International Journal of Health Geographics

RESEARCH

Open Access

Using spatial video and deep learning for automated mapping of ground-level context in relief camps

Jayakrishnan Ajayakumar[1*], Andrew J. Curtis[1], Felicien M. Maisha[2,6], Sandra Bempah[3], Afsar Ali[2,4], Naveen Kannan[1], Grace Armstrong[1] and John Glenn Morris Jr[2,5]

Abstract

Background The creation of relief camps following a disaster, conflict or other form of externality often generates additional health problems. The density of people in a highly stressed environment with questionable safe food and water access presents the potential for infectious disease outbreaks. These camps are also not static data events but rather fluctuate in size, composition, and level and quality of service provision. While contextualized geospatial data collection and mapping are vital for understanding the nature of these camps, various challenges, including a lack of data at the required spatial or temporal granularity, as well as the issue of sustainability, can act as major impediments. Here, we present the first steps toward a deep learning-based solution for dynamic mapping using spatial video (SV).

Methods We trained a convolutional neural network (CNN) model on a SV dataset collected from Goma, Democratic Republic of Congo (DRC) to identify relief camps from video imagery. We developed a spatial filtering approach to tackle the challenges associated with spatially tagging objects such as the accuracy of global positioning system and positioning of camera. The spatial filtering approach generates smooth surfaces of detection, which can further be used to capture changes in microenvironments by applying techniques such as raster math.

Results The initial results suggest that our model can detect temporary physical dwellings from SV imagery with a high level of precision, recall, and object localization. The spatial filtering approach helps to identify areas with higher concentrations of camps and the web-based tool helps to explore these areas. The longitudinal analysis based on applying raster math on the detection surfaces revealed locations, which had a considerable change in the distribution of tents over space and time.

Conclusions The results lay the groundwork for automated mapping of spatial features from imagery data. We anticipate that this work is the building block for a future combination of SV, object identification and automatic mapping that could provide sustainable data generation possibilities for challenging environments such as relief camps or other informal settlements.

Keywords Deep learning, Automated mapping, Spatial video

*Correspondence:
Jayakrishnan Ajayakumar
jxa421@case.edu
[1]Department of Population and Quantitative Health Sciences, School of Medicine, Case Western Reserve University, Cleveland, OH, USA
[2]Emerging Pathogens Institute, University of Florida, Gainesville, FL, USA
[3]Department of Geography, Kent State University, Kent, OH, USA
[4]Department of Environmental & Global Health, College of Public Health and Health Professions, University of Florida, Gainesville, FL, USA
[5]College of Medicine, University of Florida, Gainesville, FL, USA
[6]Department of Social Sciences, College of Humanity and Social Sciences, University of Goma, Goma, Democratic Republic of the Congo

Introduction

One of the greatest challenges in serving the health needs of people displaced into temporary camps after a disaster or conflict is how to continuously gauge and contextually map changing human needs and the placement of resources. These types of ephemeral locations are difficult to map in any environment, and for many developing world situations, such as in the Democratic Republic of Congo, the setting for this paper, these data challenges can be unassailable. In the greater Goma area of the DRC, temporary camps were created for a variety of reasons: in 2021, to serve the population displaced by the Nyiragongo volcanic on 22 May, and in 2022, as a result of the conflict with the M23 rebel group, with an additional 3,000 families being displaced due to flooding. These camps pose considerable health and safety challenges, leading to a variety of reported problems, including violence, rape and other sexual violence, as well as disease outbreaks [1], and even devastating fires in August 2023 [2]. In addition to being initially displaced, other vulnerable cohorts sometimes seek refuge in these camps to benefit from the assistance of NGOs. This results in a highly dynamic situation with additional internal and between-camp mobility.

From a mapping perspective, tasks of importance include the characteristics of the camp, which include where people are living, their activity spaces, and the locations of services, including (safe) food, water and sanitation. The authors of this paper have previously described and mapped the growth of one of these first camps, the Mujoga relief camp, using spatial video (SV), which is a hand-carried video camera with simultaneously collected Global Positioning System (GPS) coordinates [3–5]. This yearlong process captured the dynamic nature or shifting context within the camp. Mapping was performed because features were digitized from each spatially encoded video frame and then contextualized using mixed methods [5]. The camp in question was a highly dynamic space with refugees moving both in and out of the camp, including the growth of an "informal" tent sector within the camp. In addition, the safety of key infrastructure, such as water points and toilets, changed over the time frame of the mapping, with their visible deterioration being linked to the end of NGO funding [5]. That paper concluded that the ending of funding had created a perfect storm that might increase the likelihood of cholera occurrence. Unfortunately, that concern proved valid, especially when the conflict with the M23 rebel group led to even more displaced people and the growth of additional temporary camps in the Goma region. Figure 1 displays the approximate location for some of these camps with the total number of cholera cases recorded up to the end of January 2023, although these camps are likely to be undercounts. The figure also includes inset images of tents extracted from the SV for each camp. As further illustration of the dynamism involved, the Bushagara and Don Bosco camps opened to take the overflow of displaced people, and Munigi, which is the largest camp with a cholera treatment center (CTC), would take referrals from all other camps.

Although this mapping approach provided a useful early warning of what was to come, extending the approach to other camps would require considerable (and unsustainable) human mapping effort. SV data have been collected for approximately one year in all the other camps, but the resources are not available for the same type of mapping. One possibility is to use machine learning to automatically map the SV data collected by field epidemiological teams. To explore this possibility, we will use the longitudinal SV data collected for the Mujoga camp and attempt to automatically identify the number of tents visible in each image frame of the SV. We then utilize the spatial information from the SV, along with spatial filtering, to generate a continuous spatial distribution of visible tents. Furthermore, to quantify the change in the spatial distribution of visible tents over time, we will utilize a raster calculator to generate difference maps for the distribution of tents. While we understand that other remotely sensed options are available for the tasks presented here [6–8], what is proposed is a more on-the-ground sustainable solution that can also be used to eventually extract more detail and context from overhead imagery.

Mapping relief camps

Being able to develop sustainable, granular-scale mapping for a temporary relief camp is essential for humanitarian organizations to strategize their resources and decision-making [9]. With circumstances that often mimic other informal settlements, such as a lack of immediate family resources, overcrowding, stress, problematic sanitary services, and poor health care access, these create a complex landscape in which different diseases can thrive. While for some relief camps these challenges are initially reduced through an influx of NGO funding, the situation can quickly deteriorate once these resources end. Producing effective maps to understand where why and how to intervene in these environments is challenging because of the paucity of data [10], especially granular geographic data of the type useful for response teams. Previous attempts to fill this spatial data gap have included using cellular-mobile data [11] or satellite imagery [12]. Drawbacks might include cross-sectional approaches due to logistical limitations, which include both data acquisition and local expertise [13]. Recent research on detecting and mapping informal settlements has utilized remote sensing data, especially satellite data [10–16], which can be used for image classification, object detection and

Fig. 1 Camp locations in Goma along with cholera case details as of January 2023

semantic segmentation [17]. Combining this geographically referenced satellite imagery with crowd sourcing has also been used to map critical infrastructure, such as tents, toilets, medical care, and automated teller machines (ATMs) [18]. While the results are encouraging, the logistics involved are still beyond many typical camp situations.

Although our own mapping of informal settlements has traditionally involved intensive manual digitization, recent advances in deep learning, especially convolutional neural networks (CNNs), have provided potentially new automated or semiautomated solutions for land cover classification, digitization and cartography [19–21]. Combining remote sensing and neural network methods

Fig. 2 Ground-level image of a location in Goma from (**a**) Google Earth on November 7, 2021, and (**b**) a spatial video on November 9, 2021

such as CNNs is increasingly being used to guide humanitarian responses during conflicts, human rights violations and various disasters. For example, Wang et al. [14] used a mathematical morphology-based method to generate camp maps from high-resolution satellite imagery to estimate displaced populations. A mask region-based CNN model was developed by Gella et al. [21] for mapping refugee settlements in Cameroon using very high-resolution (VHR) satellite imagery data. Similarly, Lu et al. [15] developed a fully CNN (FCN) model to identify refugee tents along the Syria-Jordan border. A more advanced model for estimating displaced populations after a disaster involving a CNN and a generative adversarial network (GAN) was developed by Fisher et al. [22]. In this study, they utilized transfer learning from three large existing CNN models, ResNetV2, InceptionV3, and MobileNetV2, to create an object detection model and further utilized a GAN to enrich the existing dataset.

However, while the effectiveness of these AI approaches has improved with developments in image analysis using machine learning algorithms such as neural networks [23, 24], the problem of capturing multiple time periods

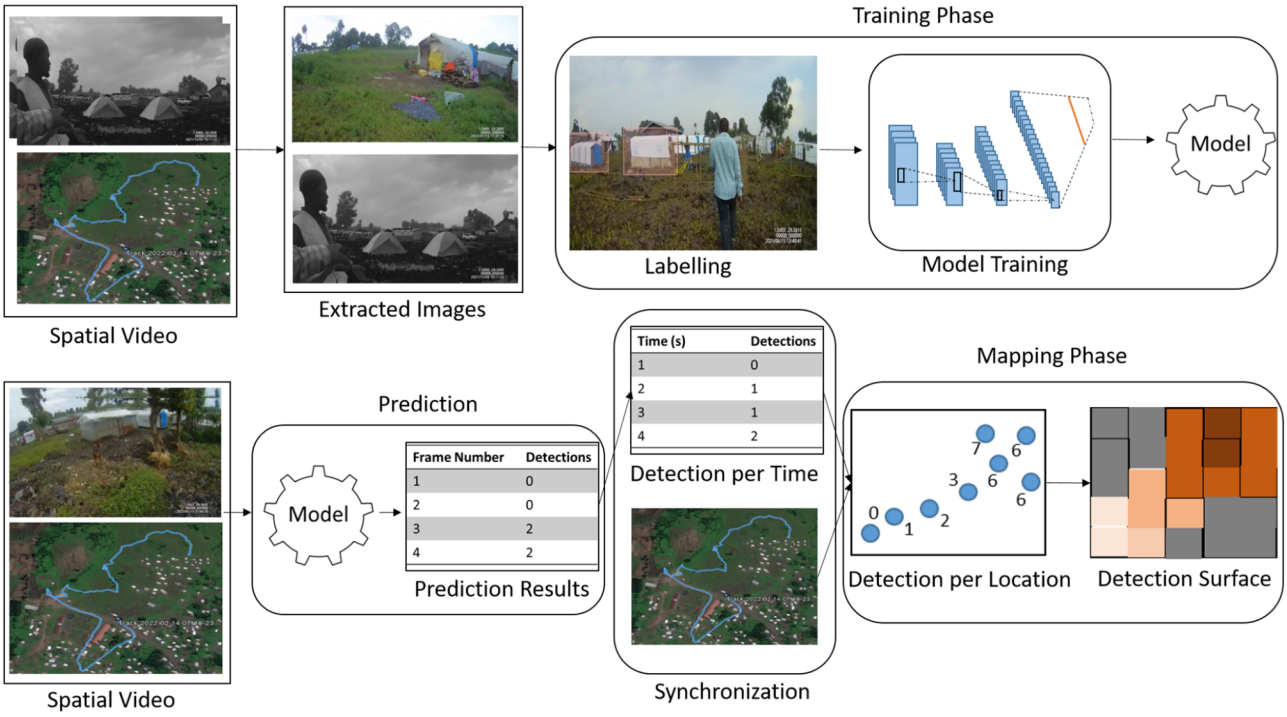

Fig. 3 Automated mapping workflow. The first task of the workflow is image extraction, followed by the training phase, which involves image labeling and model training. The prediction results from the trained model are used as the input to the mapping phase, which is then used to generate the final raster map after spatial filtering

still largely remains [25]. Spatial video (SV), which can include systematically capturing multiple time periods or collecting an immediate "snapshot" after a single event, has proven useful for responding to on-the-ground needs [26–28]. This field methodology combines a global positioning system (GPS) and video imagery, with each image frame being tagged with a location. Although the traditional version of SV still poses logistical challenges in terms of mapping, its ease of use for on-the-ground response teams and its ability to capture fine-scale spatiotemporal data, as well as its ability to add context through mixed methods [5], indicate that it is a method worthy of further development. It is a data collection strategy that conceptually puts local teams at the heart of data collection and processing. For example, the camp images shown in Fig. 1 are from SVs collected in various relief camps by local collaborative teams. This method is also appropriate for a variety of different environments and purposes, with examples of SV use in informal-type settlements, including cholera in Haiti [29], malaria in Ghana [30], environmental factors affecting dengue in Colombia [31] and Nicaragua [25], and water access in Tanzania [32]. While useful for smaller areas, SV mapping soon runs into scalability challenges when the geography expands to include numerous objects to be mapped, such as typical relief camp tents [5]. As part of an attempt to make this a more ubiquitous method that could lessen the logistical burden, the authors of this paper previously developed a neural network model to identify health risk features in SV imagery for Haiti [33]. Improving this approach could provide an efficient solution for mapping risk in informal settlement-type environments and could also be used to capture and compare geographic change over time, which could again act as a signal for potential disease outbreaks. This type of longitudinal mapping is also vital for providing ongoing support for intervention as service provision changes. As an excellent complimentary data source, granular satellite imagery may not be readily available, especially for multiple periods, for all relief camp settings. Similarly, while other spatial data collection approaches are available, such as using drones, from our experience, the technology involved and the skill needed to use this equipment limits its likely use in the camps described in this paper and for the available field team. Given these experiences, at this time, the current body camera style of the SV camera we use is ideal, small, simple to use, unobtrusive and unlikely to cause local concern, which is not the case for flying drones over a relief camp. It was tasked with new camps as they emerged with nothing more than an email exchange.

As an example of the difference between these two data sources, Fig. 2a shows the Google Earth imagery (from Maxar technologies) for a location in Goma, DRC on November 7, 2021, while Fig. 2b shows the spatial video imagery captured from the same location on November 9, 2021. The SV data revealed a change in the location and number of tents, even after just two days. In addition, the ground-level imagery provided through SV allowed for more visual context to be established, such as those that are "informally" constructed and not part of the initial relief designed to home displaced families from the volcanic eruption [5]. The objective of this paper is to develop a more sustainable approach using machine learning to automatically turn this type of SV image into an automatically generated map.

Methods

SVs were collected from a series of relief camps around Goma, DRC, some of which were set up in response to the Nyiragongo volcano eruption on 22 May 2021. The specifics of field data collection will be described in a later section, but the primary reason for the collection was to be able to map risks that could be tied to potential or actual cholera outbreaks. Typically, SV is followed by the manual digitization of features from each coordinate-enriched video frame. In this paper, we explore the possibility of automatic mapping from SV imagery. Achieving this goal requires a two-step process: model development followed by a mapping task. Model development consists of various subtasks, including frame extraction, image labeling and model training, while the mapping itself involves GPS synchronization and spatial filtering (Fig. 3).

Spatial video (SV)

SV is video imagery that has been merged with a coordinate stream from a global positioning system (GPS) receiver [34]. Typically, each video frame has a GPS coordinate connected to it, which in effect means that these media can be used as a digitizing source. For the type of camera used in this project, the Miufly body camera, the GPS coordinates are embedded in the video files through exchangeable image file format (EXIF) tags. Along with the GPS coordinates, the "media time" (the video time elapsed) is also attached to each coordinate. ExifTool [35] (an Exif parser, which can extract Exif, tags) can be used to convert the embedded GPS coordinates to a GPS Exchange Format file (GPX). In other words, the entire video path can be spatially located. Image frames from the SV are extracted for tasks such as labeling using "Frame Selector"[1], a bespoke software that was

[1] Frame Selector is a standalone bespoke web-application developed by the authors for extracting images from a video. Frame Selector utilizes Open Source Computer Vision Library (OpenCV) [36] to extract image frames from a video based on selected time. These images are further used for tasks such as labeling.

previously developed by our team for previous health risk mapping in Haiti.

Training phase

With recent advancements in CNNs, many object detection algorithms have emerged [37]. Of particular interest are R-CNN [38] and its variants [39–41], and YOLO (You Only Look Once) [42] and its variants [43, 44]. While R-CNN uses separate processes for classification and localization [38], the YOLO method [42] combines them as a unified classification and regression problem. For this study, for the object detection task (in this case, the ability to identify any tent), we used YOLOv5 [45], which has better accuracy and efficiency than older versions of the YOLO model [45].

YOLOv5 training

The training data for the YOLOv5 model consist of images and their corresponding categories, which contain the name and the normalized bounding box for each of the objects present. The normalized bounding box for an object consists of the center point, width, and height of the object normalized with respect to the dimensions of the image. To create the normalized bounding box, the width and height of the image are set to 1. Then, the center point of the object is calculated based on the ratio of the image to the object size and the location of the center point of the object with respect to the top left corner of the image. Similarly, the width and height of the object, normalized with respect to the image, can also be calculated.

To create the training dataset, we utilized Label-studio software. After all the images are labeled, the software produces two separate folders for the images as well as the labels. To gauge the performance of the algorithm with respect to the tuning of hyperparameters, a section of the data is separated into a validation set. Finally, during training, the training and validation folders are fed into the YOLOv5 algorithm, and upon completion, the YOLOv5 algorithm generates a model with updated weights, which is serialized into a hierarchical data format (HDF) file.

Mapping phase

The mapping phase begins with the model prediction task. A frame extraction process is used to separate the images from the SV, and each frame is fed to the trained model. The outputs of the model prediction task include the predicted label, the normalized centroids (x, y), the dimensions (width and height), and the confidence probability (a value between 0 and 1) for each of the detected objects. This information can further be used to include the bounding box of the detected object in the image. Since we are only classifying a single object (tent) here, the total object count for each image frame is just a single value (the total number of detections for the object). Finally, a table is generated with the frame number in sequential order and the corresponding detection totals.

GPS Synchronization

The detection objects can only be mapped after merging with the GPS data from the SV. The GPS data and the detection data cannot be directly merged due to temporal frequency variations in the GPS stream (generally 1 location per second) and the video stream (generally 30 images per second). For merging, the data are converted into detections for each second using the frame rate information for the video. For example, if the frame rate of a video is 30 frames per second (FPS), then the total object count for a second can be approximated from a chunk of 30 images. While we can use the maximum, minimum, mean, median, or sum to approximate the object count, for this study, we used the maximum, as it tends to reduce the chance of duplication[2]. The final output of this conversion process is a table with the total detection for each second (Fig. 4). This table is merged with the GPS data table (locations for each second) to generate a spatial dataset containing time (in seconds), location (from GPS data), and the total number of visible tent detections.

Spatial filtering

Although there should eventually be flexibility in mapping the identified object directly as points, for this stage of development, generating a spatial filter smoothed surface is more appropriate because it can incorporate the uncertainties associated with positional accuracy and fuzziness in the total counts, including the issue of duplication. In the spatial filtering approach, a uniform spatial grid is overlaid on top of the area of interest (Fig. 5). Then, the total count of the seen tents is calculated for each grid cell center as the total number of points (in this case, the GPS of the tent locations) falling within a circle or buffer around it. The overlaid grid can be converted to a raster (GeoTIFF) to support further spatial querying and analysis.

To maximize the utility of the automated classification and to validate the resulting patterns, a web application was developed to query the resulting spatial filter map (Fig. 6). By drawing a bounding box on top of the resulting heatmap, it is possible to identify all the intersection SVs. In this way, visible tent concentrations identified through automated mapping can easily be validated with SV images.

[2] We have used maximum as the mean and minimum underestimates the total count while sum leads to overestimation of total count due to duplication.

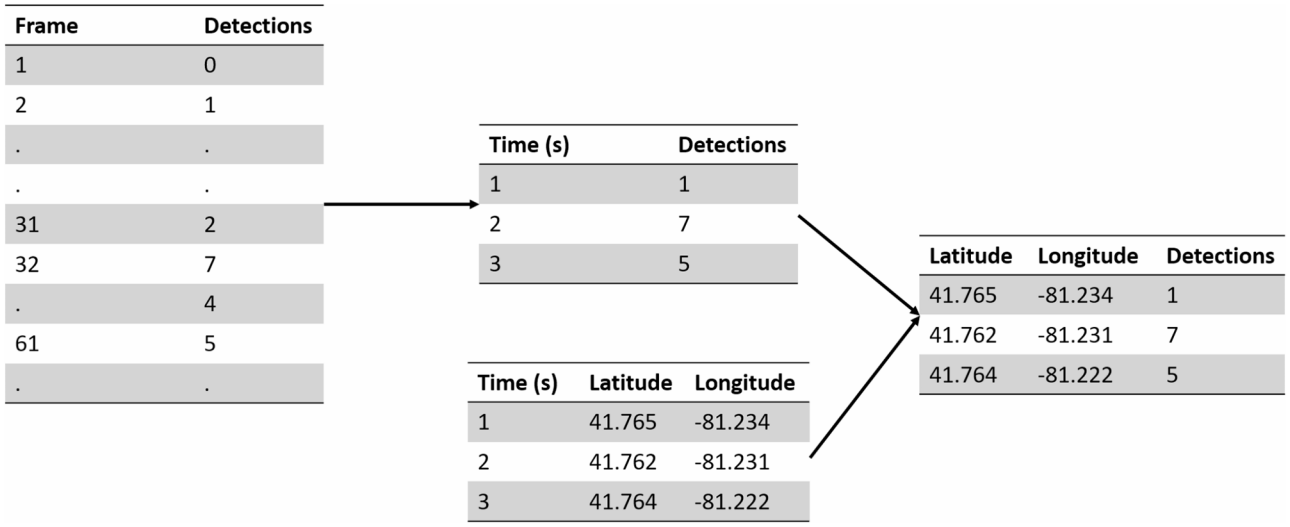

Fig. 4 Detection locations were assigned using GPS and detection data. The visible tent detections for each frame are converted into the number of detections for each second using frames per second (FPS) information of the video, which is then joined with the timestamped GPS path to generate detections per location

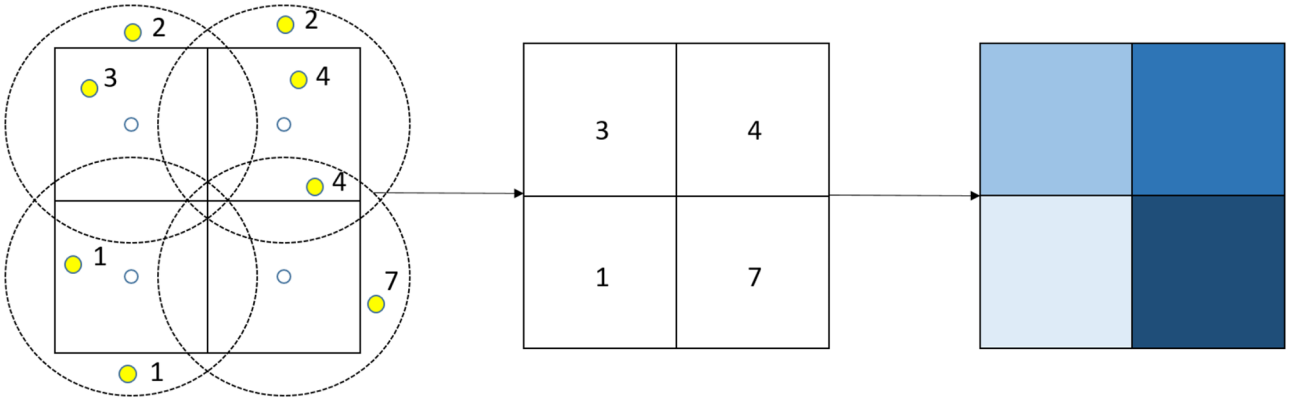

Fig. 5 Generate a detection surface using a spatial filtering approach. The yellow dots are the GPS points with detection counts, and the blue unfilled dots are the grid centers. The dotted circle represents the bandwidth of the spatial filtering approach

Data and experimental setup

After the Nyiragongo volcano eruption on 22 May 2021, a ground epidemiological team that had been working in Goma to assess the area's susceptibility to Cholera redirected time considered the impact on the Mujoga Relief Camp. Along with collecting water samples for testing Vibrio cholerae, the epidemiological team also surveyed the camp and its surroundings using SV. Once a month, for the period June 2021 to June 2022, a member from the epidemiological team walked around the camp, recording the environment with a small high-resolution Miufly body camera. A portion of these data have previously been used to manually create maps of tents, bathrooms, water points and other key features [5]. These data were used to train and test the automated mapping methodology.

Model training

To train and test the model, we utilized image fames from four videos. From 524 relevant frames containing images of tents, 420 frames (80%) were used for training, and 104 (20%) images were used for testing. The label-studio software shown in Fig. 4 was used to create the bounding box around each tent object in the video frame. The model and the corresponding code were downloaded from https://github.com/ultralytics/yolov5, which is a PyTorch-based implementation of the YOLOv5 model pretrained on Common Objects in Context (COCO). The YOLOv5 model was trained for 300 epochs with a mini-batch size of 16 and an image size of 224×224. An early-stopping regularization strategy was used to avoid overfitting. To show the spatial distribution of visible tents for a relatively large area, we utilized a set of 10 videos collected by the epidemiological team (which includes the four videos that were used for training the

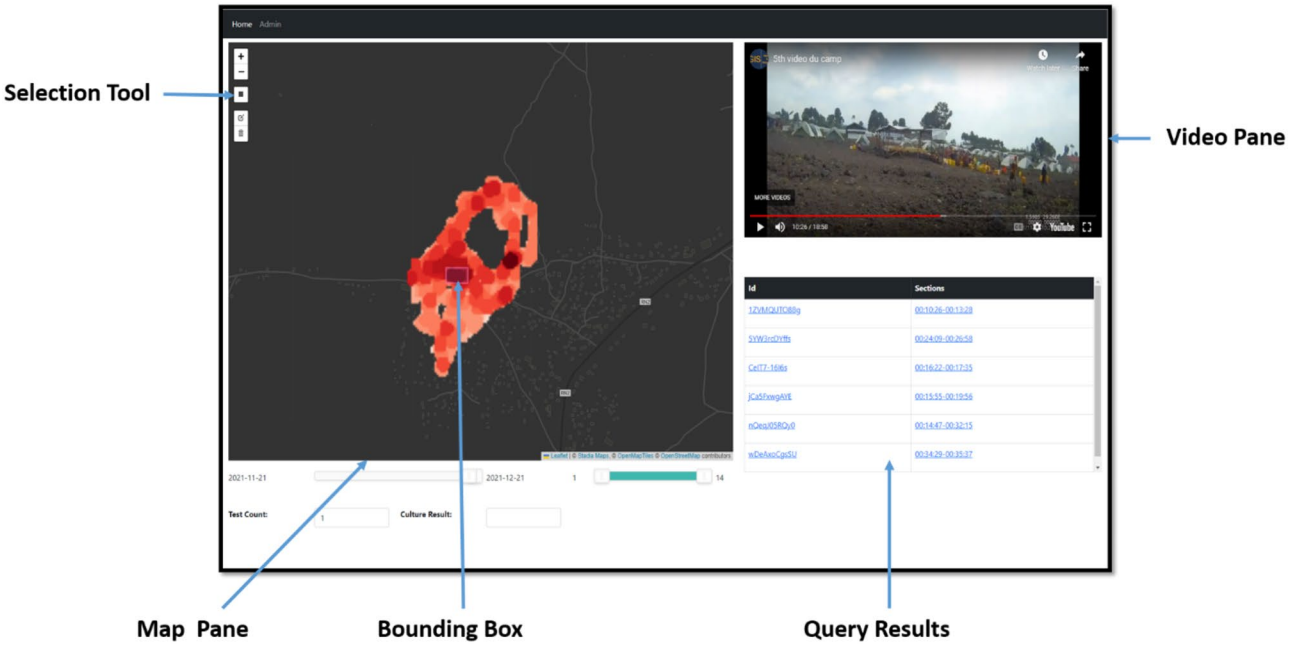

Fig. 6 Web application to query tent surfaces generated using a spatial filter. The selection tool is used to create a bounding box over the tent surface, and the application retrieves the corresponding videos with the respective segments as intervals

model). The model with trained weights was used to predict the bounding box for tents in each video frame. The prediction outputs were written to a table with the total number of tents predicted for each frame. The frame number was converted to seconds using the frame rate parameter of the video, and the max operation was used to aggregate the number of frames for each second (Fig. 4). This table was merged with the GPS data from the SV using video time. The GPS coordinates with the total seen tent information were used to generate raster maps (continuous surface) using the spatial filter operation, and the maps were compared using a raster calculator.

Longitudinal analysis of the spatial distribution of visible tents

A major reason for exploring the automatic mapping of SV is to find a more sustainable approach for capturing longitudinal change [46]. It would be helpful for response teams if the same area could be regularly resurveyed to capture highly dynamic changes such as the growth or shrinking of the camp or changes to the location or quality of key features such as water access or toilets. As a first step in that process, SVs for the months of June, July, September and November 2021 were selected. Figure 10 shows the GPS paths for the SVs. A rectangular region that intersects all the GPS paths was selected (Fig. 7). The SVs for the months of August and October were not selected because their data collection paths did not overlap with those of the four selected months. The same spatial filter method is used to generate raster surfaces for the four videos, and sample images for the region are extracted using the web application (Fig. 6). The differences in the raster values show either an increase or decrease in the number of visible tents. In this way, changes can be mapped over time for any area covered by multiple SVs. To do this, it is essential to resample the rasters and align them before applying the raster calculator so that each raster has the same extent and dimensions. Any cell value after the change calculation is assigned a Nodata (null) value if there is no associated visible tent calculation for any period (or there is no overlap among all the videos). For example, if raster A has its first three cell values of 5, 4, and nodata (null) and raster B has the values of nodata (null), 5, and 3, then the resultant raster (B-A) will have nodata (null), 1, and nodata (null).

Results

Model training results

All SV routes of varying lengths collected between November 2020 and June 2022 were used to generate the raster map of tents. The tent detection model was run across all the videos, and the resulting detection count per frame was combined with the associated coordinates to generate a detection attribute for each of the GPS track points.

To evaluate the performance of the object detection model, we used the F score, which is the harmonic mean of the precision and recall. The minimum and maximum

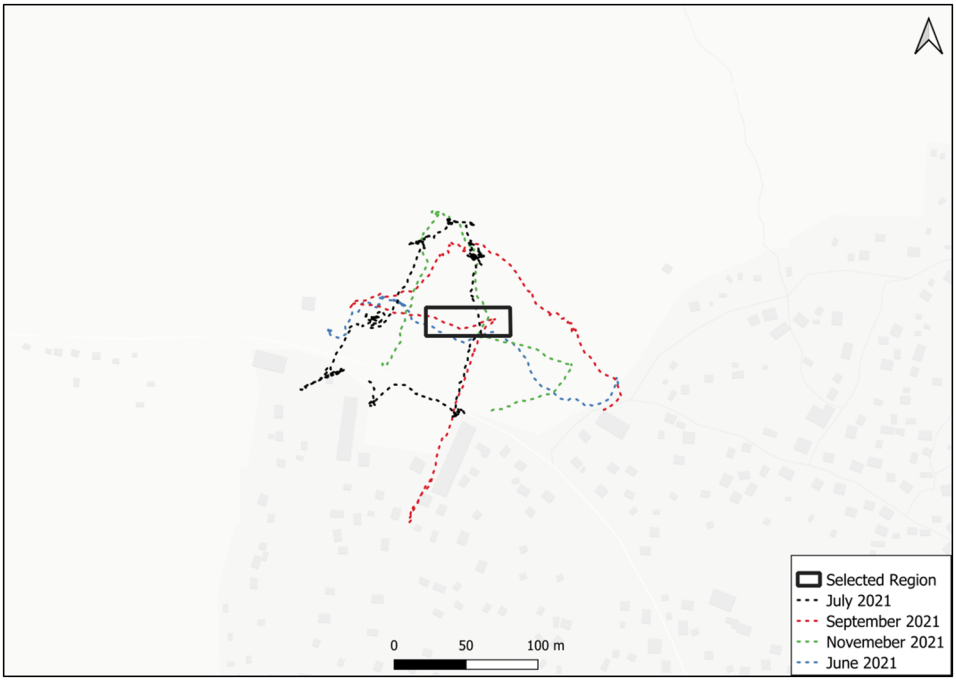

Fig. 7 GPS paths for June, July, September, and November 2021. The rectangular box represents the region selected for change analysis

Table 1 Model training results at a confidence threshold of 0.320

Metric	Value
F1-Score	0.82
Precision	0.84
Recall	0.79
mAP@0.5	0.84
mAP@0.5:0.95	0.45

values of precision, recall and F score are 0 (worst) and 1 (best), respectively. The tent detection model has a maximum F score of 0.82 (Table 1) at a confidence threshold of 0.320, while the precision and recall values are 0.84 and 0.79, respectively, at the same confidence threshold. The threshold confidence value is a combination of the class confidence threshold (the threshold value at which an object is considered to be of a particular class) and the object confidence threshold (the threshold value at which a bounding box is considered to have an object).

To determine how well the model localizes the object, the mean average precision (mAP) metric was utilized. mAP helps to determine the detection accuracy of the model (the real position of the object in the image) by considering the intersection over union (IoU) criterion, which is the overlap between the bounding boxes of the real and detected object. The IoU threshold was set at 0.5 and above for the validation runs. For our model, the mAP at the 0.5 threshold was 84% (Table 1). To further evaluate the model for object localization, we also calculated the mAP at various steps of the IoU from 0.5 to 0.95 (in steps of 0.05), and the average mAP was 45%.

Generating raster maps for tents

To illustrate SV as a source for automatic mapping, spatial filtering (Fig. 5) was used to generate a raster map of visible tent intensity. For this study, a raster cell size of 5 m and a filter radius of 10 m were chosen.

Spatial distribution of visible tents

Figure 8 shows a section of the surface visualized as a heatmap along with associated SV imagery from a few sample locations. The pixels with a higher intensity represent more detections. There were also a few instances of false positive classifications, as shown in Fig. 9.

To assess how similar the visualized heatmap is to the more standardized cartographic approach of digitizing features from the imagery, Fig. 10 displays the output raster maps for the first two time periods (June and July 2021) with manually digitized tent locations overlaid on top.

Longitudinal analysis

The months of June, July, September and November 2021 were chosen for change analysis. The same spatial filter method was used to generate raster surfaces for the four videos, which were further visualized using the web application (Fig. 6). While this is useful as an initial comparative explorative analysis to visually gauge the change in the distribution of visible tents, our conceptual

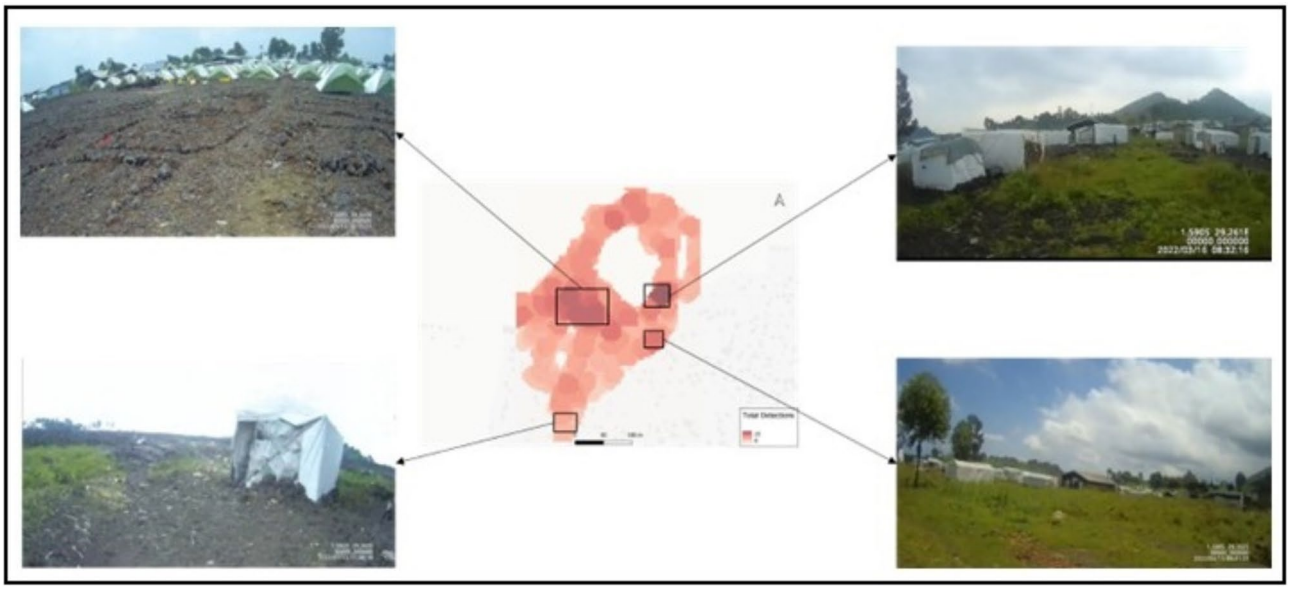

Fig. 8 High-intensity sample locations from the tent surface generated using a spatial filter

Fig. 9 False positives for tent detection. When the lean-too covered with fabric is falsely detected as a tent, it is easy to see why the mistake was made, as it has several tent-like properties

goal is to create an automatic means of mapping temporal change, which, in this case, using these data, is best exemplified using a raster calculator.

The spatial distribution of the visible tents along the GPS path and the selected region is shown in Fig. 11. Along with the maps, sample images from the locations are also provided as visual examples of change. The map classification breaks are kept the same across all the maps for comparability. A greater intensity in color indicates an increase in the number of visible tents. From September

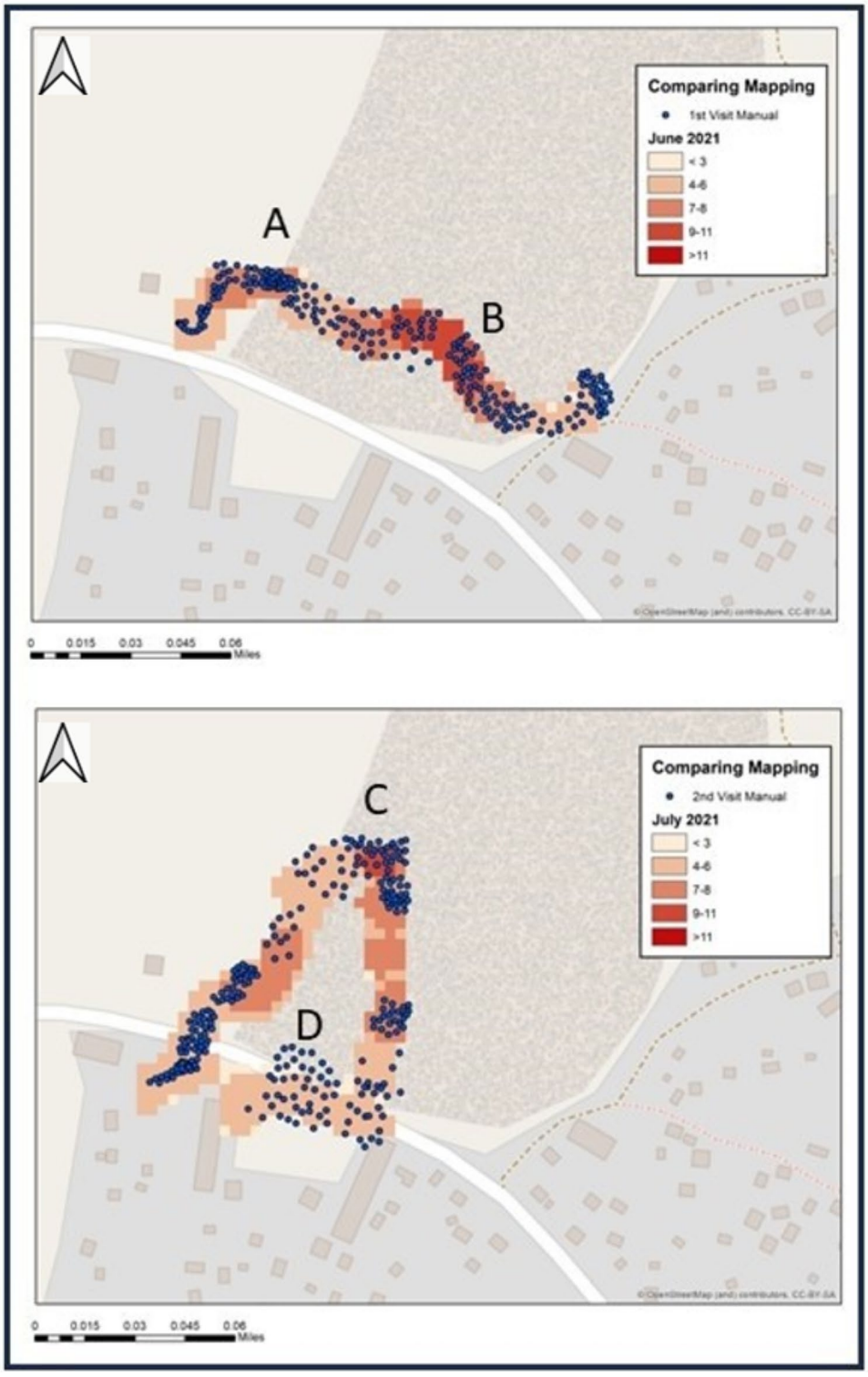

Fig. 10 Comparison of model outputs for June and July 2021 against manually digitized shelter locations. In general, the locations and intensities are similar between the maps, especially in map area **A**, although there is lower intensity in the output for sections of map area **B**, which is explained by the difference between mapping actual tent locations and creating a smoothed density of visible tents. In the lower map, the high intensity of map area **C** is captured in both outputs, although there is again some variation in the area immediately to the south. Map area **D** shows one of the few locations where the manual mapping covers areas not predicted by the model, with one explanation being that there can be more visible extrapolation from the manual mappers if tents are seen to extend beyond what might be recognizable as a tent shape. This area also contains other wooden shelters, and some of these differences are expected

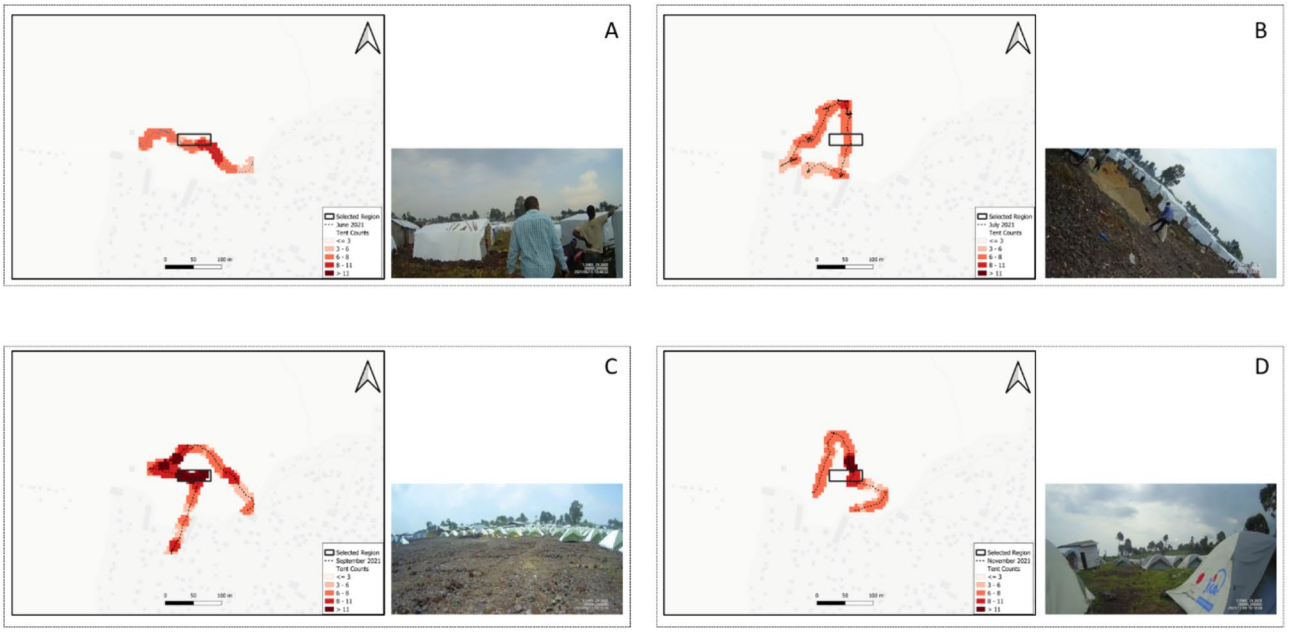

Fig. 11 Longitudinal analysis of videos from June 2021 to November 2021 for a selected area of the camp falling inside the black rectangle. The maps show the spatial distribution of visible tents for **a**) June **b**) July **c**) September and d) November 2021. The images adjacent to the map are sample frames from the spatial video for the same location

Fig. 12 Changes in visible tents for the period between June and September 2021. A divergent color scheme from red to green is used, where red indicates a decrease and green indicates an increase in the number of tents. This map clearly shows that the camp had grown in these areas between the two time periods

2021 (Fig. 11C) onward (and November 2021 (Fig. 11D)), there was a considerable increase in the number of visible tents compared to that in June (Fig. 11A) and July (Fig. 11B) 2021.

The difference raster maps (Fig. 12) for September (Fig. 11C) and June 2021 (Fig. 11A) display an apparent increase in the number of visible tents for the selected region. The negative and positive values indicate a

decrease and increase in the number of visible tents, respectively.

Discussion

Relief camps are designed to house a displaced population after a traumatic event. The size, location, quality, longevity, safety, security and health challenges of these camps can vary tremendously. Indeed, we have previously found that in the DRC, the organizational structure and internal longitudinal changes in services vary from camp to camp. Providing sustainable fine-scale mapping of the type useful for response teams is vital. While overhead remotely sensed data are certainly beneficial, there are some deficiencies with that mode that ground-level surveying can be used to address, including addressing local context. Spatial video provides a potentially sustainable option for local collaborators. Building on previous work in Haiti [33], in this paper, we have advanced a key aspect of ground-level imagery sustainability by automatically mapping key camp features through a combination of machine learning and spatial analysis. This type of detection and mapping pipeline demonstrates the potential of deep learning and mobile data collection to efficiently capture an ever-evolving environment in near real time.

While the detection model yielded encouraging results in terms of classification (F1-score of 82%) and object localization (mAP of 84%) (Table 1), it should also be remembered that these data were collected as part of a field team's primary function to monitor cholera. In this regard, it mirrors the method's utility for other settings, as it is likely that local resources have to be repositioned to collect spatial data. The relatively high localization score is probably because tents are relatively easy to localize in images [33]. To explore the potential of machine learning as part of an ongoing surveillance tool, it is vital to consider how to improve data collection approaches and whether this method produces better data. For example, while comparison with manual digitization is possible using the same SV data source, a fair comparison between automated mapping using satellite and SV imagery is only possible when the images are available for the same location across a similar timeframe.

The data collected for this paper were not the result of a scientific effort designed to investigate the stated mapping purpose, which means that there are some inherent challenges and limitations. One limitation is the inability to accurately map out individual features. Mapping out specific features based on footprints is challenging because SV cameras do not capture overhead imagery, unlike other sources such as satellites or drones. The spatial filter approach applied in this paper should be considered as a first step to extract meaning from ground-level imagery in terms of visible change. However, with additional data collection parameters applied, a more useful application would be to capture the change in the context of those features. For example, different types of tents (official or informal) or changing-quality tents (organized to disorganized) may exist.

A further challenge is that SV cameras generate only a single GPS location for a single interval compared to satellites, which can generate locations for the bounding box of the viewport. As there is inherent uncertainty regarding the distance from the GPS location to the detected object, which is heavily dependent on the focal length of the camera and the vantage point, the exact position of the detected features can be difficult to determine. Simply put, even if two cameras are at the exact same location (with the same latitude and longitude), the camera-viewing angle and vantage point determine the number of objects detected at the location. To explore this further, we plan to incorporate detailed metadata from the camera, such as pitch, yaw and roll information, to reduce positional uncertainty. While we have used the grid cell size and filter radius as 5 and 10 m respectively, it will also be interesting to identify the ideal filter size based on details such as scope of the camera and GPS accuracy. Ideally, these technical improvements could occur simultaneously while serving the immediate humanitarian need. Even now, data collection suggestions such as imposing a reasonable systematic, replicable frame by re-walking the exact same routes and angling cameras toward certain features at key points on the path are frequently realized to field teams. Further work is also needed on how to determine the exact count of tents (features) as identical frames for the same location, which makes deduplication of features more difficult. We have tried to bypass this issue using the maximum count approach for a fixed bandwidth, but again, this approach might not be accurate due to the challenges associated with the viewing range of the camera. Again, from a technical perspective, variability in camera zoom and viewing angle can also play a major role in counting/undercounting tents. For example, even though the location of the camera is similar in both Fig. 11C and D, the increased magnification of the camera in Fig. 11D might lead to an undercounting of tents. The occlusion of objects will also reduce the accuracy of the object detection algorithm; currently, we annotate objects that are only partially visible. Another issue that has not been addressed here is the correlation of image frames within a video, which could be problematic while generating test/train splits and could inflate the accuracy. While we used image frames from separate videos for test/train splitting, the frames could be similar based only on the similarity of the location. It would be interesting to look at how such nuances affect the overall performance of the model. While we have used the YOLOv5 model here,

there are other detection models such as the RCNN and its variants, which are known for its accuracy while sacrificing detection speed. Our experiments with two other models, YOLOv8, and Faster R-CNN indicated no drastic improvement in detection accuracy when compared to the YOLOv5 model. However, it will be interesting to see the variations in model accuracy when the object detection model is applied to video data captured from other environments.

It is also likely that the technology will change quickly. Future video devices might provide hand-carried 360-degree coverage at an affordable price. Video imagery from GPS-enabled drones could be used for generating automated maps; however, from our experience in the DRC, the logistics of getting a drone to the camp and having the skill to fly one were too difficult.

It is vital to stress that SVs are not designed to replace satellite imagery. There are obvious situations where the combination of imagery and a skillset would make remote sensing the best means to capture relief camp morphology. However, while these data and skillsets may be available for large, well-documented situations, as we have seen in the DRC, multiple relief camps have arisen for different reasons. In our experience, there is no readily available local skillset to capture change. Even if available, SV imagery still provides a local context that overhead imagery will miss. Therefore, SV should complement traditional remote sensing approaches, which still provide excellent coverage for land-cover change, vegetation, elevation, slope, etc., and in some cases constitute a local only option for data collection. Future research exploring this combination of data sources might consider how elevation data could be combined with SVG [47–49] data to better understand how to combine elevation, runoff, tent quality and associated disease risk. Another application might be to combine the model output with flood zone maps to identify where new tents are more vulnerable to floods. In these instances, SV can be employed by local teams to help fill in gaps whenever needed, providing vital temporal granularity along with on-the-ground images for subsequent contextual investigations.

While the driving factor for the work presented in this paper was identifying the conditions that might lead to cholera outbreaks in the camps, other risks can also be addressed through the collection of these data, with one example being the fire that claimed several lives in the Goma, DRC region, during August 2023 [2]. Although this fire-impacted camp was not included as one of our surveyed areas, if it had been, or if a similar fire had occurred in one of our studied areas, it would have been possible to map the immediate and subsequent impact on the morphology of the camp [7]. Questions could be asked such as, how were existing services now being utilized, and had the fire led to an additional surge in more "informal" tents? Overlaying health data, such as cholera case locations, or guiding preemptive water sampling, just as the team has been doing in both Haiti and the DRC, would then benefit from understanding these types of changes [29, 50]. As previously mentioned, our manual mapping of one camp predicted the likelihood of increased cholera susceptibility. Eventually, we hope to extend that same predictive capacity to all camps in the region, but to do so requires the type of automation described here.

The ethical dimensions of this form of data collection must also be mentioned. At the most basic level, no IRB should be needed, as these environmental surveys do not include interviews. Precautions regarding capturing and sharing images of faces can be made for publications and presentations. As data tend to be collected by local response teams, there is less concern regarding the Western trampling of local attitudes and customs, especially as data collection occurs in tandem with a (usually) perceived task of importance such as cholera prevention. The small size of the camera results in relatively unobtrusive data collection, which again limits on-the-ground concern and is a further reason why more sophisticated methods such as the use of 360-degree cameras and drones may not always be suitable. There could also be an argument made regarding not "negatively glamorizing" or perpetuating myths of the conditions found in these types of camps. However, the primary task of the data collector is to stop disease outbreaks and a loss of life. From experience, when questions are raised about why these data are being collected, local attitudes tend to be supportive rather than combative.

Conclusions

Tracking and mapping of temporary camps or shelters as a proxy for internally displaced population is imperative for monitoring, mitigating and guiding humanitarian responses to conflict, human rights violations and man-made or natural disasters. While video based surveying strategies such as SVs offer an exciting option for mapping such highly dynamic environments, spatially tagging and mapping numerous camps and shelters is manually infeasible. In this paper, we have addressed this challenge by developing a machine-learning model to automatically detect camps and generate raster surfaces of the detection for mapping. We have also shown how the generated surfaces can be used to capture changes in the distribution of tents across time, which is quintessential for tracking internally displaced population.

Acknowledgements

The authors would like to thank David Kaanda Kamundu, Diane Furaha Kanyere, Eduige Mutsoro, Élodie Siku, and Mpigirwa Mulolo for their help with data collection. The authors would also like to thank HEAL Africa.

Author contributions
JA designed and tested the deep learning algorithm as well as the spatial filtering method. AC supervised the spatial aspects of the project, and SB, NK, and GA provided SV data manipulation and coding support. FM supervised field data collection and AA and GM provided epidemiological guidance and field data support. GM supervised the overall project. JA and AC wrote the initial versions of the manuscript. All authors read and approved the final manuscript.

Funding
This research was funded by the National Institutes of Health RO1 AI138554.

Data availability
The dataset and the code supporting the conclusions of this article is available in the Automated Mapping Using Spatial Video and Deep Learning repository (https://figshare.com/s/32bd8a1883f5ea6d5134).

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 11 April 2024 / Accepted: 22 October 2024
Published online: 05 November 2024

References

1. Farmer B, Townsley S. Inside the besieged city of Goma, where food is fast running out – and sexual violence is rife. The Telegraph [Internet]. 2023 Jun 17 [cited 2023 Aug 7]; https://www.telegraph.co.uk/global-health/women-and-girls/drc-conflict-congo-refugee-camps-sexual-violence-goma-m23/
2. Kyala C, Prentice A, Williams A. Seven children killed in fire at Congolese camp for displaced flood victims. Reuters [Internet]. 2023 Aug 19 [cited 2023 Aug 23]; https://www.reuters.com/world/africa/seven-children-killed-fire-congolese-camp-displaced-flood-victims-2023-08-19/
3. Curtis A, Mills JW. <ArticleTitle Language="En">Spatial video data collection in a post-disaster landscape: the Tuscaloosa Tornado of April 27th 2011. Appl Geogr. 2012;32(2):393–400.
4. Curtis A, Fagan WF. Capturing Damage Assessment with a Spatial Video: An Example of a Building and Street-Scale Analysis of Tornado-Related Mortality in Joplin, Missouri, 2011. Ann Assoc Am Geogr. 2013;103(6):1522–38.
5. Curtis AJ, Maisha F, Ajayakumar J, Bempah S, Ali A, Morris JG. The Use of Spatial Video to Map Dynamic and Challenging Environments: A Case Study of Cholera Risk in the Mujoga Relief Camp, D.R.C. Trop Med Infect Disease. 2022;7(10):257.
6. Bjorgo E. Refugee Camp Mapping Using Very High Spatial Resolution Satellite Sensor Images. Geocarto Int. 2000;15(2):79–88.
7. Hassan MM, Hasan I, Southworth J, Loboda T. Mapping fire-impacted refugee camps using the integration of field data and remote sensing approaches. Int J Appl Earth Obs Geoinf. 2022;115:103120.
8. Fan R, Li J, Song W, Han W, Yan J, Wang L. Urban informal settlements classification via a transformer-based spatial-temporal fusion network using multimodal remote sensing and time-series human activity data. Int J Appl Earth Obs Geoinf. 2022;111:102831.
9. Tarnas MC, Ching C, Lamb JB, Parker DM, Zaman MH. Analyzing Health of Forcibly Displaced Communities through an Integrated Ecological Lens. Am J Trop Med Hyg. 2023;108(3):465–9.
10. Gram-Hansen BJ, Helber P, Varatharajan I, Azam F, Coca-Castro A, Kopackova V et al. Mapping Informal Settlements in Developing Countries using Machine Learning and Low Resolution Multi-spectral Data. In: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society [Internet]. New York, NY, USA: Association for Computing Machinery; 2019 [cited 2023 Jul 24]. pp. 361–8. (AIES '19). https://doi.org/10.1145/3306618.3314253
11. Deville P, Linard C, Martin S, Gilbert M, Stevens FR, Gaughan AE et al. Dynamic population mapping using mobile phone data. Proceedings of the National Academy of Sciences. 2014;111(45):15888–93.
12. Azar D, Engstrom R, Graesser J, Comenetz J. Generation of fine-scale population layers using multi-resolution satellite imagery and geospatial data. Remote Sens Environ. 2013;130:219–32.
13. Yuan J, Roy Chowdhury PK, McKee J, Yang HL, Weaver J, Bhaduri B. Exploiting deep learning and volunteered geographic information for mapping buildings in Kano, Nigeria. Sci Data. 2018;5(1):180217.
14. Wang S, So E, Smith P. Detecting tents to estimate the displaced populations for post-disaster relief using high resolution satellite imagery. Int J Appl Earth Obs Geoinf. 2015;36:87–93.
15. Lu Y, Koperski K, Kwan C, Li J. Deep Learning for Effective Refugee Tent Extraction Near Syria–Jordan Border. IEEE Geosci Remote Sens Lett. 2021;18(8):1342–6.
16. Jochem WC, Bird TJ, Tatem AJ. Identifying residential neighbourhood types from settlement points in a machine learning approach. Comput Environ Urban Syst. 2018;69:104.
17. Quinn JA, Nyhan MM, Navarro C, Coluccia D, Bromley L, Luengo-Oroz M. Humanitarian applications of machine learning with remote-sensing data: review and case study in refugee settlement mapping. Philosophical Trans Royal Soc A: Math Phys Eng Sci. 2018;376(2128):20170363.
18. Tingzon I, Orden A, Go KT, Sy S, Sekara V, Weber I, MAPPING POVERTY IN THE PHILIPPINES USING MACHINE LEARNING, SATELLITE IMAGERY, AND CROWD-SOURCED GEOSPATIAL INFORMATION. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. 2019;XLII-4-W19:425–31.
19. Mboga N, Persello C, Bergado JR, Stein A. Detection of Informal Settlements from VHR Images Using Convolutional Neural Networks. Remote Sens. 2017;9(11):1106.
20. Qiu C, Schmitt M, Geiß C, Chen THK, Zhu XX. A framework for large-scale mapping of human settlement extent from Sentinel-2 images via fully convolutional neural networks. ISPRS J Photogrammetry Remote Sens. 2020;163:152–70.
21. Gella GW, Wendt L, Lang S, Tiede D, Hofer B, Gao Y, et al. Mapping of Dwellings in IDP/Refugee Settlements from Very High-Resolution Satellite Imagery Using a Mask Region-Based Convolutional Neural Network. Remote Sens. 2022;14(3):689.
22. Fisher T, Gibson H, Liu Y, Abdar M, Posa M, Salimi-Khorshidi G, et al. Uncertainty-Aware Interpretable Deep Learning for Slum Mapping and Monitoring. Remote Sens. 2022;14(13):3072.
23. Mahmon NA, Ya'acob N. A review on classification of satellite image using Artificial Neural Network (ANN). In: 2014 IEEE 5th Control and System Graduate Research Colloquium. 2014. pp. 153–7.
24. Pan Z, Xu J, Guo Y, Hu Y, Wang G. Deep Learning Segmentation and Classification for Urban Village Using a Worldview Satellite Image Based on U-Net. Remote Sens. 2020;12(10):1574.
25. Curtis A, Quinn M, Obenauer J, Renk BM. Supporting local health decision making with spatial video: Dengue, Chikungunya and Zika risks in a data poor, informal community in Nicaragua. Appl Geogr. 2017;87:197–206.
26. Curtis A, Mills JW, Kennedy B, Fotheringham S, McCarthy T. Understanding the Geography of Post-Traumatic Stress: An Academic Justification for Using a Spatial Video Acquisition System in the Response to Hurricane Katrina. J Contingencies Crisis Manag. 2007;15(4):208–19.
27. Mills JW, Curtis A, Kennedy B, Kennedy SW, Edwards JD. Geospatial video for field data collection. Appl Geogr. 2010;30(4):533–47.
28. Lewis P, Fotheringham S, Winstanley A. Spatial video and GIS. Int J Geogr Inf Sci. 2011;25(5):697–716.
29. Curtis A, Squires R, Rouzier V, Pape JW, Ajayakumar J, Bempah S, et al. Micro-Space Complexity and Context in the Space-Time Variation in Enteric Disease Risk for Three Informal Settlements of Port au Prince, Haiti. Int J Environ Res Public Health. 2019;16(5):807.
30. Bempah S, Curtis A, Awandare G, Ajayakumar J. Appreciating the complexity of localized malaria risk in Ghana: Spatial data challenges and solutions. Health Place. 2020;64:102382.
31. Krystosik AR, Curtis A, Buritica P, Ajayakumar J, Squires R, Dávalos D, et al. Community context and sub-neighborhood scale detail to explain dengue, chikungunya and Zika patterns in Cali, Colombia. PLoS ONE. 2017;12(8):e0181208.
32. Smiley SL, Curtis A, Kiwango JP. Using Spatial Video to Analyze and Map the Water-Fetching Path in Challenging Environments: A Case Study of Dar es Salaam, Tanzania. Trop Med Infect Disease. 2017;2(2):8.

33. Ajayakumar J, Curtis AJ, Rouzier V, Pape JW, Bempah S, Alam MT, et al. Exploring convolutional neural networks and spatial video for on-the-ground mapping in informal settlements. Int J Health Geogr. 2021;20(1):5.
34. Curtis A, Bempah S, Ajayakumar J, Mofleh D, Odhiambo L. Spatial Video Health Risk Mapping in Informal Settlements: Correcting GPS Error. Int J Environ Res Public Health. 2019;16(1):33.
35. Harvey P, Körtner G. ExifTool. Kingston, Ontario, Canada) Available at http://owl phy queensu ca/∼ phil/exiftool/[Verified 29 November 2018]. 2016.
36. Bradski G, Kaehler A. others. OpenCV. Dr Dobb's journal of software tools. 2000;3(2).
37. Wang K, Liew JH, Zou Y, Zhou D, Feng J, PANet. Few-Shot Image Semantic Segmentation With Prototype Alignment. In 2019 [cited 2023 Jul 25]. pp. 9197–206. https://openaccess.thecvf.com/content_ICCV_2019/html/Wang_PANet_Few-Shot_Image_Semantic_Segmentation_With_Prototype_Alignment_ICCV_2019_paper.html
38. Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2014. pp. 580–7.
39. Girshick R, Fast R-CNN. In 2015 [cited 2023 Jul 25]. pp. 1440–8. https://openaccess.thecvf.com/content_iccv_2015/html/Girshick_Fast_R-CNN_ICCV_2015_paper.html
40. Ren S, He K, Girshick R, Sun J, Faster R-CNN. Towards Real-Time Object Detection with Region Proposal Networks. In: Advances in Neural Information Processing Systems [Internet]. Curran Associates, Inc.; 2015 [cited 2023 Jul 25]. https://proceedings.neurips.cc/paper_files/paper/2015/hash/14bfa6bb14875e45bba028a21ed38046-Abstract.html
41. He K, Gkioxari G, Dollar P, Girshick R, Mask R-CNN. In 2017 [cited 2023 Jul 25]. pp. 2961–9. https://openaccess.thecvf.com/content_iccv_2017/html/He_Mask_R-CNN_ICCV_2017_paper.html
42. Redmon J, Divvala S, Girshick R, Farhadi A. You Only Look Once: Unified, Real-Time Object Detection. In 2016 [cited 2023 Jul 25]. pp. 779–88. https://www.cv-foundation.org/openaccess/content_cvpr_2016/html/Redmon_You_Only_Look_CVPR_2016_paper.html
43. Redmon J, Farhadi A. YOLO9000: better, faster, stronger. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. pp. 7263–71.
44. Redmon J, Farhadi A. YOLOv3: An Incremental Improvement [Internet]. arXiv; 2018 [cited 2023 Jul 25]. http://arxiv.org/abs/1804.02767
45. Jocher G, Chaurasia A, Stoken A, Borovec J, Kwon Y, Michael K et al. ultralytics/yolov5: v7. 0-yolov5 sota realtime instance segmentation. Zenodo. 2022.
46. Curtis A, Blackburn JK, Widmer JM, Morris JG Jr. A ubiquitous method for street scale spatial data collection and analysis in challenging urban environments: mapping health risks using spatial video in Haiti. Int J Health Geogr. 2013;12(1):21.
47. Curtis A, Curtis JW, Shook E, Smith S, Jefferis E, Porter L, et al. Spatial video geonarratives and health: case studies in post-disaster recovery, crime, mosquito control and tuberculosis in the homeless. Int J Health Geogr. 2015;14(1):22.
48. Curtis A, Curtis JW, Porter LC, Jefferis E, Shook E. Context and Spatial Nuance Inside a Neighborhood's Drug Hotspot: Implications for the Crime–Health Nexus. Annals Am Association Geographers. 2016;106(4):819–36.
49. Curtis A, Felix C, Mitchell S, Ajayakumar J, Kerndt PR. Contextualizing Overdoses in Los Angeles's Skid Row between 2014 and 2016 by Leveraging the Spatial Knowledge of the Marginalized as a Resource. Annals Am Association Geographers. 2018;108(6):1521–36.
50. Curtis A, Blackburn JK, Smiley SL, Yen M, Camilli A, Alam MT, et al. Mapping to Support Fine Scale Epidemiological Cholera Investigations: A Case Study of Spatial Video in Haiti. Int J Environ Res Public Health. 2016;13(2):187.

Publisher's note