Research

Check for updates

Humanitarian applications of machine learning with remote-sensing data: review and case study in refugee settlement mapping

John A. Quinn[1], Marguerite M. Nyhan[1], Celia Navarro[2], Davide Coluccia[2], Lars Bromley[2] and Miguel Luengo-Oroz[1]

[1]United Nations Global Pulse, New York, NY, USA
[2]United Nations Institute for Training and Research—Operational Satellite Applications Programme, Geneva, Switzerland

(iD) JAQ, 0000-0002-6485-2945

The coordination of humanitarian relief, e.g. in a natural disaster or a conflict situation, is often complicated by a scarcity of data to inform planning. Remote sensing imagery, from satellites or drones, can give important insights into conditions on the ground, including in areas which are difficult to access. Applications include situation awareness after natural disasters, structural damage assessment in conflict, monitoring human rights violations or population estimation in settlements. We review machine learning approaches for automating these problems, and discuss their potential and limitations. We also provide a case study of experiments using deep learning methods to count the numbers of structures in multiple refugee settlements in Africa and the Middle East. We find that while high levels of accuracy are possible, there is considerable variation in the characteristics of imagery collected from different sensors and regions. In this, as in the other applications discussed in the paper, critical inferences must be made from a relatively small amount of pixel data. We, therefore, consider that using machine learning systems as an augmentation of human analysts is a reasonable strategy to transition from current fully manual operational pipelines to ones which are both more efficient and have the necessary levels of quality control.

This article is part of a discussion meeting issue 'The growing ubiquity of algorithms in society: implications, impacts and innovations'.

1. Introduction

Humanitarian relief is required in response to many types of crises, including both natural and man-made disasters. It is typically a short-term intervention aimed at the immediate saving of lives and reduction of suffering until longer term provisions can be made.

For this response to be effective, reliable and comprehensive information is critical, as early as possible, about the effects of the crisis: how many people have been affected and (as the response gets underway) how many of those have actually received relief, out of the total number targeted for assistance.

In a crisis situation, collecting this information may be difficult. For example, in a natural disaster such as a flood or earthquake, roads may become impassable, or in man-made emergencies there may be conflict; furthermore, the areas affected might be large, making it impractical to survey the situation thoroughly from the ground. In some cases, the numbers above might be arrived at essentially through a series of informed guesses, particularly during the initial hours or days after the onset of an emergency, where the only available information could be anecdotal reports from eyewitnesses. Because of this, remote-sensing data—particularly from satellites—is useful. Satellites can be tasked to collect images of the affected area, and it can be possible to obtain high-resolution imagery (50 cm resolution or less) within a matter of days, depending on certain factors including cloud cover.

With this imagery, various analysis tasks can be carried out depending on the crisis. In situations involving refugees or internally displaced people (IDPs), the numbers of people in settlements (either planned or informal) can be estimated from the numbers of different types of structures. Damage assessment can be done, e.g. by counting and mapping numbers of destroyed buildings. Other humanitarian analysis tasks may not be directly related to the provision of relief—for example, conflict documentation for naming and shaming, or collection of evidence for litigation. These procedures generally involve skilled human analysts visually interpreting the imagery. Increasing the degree of automation would have potential benefits in terms of providing results more quickly, and in being able to take advantage of the increasing availability of high-resolution satellite imagery to provide more frequent updates. However, routine deployment of machine learning-based systems for these purposes has so far been elusive. Algorithmic analysis is currently done for some types of analysis, though generally on lower resolution data (e.g. from Landsat or Sentinel sensors) in fire detection, standing water analysis or land cover mapping, for instance.

The structure of the rest of the paper is as follows. In §2, we briefly review applications of machine learning for remote-sensing data in general, and in §3 discuss the principal existing work specifically on humanitarian applications. In §4, we give experimental details of a case study on counting structures in refugee settlements, then discuss and conclude in §§5 and 6, respectively.

2. Machine learning and remote-sensing data

(a) Remote-sensing data

Advances in aerospace engineering and remote-sensing technologies have resulted in an increasingly diverse array of earth observation systems; these capturing unprecedented quantities of imagery, measuring a range of geophysical parameters and operating in a range of satellite orbits. The distribution of the data captured by these is managed by a number of satellite operators and data providers. The application and suitability of remote-sensing imagery in

humanitarian efforts depends on satellites' spatial resolution, revisit capability, spectral resolution and radiometric resolution.

There is no universal consensus on how to categorize resolution in remote sensing, though for the purposes of this article we define high resolution as 1 m or less, medium resolution as 1–10 m and low resolution as more than 10 m. We note that these are non-standard categories: for example, in the Copernicus data catalogue,[1] the threshold for 'high resolution' is 30 m. We use these divisions here because of the scope of machine learning applications possible within each category. With sub-metre resolution, it is generally possible to identify commonplace objects of interest such as buildings and vehicles. People can be seen in drone imagery, where resolution can be below 10 cm, but are not generally visible even in high-resolution satellite images (although crowds can be visible and shadows cast by individual persons may be visible depending on conditions). For resolutions of a few metres, all the above types of objects are more difficult to identify, and, therefore, methods such as object detection are more limited in their applications, and super-resolution algorithms can be applicable. For low-resolution data, machine learning methods are even more restricted, for example to land cover mapping or macro-scale analysis, e.g. of climate or ecology.

High-resolution imagery sources include Ikonos-2 (0.8 m panchromatic or 3.2 m RGBiR), Orbview-3, IRS-P6, EROS A&B (0.7 m), QuickBird (0.61 m panchromatic or 2.44 m RGBiR), Pleiades (0.5 m panchromatic or 2 m multispectral), GEOEYE-1 (0.46 m panchromatic and 1.84 multispectral (RGB and NiR) resolution) and Kompsat (0.7 m). Moderate- or low-resolution imagery includes imagery captured by satellite sensors such as Planet (3–5 m), SPOT (1.5–2.5 m), Sentinel-1 (5 m in Stripmap mode), Sentinel-2 (10–60 m with 13 spectral bands), AVHRR/3 (1090 m), Geostationary Operational Environmental Satellite (GOES, 1000 m), Moderate Resolution Imaging Spectroradiometer (MODIS, 250–1000 m), Landsat (15–120 m) and ASTER (15–90 m). Very High Spatial Resolution (VHSR) imagery sources include WorldView (0.3–0.5 m).

The number of spectral bands provided by each sensor is significant for the types of applications which are possible. For example, the multispectral WorldView sensors have enabled land use and land cover mapping to be conducted at an unprecedented level of spatial detail. Image processing techniques including RGB-pansharpening and multispectral-pansharpening also have broadened applications.

Another issue with regards to machine learning application is the consistency and calibration of different sensors. The sources that we refer to above as 'low resolution' in the list above are all sensors which image the entire globe at regular intervals, and where the corresponding pixels at different time frames are produced by the same measurement process; therefore, machine learning models can be transposed across space and time on this data relatively easily. The 'high-resolution' sources above are limited by capacity, as satellites such as WorldView and Pleiades are unable to transmit all of the imagery they capture back to the Earth. They are operated on a tasking basis, where images for certain areas are requested, and, therefore, the available data from these sources is a patchwork of images at different places and times. For some areas, e.g. remote parts of developing countries, the most recent high-resolution imagery may be several years old, so that tasking new imagery is necessary if such data are needed, and historical analyses are limited. In addition, the cameras on the satellites are rotated in order to capture the areas of interest, so that the angle of incidence varies and even images of the same place taken by the same satellite at different times may not be directly comparable. Hence, in machine learning terms, high-resolution imagery not only provides greater scope in terms of recognizing and segmenting objects on the ground, but also makes it is more necessary to consider dataset shift and model generalization issues.

Among all sources, satellite location revisit cycles vary widely. For example, while the GOES system can provide continuous and timely environmental and atmospheric observations over the Earth's surface, MODIS has a revisit cycle of 1–2 days, Landsat-7 has a revisit cycle of 16 days

---

[1]https://spacedata.copernicus.eu/documents/12833/14545/DAP_Release_Phase_2_1.0_final

and Sentinel-1 6 days. As the operational use of satellite systems can be hindered due to limited revisit cycles, complementary systems such as constellations of CubeSats have been suggested as a means of overcoming these limitations [1].

(b) Machine learning with remote-sensing data

Machine learning methods have been in routine use for the analysis of remote-sensing data for some time. One of the earliest applications was land cover classification with multi-spectral data (e.g. Landsat), often using a random forests or support vector machines, and this has been standard practice for around two decades. The application of machine learning methods for the efficient detection and classification of remote-sensing imagery have been reviewed previously and have focused on neural networks, support vector machines, decision trees and random forests and *k*-nearest neighbours [2–5].

Most machine learning approaches to image analysis currently are variations of deep learning methods, which have substantially improved the state of the art in various application domains, and for which remote-sensing applications are now emerging. Whereas previously it was normal to split up an analysis task into separate steps (e.g. selection or hand-coding of features, followed by application of a machine learning model, followed by post-processing), one important aspect of deep learning has been increased 'end-to-end learning'. In this type of set-up, the model simultaneously learns a feature representation, intermediate processing steps and tunes parameters for generating the final output. This requires large training datasets and computational power, but often results in strikingly better performance than was possible with previous methods. Specific discussion about deep learning methods in remote sensing can be found in [6,7].

Deep learning autoencoders are a type of network structure of particular significance in remote sensing. Autoencoders for images are models which have the ability to map each pixel in an image to a new value. Thus, they are useful for segmentation tasks such as land-cover mapping, in which we want to categorize each pixel as belonging to the class of forest, water, urban area and so on.

Object detection methods are another area of deep learning which has an important impact in terms of remote-sensing applications. Whereas autoencoders are generally used for mapping the spatial extent of 'stuff' (such as water, road surface or crop land), object detection methods are used for mapping the location of 'things' (such as cars or buildings). Object detection methods generally output a bounding box—i.e. the top, bottom, leftmost and rightmost limits—of each object detected in an image, for example with region-proposal convolutional neural networks (RCNNs), of which Faster-RCNN is a common method [8]. Other methods are able to do instance segmentation, in which for each detected object the model outputs which pixels in the image are assigned to that object. We carry out such experiments below with the Mask RCNN model [9].

Deep learning applications to satellite imagery include the use of convolutional neural networks for high-precision land-cover mapping [10] and scene classification [11]. One issue with deep learning models is that they are rarely practical to train from scratch for a new problem, unless there is a training dataset of significant size and the corresponding computational power available to train a network to convergence. Instead, it is usually necessary to take a network pre-trained on another dataset and use transfer learning to adapt it to the new problem. An example of transfer between remote-sensing scene classification problems is given in [12].

3. Humanitarian applications

Remote-sensing technologies are increasingly being used to monitor, mitigate and guide humanitarian responses to conflict, human rights violations and man-made or natural disasters [13–16]. This includes the monitoring and documentation of large-scale displacement and destruction caused by conflicts and the early warning of imminent hostilities or border conflicts.

The use of remote-sensing technology to study violent conflict and human rights has increased considerably over the last decade, and is especially valuable in difficult-to-reach or dangerous conflict zones where field observations are sparse or non-existent [17].

Fire detection products derived from satellite imagery may be used as an input into early warning systems to flag potential human-rights violations or humanitarian emergencies. For instance, MODIS imagery has been used to identify burning campaigns in human settlements in Darfur in Sudan during periods of ethnic violence [18–20]. In Kenya, the United Nations used satellite imagery to locate areas where violence had potentially occurred [21].

Night-time lights, as measured using satellite imagery, can be used for monitoring unfolding humanitarian crises [22,23]. Using satellite images acquired from the Defense Meteorological Satellite Programme's (DMPS) Operational Linescan System (OLS), Li & Li [22] investigated the spatial and temporal patterns of night-time light in Syria, its international border and surrounding regions. They also observed a moderate correlation ($R^2 = 0.52$) between night-time light loss and numbers of IDPs in districts. In the Caucasus region of Russia and Georgia, fluctuations in the night-time lights of cities were evaluated from 1992 to 2009, to detect conflict-related events such as large fires and large-scale movements of populations [23]. A review of research and applications using remote sensing in conflict and human rights scenarios can be found in Marx & Goward [24] and Witmer [17].

Remote sensing has been used widely to map the effects of conflict, for example determining structural damage to buildings and critical facilities; and damage to transportation networks which in turn may affect humanitarian access [3,25]. For example, high-resolution satellite imagery has been used to rapidly assess damage to agriculture in the Gaza strip [14], impact craters, debris and damaged structures in Eastern Ghouta in Syria [15].

Images acquired using remote-sensing technologies have been employed for monitoring and guiding humanitarian responses to natural disasters including floods, earthquakes, volcanoes, tropical cyclones and landslides. For example, in response to tropical cyclone GITA-18, which affected the Tongatapu Island in Tonga, building damage density was assessed using Pleiades and WorldView-2 satellite imagery [26]. Cooner *et al.* [27] examined using high-resolution mulitspectral and panchromatic remote-sensing data to detect urban damage following the earthquake event near Port-au-Prince in Haiti in 2010. In this study, Cooner *et al.* [27] examined machine learning algorithms including various neural network architectures and random forests to detect damage caused by the earthquake. Complementary systems have been proposed for monitoring areas being affected by natural disasters, such as constellations of nanosatellites or CubeSats [1]. In this study, CubeSat constellations are proposed as a way of overcoming revisit time limits of VHSR satellite systems, which reduce their operational use for the management of disasters.

The number of refugees and internally displaced persons (IDPs) is rapidly increasing, due to conflict situations, man-made or natural disasters, and other crises [28]. According to the United Nations High Commissioner for Refugees (UNHCR), there were approximately 65.6 million forcibly displaced people at the end of 2016, including 40.3 million IDPs, 22.5 million refugees and 2.8 million asylum seekers [28].

IDPs who are not considered urban IDPs [29] usually reside in self-settled or planned settlements, where essential facilities may be provided by national or international humanitarian relief organizations [28]. As such, accurately estimating refugee occupancy rates in settlements is essential for planning and managing efficient relief operations, and enhancing logistical support for allocating survival contingencies. Refugee population numbers can be inferred from the number and size of structures within refugee settlements including tents and improvised shelters [30]. The immense size and complexity of refugee settlements, the potential number of structures and the different types, are challenges when producing accurate estimates. On-the-ground surveys of settlements including structures can be labour-intensive, time-consuming, costly and dangerous. However, these *in situ* measurements offer advantages in terms of assessing whether structures are occupied or not.

With the continuous emergence of satellite sensors providing data of increasing spatial and temporal resolution, the role of remote-sensing-based applications has become increasingly important for supporting humanitarian relief operations, especially in remote or difficult to access areas [31,32]. Remote sensing including high-resolution satellite imagery has been used to provide evidence of new refugee settlements [33], and to develop detailed maps of settlements by detecting and classifying infrastructure and structures within these [34,35].

Current practice focused on detecting and classifying structures in settlements relies on the manual analysis of remote-sensing data, requiring the identification and interpretation of high-resolution satellite imagery by trained analysts [30,36–38]. Although highly accurate, this work is time consuming and is labour-intensive, which may limit its applicability in response to crises situations, or in large areas that require regular monitoring over long periods of time. Decentralized approaches, for example crowd-sourcing or distributing the manual analyses of satellite imagery among several analysts, have been used to improve the scale and timeliness of these tasks. They can also create additional management and quality-control challenges, though.

Other more efficient automatic or semi-automated methods including machine learning are showing promise in improving the efficiency of analyst workflows (e.g. [30,34,38–41]). In terms of computer-assisted building detection, several case studies have been conducted, included pixel-based classification (e.g. [30]), object-based classification rule sets (e.g. [38,39,41]) and approaches based on mathematical morphology methods incorporating morphology thresholds (e.g. [30,34, 40,42]). However, automating the detection and classification of structures using satellite imagery, with sufficient accuracy to be practical, is still an open problem.

(a) Public data science challenges in humanitarian remote sensing

In recent years, numerous challenges have been launched among the global data science community, the object of these being to crowd-source the development of machine learning techniques for automating the analysis of remote-sensing imagery, with some of them, in particular, in the context of humanitarian or sustainable development efforts.

The theme of the DIUx xView 2018 Detection Challenge was to detect emerging natural disasters.[2] The DeepGlobe CVPR 2018-Satellite Challenge focused on detecting roads, buildings and land cover,[3] while a DigitalGlobe challenge focused generally on the creation of accurate maps for potential use in future disaster response situations.[4] A recent challenge set by the Defence Science and Technology Laboratory (Dstl) requested competitors to develop machine learning methods to automatically detect and label significant features such as waterways, buildings and vehicles using multi-spectral satellite imagery,[5] for which the winning entries were all autoencoder models. The Crowd AI mapping challenge[6] is aimed at the detection of buildings, for use in humanitarian response in areas which are otherwise not mapped in detail.

Other crowd-sourced challenges have offered remote-sensing data for monitoring adverse anthropogenic impacts on the Amazon rainforest, including deforestation, biodiversity losses and habitat losses.[7] The Data for Climate Action Challenge offered satellite imagery (3–5 m resolution) as part of a broader pool of data resources to research insights and solutions for climate-change mitigation and adaptation, and sustainable development efforts.[8] Another challenge offered very high-resolution imagery (less than 10 cm) from the UAVs for Disaster Resilience Program to accelerate and improve humanitarian and development efforts of the South Pacific Island.[9] The

[2]https://xviewdataset.org

[3]https://deepglobe.org

[4]https://blog.digitalglobe.com/developers/the-spacenet-challenge-round-2-has-launched

[5]https://www.kaggle.com/c/dstl-satellite-imagery-feature-detection

[6]https://www.crowdai.org/challenges/mapping-challenge

[7]https://www.planet.com/pulse/forest-recognition-planet-launches-kaggle-competition

[8]https://www.dataforclimateaction.org

[9]https://werobotics.org/blog/2018/01/10/open-ai-challenge

objective of this challenge was to develop machine learning classifiers to automate the analyses of the imagery, including developing baseline maps and conducting damage assessment. Examples of campaigns aiming to crowd-source the analysis of satellite imagery during crises include Amnesty International's Decode Darfur[10] and Decode the Difference[11] projects.

4. Case study: structure counting in refugee settlements

As discussed above, counting the numbers of different types of structures in a refugee or IDP settlement is a common analysis task. In practice, this is currently routinely done by human expert analysts, though the repetitive nature of this task makes it a natural candidate for automation. A single settlement may have tens of thousands of structures, and identifying them from imagery is a task which can take an experienced analyst some days, with quality control checks by a second analyst adding further time. We show examples of structures in a variety of refugee/IDP settlements mapped by UNOSAT in figure 1.

As an object detection problem, there are three main difficulties in practice. First, there is a high degree of variation between different settlements, making it difficult to train a model on one settlement and have it generalize well to others. This variation can be due to the settlements being in regions with different terrain (e.g. desert or savannah), with different types of structures (e.g. tents, semi-permanent structures or improvised shelters made of any materials to hand), or because imagery is collected by different sensors, in different weather conditions or time of day.

Second, the objects being detected are small and sometimes clustered close together. An improvised shelter 3 m across manifests in 50 cm resolution imagery as a blob 6 pixels across. Such shelters may be made from natural materials to hand in the area, making them appear similar to the rest of the terrain. Distinguishing such structures from other objects (including rocks, bushes, small uninhabited structures and so on) can be a matter of judgement requiring domain knowledge and reasoning about the context—which, despite recent advances in object detection with deep learning algorithms, is often difficult even for current state-of-the-art methods.

Third, for this task a high degree of accuracy is needed. Because the results of the analysis are used to inform critical decisions about the resources needed to maintain a settlement, current quality control procedures are rigorous. Even if we view the model outputs as not the final product of themselves, but as sets of candidate structure detections which could speed up the work of an analyst, unless precision and recall are high enough it can take more work for the analyst to correct an incomplete set of detections than to start from scratch. False detections are more of a problem than missed detections, since identifying them and then cleaning them up (within an interface such as ArcGIS) is a lengthier process.

(a) Data

The data used in this case study was annotated high-resolution imagery from 13 refugee/IDP settlements listed in table 1. These images were in some cases composites of separate images collected by different satellites and/or at different times, in the cases that there was no image available covering the entire settlement at once. Three bands were used where possible, though for some settlements only a panchromatic channel was available. Accompanying these images were longitude/latitude point locations of structures within those settlements, identified by experienced analysts with quality-control checks done by a secondary analyst. The points data for most of the settlements we analysed in this study, as well as maps giving more context for the settlements, is publically available online.[12]

---

[10]https://decoders.amnesty.org/projects/decode-darfur

[11]https://decoders.amnesty.org/projects/decode-the-difference

[12]http://www.unitar.org/unosat/maps and https://data.humdata.org/organization/un-operational-satellite-appplicati ons-programme-unosat.

(*a*) Doro, South Sudan ©2018 CNES, airbus

(*b*) Dalakaleri, Nigeria ©2018 DigitalGlobe
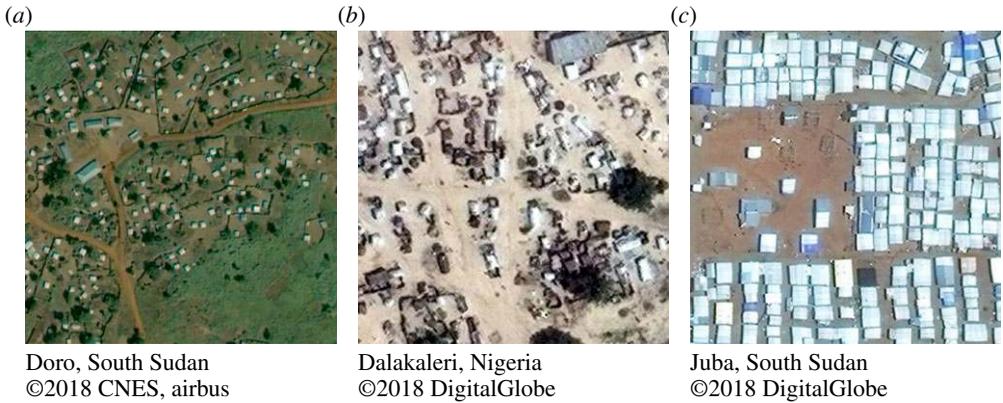
(*c*) Juba, South Sudan ©2018 DigitalGlobe

Figure 1. Examples of satellite imagery of refugee/IDP settlements in which dwelling structures are visible.

Table 1. Refugee/IDP settlements analysed in these experiments. Doro 1 and Doro 2 refer to the same location at different times (analysis was carried out for snapshots in 2014 and 2017, respectively).

settlement	country	structures	area (km^2)	image bands
Ajuong	South Sudan	13 395	18.2	1
Doro 1	South Sudan	16 463	8.0	3
Doro 2	South Sudan	12 819	8.0	3
Ganyel	South Sudan	3415	29.3	3
HTC	Iraq	4253	2.1	1
Juba	South Sudan	11 096	0.8	3
K-18	Iraq	1247	0.5	1
Khaldiyah	Iraq	3844	1.5	1
Muna	Nigeria	2822	0.7	3
Ngala	Nigeria	4488	1.3	3
Nyal	South Sudan	5249	54.0	3
Yida	South Sudan	17 064	26.9	3
Wau	South Sudan	5177	0.13	3

The images were split into 300×300 pixel tiles, and in order to train segmentation models we manually traced out polygons corresponding to each structure, for a total of 87 137 structures. Preprocessing code was written to augment these data for use while training, varying the brightness and scale, and randomly rotating and flipping the tiles. This augmentation step was an important stage of data preparation since in contrast to general computer vision problems, here we have data from a relatively small number of contexts; tiles look relatively similar within each settlement, increasing the risk of overfitting.

(b) Model

To carry out object detection, we use the Mask-RCNN model [9], which has shown good performance on detection and segmentation problems in other domains. This model is constructed to simultaneously predict the bounding boxes of objects in an input image, the class

of each of those objects and a pixel segmentation mask. The ability to provide pixel segmentations is a particular attraction in this application, as it allows computations about the total roof area of structures in a settlement.

The structure of the Mask-RCNN is that an input layer is first connected to a feature extraction stage, typically using a different pre-trained network. We use ResNet101 [43] pre-trained on ImageNet and with the head layers removed. Connected to the upper feature extraction layers, there are three output sub-networks, for object bounding boxes, classes and segmentation masks, respectively. Bounding boxes are identified by associating objects with *anchor boxes*, which are a set of overlapping rectangular regions in the image with varying scale and aspect ratio, and then outputting offsets from the anchor box. Pixel segmentations are predicted in the form of a fixed-size grid (we use 28×28) aligned over the bounding box. Further details of the Mask-RCNN model structure can be found in [9]. We used a modified version of an open source implementation.[13]

(c) Experiments

We carried out two sets of experiments with this model and data. The first was to test the extent to which structures in a settlement can be detected and enumerated by a model trained only with images from other settlements. This corresponds to the situation that imagery for a new refugee settlement is available, and we require an immediate count using a pre-trained model (trained in the past on other refugee settlements). We refer to models trained in this way as *base nets* for each settlement; the base net for some settlement is trained with all available data from other settlements. The second set of experiments looked at the change in performance when a small amount of training data from the target settlement is available, so that we can carry out transfer learning to improve the fit of the base net to the specific settlement that needs to be assessed. This would correspond to the case that a new settlement is to be analysed, but that there is a little time available to provide locations of some reference structures and update the model, in order to approve accuracy. In these experiments, we used different proportions of the area of the test settlement as adaptation data. Adaptation data were randomly selected from the settlement, though we note that more effective strategies may be possible, for example, ensuring that there is adaptation data with a balanced selection of the different types of structures visible in the settlement. For a particular adaptation dataset, we evaluate on the remainder of the area of the settlement. Training of networks took 2–3 h on a machine with 4 GPUs, whereas each adaptation phase took approximately 20 min of training on a single GPU. Detection using a trained model on a single GPU took around 400 ms per 300×300 tile, or approximately $18 \, \text{s} \, \text{km}^{-2}$ of imagery (at 40–50 cm resolution). This is already fast enough for practical application and could likely be considerably speeded up further with model compression.

Results are shown in figure 2, with precision and recall calculated for each settlement with the basenet and the adapted network using from 10 to 50% of the true structures for training. We evaluate precision and recall by considering any detected bounding box that coincides with a true bounding box with intersection over union (IoU) greater than 0.25 as a true positive. Average precision (AP) figures are given, which are the areas under these curves. In general, the accuracy increases when adaptation data are available, as we would expect; though the extent to which this is true depends on how unique each settlement was with respect to the other settlements used for training each network. We also note that the lowest performance was for the settlements in which the density of structures was very high. The two settlements with significantly lower AP than the others, Juba (pictured in figure 1 and Wau, have on average 17 497 structures per km². The best-scoring two settlements for the adapted models were K18 and Khaldiyah, with average 2545 structures per km². The model has trouble distinguishing individual structures when they are dense and even adjoining each other. Interestingly, these two best-scoring settlements had
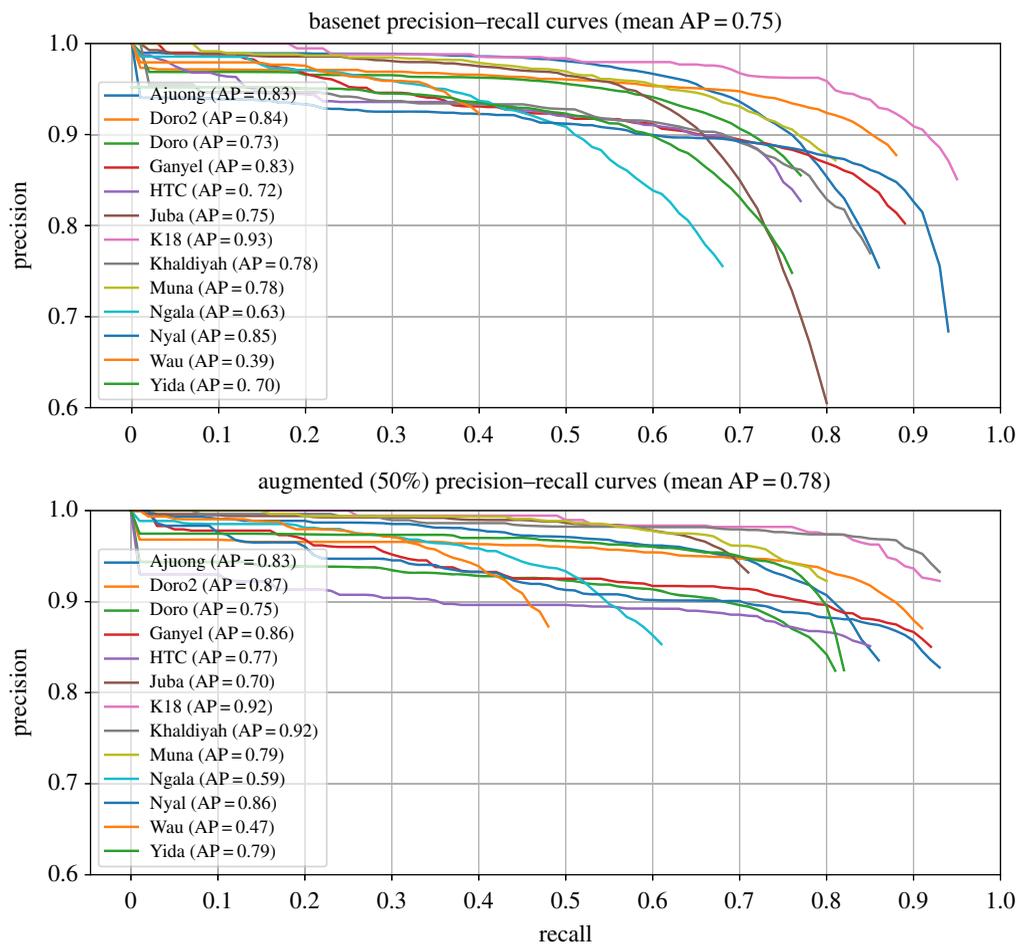
---

[13]https://github.com/facebookresearch/Detectron

Figure 2. Precision–recall curves for basenets (tested on entire settlement) and models augmented with 50% of settlement data (tested on remaining 50%).

only greyscale (panchromatic) imagery, suggesting that the characteristics of the settlement are a much more important factor for accuracy of structure detection than the availability of multi-band colour images.

The extent to which adaptation data helped detection accuracy is shown for sample settlements in figure 3, where we zoom in on the high-precision, high-recall area of each curve and show the different between the basenet and each of the adapted nets. For some settlements, adaptation data give a strong improvement; other less so, and sometimes with accuracy in fact decreasing. Note that for each quantity of adaptation data, the test set changes each time (as we only evaluate on parts of the settlement which were not used for adaptation training). Hence the performance is not guaranteed to increase with increasing adaptation data.

Figure 4 illustrates these results by showing detections for one sample tile in Doro settlement. The basenet is unable to detect several of the structures, because there are structures with appearances particular to this settlement. With a small amount of adaptation data, these false negatives are corrected. Figure 5 shows detections across the extent of an entire settlement, compared to ground truth locations. Although some false positives are evident, the overall structure of the settlement is clearly detected. Post-processing to remove outliers is likely to improve results further.
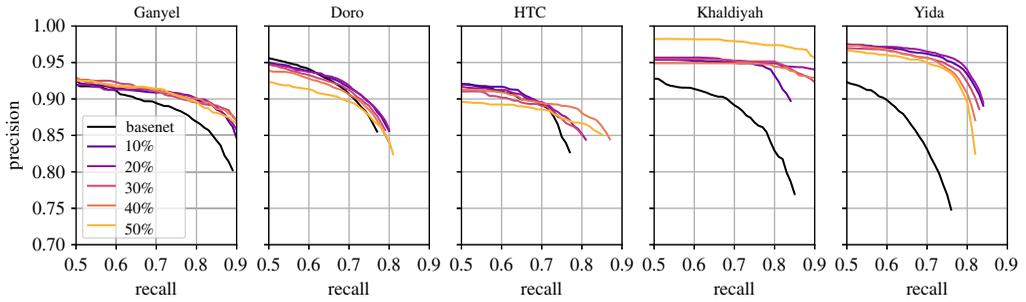
Figure 3. Detail of precision–recall curves showing changes when varying levels of adaptation data are available for specific settlements.
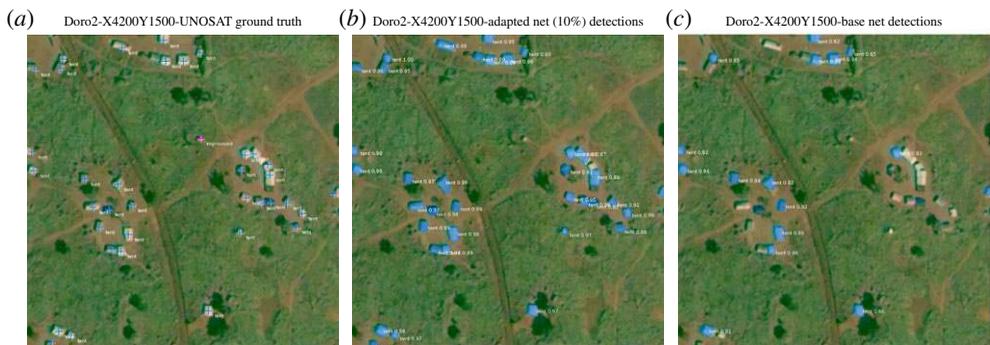

Figure 4. Examples of structure detections in Doro settlement. (*a*) Ground truth locations of structures; (*b*) detected structure polygons with adapted network and (*c*) detected structure polygons with basenet. Numbers in white denote detection confidences. Imagery © 2018 DigitalGlobe.
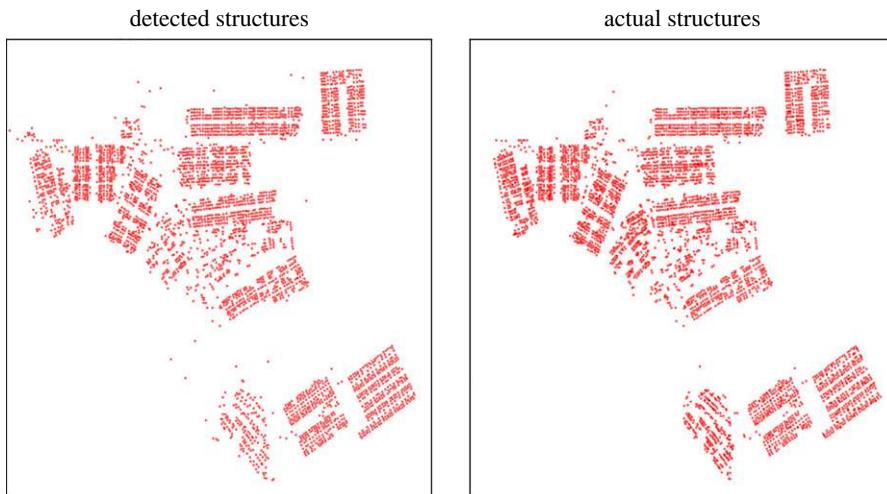

Figure 5. Detected structure locations across the extent of HTC settlement, compared to actual structure locations. (Online version in colour.)
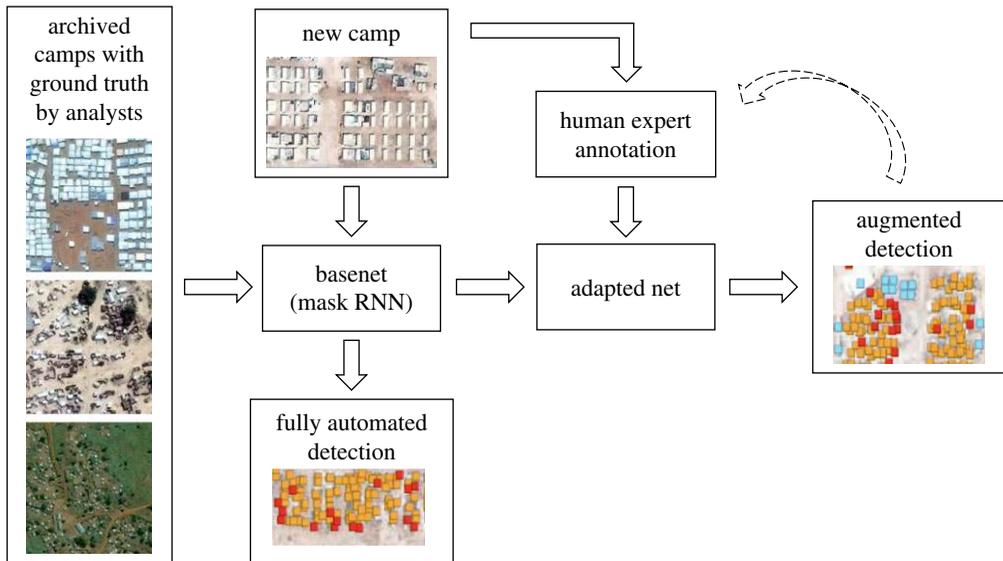
Figure 6. Schematic of the assisted mapping process with basenets and adapted nets. At each iteration, a human expert corrects some proportion of the latest output, which is then used for re-training the adapted net to improve accuracy.

5. Discussion

In this paper, we reviewed remote-sensing technologies and machine learning methods for guiding humanitarian responses to conflict, human rights violations and man-made or natural disasters. We also reviewed previous work in automating the development of detailed maps of refugee settlements for estimating their populations. We provided details of a case study in which we applied deep learning methods to detect structures on multiple refugee settlements in a number of locations in Africa and the Middle East.

In our analyses, we demonstrated that it is possible to detect a large proportion of structures within the settlements studied. However, there is still considerable variation in the characteristics of imagery collected from different satellite sensors, geographical regions and settlements types, and this was reflected in our results of our study. As such, this necessitated a semi-automated, interactive learning approach in order to reach usable levels of accuracy when translating generally trained models to specific locations. In a leave-one-out training and testing strategy for a base network, the accuracy varies considerably, and is generally best for settlements which have similar characteristics to those in the training data and for those where the structures are not densely packed (in which case the segmentation of distinct structures is a difficulty). The adapted results are significantly closer to having practicable accuracy, as we might expect. Also as we would expect, the improvement from adaptation data varies according to how unusual the test settlement is compared to the training data with which the basenet was trained with.

As the detection of structures is typically conducted through the manual analyses and visual interpretation of satellite imagery, there is considerable potential for automatic analyses to augment human analyses tasks, therefore reducing the amount of work needed to provide an accurate assessment of the images in practice [44,45]. However, there are still multiple challenges to be addressed when incorporating such methods into practical work flows. Figure 6 shows a schematic of how we envisage an assisted mapping process using basenets and adaptive nets. At each iteration stage, a human analyst would provide corrections to a certain proportion of the most recent output. This corrected information is then used as extra training data to further adapt the network. After some number of iterations, when the analyst is satisfied that the accuracy is at an acceptable level, the latest output is used as the final structure mapping.

A number of methodological improvements may fine-tune the performance of detection models in future work. These may include more complex or varied network structures, data augmentation strategies or the inclusion of post-processing techniques incorporating contextual knowledge relevant to the location being examined [46]. It is also important to understand the rates of deterioration of results when transferring models to unknown scenarios and geographies.

Technical difficulties with regards to computational capacity may be influenced by the fact that analyst interaction cycles will need to be integrated with neural network training and classification in as close to real time as possible. As an analyst manually detects structures in a settlement, a machine learning model, computing in parallel, will use this information to automatically detect the remaining structures in the settlement. It is possible that future work will include the development of an evolving evaluation framework, incorporating manually tagged structures that will inform the stop criteria for the adaptation, i.e. the proportion of adaptation data for a new location that would be required as input from an analyst to the machine learning model, so that it would automatically detect the remaining structures.

One of the key barriers to the adoption of computer-assisted methods to practical work flows is related to trust. In order to facilitate a transition from current fully manual operational pipelines to more efficient methods, the user needs to be aware of the performance metrics of the technology in certain scenarios and become accustomed to it. For automated or semi-automated methods to be adopted, it is necessary for the technology to be seamlessly integrated into the manual processing work flow. Even though there are some design criteria that might not optimize the algorithmic performance, it may still provide opportunity for greater synergy between humans and machines. For instance, producing false positives carries a psychological weight, since deleting false positives may seem like an arduous task which hampers progress.

Human experts apply contextual knowledge, such as reasoning about structures from their shadows, for example, that a structure with internal shadows may not have a roof. Although there is some research on how to include context in automatic image analysis [46], it is difficult to encode this type of knowledge or reasoning into current object detection methods. Sometimes what can be complicated for the machine is trivial for humans and what can be tedious for humans might be easy for machines. In this context, the key for further systems might be to evaluate the collective performance of the humans assisted by algorithms as a whole—and do not evaluate independently humans and algorithms. Further research should therefore consider the entire workflow, as the assessment of satellite imagery involves some degree of interpretation and use of contextual information which can pose difficulties for automated processing.

Privacy, sensitivity and ethical dilemmas are major considerations for this work. With the emergence of high-resolution, high-frequency imagery that is easily accessible, it is imperative that we seriously consider the privacy implications and the potential unintended consequences of sharing or using satellite imagery. In humanitarian contexts, vulnerable populations are particularly exposed, and any experimental uses and new methodologies should occur under agreed normative frameworks which follow ethical and responsible use principles.

6. Conclusion

In this paper, we have discussed a number of applications of machine learning to remote sensing in humanitarian emergencies. For the majority of these applications, there are some prototype-stage results, but as yet there are few systems with the accuracy and robustness to be deployed in a crisis situation. However, given the pace of advances in computer vision, where in other domains neural networks are matching or even surpassing human performance, it is likely that more practical systems will emerge. In our own work looking at remote sensing structure identification, of which the experiments reported in this paper are the most recent, we have obtained significantly better results in the last 1–2 years with the availability of better models and tools. This, combined with the increasing availability of remote-sensing image data and even video from new satellites and drones, is likely to increase the possibilities for automation.

At the same time, humanitarian scenarios are particularly challenging in that a very high degree of accuracy is needed, and in some cases approximative results may be worse than useless. Time also matters. In this regard, machine learning in these applications is somewhat similar to medical imaging. Important decisions must be made on the basis of what may sometimes be small and indistinct features in an image. However, whereas medical imaging normally involves carefully calibrated equipment, high-resolution remote-sensing imagery is affected by many factors including the sensor, climate, time of day and other factors.

We conclude that while there are many promising lines of research into humanitarian applications of machine learning on remote-sensing data, fully automated processing is not yet practical in the majority of cases. Structure counting in refugee settlements is an example, though one which also illustrates the possibilities of combined human–machine analysis, i.e. where human experts help to calibrate a model and also to post-process the model's output. Augmentation of human capabilities is, therefore, a good strategy, to aim for human experts aided by machine learning systems to be able to carry out analysis with high throughput and yet maintaining the necessary levels of quality control.

References

1. Santilli G, Vendittozzi C, Cappelletti C, Battistini S, Gessini P. 2018 CubeSat constellations for disaster management in remote areas. *Acta Astronaut.* **145**, 11–17. (doi:10.1016/j.actaastro.2017.12.050)
2. Maxwell AE, Warner TA, Fang F. 2018 Implementation of machine learning classification in remote sensing: an applied review. *Int. J. Remote Sens.* **39**, 2784–2817. (doi:10.1080/01431161.2018.1433343)
3. Pagot E, Pesaresi M. 2008 Systematic study of the urban post-conflict change classification performance using spectral and structural features in a support vector machine. *IEEE J. STARS* **1**, 120–128. (doi:10.1109/jstars.2008.2001154)
4. Mountrakis G, Im J, Ogole C. 2011 Support vector machines in remote sensing: a review. *ISPRS J. Photogramm. Remote Sens.* **66**, 247–259. (doi:10.1016/j.isprsjprs.2010.11.001)
5. Belgiu M, Dragut L. 2016 Random forests in remote sensing: a review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **114**, 24–31. (doi:10.1016/j.isprsjprs.2016.01.011)
6. Zhu XX, Tuia D, Mou L, Xia GS, Zhang L, Xu F, Fraundorfer F. 2017 Deep learning in remote sensing: a comprehensive review and list of resources. *IEEE Geosci. Remote Sens. Mag.* **5**, 8–36. (doi:10.1109/mgrs.2017.2762307)
7. Zhang L, Zhang L, Du B. 2016 Deep learning for remote sensing data: a technical tutorial on the state of the art. *IEEE Geosci. Remote Sens. Mag.* **4**, 22–40. (doi:10.1109/MGRS.2016.2540798)

8. Ren S, He K, Girshick R, Sun J. 2017 Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans. Patt. Anal. Mach. Intell.* **6**, 1137–1149.

9. He K, Gkioxari G, Dollar P, Girshick R. 2018 Mask R-CNN. See https://arxiv.org/pdf/1703.06870.pdf.

10. Minetto R, Pamplona Segundo M, Sarkar S. 2018 Hydra: an emsemble of convolutional neural networks for geospatial land classification. See http://arxiv.org/abs/1802.03518 (accessed March 2018).

11. Zou Q, Ni L, Zhang T, Wang Q. 2015 Deep learning based feature selection for remote sensing scene classification. *IEEE Geosci. Remote Sens.* **12**, 2321–2325. (doi:10.1109/LGRS.2015.2475299)

12. Hu F, Xia GS, Hu J, Zhang L. 2015 Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. *Remote Sens.* **7**, 14 680–14 707.

13. Van de Walle B, Van Den Eede G, Muhren W. 2009 Humanitarian information management and systems. In *Mobile response* (eds J Löfler, M Klann), pp. 12–21. (LNCS 5424). Berlin, Germany: Springer.

14. UNOSAT-UNITAR. 2015 Impact of the 2014 conflict in the Gaza Strip: UNOSAT satellite derived geospatial analysis. See https://unosat.web.cern.ch/ (accessed March 2018).

15. UNOSAT-UNITAR. 2018 Rapidly assessed damage occurring between 23rd February 2018 and 6 March 2018 in Eastern Ghouta Area, Damascus. See http://www.unitar.org/unosat/ (accessed March 2018).

16. UNOSAT-UNITAR. 2018 Damage assessment of eastern part of Nuku-alofa, Tongatapu Island, Tonga. See http://www.unitar.org/unosat/ (accessed March 2018).

17. Witmer FDW. 2015 Remote sensing of violent conflict: eyes from above. *Int. J. Remote Sens.* **36**, 2326–2352. (doi:10.1080/01431161.2015.1035412)

18. Bromley L. 2010 Relating violence to MODIS fire detections in Darfur, Sudan. *Int. J. Remote Sens.* **31**, 2277–2292. (doi:10.1080/01431160902953909)

19. Marx A, Loboda T. 2013 Landsat-based early warning system to detect the destruction of villages in Darfur, Sudan. *Remote Sens. Environ.* **136**, 126–134. (doi:10.1016/j.rse.2013.05.006)

20. Prins E. 2008 Use of low cost landsat ETM+ to spot burnt villages in Darfur, Sudan. *Int. J. Remote Sens.* **29**, 1207–1214. (doi:10.1080/01431160701730110)

21. Anderson D, Lochery E. 2008 Violence and exodus in Kenya's rift valley, 2008: predictable and preventable?. *J. East. Afr. Stud.* **2**, 328–343. (doi:10.1080/17531050802095536)

22. Li X, Li D. 2014 Can night-time light images play a role in evaluating the Syrian Crises? *Int. J. Remote Sens.* **35**, 6648–6661. (doi:10.1080/01431161.2014.971469)

23. Witmer FDW, O'Loughlin J. 2011 Detecting the effects of wars in the Caucasus regions of Russia and Georgia using radiometrically normalized DMSP-OLS nighttime lights imagery. *Gisci. Remote Sens.* **48**, 478–500. (doi:10.2747/1548-1603.48.4.478)

24. Marx A, Goward S. 2013 Remote sensing in human rights and international humanitarianlaw monitoring: concepts and methods. *Geographic. Rev.* **103**, 100–111. (doi:10.1111/j.1931-0846.2013.00188.x)

25. Knoth C, Pebesma E. 2017 Detecting dwelling destruction in Darfur through object-based change analysis of very high-resolution imagery. *Int. J. Remote Sens.* **38**, 273–295. (doi:10.1080/01431161.2016.1266105)

26. UNOSAT-UNITAR 2018c. Damage assessment of eastern part of Nuku-alofa, Tongatapu Island, Tonga. See http://www.unitar.org/unosat/map/2775 (accessed July 2018).

27. Cooner AJ, Shao Y, Campbell JB. 2016 Detection of urban damage using remote sensing and machine learning algorithms: revisiting the 2010 Haiti earthquake. *Remote Sens.* **8**, 868. (doi:10.3390/rs8100868)

28. UNHCR (United Nations High Commission for Refugees). 2017 *Global trends: forced displacement in 2016*. Geneva, Switzerland: UNHCR.

29. Taubenbock H, Kraff NJ, Wurm M. 2018 The morphology of the arrival city—a global categorization based on literature surveys and remotely sensed data. *Appl. Geogr.* **92**, 150–167. (doi:10.1016/j.apgeog.2018.02.002)

30. Giada S, DeGroeve T, Ehrlich D, Soille P. 2003 Information extraction from very high resolution satellite imagery over Lukole refugee camp, Tanzania. *Int. J. Remote Sens.* **24**, 4251–4266. (doi:10.1080/0143116021000035021)

31. Kranz O, Zeug G, Tiede D, Clandillon S, Bruckert D, Kemper T, Lang S, Caspard M. 2010 Monitoring refugee/IDP camps to support international relief action. In *Geoinformation for*

disaster and risk management—examples and best practices. Joint Board of Geospatial Information Societies (JB GIS), United Nations Office for Outer Space Affairs (UNOOSA) (eds O Altan, R Backhaus, P Boccardo, S Zlatanova), pp. 51–56.

32. Kuffer M, Pfeffer K, Sliuzas R. 2016 Slums from space-15 years of slum mapping using remote sensing. *Remote Sens.* **8**, 455. (doi:10.3390/rs8060455)

33. UNITAR. 2011 UNOSAT Brief 2011—satellite applications for human security. Geneva, Switzerland: United Nations Institute for Training and Research.

34. Wang S, So E, Smith P. 2015 Detecting tents to estimate displaced populations for post-disaster relief using high resolution satellite imagery. *Int. J. Appl. Earth Observ. Geoinf.* **36**, 87–93. (doi:10.1016/j.jag.2014.11.013)

35. Aravena Pelizari P, Sprohnle K, Geib C, Schoepfer E, Plank S, Taubenbock H. 2018 Multi-sensor feature fusion for very high spatial resolution built-up area extraction in temporary settlements. *Remote Sens. Environ.* **209**, 793–807. (doi:10.1016/j.rse.2018.02.025)

36. Bjørgo E. 2000 Using very high spatial resolution multispectral satellite sensor imagery to monitor refugee camps. *Int. J. Remote Sens.* **21**, 611–616. (doi:10.1080/014311600210786)

37. Checchi F, Stewart BT, Palmer JJ, Grundy C. 2013 Validity and feasibility of a satellite imagery-based method for rapid estimation of displaced populations. *Int. J. Health Geogr.* **12**, 1–12. (doi:10.1186/1476-072X-12-1)

38. Spröhnle K, Tiede D, Schoepfer E, Füreder P, Svanberg A, Rost T. 2014 Earth observation-based dwelling detection approaches in a highly complex refugee camp environment—a comparative study. *Remote Sens.* **6**, 9277–9297. (doi:10.3390/rs6109277)

39. Lang S, Tiede D, Holbling D, Fureder P, Zeil P. 2010 Earth observation (EO)-based ex-post assessment of IDP camp evolution and population dynamics in Zam Zam, Darfur. *Int. J. Remote Sens.* **31**, 5709–5731. (doi:10.1080/01431161.2010.496803)

40. Kemper T, Jenerowicz M, Soille P, Pesaresi M. 2011 Enumeration of dwellings in Darfur camps from GeoEye-1 satellite images using mathematical morphology. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **4**, 8–15. (doi:10.1109/JSTARS.2010.2053700)

41. Spröhnle K, Fuchs EM, Aravena Pelizari P. 2017 Object-based analysis and fusion of optical and SAR satellite data for dwelling detection in refugee camps. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **10**, 1780–1791. (doi:10.1109/JSTARS.2017.2664982)

42. Heinzel J, Kemper T. 2014 Use of new coastal spectral band for precise dwelling extraction in the Hagaderea refugee camp. In *Proc. of ESA-EUSC-JRC 9th Conf. on Image Information Mining*.

43. He K, Zhang X, Ren S, Sun J. 2016 Deep residual learning for image recognition. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 770–778.

44. Knoth C, Slimani S, Appel M, Pebesma E. 2018 Combining automatic and manual image analysis in a web-mapping application for collaborative conflict damage assessment. *Appl. Geogr.* **97**, 25–34. (doi:10.1016/j.apgeog.2018.05.016)

45. Hu Y, Janowicz K, Couclelis H. 2016 Prioritising disaster mapping tasks for online volunteers based on information value theory. *Geogr. Anal.* **49**, 175–198. (doi:10.1111/gean.12117)

46. Tiede D, Krafft P, Fureder P, Lang S. 2017 Stratified tempplate matching to support refugee camp analysis in OBIA workflows. *Remote Sens.* **9**, 326. (doi:10.3390/rs9040326)