the International ISCRAM Conference, May 2016.

Horn, G.V., Aodha, O.M., Song, Y., Cui, Y., Sun, C., Shepard, A., Adam, H., Perona, P., Belongie, S., 2018. The iNaturalist Species Classification and Detection Dataset. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 8769–8778. https://doi.org/10.1109/CVPR.2018.00914 arXiv:1707.06642.

Hui, A., Walker, G., 2018. Concepts and methodologies for a new relational geography of energy demand: Social practices, doing-places and settings. Energy Res. Soc. Sci. 36 (August 2017), 21–29. https://doi.org/10.1016/j.erss.2017.09.032.

Humanitarian OpenStreetMap Team, 2021. Humanitarian OpenStreetMap Team (HOT). URL: https://www.hotosm.org/.

Jones, F.M., Allen, C., Arteta, C., Arthur, J., Black, C., Emmerson, L.M., Freeman, R., Hines, G., Lintott, C.J., Macháčková, Z., et al., 2018. Time-lapse imagery and volunteer classifications from the Zooniverse Penguin Watch project. Scient. Data 5 (1), 1–13. https://doi.org/10.1038/sdata.2018.124.

Kes, A., Swaminathan, H., 2006. Gender and Time Poverty in Sub-Saharan Africa. In: Blackden, C.M., Wodon, Q. (Eds.), Gender, time use, and poverty in Sub-Saharan Africa, no. World Bank Working Paper No. 73, The International Bank for Reconstruction and Development/The World Bank, Washington, D.C., 2006, Ch. 2. doi:10.1596/978-0-8213-6561-8.

Lam, D., Kuzma, R., McGee, K., Dooley, S., Laielli, M., Klaric, M., Bulatov, Y., McCord, B., 2018. xview: Objects in context in overhead imagery, CoRR abs/1802.07856. arXiv:1802.07856. URL: http://arxiv.org/abs/1802.07856.

Lee, S.J., 2018. Adaptive electricity access planning, Master's thesis, Massachusetts Institute of Technology, Institute for Data, Systems, and Society and the Department of Electrical Engineering and Computer Science.

Leonard, A., Wheeler, S., McCulloch, M., 2022. Rural Home Annotation Dataset Mapped by Citizen Scientists in Satellite Imagery (2022). doi:10.17632/xw6gr8p2cn.1.

Leonard, A., Wheeler, S., McCulloch, M., 2022. Evaluating "Power to the People": Best practices for positive community impact in remote mapping citizen science projects, sSRN Preprint (2022). URL: http://ssrn.com/abstract=4052549.

Maggiori, E., Tarabalka, Y., Charpiat, G., Alliez, P., 2017. Can semantic labeling methods generalize to any city? the inria aerial image labeling benchmark, in: IEEE International Geoscience and Remote Sensing Symposium (IGARSS), IEEE.

Mnih, V., 2013. Machine learning for aerial image labeling, Ph.D. thesis, University of Toronto.

Mundhenk, T.N., Konjevod, G., Sakla, W.A., Boakye, K., 2016. A large contextual dataset for classification, detection and counting of cars with deep learning. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (Eds.), Computer Vision – ECCV 2016. Springer International Publishing, Cham, pp. 785–800.

OpenStreetMap Contributors, 2021. OpenStreetMap. URL: https://www.openstreetmap.org/ (2021).

Pan, X., Yang, F., Gao, L., Chen, Z., Zhang, B., Fan, H., Ren, J., 2016. Building extraction from high-resolution aerial imagery using a generative adversarial network with spatial and channel attention mechanisms. Remote Sens. 11 (8) https://doi.org/10.3390/rs11080917. URL: https://www.mdpi.com/2072-4292/11/8/917.

Qin, Y., Wu, Y., Li, B., Gao, S., Liu, M. Y. Zhan, 2019. Semantic segmentation of building roof in dense urban environment with deep convolutional neural network: A case study using gf2 vhr imagery in China. Sensors 19 (5). doi:10.3390/s19051164.

Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. In: Cortes, C., Lawrence, N., Lee, D., Sugiyama, M., Garnett, R. (Eds.), Advances in Neural Information Processing Systems, Vol. 28, Curran Associates Inc, 2015. URL: https://proceedings.neurips.cc/paper/2015/file/14bfa6bb14875e45bba028a21ed38046-Paper.pdf.

Saito, S., Yamashita, T., Aoki, Y., 2016. Multiple Object Extraction from Aerial Imagery with Convolutional Neural Networks. J. Imaging Sci. Technol. 60 (1), 104021–104029. https://doi.org/10.2352/J.ImagingSci.Technol.2016.60.1.010402.

Salary.com, Hourly Wage for Researcher I - Academic Salary in the United States, 2022. URL: https://www.salary.com/research/salary/benchmark/researcher-i-academic-hourly-wages.

Scholz, S., Knight, P., Eckle, M., Marx, S., Zipf, A., 2018. Volunteered geographic information for disaster risk reduction-the missing maps approach and its potential within the Red Cross and Red Crescent movement. Remote Sens. 10 (8). doi:10.3390/rs10081239.

Shermeyer, J., Van Etten, A., 2019. The effects of super-resolution on object detection performance in satellite imagery. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2019.

Shi, Y., Rathod, V., 2021. TensorFlow 1 Detection Model Zoo, URL: https://github.com/tensorflow/models/blob/master/research/object_detection/g3doc/tf1_detection_zoo.md.

Shove, E., Walker, G., 2014. What Is Energy For? Soc. Pract. Energy Demand, Theory, Culture Soc. 31 (5), 41–58. https://doi.org/10.1177/0263276414536746.

Swanson, A., Kosmala, M., Lintott, C., Simpson, R., Smith, A., Packer, C., 2015. Snapshot Serengeti, high-frequency annotated camera trap images of 40 mammalian species in an African savanna. Scient. Data 2 (1), 1–14. https://doi.org/10.1038/sdata.2015.26.

The Zooniverse, 2020. Fulfilling service hour requirements through Zoonvierse, URL: https://blog.zooniverse.org/2020/03/26/fulfilling-service-hour-requirements-through-zooniverse/.

The Zooniverse Team, 2021. Welcome to the Zooniverse, URL: https://www.zooniverse.org/.

Van Etten, A., 2018. You only look twice: Rapid multi-scale object detection in satellite imagery, CoRR abs/1805.09512. arXiv:1805.09512. URL: http://arxiv.org/abs/1805.09512.

Van Etten, A., Lindenbaum, D., Bacastow, T.M., 2018. Spacenet: A remote sensing dataset and challenge series, CoRR abs/1807.01232. arXiv:1807.01232. URL: http://arxiv.org/abs/1807.01232.

Wilson, J.S., Pan, A.D., General, D.E.M., Koch, J.B., 2020. More eyes on the prize: an observation of a very rare, threatened species of Philippine Bumble bee, Bombus irisanensis, on iNaturalist and the importance of citizen science in conservation biology. J. Insect Conserv. 24 (4), 727–729. https://doi.org/10.1007/s10841-020-00233-3.

Yi, Y., Zhang, Z., Zhang, W., Zhang, C., Li, W., Zhao, T., 2019. Semantic segmentation of urban buildings from vhr remote sensing imagery using a deep convolutional neural network. Remote Sens., 11 (15). doi:10.3390/rs11151774.

Yuan, J., 2018. Learning building extraction in aerial scenes with convolutional networks. IEEE Trans. Pattern Anal. Mach. Intell. 40 (11), 2793–2798. https://doi.org/10.1109/TPAMI.2017.2750680.