GroundTruth: Augmenting Expert Image Geolocation with Crowdsourcing and Shared Representations

SUKRIT VENKATAGIRI, Virginia Tech, USA
JACOB THEBAULT-SPIEKER, Virginia Tech, USA
RACHEL KOHLER, Virginia Tech, USA and BNSF Railway Company, USA
JOHN PURVIANCE, Virginia Tech, USA and Bloomberg L.P., USA
RIFAT SABBIR MANSUR, Virginia Tech, USA
KURT LUTHER, Virginia Tech, USA

Expert investigators bring advanced skills and deep experience to analyze visual evidence, but they face limits on their time and attention. In contrast, crowds of novices can be highly scalable and parallelizable, but lack expertise. In this paper, we introduce the concept of shared representations for crowd–augmented expert work, focusing on the complex sensemaking task of image geolocation performed by professional journalists and human rights investigators. We built GroundTruth, an online system that uses three shared representations—a diagram, grid, and heatmap—to allow experts to work with crowds in real time to geolocate images. Our mixed-methods evaluation with 11 experts and 567 crowd workers found that GroundTruth helped experts geolocate images, and revealed challenges and success strategies for expert–crowd interaction. We also discuss designing shared representations for visual search, sensemaking, and beyond.

CCS Concepts: • **Human-centered computing → Collaborative and social computing systems and tools**; **Collaborative and social computing design and evaluation methods**;

Keywords: crowdsourcing; crowd-augmented expert work; geolocation; investigation; verification; journalism; misinformation; shared representations; real-time crowdsourcing; expert; crowd; sensemaking; visual search

ACM Reference Format:
Sukrit Venkatagiri, Jacob Thebault-Spieker, Rachel Kohler, John Purviance, Rifat Sabbir Mansur, and Kurt Luther. 2019. GroundTruth: Augmenting Expert Image Geolocation with Crowdsourcing and Shared Representations. In *Proceedings of the ACM on Human-Computer Interaction*, Vol. 3, CSCW, Article 107 (November 2019). ACM, New York, NY. 30 pages. https://doi.org/10.1145/3359209

1 INTRODUCTION

High-stakes settings like investigative journalism and human-rights advocacy increasingly leverage photo and video evidence from social media in their investigations [15, 32, 75]. For instance, in 2017, Europol launched the Stop Child Abuse: Trace An Object campaign [5] that relies on volunteer crowds to help identify the origin of objects in the backgrounds of imagery (photos and videos) involving child abuse. Similarly, Bellingcat, an online open-source investigative community, uses

Authors' addresses: Sukrit Venkatagiri[1]: sukrit@vt.edu; Jacob Thebault-Spieker[2]: jthebaultspieker@vt.edu; Rachel Kohler[3]: rkohler1@vt.edu; John Purviance[4]: jpurviance@vt.edu; Rifat Sabbir Mansur[2]: rifatsm@vt.edu; Kurt Luther[1]: kluther@vt.edu. [1–4] Department of Computer Science and Center for Human-Computer Interaction, Virginia Tech — [1] Arlington, VA, USA and [2] Blacksburg, VA, USA; [3] BNSF Railway Company, Fort Worth, TX, USA; [4] Bloomberg L.P, New York, NY, USA.

2573-0142/2019/11-ART107 $15.00
https://doi.org/10.1145/3359209

social media imagery to investigate the credibility of claims made by and about governmental or terrorist activity [36].

Uncertainty in image provenance raises questions about whether the imagery has been altered or is being reused. This uncertainty affects the credibility of the imagery and subsequently harms its efficacy as a form of evidence, whether in the court of law or of public opinion. If these images are to be used as evidence, verification is critical [32, 75].

A key step in this verification process is *image geolocation*, a complex sensemaking process that involves identifying the exact location where photo or video imagery was taken [36, 46]. If an expert investigator succeeds in geolocating an image, then it can reliably be used to make claims about an event that happened at a particular place and time. Initially, an expert inspects the image for clues, such as familiar landmarks, license plates, street signs, etc., to narrow their search. When these clues are not conclusive, they use inference and experience to conduct a manual, brute-force search through large swathes of satellite imagery, looking for the location depicted in the image [36]. This task may take hours to days, and may not prove fruitful. It also does not scale easily [7, 19], meaning that successful geolocation is limited by experts' time and attention. Computer vision attempts at automating this process [33, 85, 87] are insufficiently accurate, placing photos within within 200km of the correct location less than 30% of the time. Further, they have constraints that may not generalize well for many real-world contexts [14].

An alternative approach that has seen success is leveraging the powerful and adaptive capabilities of geographically distributed online crowds [1–5, 78]. However, crowds often lack expertise in knowing where to look or how to assess relevance, which can lead to false positive rates as high as 64.1% [46]. Undirected crowds can also lead to vigilantism [83, 93] and misidentification [63]. The question then arises, can crowds effectively augment expert work practice to geolocate images?

In this paper, we propose an approach that combines experts' deep domain knowledge and experience with the speed and scale of crowds. To enable this approach, we extend Heer's idea of *shared representations* between humans and intelligent agents [35], and use it to facilitate crowd-supported expert image geolocation. Heer describes shared representations as a common language through which both humans and intelligent agents can work in tandem to achieve a shared objective, balancing the complementary strengths and weaknesses of each. This approach is relevant to crowd-supported expert work practice because it augments but does not replace experts, while still promoting efficiency and correctness, and it requires "neither perfect accuracy nor exhaustive modeling of the user's tasks to be useful" [35].

We explore this approach through GroundTruth, a system we developed to help experts geolocate images with a crowd. GroundTruth consists of three shared representations as system components: (1) an expert-created *aerial diagram* to help share context with the crowd, focus their attention, and overcome their spatial reasoning limitations; (2) a *gridded map overlay* specified by experts that generates microtasks for crowd workers, indicating where they should search, while providing the expert an overview of crowd progress; and (3) a *heatmap* displaying expert and crowd decisions which quickly and at-scale indicates to the expert where their own time and attention is best spent.

We conducted a mixed-methods evaluation of GroundTruth involving a think-aloud protocol, log analysis, and semi-structured interviews with 11 experts working with 567 crowd workers. We find that GroundTruth effectively merges the benefits of both expertise and crowdsourcing, demonstrating the feasibility of crowd-supported expert image geolocation using shared representations. Experts worked with crowds in real-time to narrow the search area substantially, and frequently succeeded in geolocating the image. Experts were also excited by the idea of incorporating GroundTruth into their toolset since it provides features that are not currently available in other tools. Finally, we reflect on challenges and successes in designing shared representations highlighted through our evaluation.

In summary, our work makes four contributions:

(1) Our paper makes a technical contribution by introducing *shared representations* in crowd-supported visual search, allowing visual traits and context to be easily communicated between experts and novice crowds performing a complex sensemaking task: image geolocation.
(2) GroundTruth makes a system contribution as an operationalization of shared representations that enables expert investigators to geolocate images with the help of crowds. This is done using three shared representations: an aerial diagram, a gridded map overlay, and a heatmap displaying experts' decisions and real-time crowd feedback.
(3) A mixed methods evaluation with expert investigators, who drew their own aerial diagrams and worked with crowds in real time to geolocate images. Experts expressed an overall preference for GroundTruth over current expert work practice and tools. Our evaluation finds that shared representations support new collaborative work dynamics.
(4) We further develop implications for enriching expert–crowd collaboration in investigative work, and applying shared representations to complex tasks beyond visual search, which are currently only the purview of experts.

2 BACKGROUND

To provide context for our system and evaluation, we first summarize expert image verification and geolocation workflows. Although there are empirical studies of verification in general [14, 74], image geolocation has seen less scholarly attention [45, 60]. Therefore, we draw on practitioner accounts [10, 32, 36, 75] to fill gaps in the limited scholarly record. We also relate these existing practices to sensemaking theory [67] and identify challenges that motivated our design rationale and system development.

2.1 Current Expert Practice in Image Geolocation

Experts in many investigative fields, such as journalism, human rights, and military intelligence, perform image geolocation as a key step in the broader task of verifying photos and videos shared on social media. The goal is to identify the precise location where the image was captured, to help support or refute claims about its provenance and meaning.

Experts perform image geolocation using a largely manual process characterized by iteratively narrowing down possibilities to find a needle in a haystack. They start with a photo that they want to geolocate, obtained through social media or by received from clients. Next, they examine the surrounding context and metadata of the image, researching the user who posted it and the claims made about it. Most social media platforms scrub metadata, including geotags, for uploaded content as a privacy measure, which is why experts focus their attention on the actual visual content of the images [75]. Experts look for road signs, business names, phone numbers, unique landscape or architectural features, or other clues that could point to certain locations or rule out others. If this step does not sufficiently narrow down the location, experts may resort to a brute-force approach of manually searching satellite imagery in candidate locations for potential matches. Experts often draw an aerial diagram of the ground-level photo of interest to ease visual comparison [36]. This "translation" process requires expertise in spatial reasoning and mental rotation which experts develop over time [60].

2.2 Expert Image Geolocation as Sensemaking

The process of image geolocation, and image verification more generally, can be understood as a sensemaking task, in which the goal is to gather and analyze large amounts of diverse, unstructured information to arrive at a theory or conclusion [23, 67, 69]. One influential model by Pirolli and

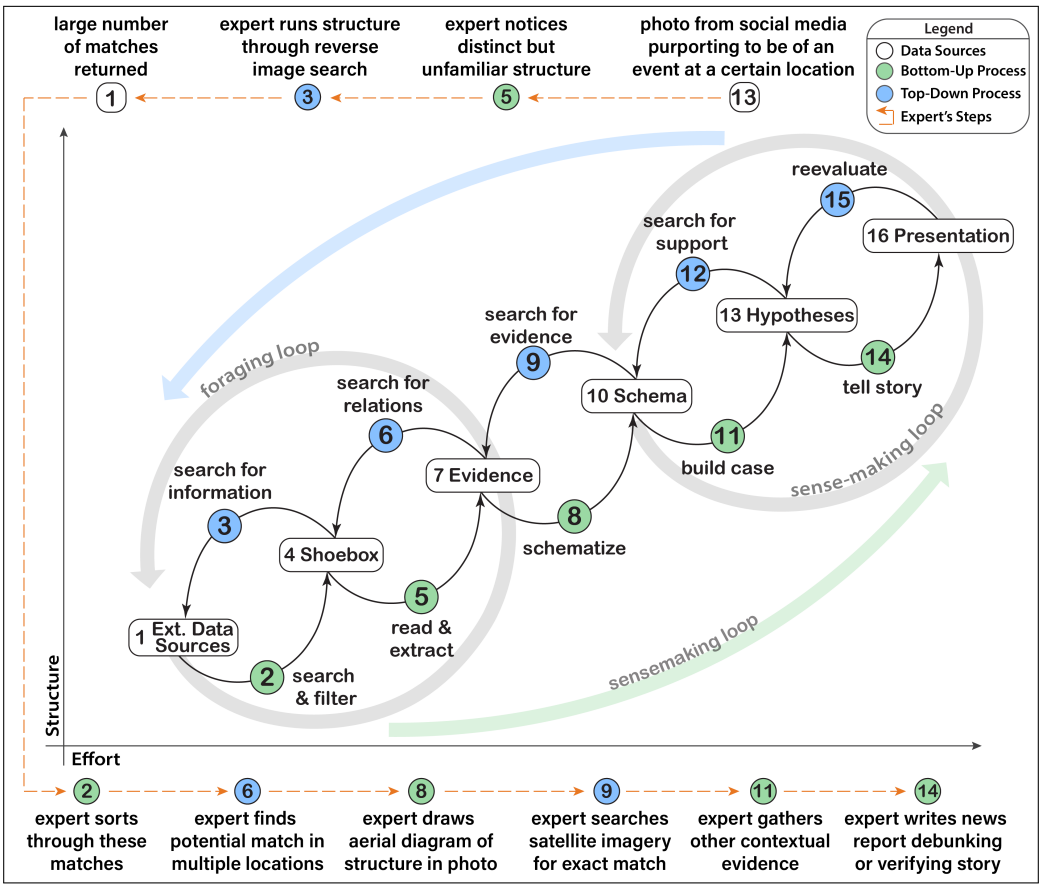

Fig. 1. Expert image geolocation as sensemaking. The orange dotted arrow indicates one possible set of steps that an image verification expert would undertake to debunk or verify visual media through image geolocation. We relate these steps to Pirolli and Card's sensemaking loop for intelligence analysts [67] (inset). See Section 2.2 for an example.

Card [67] characterizes sensemaking as a 16-step, iterative process with two key subloops, one focused on gathering, searching, and extracting information (*foraging*); the other on building a mental model that best fits the information (*synthesis*); with structure and effort increasing in later steps. In our example below, we relate the steps of geolocation back to the sensemaking steps as defined by Pirolli and Card [67] (see Fig. 1). Half the steps (15, 12, 9, 5, 3) are top-down, moving from theory to data, while the other half (2, 5, 8, 11, 14) are bottom-up, creating a "dual search" cycle between adjacent steps.

Applying this model to an image geolocation example, suppose an expert investigates a photo that they encountered on social media purporting to show evidence of a violent protest in an unidentified German city. This claim may provide they with an initial top-down hypothesis. However, as they inspect the visual content, clues in the road signage and building architecture instead suggest Austria, providing a bottom-up challenge to the initial hypothesis. On one building, the expert notices a distinctive but unfamiliar logo (*step 5: read and extract*). They run a reverse image search (3: *search for information*) which yields a large number of results (*1: external data sources*). Sorting through these (*2: search and filter*), they find a match for a business with multiple offices in three

cities in Austria (*6: search for relations*). They then draw an aerial diagram of the photo of interest (*8: schematize*) and manually search satellite imagery from each city (*9: search for evidence*), one at a time. Four hours later, they locate the matching building in one of the cities. This discovery, along with other contextual evidence (*11: build case*), allows them to debunk the original claim of Germany (*14: tell story*), based on which they writes a news report (*16: presentation*).

This paper focuses on the latter steps of the geolocation task, when the expert must search satellite imagery using a manual, brute-force process. Looking through large areas can take hours or even days, with no guarantee of success. Fatigue, coupled with the time-crunched nature of this work, is a major challenge because constant vigilance is required to find a needle in a haystack. There is some evidence that image geolocation experts [45] (along with other journalists and investigators [19, 93]) seek crowdsourced assistance, but little technological support exists, and their success rates are unknown. To address this research gap, we investigate how experts respond to crowdsourced support in image geolocation.

3 RELATED WORK

3.1 Collaborative Sensemaking and Visual Search

3.1.1 Collaborative Sensemaking. Collaboration has the potential to speed up sensemaking by dividing up foraging and search tasks, as well as providing multiple perspectives on schematizing and theorizing about connections, among other benefits [23]. However, collaboration also creates coordination challenges. For example, collaborators may have different skills and backgrounds, be geographically separated, need to externalize their thoughts for others, and have access to different parts of the dataset [22, 24, 27, 38]. Most of these efforts focus on small groups, i.e., dyads [27–29, 38] or triads [72, 92].

However, less work has explored scaling up collaborative sensemaking to a larger number of people, such as crowd workers, where coordination challenges are amplified [80]. Some projects have focused on crowdsourcing specific sensemaking steps as microtasks, such as searching and filtering [66], reading and extracting [16], or schematizing [18, 43, 56]. Others crowdsourced the *entire sensemaking loop* to perform complex tasks like solving mysteries [54, 80] or writing articles [31, 44]. While most of these efforts focus on how either novice crowds alone or crowd–AI hybrids can complete sensemaking tasks, our system differs by exploring how crowds can augment an expert's sensemaking process. Further, while the majority of crowd sensemaking research focuses on text data [e.g., 9, 16, 31, 43, 56, 80], we focus on visual data.

3.1.2 Crowdsourced Visual Search. Prior work on crowdsourced visual search has largely focused on satellite imagery. The focus is often on counting or annotating things (e.g., clouds [94], building damage [39], tanks [68]). Sometimes the focus is searching for a specific needle in a haystack, like Genghis Kahn's tomb [55] or the boat of missing computer scientist Jim Gray [2]. These projects, as well as ours, divide up an area of satellite imagery into smaller cells and ask crowds to conduct a visual search for objects, buildings, etc.

In our prior work [46], we asked crowds to help geolocate an image by searching through satellite imagery using either the ground-level photo or an aerial diagram. Likewise, our system here tasks crowds with searching through satellite imagery using aerial diagrams to support experts. While the previous study used a controlled experiment and perfect diagrams to determine an upper-bound on crowd performance, this paper explores the real-world feasibility of crowds using expert-drawn diagrams, bigger search areas, and more diverse test photos.

Furthermore, all of the above projects focus on crowd performance and did not involve actual expert investigators. In contrast, our work here focuses on an *expert-driven system*, GroundTruth, that aggregates and visualizes the crowd's results and incorporates it into an expert's workflow in

real time. Also in contrast to prior work, our evaluation focuses on how these experts respond to crowd feedback and integrate it into their own investigative activities.

3.2 Requester–Crowd Interaction Models

Early crowdsourcing focused on a relatively simple model where a requester posted a task and returned later, after the task was completed, to review results. The introduction of the waiting room or retainer model allowed requesters to receive results in near real-time [11, 49, 50]. This fast turnaround time enabled new, richer modalities of interaction between requesters and crowds, such as clarification of ambiguous task instructions [13, 58], workflows [47], or conversations with multiple exchanges [51]. Like this last class of systems, GroundTruth leverages real-time crowdsourcing (via LegionTools [25]) to quickly return crowd results to the requester, a key requirement for image verification tasks that are often time-sensitive.

Most similar to our work, a subset of real-time crowdsourcing systems support what we term *crowd-augmented expert work*. This model assumes that (1) requesters are experts in some domain (e.g., animation, innovation, design), (2) requesters are simultaneously performing the same (or a superset) of tasks as the crowd, and (3) the expert's own work is shaped and redirected based on the crowd results streaming in real-time. Crowdboard [8] explored whether online crowds could augment expert ideation in a hybrid physical–virtual studio. Apparition [48, 52] and SketchExpress [53] enabled crowds to prototype user interfaces and animations drawn or described by expert designers. Inspired by these prior works, GroundTruth extends the crowd-augmented expert work paradigm to *visual search tasks*, specifically image geolocation. Our focus on visual search, a type of analytic task, complements prior work in this area largely focused on creative and expressive tasks, which pose distinct challenges [21].

3.3 Visualization and Shared Representations

A variety of technological solutions have been explored for improving coordination in collaborative sensemaking. One theme has been to use tools like visualizations and tabletop displays to help collaborators externalize their ideas in ways that are easily shared and aggregated with others [22, 27–29, 92], though these tools are geared towards lengthy sessions with 2–3 collaborators, not crowds and microtasks. Likewise, as crowdsourcing workflows have become more complex, new tools have helped requesters to monitor and interpret crowd work, often with dashboards and visualizations [13, 42, 47, 57, 70, 90]. Inspired by these, GroundTruth also leverages visualizations to aggregate and display crowd results to the requester.

However, unlike these works, in which requesters typically wait extended periods of time to review work performed entirely by the crowd, GroundTruth's crowd-augmented expert workflow has crowds provide feedback in real-time, while an expert performs the same (and superset of) tasks. Most similar to our work, Crowdboard [8] provided two versions of an interactive workspace: "studio" for in-person experts, and "web" for online crowds; while Apparition [48] provided a "shared canvas" which showed or hid editing tools depending on roles assigned to experts and crowds. While these interfaces proved effective in their respective task domains, it is unclear how to adapt the underlying principles for an analytic task involving visual search of geographic regions.

Consequently, we adapt a concept from mixed-initiative systems, *shared representations* [35], that suggests concrete design principles for collaborative analytic work, such as data analysis and sensemaking. We detail how we adapted these principles for crowd-augmented expert image geolocation in the following section.

4 GROUNDTRUTH

4.1 Design Rationale for Shared Representations

Heer's [35] shared representations, adapted from Horvitz's guidelines for mixed-initiative systems [37], enable people and agents of varying abilities to work together to achieve a common objective. They enable a user to direct tasks, performing a superset of the agents work. In parallel, the agent can support the user's foraging tasks. This allows for the merging of the agent's scale and speed with the user's deep domain knowledge and skills.

Building on these ideas, we explored whether shared representations could inform the design of a system to support image geolocation in which expert work was augmented by human crowd workers rather than AI agents. To do so, we employed an iterative design process with expert investigators over the span of eighteen months, along with pilot studies. From these efforts, we developed three types of shared representations, each motivated by one of Heer's principles [35].

(1) Heer's first principle encourages augmentations that "provide significant value, promoting efficiency, correctness, and consideration of alternate possibilities that a user might not have otherwise considered." We adapted this idea for expert–crowd interaction as a *shared lens*. In GroundTruth, this takes the form of an expert-drawn aerial diagram. The aerial diagram is a shared lens into what the expert believes is relevant in a satellite image search, and indicates to the crowd what to look for. It bootstraps existing expert practice of aerial diagramming and overcomes the crowd's limited spatial reasoning skills, allowing them to support experts at-scale and promote consideration of alternatives.
(2) Heer's second principle emphasizes automated suggestions that "augment, but do not replace, user interaction" and "blend into the interactive experience in a nondisruptive manner and can be directly invoked or dismissed" by the user. We adapted this idea as a *shared environment* between experts and crowds that, for GroundTruth, takes the form of a gridded map overlay. The grid structures the search area, telling the crowd where to look, and divides it into smaller cells (microtasks) for them to easily provide feedback. The grid lets experts direct investigations, and work with the crowd in a nondisruptive, dismissable manner. That is, the expert performs tasks that are the same as—and a superset of—the crowd's, working alongside them in real time.
(3) Heer's third principle encourages augmentations that "require neither perfect accuracy nor exhaustive modeling of the user's task to be useful." We adapted this idea for expert–crowd interaction as *shared analysis*. In GroundTruth, this takes the form of a heatmap. The heatmap enables shared analysis between experts and crowds to locate the ground-level photo (and aerial diagram) within cells of satellite imagery. The heatmap aggregates crowd feedback to prioritize expert attention and allows experts to easily exclude cells as they conduct their search in parallel. Although crowds prioritize some false positives, they also rule out many irrelevant cells, providing value to experts despite imperfect accuracy.

4.2 System Description and Scenario

GroundTruth consists of two different interfaces for the three shared representations (aerial diagram, grid, heatmap). The expert interface allows an expert to define and manage a geolocation task, where the expert uploads the ground-level photo and aerial diagram, and specifies the search space (for both them and crowd workers) by drawing the grid. The crowd worker interface allows crowd workers to perform geolocation microtasks specified by the expert, using the aerial diagram.

We now describe the expert and crowd worker interfaces in detail using the following fictional scenario based on a real-life event [82]. Noor is an investigative journalist who works for an online intelligence group that investigates war crimes. Late last night, the Turkish Air Force bombed the

city of Aleppo, and a few seconds of footage of this bombing was released by Turkish state-run press. Located two thousand kilometers away, Noor does not have access to drone footage; and with the recentness of the event, updated satellite imagery is unavailable. To confirm the video's authenticity, Noor must verify that these air strikes did indeed occur in Aleppo. Under a time crunch, Noor decides to use GroundTruth to help her geolocate the imagery depicted in the video.

4.3 Expert Interface

Step 0: Drawing the aerial diagram. Upon logging in, Noor is asked to upload two images: a photo to be geolocated, and an aerial diagram that represents a bird's-eye view of the photo, which will help both her and crowd workers easily find matching satellite imagery. She first captures a screenshot from the video, and uploads it. Then, using Adobe Photoshop, she starts to draw an aerial diagram of the photo that she wants to geolocate (e.g., the photo on the top left in Fig. 2). She decides to include roads, the outlines of unique buildings, and trees (which are rare in Aleppo), and then uploads it. GroundTruth uses a digital photo format, thus allowing an expert to draw the aerial diagram by hand or using any number of digital tools. Then, Noor specifies the width of the diagram she has drawn in meters or feet.

Next, she is shown a two-panel interface (Fig. 2), with tooltips and interface elements that change depending on which of three steps she is currently in. The three steps are: (1) navigating to the correct location on the map, (2) drawing the search area, and (3) filtering through crowd feedback on the heatmap. On the left, placed within a tabbed view, are the ground-level photo and aerial diagram, along with rotate, pan, and zoom controls. On the right is a Google Maps map, with controls to zoom in/out, and to toggle the satellite/map view.

Step 1: Narrowing the search area. Noor needs to verify that the air strikes took place in Aleppo. She has narrowed down the location to area of several square kilometers, in the northwest corner of the city, based on news reports and contextual clues associated with the ground-level photo. Noor thus navigates to this location using the embedded search bar on the map.

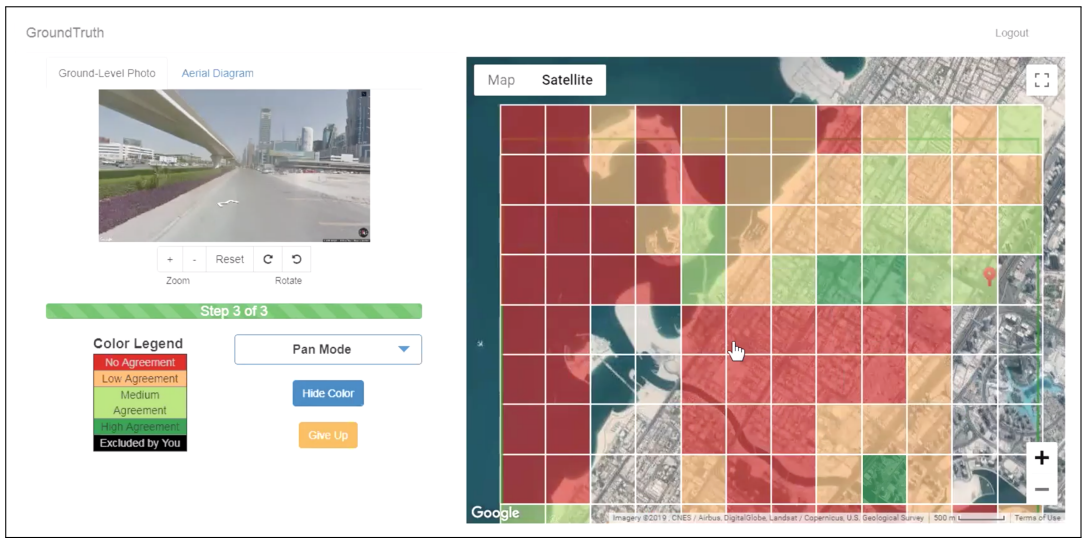

Fig. 2. Expert interface during George's session. The correct location is located in the dark green cell on the bottom right. Priority (agreement) in the color legend corresponds to the number of crowd workers (out of 3) who said that the satellite imagery in the cell matched the aerial diagram. For details, please see Section 4.5.

Fig. 3. Crowd worker interface. Note that this is a reproduction of what a real crowd worker would have seen during George's session, and includes the same aerial diagram that he created.

Step 2: Drawing the search area. Next, Noor delineates the search area that she thinks contains the photo by clicking on the Draw Search Area button, and then clicking and dragging the mouse cursor to draw opposing corners of a rectangular gridded map overlay. In practice, the area can be as large, or as small as one would like, based on constraints such as time, cost, the number of crowd workers available, etc. Next, the system automatically divides the search area into a grid of equally sized *regions*. Each region consists of a 4×4 grid of 16 *cells*, each of which are the same width as the aerial diagram provided by Noor, to allow for 1:1 comparison. Three crowd workers are asked to compare each cell to the aerial diagram provided by Noor.

Step 3: Analyzing crowd feedback. Now, Noor can begin to search through the map with the grid overlaid on top of her target search area. In parallel, she recruits people on Amazon Mechanical Turk (MTurk) to help. She takes their feedback into account as it begins to stream in and is displayed as a heatmap. On the left panel, there is a color legend that indicates the number of crowd workers that said the cell matched the aerial diagram provided. The colors correspond to the number of Yes/Maybe judgements (between 0 to 3). Next to this is text that alternates between *Pan Mode* and *Inspection Mode*. Pan Mode is the default mode where Noor can pan across and zoom into or out of the map. In Pan Mode, she can give up on the search using the Give Up button, or hide crowd feedback with the Hide Color button. If she continues to zoom in, until only one cell takes up 80% of the map, she enters Inspection Mode. Here, two new buttons are added: clicking the Exclude Cell button excludes a cell from the search, and clicking the Found It button indicates that she has found the cell that matches her image. Even when she hides crowd feedback, her decisions are still visible, in line with Heer's second principle.

The crowd, noticing that Noor's diagram contains several trees, quickly eliminate cells of satellite imagery that do not contain them. When they encounter cells that contain trees, they attempt to carefully match it to the buildings that she has represented in the aerial diagram. Within minutes, the crowd has narrowed down the search area by 50%, highlighting areas that she should pay attention to and prioritize. She inspects three cells, and finds that one of them matches the aerial diagram. She clicks on the Found It button to end her search, having successfully geolocated and verified an airstrike that took place in Aleppo, Syria.

4.4 Crowd Worker Interface

The crowd worker interface consists of three columns. The left one shows only the aerial diagram drawn by Noor, and is randomly rotated to avoid imposing an orientation. The crowd is given only the aerial diagram because crowd performance is better, in comparison to being shown only the ground level photo or both [46]. A crowd worker can rotate, zoom, and reset the image to its original position using the buttons underneath the image. On the right is a small mini-map of the crowd worker's search region consisting of 16 smaller, equally-sized cells (4×4 grid). In the middle, there is a single cell, shown in a Google Maps satellite view, and workers are prevented from panning outside this cell. The worker can click on the green Yes/Maybe button if the satellite cell potentially matches the aerial diagram, or a red No button if it does not. Their judgment is reflected on the mini-map where the corresponding cell is either colored green or red, respectively. Once sure of their choice, they can click on the Next button. Then, the satellite view advances to the next cell following a Creeping Line search pattern used by search and rescue professionals [91]. Once the crowd worker provides feedback for all 16 cells, they can submit the task to MTurk. GroundTruth also allows experts to recruit crowd workers from social media or other sources. In either case, a single URL is provided and the system automatically directs them to a region to perform the geolocation microtask.

The crowd worker interface for GroundTrtuh (Fig. 3) is based on our prior work [46], where we designed a technique and an interface to facilitate crowd-supported image geolocation and there was no expert interface present. The crowd worker interface here differs from our prior work [46] in several ways. First, to streamline visual comparison, we increased the size of the diagram to match the size of the satellite image cell. Because of this layout change, we moved the mini-map to the top right. We also added buttons to allow the user to zoom in and out of the diagram and reset its position. All other parts of the interface are visually similar.

4.5 Implementation Details

We built GroundTruth using the Django web framework, a PostgreSQL database, and the Google Maps API for displaying satellite imagery, drawing the grid, and displaying the heatmap of expert and crowd decisions. The gridded map region is drawn using Google Maps' Shapes API, by clicking and dragging the mouse cursor across the map to delineate a rectangular bounding box representing the area under investigation. Next, GroundTruth divides the bounding box into equally sized *regions*, composed of equally sized 4×4 *cells* whose height and width are defiend by the expert when they specify the cell/diagram width. However, if it cannot be divided equally, it is resized so that it can be. Each region and corresponding cells are stored in the database. Due to the map projection used by Google Maps, cells do not appear square unless they are near the equator. As workers load the crowd worker interface, they are redirected to a specific region, i.e., a subset of the grid.

Their feedback for each cell is stored individually in the database as a *judgement*. On the expert interface, the system queries the database to see if three judgements have been made for a cell (by three workers). If so, the number of Yes/Maybe judgements is calculated based on the number of workers that said the cell matched the diagram by clicking the Yes/Maybe button. Each cell is colored based on how many workers clicked Yes/Maybe, using the *one-yes* rule, described in Section 5.5. A cell is only colored once three crowd workers have provided feedback for it. The colors are red, orange, light green, and dark green that correspond to no (zero of three crowd workers said yes), low (one of three), medium (two of three), and high priority (all three), respectively. While experts were shown the color legend that used the word "agreement," for clarity, we henceforth refer to it as crowd worker "priority."

5 EVALUATION METHODS

Heer [35] states that the evaluation of systems that use shared representations must consider both "task time and quality measures, as well as more qualitative concerns regarding participants' perceived autonomy and creativity of action." Along similar lines, we seek to form a *holistic* understanding of how GroundTruth supports experts in performing image geolocation with crowds, and posed the following research questions:

RQ1 Performance: How well did experts and crowds perform image geolocation together using GroundTruth?

RQ2 Experience: What were experts' experiences of using GroundTruth to work with crowds in geolocating an image?

We conducted an exploratory evaluation in which geolocation experts worked with crowds from MTurk to geolocate images. Our mixed-methods approach included a think-aloud protocol and analysis of system log data, followed by in-depth, semi-structured interviews. Below, we describe the setup for our study, expert recruitment, the study procedure, and qualitative and quantitative data collection and analysis methods.

5.1 Image Dataset Selection

To provide a controlled "playground" for experts with different skill sets and who are used to geolocating images in different contexts, we selected a diverse set of images to geolocate, in terms of possible challenges and strategies. We randomly chose coordinates—using guidelines provided by Mehta et al. [60]—that matched each of 24 triplets (6 biomes × 2 population density types × 2 language types), and excluded those that were of prominent locations or had identifying location markers. We then obtained their respective ground-level photos using Google Street View in GeoGuessr to remove road names. The resultant set of 24 images were displayed randomly in an image gallery that experts could choose from.

Now, image geolocation can involve manually searching an area as small as a town, to as large as a country [60]. Thus, due to time constraints with these experts and to ensure that they each had a similarly sized, manageable search area, we narrowed the search area to one that was 240 to 300 times larger than the area that was depicted in the ground-level photo. This varied based on the diagram width that the expert selected. The narrowed search area was in the form of a green rectangle, that encompassed the image location.

5.2 Expert Recruitment

We recruited 11 participants (experts) with expertise in image verification and geolocation, defined as performing verification/geolocation on a weekly basis for at least one year. To reach the widest audience possible in this relatively small population, we recruited participants using purposive sampling [62]. This involved advertising our study on Twitter with a link to a screening survey. We also reached out to experts that had previously participated in our design process, as well as others that we found online via email. We paid experts $75 for taking part in a 90-minute study, which is commensurate with their specialized skills. After 11 experts, we reached theoretical saturation and concluded recruitment.

After recruitment, we asked experts to fill in an online consent form and pre-survey, and scheduled a date for the study. Experts self-reported belonging to a wide range of domains, including journalism, law enforcement, human rights/war zone investigation, and financial intelligence. Two experts identified as female and nine as male; and ranged in age from 18 to 45, with a median age of 32. Nine participants identified as White/Caucasian, one as Asian, and one as Asian American. We include seven participants' full names and affiliations, while four participants expressed wanting to

Name	Gender	Age	Country	Profession	Organization	Yrs.
Lorenzo Romani	Male	30–35	Italy	OSINT Analyst	Anon. Intelligence Firm	4
Dakota Flournoy	Female	26–30	USA	Journalist	Storyful	3
John*	Male	36–40	Europe	Senior Analyst	Anon. Military	15+
Jack*	Male	30–35	Canada	Investigator	Anon. Intelligence Org.	1
Aqwam Hanifan	Male	26–30	Indonesia	Reporter	Tirto.id	3
Alec*	Male	18–25	USA	Consultant	Anon. Intelligence Firm	1
Benjamin Decker	Male	30–35	USA	Researcher	Harvard University	5
George*	Male	41–45	USA	Investigator	Anon. Police Dept.	16
Neil Seward	Male	18–25	Canada	Manager	Anon. Bank	9
P. Kim Bui	Female	36–40	USA	Director	Anon. Newspaper	15
James O'Brien†	Male	46–50	USA	IT Consultant	Self-Employed	18

Table 1. Expert participants and demographics. *= Anonymized, †= James_FP session.

remain anonymous. We believe these experts are accurately able to assess the pros and cons of being identified because of their professions: as journalists who are often in the public eye, and as intelligence analysts who are aware of various security and privacy related issues. McGregor et al. [59] also found that journalists were able to accurately assess similar security and privacy issues.

5.3 Crowd Worker Recruitment

We used LegionTools [25] to facilitate real-time, expert–crowd interaction. The retainer mode allowed us to recruit unique crowd workers for each session from Amazon Mechanical Turk (MTurk), with no qualifications except for being USA-based. This was to ensure that a large a pool of crowd workers would be available to work in a short period of time. Prior to accepting the task, crowd workers were asked to provide consent. Once recruited, they finished a tutorial and were then pooled up at a waiting page. This pooling was necessary to facilitate a real-time experience for experts. Based on pilot studies, it takes 10–15 minutes to pool 30 crowd workers.

We paid crowd workers $7.50/hr for the time they were active in our task, i.e., completing a short 2-minute tutorial and working on a 10-minute task. While they waited for the expert to specify the search area, we paid crowd workers $2.00/hr (for up to 15 minutes). While waiting, we informed crowd workers that they could complete other tasks, and would receive a browser pop-up notification when our task was ready. The total pay per task was at most $2.00.

5.4 Procedure

Before each session, experts filled in a consent form and pre-survey asking about demographic information (Table 1) and work practice. Then, we presented each expert with a walkthrough of the system, asked them to perform an image geolocation task, and conducted a semi-structured interview on their experience. All eleven sessions were run over video chat software (Zoom).

We first demonstrated how GroundTruth worked with a walkthrough, and provided a high-level overview of the system's underpinnings. After clarifying any questions, we gave each expert a link to the aforementioned image gallery. Here, we instructed them to choose one image to geolocate, that was similar to what they would typically encounter in their work. After each session, we removed the image that they picked from the gallery, so that no two experts would geolocate the same image. This was to elicit diversity in terms of techniques and challenges.

Next, we asked them to rate how difficult the image would be to geolocate using a seven-point Likert scale (Table 2). Then, we provided a narrowed location (240 to 300 times larger than the area depicted in the photo) to reduce the search area to a manageable size, and due to time constraints.

We used a think-aloud protocol [20] for the rest of the session, where we asked experts to externalize their thoughts. To familiarize experts with the protocol, we asked them to practice using the protocol on an unrelated task (using a search engine to find a desktop wallpaper).

We then asked experts to spend at most 10 minutes drawing the aerial diagram that the crowd would use. They went through the four-step process on the expert interface (as in Section 4.3), using the tooltips provided. Once they had specified the search area, the system divided it into smaller cells (based on the width chosen), and were given to the crowd to provide feedback.

In parallel, crowd workers were hired using LegionTools, who completed a tutorial and were sent to a waiting page. We hired crowd workers when the expert began to draw the aerial diagram. Once the expert had specified the search area, we routed crowd workers from the waiting page to the task page to work on the task.

As the heatmap populated crowd feedback, experts inspected each cell of satellite imagery, and provided their judgments on whether or not it contained the correct location. We gave each expert a time limit of 30 minutes, at the end of which we asked them to pick a cell that they believed contained the image, if they hadn't already. They were informed that the narrowed search area they were provided contained the correct location. We also provided one expert (James) with an area that did not contain the correct location. We will refer to James and his session as *James_FP*.

After they picked a cell that they thought contained the correct location, we conducted in-depth, semi-structured interviews with each expert, asking them about their overall experience using GroundTruth, their use and perceptions of the heatmap, how using the system compares to current practice, how they see it fitting into their work, among other questions.

5.5 Data Collection and Analysis

The audio and video of each session was recorded using Zoom, and fully transcribed. Sessions ranged from 68 to 112 minutes (median = 80). The first author of this work conducted each session and took detailed notes [76] both during and after each session on how the expert utilized GroundTruth. The notes recorded every time the expert commented on, or reacted to crowd feedback; what cells they looked at; any issues they faced with the user interface; and other points of interest. Notes were subsequently incorporated into each transcript.

Next, we conducted a deductive thematic analysis [15] of the transcripts, based on themes relevant to our research questions and three system components. These themes describe experts' behavior while using GroundTruth and their reflective experiences afterwards (RQ2): drawing the aerial diagram, using the grid and the heatmap with crowd feedback, among others.

To analyze expert performance (RQ1), we determined how far the location that the expert selected was from the location in the ground-level photo, in terms of distance and number of cells. We also calculated how long experts took to do so, from the moment they specified the search area. We also asked experts to rate how difficult the image would be to geolocate.

To analyze each crowd's performance (RQ1), we determined how long they took to provide on the cell that contained the ground-level photo, and on the entire expert-specified search area. We also calculated what percentage of the search area they were able to rule out while retaining the correct cell, and what percentage of cells they indicated as low, medium, and high priority. Here, the correct cell is the cell that contains the location in the ground-level photo. To analyze their feedback, we relied on the one-yes rule proposed by Kohler et al. [46] for calculating aggregated crowd worker performance in image geolocation tasks. The one-yes rule states that if at least one crowd worker said that a cell and the aerial diagram match, then that particular cell would be classified as a possible match. Conversely, a cell will turn red if all three workers said that there is no match to the aerial diagram. This helps ensure that a needle-in-a-haystack problem like geolocation—where there is only one correct answer—will not be overly aggressive in ruling

	Name	Lorenzo	Dakota	John	Jack	Aqwam	Alec	Ben	George	Neil	Kim	James†
Expert	Image Location	Albuquerque, NM, USA	Ft. McPherson, Canada	Bangkok, Thailand	Stockholm, Sweden	Honolulu, HI, USA	Simav, Turkey	Capetown, South Africa	Dubai, UAE	Southampton, UK	Majorca, Spain	Jambi, Indonesia
	Image Difficulty* (-3,+3)	3	1	1	0	-2	2	3	3	2	1	2
	Cell Width (m)	100	200	100	200	100	100	100	500	100	200	100
	Regions Specified (size in km^2)	12 (1.92)	12 (7.68)	9 (1.44)	18 (11.52)	9 (1.44)	9 (1.44)	12 (1.92)	9 (36)	15 (2.4)	12 (7.68)	9 (1.44)
	Cells From Correct Location (number)	**1**	**0**	**7**	**0**	**0**	**2**	**0**	**0**	**0**	**2**	**N/A**
	Time to Find Image Location (min.)	23	6	29	2	10	14	3	27	6	25	6
Crowd	Time to Cover Search Area (min.)	22	16	7.5	22	10	11	10	8	17.5	20.5	7.5
	Included Cell with Image?	**Yes**	**Yes**	**Yes**	**Yes**	**No**	**Yes**	**No**	**Yes**	**Yes**	**No**	**N/A**
	Time to Include Cell with Image (min.)	11	13.75	5	22	N/A	5	N/A	7	16.5	N/A	N/A
	Search Area Reduced (%)	66.5	40.8	40.6	42.9	27.3	39.9	53.9	30.1	41.8	47.1	41.7
	Low Priority (%)	73.4	81.4	49.4	60.4	57.7	43	68.2	46	57.6	67.3	48.8
	Medium Priority (%)	20.3	15.9	29.4	32.3	27.9	36	21.6	32	31.7	26.7	41.7
	High Priority (%)	6.3	2.7	21.2	7.3	14.4	21	10.2	22	10.8	5.9	9.5

Table 2. Expert and crowd session details and performance. *= Image difficulty was assessed using a seven-point Likert Scale ranging from from extremely easy (-3) to extremely difficult (+3). † = James_FP session.

out possible cells. We made this conservative decision because incorrectly discarding the correct location is worse than keeping an incorrect one. A long tail distribution among the three priorities (with fewer high priority cells) is indicative of better performance since the crowd is able to better direct an expert's attention.

5.6 Limitations

We are unable to draw conclusions across participants because we recruited experts from different fields that involve image geolocation, and no two experts geolocated the same image. However, this allowed us to highlight the diversity of challenges faced and strategies employed by experts in geolocating images with crowds.

Further, geolocation involves considering context, searching social media for corroborating sources, and inspecting visual clues, only resorting to brute-force satellite image search when earlier steps are insufficient. In our study, images contained no context within them, which is not typical of how experts encounter images in their work. However, this allowed us to simulate the brute-force step of image geolocation, when other approaches have been exhausted.

Finally, due to experts' time constraints, we limited each session to about 90 minutes and experts did not go through the full image geolocation process, focusing only on the brute-force step. However, this created a time-compressed situation where experts might normally seek others' help. Experts' time constraints, coupled with the typical duration of an image geolocation task (hours to days [32, 36]) also meant that it was infeasible to obtain a baseline. Future work should also study expert performance without crowd support, to obtain baseline performance metrics.

6 RQ1: HOW WELL DID EXPERTS AND CROWDS PERFORM?

In this section, we summarize how experts rated the images in terms of difficulty to geolocate, experts' and crowds' overall performance, an analysis of how expert-drawn aerial diagrams affected crowd performance, and how crowd feedback affected expert performance. The correct cell is

defined as the cell that contains the location depicted in the ground-level photo. Table 2 includes detailed information about each expert's session.

6.1 Image Difficulty

Experts evaluated image difficulty on a seven-point Likert scale ranging from extremely easy (-3) to extremely difficult (+3), with a median rating of moderately difficult (+2), and only two as easy (<0).

Experts said that the images we presented to them in the image gallery did not have any clues within the images that would have helped them find it easily, and that it would have been difficult to geolocate without the initial clue that we gave the experts (typically, they would rely on clues associated with or present within the image, see Section 2). Once we provided experts with a hint that narrowed the search location to an area that was 240–300 times the width they estimated for the ground-level photo, some experts thought the image was easier to find than they initially thought, while others thought it was harder to find.

6.2 Overall Performance

Overall, experts ranged from identifying the single correct cell to selecting a cell no more than seven away. While six experts identified the correct cell, the other four were an average of three cells away. Note that the cell widths chosen by experts here varied from 100 to 500 meters (the system allows for cell widths between 100 to 1000m.). They took between 2 and 29 minutes (mean = 13.7), with the session capped at 30 minutes.

Crowd workers took between 8m 20s and 22m 10s to provide feedback on the entire search area (mean = 13m 5s). They narrowed the search area by 27.3 to 66.5% (mean = 43%) for an area that ranged from 144 to 288 cells (1.44 to 36km^2, mean = 6.81km^2). In 7 out of 10 cases, crowd workers were able to identify the correct cell in a search area of 144 to 288 cells, taking between 5m 16s to 22m 10s (mean = 11m 37s). Crowd workers ruled out the correct cell in three instances, excluding the James_FP session where the search area intentionally did not contain the correct location.

For the three sessions where the crowd did not identify the correct cell, two experts (Aqwam, Ben) pinpointed the correct location within 10 minutes. On the other hand, there were three cases (Lorenzo, John, Alec) where experts were, on average, 350m away from the correct location. Here, the crowd marked the correct location in green within 11 minutes. In other words, there was only one session (out of ten)—Kim's—where both the crowd and expert did not pinpoint the correct location.

6.3 Effect of Aerial Diagram on Crowd Performance

The characteristics of the aerial diagram, such as how easy it was for crowds to interpret, whether it incorporated real or imagined features, and how common those features were in the satellite imagery, affected how well the crowd performed.

In sessions where experts depicted unique architectural and structural features in their aerial diagrams, and/or where the search area consisted of satellite imagery that was easy to rule out, the crowd performed well. For example, in Lorenzo's session, the crowd was able to reduce the search area by 66.5% because the diagram contained roads, while large parts of the search area consisted of shrubland without roads. In addition, 73.4% of the remaining search area was marked as low priority (yellow, one crowd worker said the cell matched the diagram), 20.3% as medium (light green, two), and 6.3% as high (dark green, three). In Dakota's session, although the crowd eliminated 40.8% of the search area from the search, we still observe a long tail in terms of priority: 2.7% of cells were marked as high priority, one of which was the correct cell.

When the expert's aerial diagram was unclear, or when multiple cells of satellite imagery appeared to match features depicted in the aerial diagram, crowd performance was lower than average. For

example, in Aqwam's session, the crowd reduced 27% of the search area (the average was 43%), perhaps because his diagram depicted features that were not actually in the ground-level photo or the satellite imagery. In this case, no crowd worker thought the correct cell matched the aerial diagram. In George's session, the crowd marked the correct cell as high priority. However, the crowd eliminated just 30% of the search area. This may be because in George's aerial diagram he annotated two rectangles as "high-rise buildings" in the center of Dubai, a dense urban area with nearly 200 skyscrapers. Similarly, in John's session, crowd workers had marked the correct cell as high priority. However, there were more high-priority cells than average (21% vs. 12%) because the building he depicted in his aerial diagram had a structural feature that was common in the dense urban area he was searching through.

6.4 Effect of Heatmap on Expert Performance

Experts used and interpreted the crowd feedback displayed as a heatmap in a variety of ways, which affected their performance. Four experts (Lorenzo, Aqwam, Alec, George) found the correct location using crowd feedback. Three others (John, Alec, Kim) chose cells that looked visually similar to the correct location but were not. Here, the crowd also indicated that they looked similar (medium/high priority). Three (Jack, Ben, Dakota) found the correct location before the crowd could finish giving feedback for that cell. Finally, James_FP is in a separate category.

Lorenzo, Aqwam, Alec, and George found the correct location by inspecting areas that the crowd had marked as high priority. John also inspected the correct cell that the crowd had marked as high priority, and thought it was a possible match but then excluded it from the search. He said this was because some features in the satellite imagery did not match due to the angle from which it was taken. Towards the end of the time limit given, John picked a location that was 700m (seven cells) from the correct one; crowd workers had marked the cell he chose as high priority.

Alec and Kim identified locations that looked similar to the correct one, but were not, and the two crowds had marked them as medium priority. In both sessions, Alec and Kim directed their search based on crowd feedback, by first inspecting high and medium priority cells. The location Alec picked was within 200m (two cells) of the correct one. Alec said that the satellite imagery was not detailed enough to be 100% certain of his decision. Kim managed to narrow the location to within 400m (four cells) of the correct one. She initially disagreed with the crowd's medium priority level, but then changed her mind. Although Kim took 25 minutes to search through 7.68km^2, she expressed desire for another hour to verify her choice and continue searching. In Alec's session the crowd marked the correct cell as medium priority, while in Kim's session, they ruled it out.

On the other hand, Jack, Ben, and Dakota knew exactly what features to look for and were able to quickly find the correct location within 0m, before crowd feedback came in for the correct cell. They took two, three, and six minutes, respectively. Both Jack and Dakota relied on crowd feedback initially rule out locations, which let them quickly find the correct location. In Dakota's session, a large portion of the search area consisted of forests, while the ground-level photo was that of a cylindrical structure, which is possibly why she was able to find it quickly. Jack also knew exactly what features to look for (a bridge-like structure along a canal), and in only a portion of the search area contained canals, making it easier to filter through. Similarly, Ben said that there were many distinctive features in the ground-level photo that made it easy for him to geolocate.

Finally James_FP thought that he had found the correct location after six minutes of searching, even though it was not contained within the search area that he drew. James_FP mentioned that without recent, high-quality satellite imagery, it would be difficult to directly confirm whether or not it was the correct location.

7 RQ2: WHAT WERE EXPERTS' EXPERIENCES OF USING GROUNDTRUTH?

Overall, experts said that they were excited by the potential for GroundTruth to scale up expertise in fields such as journalism, human rights advocacy, and criminal investigations. They mentioned specific benefits such as training new investigators, and enabling quicker and better support from crowds, and that it may even save lives.

7.1 Aerial Diagram

7.1.1 Determining the Width. Experts varied in their ability to estimate the width of the ground-level photo. Some, like Ben and Lorenzo, were able to quickly and accurately make a reasonable guess. For example, Ben's strategy involved extrapolating from smaller objects of known size: "*so if it's one lane, [it's] essentially one car, which means somewhere between four and seven feet across approximately.*" Other experts, like Kim, found the task more challenging. Kim said, "*the width is hard for me. I'm not a good judge of distance whereas many other people are. I'm pretty terrible at knowing what 300 feet is.*"

In designing GroundTruth, we included the width estimation step for the practical purpose of leveraging expert knowledge to determine an appropriate cell width for crowd workers. However, some experts found that the step also supported their own spatial reasoning by prompting them to reflect on the relative proportions of key objects. For example, Dakota said, "*that's very helpful… [to] estimate the proportions of the area. You know, in reference, I didn't know it was next to a river and that helped me get a sense of how big of an area I need to be looking in.*"

7.1.2 Drawing the Diagram. While drawing the diagram, experts highlighted salient features in the photo that they thought would be visible on satellite imagery. Some experts chose to label their diagrams, while others only drew the outlines of buildings and roads. Many noted that drawing aerial diagrams is a skill that is developed over time. For example, Kim said, "*I know to do buildings and big roads… but they don't always think about trees, driveways, the color of the roof, the smaller details. I always tell people not to look in the foreground but to look in the background of images, and that's a bit of a training process.*" In her diagram, Kim provided crowd workers key features to look for, such as "*white shutters, roof tiles, and a large, green lamp.*"

Eight out of eleven experts elected to draw the diagram by hand and upload a digital photo, while three used digital drawing tools: Microsoft PowerPoint (George), Microsoft Paint (Lorenzo), and Adobe Photoshop (Aqwam). Experts who expressed the preference for drawing it by hand explained that it was quicker, as well as easier to represent certain features and make annotations.

Reflecting on their experiences using the diagram, and viewing crowd feedback on the heatmap, all experts felt that they would have drawn the diagram differently and taken more time to do so. Since the GroundTruth diagram is dual-purpose in nature, experts grappled with drawing the diagram just for themselves versus also drawing for crowd workers. Experts said that determining what features to include within their diagrams, and what to exclude, was crucial. George explained:

> *I probably would have taken about an hour to do it versus 10 minutes because I think crap in equals crap out… If you don't have a quality diagram that really conveys what you're seeing, you're going to get those false positives… and I take full ownership for that just based on the poor accuracy of my diagram.*

Kim and Ben wanted to communicate more detail to the crowd on what to look for in the form of a list or detailed notes. Kim said, "*I would have said, 'I don't think this is in an extremely rural area, or in the middle of a city. It seems like it's in the outskirts of a city, keep an eye out for X, Y, Z.'*"

John and Dakota both highlighted their "*extremely limited artistic ability*" which was sufficient for their own purposes but, they worried, not for crowds. John proposed that getting feedback from

crowd workers on his diagram would help him improve. Alternatively, experts expressed a desire to have a built-in tool to draw the diagram with pre-specified symbols to reduce uncertainty for crowds and speed up the process. For example, Alec said, "*It's a good process, but to make it more streamlined you might want have a key of labels you can put on it…Maybe set symbols for certain things, because I wasn't sure what to put for a tree.*" Other possibilities suggested by experts included outsourcing this task to a graphic design expert or even another set of crowd workers.

7.1.3 Using the Diagram. Many experts made references to the diagrams they had drawn. John, Dakota, Alec, and Neil looked at the diagram on the interface intermittently to compare it to feedback that crowd workers had provided, whereas Ben, Aqwam, and James_FP frequently switched between using the diagram and the ground-level photo. Kim preferred to use the physical diagram she had drawn because "*I can rotate it [on the interface]. But this is literally what I'm doing on paper…I'm a super tactile person.*" Lorenzo said that he uses aerial diagrams in his work practice, but did not use the digital diagram during his session, relying solely on the ground-level photo.

7.1.4 Diagram as Privacy Protector. Experts also identified privacy benefits related to using an aerial diagram. Because the diagram can abstract away or hide certain sensitive details, experts believed it allowed for more sensitive investigations to be crowdsourced than would typically be possible. John explained:

> *There wasn't much that was actually shared because it was only the diagram…There was no context that was actually given to the crowd worker. Whereas if you give someone a photograph, let's say for a war zone or something like that, then that can contain the kinds of information that you necessarily don't want to give out to other people.*

George compared the diagram to police sketches of suspects drawn with the help of eyewitnesses:

> *It's not a photograph of the actual suspect. It's a rendering of the interpretation from both victim and artist working cooperatively. This is the mapping equivalent of that where you've got an analyst who's rendering based on a slice of imagery but then it gets provided to the crowd.…I don't think we'd run into any privacy issues with that.*

7.2 Gridded Map Overlay

7.2.1 Grid as an Organizational Tool. Some experts would search in a linear fashion through the cells, ruling them out one by one, only deviating when they saw potential matches along the way. Others would search through the grid based on where crowds had (or had not) provided feedback. George and Kim said that the grid allowed them to keep track of where they had searched before, because without it, there would have been a higher tendency to wander across the map haphazardly.

Lorenzo, John, Jack, Aqwam, and Kim extensively utilized the Exclude button to mark cells they had examined and ruled out. Beyond excluding cells, Lorenzo, John, and Kim expressed the desire to have a Possibly button for them to mark cells that they had examined but not fully ruled out.

7.2.2 Grid as a Coordination Tool. Experts suggested that the grid can be used not only to structure their search pattern, but also to aid in coordination between an expert and crowds. Kim and John proposed that another potential use of the Exclude button could be to rule out areas so that crowd workers would not have to review them, saving them effort and time. Dakota and John highlighted how GroundTruth's grid could allow them to geolocate images with their colleagues in a collaborative manner. Dakota explained:

> *[I would say,] "I need your help geolocating, let's go." Then all of us focus on it, and really just to narrow it down, and they say, "Okay, now we've got, you know, 70% or whatever*

> *of this grid looked at a cursory glance." I can take it from [t]here. I think that would be fantastic.*

However, Neil pointed out that because of these experts' limited time, and difficulties in coordination, it may be difficult to find experts.

John and Kim said that in the past, they had divided up a search area into a grid to collaborate with others, either by sharing coordinates or printing out an entire map and indicating who should search where. John described how he and his colleagues were trying to find Austin Tice, a US journalist and former Marine who was abducted in Syria, based off a video that was released of him where his captors were driving through a valley. He mentioned how GroundTruth's grid would have been useful in delegating work in that investigation:

> *It's several tens of square kilometers, maybe a mountain range, between Syria and Lebanon and that's what we actually did on Google Earth. We started to draw this grid of search areas and started to actually go through bit by bit... We've talked about it several times, "Oh we wish that we had something like this, that would actually help us structure the delegation of work in a smarter way."*

John added that experts do not necessarily perform the same tasks with one another, but that "*it's one person finding something, and then they call on others to actually verify that that they agree with that finding*." Therefore, he wished that each cell was labeled with an identifier so that it would be easier to refer to when coordinating a search with colleagues: "*Hey go look at B1 or B30 or whatever.*"

7.2.3 Grid Non-use. Three experts (Jack, Ben, Neil) did not use the grid to structure their search. In one session, Jack, thinking that the correct location was located near a water body, immediately began searching along the length of a canal. Similarly, Ben and Neil directed their search by first looking at areas that they thought had matching architectural and geographical features.

Furthermore, Jack wished that the grid overlay could be turned off completely, while Neil, George wished that the opacity of crowd worker feedback could be varied. Jack stated a preference for searching through the map without any labels or satellite imagery, which he found distracting at times, and view only the outlines of buildings and roads. He often looks for matching shapes in the map first, and then looks to verify the location further once he has a potential match.

7.3 Heatmap of Crowd Feedback

7.3.1 Feedback Usage Patterns. Most experts directed their search based on crowd feedback, and worked in real-time alongside crowd workers. Some experts (Dakota, Kim, John, Alec) described prioritizing cells with high crowd agreement. Others (Lorenzo, James_FP, Aqwam) focused on areas that the crowd had not yet provided feedback on.

Most experts also avoided looking at areas with low crowd agreement, other than to check how crowd workers were performing. For example, Alec said:

> *They have all the red squares at the bottom, I didn't waste my time looking at that. With the middle area, they said it was more likely to be in there so I looked more in there. It gave me more of a baseline to look at, we weren't starting from scratch.*

All experts used the Hide/Show Color buttons when closely inspecting cells in order to hide crowd feedback. Unexpectedly, some experts, such as Ben, John, and Kim, also used these buttons to deliberately—but temporarily—hide crowd feedback to avoid being "biased" by it. These experts elaborated that if they found a likely match first, and others independently agreed with them, it could provide valuable confirmation. Ben described a prior experience geolocating an Instagram video of an Islamic State fighter with his colleagues:

> *We generally double blind each other and we'll then measure up against work…So nobody even discussed what they were even hypothesizing or looking at until we were all finished…It could lead to some bias or something being overlooked.*

On reflection, these experts noted that the nature of their work shifted from actually doing geolocation towards verifying crowd feedback. This change was described as more of a spectrum rather than a dichotomy. For example, John reflected that "*I was waiting for the inputs to come from others and then start verifying their findings…that kind of changed the way that I would normally run the process.*"

Jack, Ben, Dakota, and Neil found the correct location before crowd workers returned feedback for that location. Jack and Ben did not rely on crowd feedback, while Neil sometimes made use of it. Dakota relied on crowd feedback initially, but once she had inspected all cells that crowd workers thought matched, she went on to inspect cells that had no feedback yet.

7.3.2 Feedback Accuracy and Utility. Overall, experts said that crowd workers performed well with the diagrams they were provided. Neil and Kim were initially skeptical of crowd feedback, but then grew to trust it. Others (Lorenzo, Alec, Dakota) trusted crowd feedback from the beginning.

Dakota, Ben, George, and John said that the most helpful part of crowd feedback was that it helped to rule out locations, so that these experts did not have to search there. For example, Dakota said, "*The thing that [the crowd] did the best was specifically blocking out areas that didn't include these features, or it didn't include them all together, which is super helpful.*"

7.3.3 Feedback Speed. Most experts were impressed with how quickly crowd feedback streamed in. However, experts had mixed opinions about whether they wanted crowd feedback to appear in real time or to be asynchronous. Dakota and George identified certain time-sensitive applications in their work that would benefit from real-time crowd feedback. Reflecting on his use of GroundTruth, George said:

> *The most obvious benefit still is from saving time, being able to distill that larger map into regions where I could direct my attention much more quickly than I would have been able to do on my own…I think it's very exciting what you guys are doing…Let me take this opportunity to tell you that what you're doing could save lives in that type of [time-compressed] scenario and never forget that.*

In contrast, Jack enjoys the challenge of geolocating images himself, and felt that he would only resort to using GroundTruth when he is stumped and wants to take a break. He explained, "*I could see myself like sort of turning this on, drawing the square, going to get a cup of coffee or something, coming back and then seeing where can I start looking, where are the dark greens, where are the light greens.*" Similarly, Ben juggles multiple projects at different stages, and said that it would be useful if he could submit images to be geolocated, and then review crowd feedback at a later point of time.

7.3.4 Trusting the Crowd. Experts voiced several considerations that affected their willingness to trust crowd feedback provided by GroundTruth. Building on the above discussion of cost and incentives, Kim questioned whether paid crowds would complete the task in good faith, saying, "*Some people might be just like, 'Let me do this as fast as I can so I get money,' and then some people might be legitimately enjoying the experience. So, that's generally the plus and minus of using [Amazon Mechanical] Turk workers.*" Relatedly, Jack voiced concerns about Mechanical Turk workers providing crowd feedback because he said their anonymity limits accountability.

Taking this idea further, Ben worried about the potential for adversarial crowd workers to hijack or mislead a GroundTruth investigation:

> *... oftentimes the geolocation happens to be around something political. Particularly if it's a war crimes investigation, or some sort of criminal act that's being investigated. Inherently, there's going to be opinions that attempt to [alter] the perception of the outcome, whether that means brushing it under the rug, whether it means hyperbolizing it, etc.*

John wished that the system could be set up such that he could work only with people that he trusts and have been previously vetted:

> *That we could actually split the works work amongst ourselves so not necessarily outsource to anybody else. But we would need a controlled platform where we can bring in trusted partners and people we know, [and] people who know what they're doing, basically, and then actually split the tasks between them....*

7.3.5 Cost and Incentives of the Crowd. Two experts who worked for the private sector (Alec, Neil) said they felt that the actual cost of running their sessions ($2 per crowd worker, approximately $60–120 per search area) was very reasonable given the speed with which feedback streamed in. Kim, a journalist, said that while paying workers would be useful for breaking news desks, it would depend upon each organization's budget.

Furthermore, although several experts felt that workers should be paid a fair wage for their time, John and Jack, who work with a volunteer-based group of open-source investigators, said they doubted that paid crowds would be necessary for their work. They had the ability to mobilize a large number of volunteer crowds on social media, and had already done so for investigations in the past. Building on this idea, Ben was enthusiastic about GroundTruth's ability to democratize open-source investigations by allowing non-experts to make valuable contributions, as well as empowering experts:

> *... it demonstrates the opportunities to do verification at scale, that is truly unique. And I think that really empowering the much larger investigations, is a really significant value add for the research and journalism communities. Especially [when there] are a limited number of people who have these skills.*

8 DISCUSSION

8.1 Designing Shared Representations for Image Geolocation

We designed GroundTruth to help crowds augment experts' complex sensemaking task of image geolocation, drawing inspiration from Heer's notion of shared representations in mixed-initiative systems [35]. In this section, we reflect on the successes and challenges in adapting the principles of shared representations from AI to crowds within the context of image geolocation.

8.1.1 Shared Lens. GroundTruth provided a shared lens between experts and crowds to support Heer's first principle of providing the user with significant value and promoting efficiency, correctness, and consideration of alternatives. The shared lens took the form of an aerial diagram drawn by the expert and shared with the crowd.

Diagramming for a crowd. Our prior work [46] showed that crowds using just a ground-level photo to search satellite imagery lead to unacceptably high false negatives, whereas a perfectly drawn aerial diagram dramatically improved crowd performance. However, little was known about how well diagrams drawn by real geolocation experts would fare. The positive results of our evaluation suggest that, following Heer's first principle, real diagrams do enable valuable crowd performance, helping the expert work more efficiently and correctly by prioritizing high-agreement cells. Aerial diagrams provided a shared lens that helped close the expertise gap between experts and crowds in two ways. First, they bootstrapped the expert's traditional process to leverage spatial

reasoning and mental rotation skills that novice crowds lack. Second, experts drew on experience to focus the crowd's attention on key elements (i.e., permanent, unique), while omitting the rest.

More broadly, our evaluation illuminated how drawing an aerial diagram for one's self differs from one intended to be a communication tool for an unknown crowd of novices. Experts also described additional benefits of the diagram we had not considered, such as its privacy-preserving attributes that can protect both the investigation and the crowd. However, the dual goals of a diagram in the context of GroundTruth also raised new challenges and concerns. Some wanted to include additional details and context that would not benefit themselves, but might help clarify meanings for the crowd. Prior work has found that such shared context is important when dividing a task into microtasks [73], even when performed by the same person (e.g., selfsourcing [81]).

Providing these details would require extra time and labor, so experts also suggested better tools for rapidly drawing diagrams, such as a library of symbols for common objects (e.g., buildings, intersections, bridges), as seen in creative contexts [48]. Besides speeding up the drawing task, regular use of these tools could further streamline communication between experts and crowds.

Benefits of drawing experience. While all experts had deep experience with image geolocation, not all of them were familiar with aerial diagramming, resulting in lower-quality diagrams. This gap speaks to the multi-step process of traditional image geolocation, where experts only resort to brute-force satellite imagery analysis when previous steps are insufficient [14, 46]. While our diversity of experts and images prevents direct comparisons, it is suggestive that the three experts who voiced the least confidence and experience with diagramming (John, Kim, and Alec, see Section 7.1.2) also returned the largest distances from the target location (700m, 400m, and 200m, respectively, see Table 2). These results point to the benefits of expert training or experience in drawing diagrams for themselves as well as the crowd.

8.1.2 Shared Environment. We designed GroundTruth to provide a shared environment in the form of a gridded map overlay. Heer's second principle specifies that agents should augment but not replace user interaction, blend in nondisruptively, and be easily invoked or dismissed. Along these lines, crowd feedback is visualized on the same grid that the expert uses to search the imagery, but can be toggled on and off, while still retaining the experts' own decisions to exclude cells.

Support selfsourcing. We designed GroundTruth's gridded map overlay primarily to support expert–crowd interaction, explicitly prioritizing experts' decisions over the crowd's. Most experts emphasized how it enabled them to visualize and act on crowd feedback while retaining agency [35]. However, we were surprised by how many experts found the grid to be helpful per se in enabling them to systematically search a region and mark off cells using the Exclude button. For these experts, the grid supported a selfsourcing [81] practice not readily available in existing tools. One expert further suggested the ability to judge a cell as Possibly a match, in addition to the current Exclude option, to flag it for closer inspection after an initial pass.

Symbols and structures. Experts suggested mechanisms within the grid that could more effectively support collaborative search. One suggested unique identifiers for each cell to enable easy reference when communicating with colleagues. If implemented, such a feature could leverage existing geographic identifier schema such as What3words [6] for integration with broader volunteered geographic information (VGI) efforts. Another expert suggested the ability to label or add notes to individual cells, either for themselves or direct the crowd in a more nuanced manner.

8.1.3 Shared Analysis. Heer's third principle embraced augmentations that required neither perfect accuracy nor exhaustive modeling of the user's task to be useful. Likewise, GroundTruth supported a shared analysis where crowds would contribute to one module of the broader image

geolocation pipeline: satellite image search. Further, as our prior work suggested that an ideal setup yields a 50% false positive rate [46], we anticipated crowd analysis would be useful, but imperfect.

Crowd analysis with real experts. Our evaluation found that crowds augmented experts' work in ways that were useful without being completely accurate or exhaustive. Even with real experts, authentic diagrams, larger search areas, and more diverse images, crowds reduced the search areas by an average of 43%, comparable to the false positives for an ideal setup seen in prior work. Indeed, there was only one session (of 10) where both the crowd and expert missed the correct location (see Section 6.2 for more details). Taken together, our results not only show that the crowd's shared analysis augmented expert performance in realistic settings, but also suggest that experts might do better to pay closer attention to the crowd's feedback.

Optimizing crowd allocation. While experts valued crowd feedback, they also suggested ways that GroundTruth could better allocate worker effort. For example, the system hires multiple crowd workers to review every cell, regardless of the expert's judgements. However, if an expert excludes a cell not yet reviewed by the crowd, the system could remove those tasks from the worker queue as redundant, allowing workers to focus on other cells. While prior work have also implemented multi-step aggregation and review pipelines [12, 41, 50, 57], in this scenario, workers released from redundant cells would be dynamically reallocated to high priority cells to provide another layer of review prior to expert verification. An expert also requested the ability to define the initial investigation area as a polygon rather than a square, to capture nuances of geography and avoid low-probability areas like mountains or bodies of water.

8.2 Enriching Expert–Crowd Collaboration in Investigations

The previous section discussed how we adapted three of Heer's principles of shared representations to support crowd-augmented expert image geolocation, a subset of image verification investigations. However, Heer [35, p. 2] actually enumerates four principles inspired by Horvitz's guidelines for mixed-initiative user interfaces [37]. The fourth principle states that "through interaction, both people and machines can incrementally learn and adapt." As with the other three, this principle requires some translation to apply it to expert–crowd interactions, rather than users and AI agents, but the effort yields a valuable perspective. We propose three areas of future work inspired by our experiences with GroundTruth that suggest how learning and adaptation between experts and crowds can enable more productive collaboration on other types of investigative and analytic tasks.

8.2.1 Enrich Expert–Crowd Interaction. Above, we reviewed prior work on requester–crowd interaction models, suggesting that innovations in real-time crowdsourcing have enabled richer interactions between requesters and crowds. These interactions can be imagined along a spectrum of increased requester participation. One class of projects employs a hand-off model where requesters specify some initial criteria and then a crowd rapidly completes the tasks largely on its own [11, 50]. A second class has a requester interact sporadically with the crowd, providing facilitation and guidance as required, either with the meta-workflow [13, 58] or the task itself [12, 17, 51]. A third class, which we called crowd-augmented expert work, has experts (i.e., requesters with valuable task-specific skills) working alongside the crowd and performing a superset of the crowd's tasks, with crowd results streaming in and influencing the expert's own work [8, 48, 53].

While GroundTruth embodied the crowd-augmented expert work model within the domain of visual search, the flow of information between experts and crowds was nevertheless highly constrained. Experts defined a search area and diagram for crowds, while crowds returned search results to experts for review. A two-way flow of information might allow experts and crowds better adapt to each other's progress, increasing efficiency and accuracy, albeit with higher collaboration

costs. Prior work such as Crowdboard's virtual sticky notes [8], Apparition's synchronized canvas [48], and structured communication for writing tasks [73] suggest domain-specific mechanisms for richer expert–crowd interactions that could be adapted for analytic tasks like visual search.

Taking this idea further, a growing body of research explores how crowd collectives could address power imbalances between requesters and workers on crowdsourcing platforms [30, 71, 89]. On the other hand, crowdsourced investigations lacking expert oversight are often problematic [63, 93]. Richer two-way communication between experts and crowds could better harness their complementary strengths. Workers might contribute insights and voice concerns about an investigation, while experts bring leadership, professionalism, and investigative expertise to guide the crowd.

8.2.2 Help Crowds Learn Valuable Skills. Heer's idea of incremental learning aligns with recent crowdsourcing scholarship on supporting more complex and creative tasks, and providing more meaningful work experiences for crowd workers beyond the scope of the current task. One thread of research has explored providing just-in-time learning for novice workers to gain task-specific skills [57, 65], with recent work suggesting that crowds can gain more generalized skills and knowledge within the context of microtasks [30, 86]. Crowd workers can also learn from previous workers through a shared memory space [26, 51]—a type of shared representation between crowd workers across instances—that ensures that new crowd workers can quickly adapt to a particular expert's diagrams and working style. Another thread of research seeks to offer novice crowds career ladders and paths to more financially and intrinsically rewarding work [79]. Enriching expert–crowd interaction requires expanding the typical dichotomy of experts and novice crowds, creating intermediate roles that workers can claim as they learn skills and gain knowledge that helps them participate both in the task at hand, and the broader labor market. For instance, what would it mean to train crowd workers to help with other steps in the investigation pipeline?

8.2.3 Engage Diverse Crowds. A focus on crowd learning also directs attention to the composition and diversity of crowds in expert–crowd collaboration. Most real-time crowdsourcing research hires paid crowd workers, who are assumed to be novices and largely transient beyond one task session [72]. Our work with GroundTruth was no exception, though our evaluation revealed that many experts preferred other types of crowds. Some experts worked as freelancers or in cash-strapped organizations where budgeting regular crowdsourcing payments seemed infeasible. Others had access to volunteer novice crowd labor, either through institutional programs or social media followers, that mitigated the need for payment. Still others proposed using their colleagues as crowd workers, drawing on their expertise and mutual trust to speed up searches and improve accuracy. Beyond this diversity of incentives motivating crowd work, a diversity of demographics, experiences, and geographic locations may streamline investigations by allowing experts to solicit specialized knowledge from the crowd.

8.3 Generalizing Shared Representations

While this paper focuses on image geolocation, we suggest that shared representations can be adapted to other complex tasks and domains where crowds support expert investigative work. Most similarly, shared representations could enable other types of crowdsourced satellite image analysis, such as natural disaster damage assessment [77]. They could also help with other types of "needle-in-a-haystack"-type visual search tasks, such as identifying objects to combat human trafficking [5], identifying people in historical photos [61], or finding missing pets after a crisis [88].

More broadly, shared representations may benefit sensemaking activities for which searching is only one of many foraging and synthesis tasks. As discussed in Section 2.2, image verification is a sensemaking activity encompassing not only image geolocation, but also consideration of other types of visual clues in the photo of interest, as well as the broader context about who posted the

photo, when, and why. This and other sensemaking domains also require analysis of non-visual material, expanding the notion of a shared lens to other media. For example, crowds could augment expert analysis of textual data to perform bottom-up qualitative content analysis [9], to investigate evidence documents to solve a crime [27], or to compare features when shopping for an unfamiliar product [43].

Finally, it may be possible to extend the utility of shared representations beyond the analytic tasks addressed by this paper and by Heer [35], which pose unique constraints [21]. Systems like Crowdboard [8] and Apparition [48] illustrate how crowd-augmented expert work can support generative tasks like ideation and prototyping, while other models of expert–crowd interaction support creativity in domains like writing [40]. Designing such shared representations, and understanding how they must differ from those presented here, requires a detailed understanding of the activities that a user engages in during the sensemaking process [64]. We can gain this knowledge, as in GroundTruth's example, by working closely with real experts to understand their current practice, needs, and attitudes.

8.4 Broader Impacts

GroundTruth has the potential to save expert investigators time and scale up their expertise, as well as possibly save lives. It contributes to democratizing the field of visual investigation, empowering newcomers and novices to help debunk misinformation, and responds to a growing need for tools to support information credibility assessment [84]. However, technology can also have negative impacts [34]. Indeed, GroundTruth could be used by oppressive governments to geolocate images, or hijacked by troll farms to skew investigations. Further, by supporting visual investigations, we may contribute to the growing atmosphere of surveillance and indifference towards privacy. Overwhelmingly, however, we believe GroundTruth can be used to do more good than harm.

9 CONCLUSION

In this work, we extended the concept of shared representations to crowd-augmented visual search, in which experts guide and work with a crowd to perform image geolocation. We explore this through our system, GroundTruth, that consists of the shared lens of an aerial diagram, the shared environment of a gridded map overlay, and the shared analysis of crowd feedback. Experts were successfully able to work with crowds on an image geolocation task, and we draw implications for enriching expert–crowd interaction in investigative work, and designing shared representations in domains ranging from creative writing to humanitarian crisis response, and beyond.

ACKNOWLEDGMENTS

We would like to thank Crowd Intelligence Lab members Anne Hoang, Daniel Ocheltree, and Puriwat Lahpong for their contributions to this paper. We would also like to thank Sang Won Lee, Aakash Gautam, and the anonymous reviewers for insightful and detailed comments. Lastly, we wish to thank our expert participants and crowd workers for their involvement, without whom this work would not have been possible. This work was supported by NSF awards IIS-1651969 and IIS-1527453.

REFERENCES

[1] 2019. Amnesty Decoders Strike Tracker. https://decoders.amnesty.org/projects/strike-tracker
[2] 2019. Help Find Jim Gray With Web 2.0. https://techcrunch.com/2007/02/03/help-find-jim-gray-with-web-20/
[3] 2019. Humanitarian UAV Network. http://uaviators.org/map
[4] 2019. Quiztime – Medium. https://medium.com/quiztime
[5] 2019. Stop Child Abuse – Trace an Object. https://www.europol.europa.eu/stopchildabuse
[6] 2019. what3words | Addressing the world. https://map.what3words.com/

[7] Tanja Aitamurto. 2015. Motivation Factors in Crowdsourced Journalism: Social Impact, Social Change, and Peer Learning. *International Journal of Communication* 9, 0 (2015). https://ijoc.org/index.php/ijoc/article/view/3481

[8] Salvatore Andolina, Hendrik Schneider, Joel Chan, Khalil Klouche, Giulio Jacucci, and Steven Dow. 2017. Crowdboard: Augmenting In-Person Idea Generation with Real-Time Crowds. In *Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition (C & C '17)*. ACM, New York, NY, USA, 106–118. https://doi.org/10.1145/3059454.3059477

[9] Paul André, Aniket Kittur, and Steven P. Dow. 2014. Crowd Synthesis: Extracting Categories and Clusters from Complex Data. In *Proceedings of CSCW 2014*.

[10] Trushar Barot. 2014. Verifying Images. In *Verification Handbook: A Definitive Guide to Verifying Digital Content for Emergency Coverage*. http://verificationhandbook.com/book/chapter4.php

[11] Michael S Bernstein, Joel Brandt, Robert C Miller, and David R Karger. 2011. Crowds in two seconds: Enabling realtime crowd-powered interfaces. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*. ACM, 33–42.

[12] Michael S Bernstein, Greg Little, Robert C Miller, Björn Hartmann, Mark S Ackerman, David R Karger, David Crowell, and Katrina Panovich. 2010. Soylent: a word processor with a crowd inside. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology*. ACM, 313–322.

[13] Jonathan Bragg, Mausam, and Daniel S. Weld. 2018. Sprout: Crowd-Powered Task Design for Crowdsourcing. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology (UIST '18)*. ACM, New York, NY, USA, 165–176. https://doi.org/10.1145/3242587.3242598

[14] Petter Bae Brandtzaeg, Marika Lüders, Jochen Spangenberg, Linda Rath-Wiggins, and Asbjørn Følstad. 2016. Emerging Journalistic Verification Practices Concerning Social Media. *Journalism Practice* 10, 3 (2016), 323–342. https://doi.org/10.1080/17512786.2015.1020331 arXiv:https://doi.org/10.1080/17512786.2015.1020331

[15] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[16] Joel Chan, Joseph Chee Chang, Tom Hope, Dafna Shahaf, and Aniket Kittur. 2018. SOLVENT: A Mixed Initiative System for Finding Analogies Between Research Papers. 2 (2018), 31:1–31:21. Issue CSCW. https://doi.org/10.1145/3274300

[17] Joel Chan, Steven Dang, and Steven P. Dow. 2016. Improving Crowd Innovation with Expert Facilitation. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16)*. ACM, 1223–1235. https://doi.org/10.1145/2818048.2820023

[18] Lydia B. Chilton, Greg Little, Darren Edge, Daniel S. Weld, and James A. Landay. 2013. Cascade: crowdsourcing taxonomy creation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, 1999–2008. https://doi.org/10.1145/2470654.2466265

[19] Dharma Dailey and Kate Starbird. 2014. Journalists as Crowdsourcerers: Responding to Crisis by Reporting with a Crowd. *Computer Supported Cooperative Work (CSCW)* 23, 4 (01 Dec 2014), 445–481. https://doi.org/10.1007/s10606-014-9208-z

[20] K Anders Ericsson and Herbert A Simon. 1984. *Protocol analysis: Verbal reports as data*. the MIT Press.

[21] Gregory J. Feist. 1991. Synthetic and analytic thought: Similarities and differences among art and science students. *Creativity Research Journal* 4, 2 (Jan. 1991), 145–155. https://doi.org/10.1080/10400419109534382

[22] Gerhard Fischer, Elisa Giaccardi, Hal Eden, Masanori Sugimoto, and Yunwen Ye. 2005. Beyond binary choices: Integrating individual and social creativity. 63, 4 (2005), 482–512. https://doi.org/10.1016/j.ijhcs.2005.04.014

[23] Kristie Fisher, Scott Counts, and Aniket Kittur. 2012. Distributed Sensemaking: Improving Sensemaking by Leveraging the Efforts of Previous Users. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, 247–256. https://doi.org/10.1145/2207676.2207711

[24] Susan Gasson. 2005. The Dynamics of Sensemaking, Knowledge, and Expertise in Collaborative, Boundary-Spanning Design. 10, 4 (2005). https://doi.org/10.1111/j.1083-6101.2005.tb00277.x

[25] Mitchell Gordon, Jeffrey P Bigham, and Walter S Lasecki. 2015. LegionTools: a toolkit+ UI for recruiting and routing crowds to synchronous real-time tasks. In *Adjunct Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*. ACM, 81–82.

[26] Sai R Gouravajhala, , Youxan Jiang, Preetraj Kaur, Jarir Chaar, and Walter S Lasecki. 2018. Finding Mnemo: Hybrid Intelligence Memory in a Crowd-Powered Dialog System. (2018).

[27] Nitesh Goyal and Susan R. Fussell. 2016. Effects of Sensemaking Translucence on Distributed Collaborative Analysis. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16)*. ACM, 288–302. https://doi.org/10.1145/2818048.2820071

[28] Nitesh Goyal, Gilly Leshed, Dan Cosley, and Susan R. Fussell. 2014. Effects of Implicit Sharing in Collaborative Analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, 129–138. https://doi.org/10.1145/2556288.2557229

[29] Nitesh Goyal, Gilly Leshed, and Susan R. Fussell. 2013. Effects of Visualization and Note-taking on Sensemaking and Analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, 2721–2724. https://doi.org/10.1145/2470654.2481376 event-place: Paris, France.

[30] Mary L. Gray and Siddharth Suri. 2019. *Ghost Work: How to Stop Silicon Valley From Building a New Global Underclass*. Houghton Mifflin Harcourt.

[31] Nathan Hahn, Joseph Chang, Ji Eun Kim, and Aniket Kittur. 2016. The Knowledge Accelerator: Big Picture Thinking in Small Pieces. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, 2258–2270. https://doi.org/10.1145/2858036.2858364

[32] Michael Edison Hayden. 2019. A Guide to Open Source Intelligence (OSINT). https://www.cjr.org/tow_center_reports/guide-to-osint-and-hostile-communities.php#four

[33] James Hays and Alexei A Efros. 2008. IM2GPS: estimating geographic information from a single image. In *2008 ieee conference on computer vision and pattern recognition*. IEEE, 1–8.

[34] B. Hecht, L. Wilcox, Bigham, J.P., J. Schöning, E. Hoque, Bisk Y. Ernst, J., L. De Russis, L. Yarosh, D. Anjum, B.and Contractor, and C. Wu. 2019. It's Time to Do Something: Mitigating the Negative Impacts of Computing Through a Change to the Peer Review Process. https://acm-fca.org/2018/03/29/negativeimpacts/

[35] Jeffrey Heer. 2019. Agency plus automation: Designing artificial intelligence into interactive systems. *Proceedings of the National Academy of Sciences* 116, 6 (2019), 1844–1850. https://doi.org/10.1073/pnas.1807184115 arXiv:https://www.pnas.org/content/116/6/1844.full.pdf

[36] Eliot Higgins. 2014. A Beginner's Guide to Geolocation. https://www.bellingcat.com/resources/how-tos/2014/07/09/a-beginners-guide-to-geolocation/

[37] Eric Horvitz. 1999. Principles of Mixed-initiative User Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '99)*. ACM, New York, NY, USA, 159–166. https://doi.org/10.1145/302979.303030

[38] Ruogu Kang, Aimee Kane, and Sara Kiesler. 2014. Teammate Inaccuracy Blindness: When Information Sharing Tools Hinder Collaborative Analysis. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '14)*. ACM, 797–806. https://doi.org/10.1145/2531602.2531681 event-place: Baltimore, Maryland, USA.

[39] N. Kerle and R. R. Hoffman. 2013. Collaborative damage mapping for emergency response: the role of Cognitive Systems Engineering. 13, 1 (2013), 97–113. https://doi.org/10.5194/nhess-13-97-2013

[40] Joy Kim, Justin Cheng, and Michael S. Bernstein. 2014. Ensemble: Exploring Complementary Strengths of Leaders and Crowds in Creative Collaboration. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '14)*. ACM, New York, NY, USA, 745–755. https://doi.org/10.1145/2531602.2531638

[41] Joy Kim, Sarah Sterman, Allegra Argent Beal Cohen, and Michael S. Bernstein. 2017. Mechanical Novel: Crowdsourcing Complex Work Through Reflection and Revision. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. ACM, New York, NY, USA, 233–245. https://doi.org/10.1145/2998181.2998196

[42] Aniket Kittur, Susheel Khamkar, Paul André, and Robert Kraut. 2012. CrowdWeaver: visually managing complex crowd work. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work (CSCW '12)*. ACM, 1033–1036. https://doi.org/10.1145/2145204.2145357

[43] Aniket Kittur, Andrew M. Peters, Abdigani Diriye, and Michael Bove. 2014. Standing on the Schemas of Giants: Socially Augmented Information Foraging. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '14)*. ACM, 999–1010. https://doi.org/10.1145/2531602.2531644

[44] Aniket Kittur, Boris Smus, Susheel Khamkar, and Robert E. Kraut. 2011. CrowdForge: crowdsourcing complex work. In *Proceedings of the 24th annual ACM symposium on User interface software and technology (UIST '11)*. ACM, 43–52. https://doi.org/10.1145/2047196.2047202

[45] Rachel Kohler and Kurt Luther. 2017. Crowdsourced Image Geolocation as Collective Intelligence. *Collective Intelligence 2017* (2017).

[46] Rachel Kohler, John Purviance, and Kurt Luther. 2017. Supporting Image Geolocation with Diagramming and Crowdsourcing. In *Fifth AAAI Conference on Human Computation and Crowdsourcing*.

[47] Anand Kulkarni, Matthew Can, and Björn Hartmann. 2011. Turkomatic: Automatic, Recursive Task and Workflow Design for Mechanical Turk. In *Proceedings of the 11th AAAI Conference on Human Computation (AAAIWS'11-11)*. AAAI Press, 91–96. http://dl.acm.org/citation.cfm?id=2908698.2908716

[48] Walter S. Lasecki, Juho Kim, Nick Rafter, Onkur Sen, Jeffrey P. Bigham, and Michael S. Bernstein. 2015. Apparition: Crowdsourced User Interfaces That Come to Life As You Sketch Them. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 1925–1934. https://doi.org/10.1145/2702123.2702565

[49] Walter S. Lasecki, Christopher D. Miller, Raja Kushalnagar, and Jeffrey P. Bigham. 2013. Legion Scribe: Real-time Captioning by the Non-experts. In *Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility (W4A '13)*. ACM, New York, NY, USA, Article 22, 2 pages. https://doi.org/10.1145/2461121.2461151

[50] Walter S. Lasecki, Kyle I. Murray, Samuel White, Robert C. Miller, and Jeffrey P. Bigham. 2011. Real-time Crowd Control of Existing Interfaces. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*

(UIST '11). ACM, New York, NY, USA, 23–32. https://doi.org/10.1145/2047196.2047200

[51] Walter S. Lasecki, Rachel Wesley, Jeffrey Nichols, Anand Kulkarni, James F. Allen, and Jeffrey P. Bigham. 2013. Chorus: A Crowd-powered Conversational Assistant. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology (UIST '13)*. ACM, 151–162. https://doi.org/10.1145/2501988.2502057

[52] Sang Won Lee, Rebecca Krosnick, Sun Young Park, Brandon Keelean, Sach Vaidya, Stephanie D. O'Keefe, and Walter S. Lasecki. 2018. Exploring Real-Time Collaboration in Crowd-Powered Systems Through a UI Design Tool. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 104 (Nov. 2018), 23 pages. https://doi.org/10.1145/3274373

[53] Sang Won Lee, Yujin Zhang, Isabelle Wong, Yiwei Yang, Stephanie D. O'Keefe, and Walter S. Lasecki. 2017. SketchExpress: Remixing Animations for More Effective Crowd-Powered Prototyping of Interactive Interfaces. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology (UIST '17)*. ACM, New York, NY, USA, 817–828. https://doi.org/10.1145/3126594.3126595

[54] Tianyi Li, Kurt Luther, and Chris North. 2018. CrowdIA: Solving Mysteries with Crowdsourced Sensemaking. 2 (2018), 105:1–105:29. Issue CSCW. https://doi.org/10.1145/3274374

[55] Albert Yu-Min Lin, Andrew Huynh, Gert Lanckriet, and Luke Barrington. 2014. Crowdsourcing the Unknown: The Satellite Search for Genghis Khan. 9, 12 (2014), e114046. https://doi.org/10.1371/journal.pone.0114046

[56] Kurt Luther, Nathan Hahn, Steven P. Dow, and Aniket Kittur. 2015. Crowdlines: Supporting Synthesis of Diverse Information Sources through Crowdsourced Outlines. In *Third AAAI Conference on Human Computation and Crowdsourcing*. https://www.aaai.org/ocs/index.php/HCOMP/HCOMP15/paper/view/11603

[57] Kurt Luther, Amy Pavel, Wei Wu, Jari-lee Tolentino, Maneesh Agrawala, Björn Hartmann, and Steven P. Dow. 2014. CrowdCrit: Crowdsourcing and Aggregating Visual Design Critique. In *Proceedings of the Companion Publication of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW Companion '14)*. ACM, New York, NY, USA, 21–24. https://doi.org/10.1145/2556420.2556788

[58] V. K. Chaithanya Manam and Alexander J. Quinn. 2018. WingIt: Efficient Refinement of Unclear Task Instructions. In *Sixth AAAI Conference on Human Computation and Crowdsourcing*. https://aaai.org/ocs/index.php/HCOMP/HCOMP18/paper/view/17931

[59] Susan E. McGregor, Elizabeth Anne Watkins, and Kelly Caine. 2017. Would You Slack That?: The Impact of Security and Privacy on Cooperative Newsroom Work. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW, Article 75 (Dec. 2017), 22 pages. https://doi.org/10.1145/3134710

[60] Sneha Mehta, Chris North, and Kurt Luther. 2016. An exploratory study of human performance in image geolocation tasks. In *HCOMP 2016 GroupSight Workshop on Human Computation for Image and Video Analysis*.

[61] Vikram Mohanty, David Thames, and Kurt Luther. 2018. Photo Sleuth: Combining Collective Intelligence and Computer Vision to Identify Historical Portraits. In *ACM Conference on Collective Intelligence (CI 2018)*.

[62] Janice M Morse. 2004. Purposive sampling. *Encyclopedia of social science research methods* (2004), 885–886.

[63] Johnny Nhan, Laura Huey, and Ryan Broll. 2015. Digilantism: An Analysis of Crowdsourcing and the Boston Marathon Bombings. *The British Journal of Criminology* 57, 2 (12 2015), 341–361. https://doi.org/10.1093/bjc/azv118 arXiv:http://oup.prod.sis.lan/bjc/article-pdf/57/2/341/10461339/azv118.pdf

[64] Donald A. Norman. 2005. Human-centered Design Considered Harmful. *Interactions* 12, 4 (July 2005), 14–19. https://doi.org/10.1145/1070960.1070976

[65] Vineet Pandey, Amnon Amir, Justine Debelius, Embriette R. Hyde, Tomasz Kosciolek, Rob Knight, and Scott Klemmer. 2017. Gut Instinct: Creating Scientific Theories with Online Learners. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 6825–6836. https://doi.org/10.1145/3025453.3025769

[66] Alexandra Papoutsaki, Hua Guo, Danae Metaxa-Kakavouli, Connor Gramazio, Jeff Rasley, Wenting Xie, Guan Wang, and Jeff Huang. 2015. Crowdsourcing from Scratch: A Pragmatic Experiment in Data Collection by Novice Requesters. In *Third AAAI Conference on Human Computation and Crowdsourcing*. https://www.aaai.org/ocs/index.php/HCOMP/HCOMP15/paper/view/11582

[67] Peter Pirolli and Stuart Card. 2005. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of international conference on intelligence analysis*, Vol. 5. McLean, VA, USA, 2–4.

[68] Nathaniel A. Raymond, Benjamin I. Davies, Brittany L. Card, Ziad Al Achkar, and Isaac L. Baker. 2013. While We Watched: Assessing the Impact of the Satellite Sentinel Project. (2013). https://www.georgetownjournalofinternationalaffairs.org/online-edition/while-we-watched-assessing-the-impact-of-the-satellite-sentinel-project-by-nathaniel-a-raymond-et-al

[69] Daniel M Russell, Mark J Stefik, Peter Pirolli, and Stuart K Card. 1993. The cost structure of sensemaking. In *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems*. ACM, 269–276.

[70] Jeffrey Rzeszotarski and Aniket Kittur. 2012. CrowdScape: interactively visualizing user behavior and output. In *Proceedings of the 25th annual ACM symposium on User interface software and technology (UIST '12)*. ACM, 55–62. https://doi.org/10.1145/2380116.2380125

[71] Niloufar Salehi, Lilly C. Irani, Michael S. Bernstein, Ali Alkhatib, Eva Ogbe, Kristy Milland, and Clickhappier. 2015. We Are Dynamo: Overcoming Stalling and Friction in Collective Action for Crowd Workers. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 1621–1630. https://doi.org/10.1145/2702123.2702508 event-place: Seoul, Republic of Korea.

[72] Niloufar Salehi, Andrew McCabe, Melissa Valentine, and Michael Bernstein. 2017. Huddler: convening stable and familiar crowd teams despite unpredictable availability. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, 1700–1713.

[73] Niloufar Salehi, Jaime Teevan, Shamsi Iqbal, and Ece Kamar. 2017. Communicating Context to the Crowd for Complex Writing Tasks. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. ACM, New York, NY, USA, 1890–1901. https://doi.org/10.1145/2998181.2998332

[74] Ivor Shapiro, Colette Brin, Isabelle Bédard-Brûlé, and Kasia Mychajlowycz. 2013. Verification as a Strategic Ritual. 7, 6 (2013), 657–673. https://doi.org/10.1080/17512786.2013.765638

[75] Craig Silverman. 2013. Verification Handbook: A Definitive Guide to Verifying Digital Content for Emergency Coverage. http://verificationhandbook.com/

[76] James P Spradley. 2016. *Participant observation*. Waveland Press.

[77] Kevin Stowe, Martha Palmer, Jennings Anderson, Marina Kogan, Leysia Palen, Kenneth M. Anderson, Rebecca Morss, Julie Demuth, and Heather Lazrus. 2018. Developing and Evaluating Annotation Procedures for Twitter Data during Hazard Events. In *Proceedings of the Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions (LAW-MWE-CxG-2018)*. Association for Computational Linguistics, Santa Fe, New Mexico, USA, 133–143. https://www.aclweb.org/anthology/W18-4915

[78] Andrew Sullivan. 2015. Category: View From Your Window. http://dish.andrewsullivan.com/category/view-from-your-window/

[79] Ryo Suzuki, Niloufar Salehi, Michelle S. Lam, Juan C. Marroquin, and Michael S. Bernstein. 2016. Atelier: Repurposing Expert Crowdsourcing Tasks As Micro-internships. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 2645–2656. https://doi.org/10.1145/2858036.2858121

[80] Yla Tausczik and Mark Boons. 2018. Distributed Knowledge in Crowds: Crowd Performance on Hidden Profile Tasks. In *Twelfth International AAAI Conference on Web and Social Media*. https://aaai.org/ocs/index.php/ICWSM/ICWSM18/paper/view/17817

[81] Jaime Teevan, Daniel J Liebling, and Walter S Lasecki. 2014. Selfsourcing personal tasks. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*. ACM, 2527–2532.

[82] Ece Toksabay and Angus McDowall. 2016. Turkey bombs Syrian Kurdish militia allied to U.S.-backed force. https://www.reuters.com/article/us-mideast-crisis-syria-kurds/turkey-bombs-syrian-kurdish-militia-allied-to-u-s-backed-force-idUSKCN12K0ER

[83] Daniel Trottier. 2017. Digital Vigilantism as Weaponisation of Visibility. *Philosophy & Technology* 30, 1 (01 Mar 2017), 55–72. https://doi.org/10.1007/s13347-016-0216-4

[84] Sukrit Venkatagiri and Amy X Zhang. 2018. Response to "Heuristics for the Online Curator", Brian G Southwell and Vanessa Boudewyns (Eds.). *Curbing the Spread of Misinformation: Insights, Innovations, and Interpretations From the Misinformation Solutions Forum*, 9–10. https://doi.org/10.3768/rtipress.2018.cp.0008.1812

[85] Nam Vo, Nathan Jacobs, and James Hays. 2017. Revisiting im2gps in the deep learning era. In *Proceedings of the IEEE International Conference on Computer Vision*. 2621–2630.

[86] Nai-Ching Wang, David Hicks, and Kurt Luther. 2018. Exploring Trade-Offs Between Learning and Productivity in Crowdsourced History. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW (Nov. 2018), 178:1–178:24. https://doi.org/10.1145/3274447

[87] Tobias Weyand, Ilya Kostrikov, and James Philbin. 2016. Planet-photo geolocation with convolutional neural networks. In *European Conference on Computer Vision*. Springer, 37–55.

[88] Joanne I White, Leysia Palen, and Kenneth M Anderson. 2014. Digital mobilization in disaster response: the work & self-organization of on-line pet advocates in response to hurricane sandy. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. ACM, 866–876.

[89] Mark E. Whiting, Dilrukshi Gamage, Snehalkumar (Neil) S. Gaikwad, Aaron Gilbee, Shirish Goyal, Alipta Ballav, Dinesh Majeti, Nalin Chhibber, Angela Richmond-Fuller, Freddie Vargus, Tejas Seshadri Sarma, Varshine Chandrakanthan, Teogenes Moura, Mohamed Hashim Salih, Gabriel Bayomi Tinoco Kalejaiye, Adam Ginzberg, Catherine A. Mullings, Yoni Dayan, Kristy Milland, Henrique Orefice, Jeff Regino, Sayna Parsi, Kunz Mainali, Vibhor Sehgal, Sekandar Matin, Akshansh Sinha, Rajan Vaish, and Michael S. Bernstein. 2017. Crowd Guilds: Worker-led Reputation and Feedback on Crowdsourcing Platforms. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. ACM, New York, NY, USA, 1902–1913. https://doi.org/10.1145/2998181.2998234 event-place: Portland, Oregon, USA.

[90] Wesley Willett, Shiry Ginosar, Avital Steinitz, BjÃűrn Hartmann, and Maneesh Agrawala. 2013. Identifying Redundancy and Exposing Provenance in Crowdsourced Data Analysis. (2013).

[91] Helen Wollan. 2004. Incorporating heuristically generated search patterns in search and rescue. *University of Edinburgh* (2004).

[92] Anna Wu, Gregorio Convertino, Craig Ganoe, John M. Carroll, and Xiaolong (Luke) Zhang. 2013. Supporting collaborative sense-making in emergency management through geo-visualization. 71, 1 (2013), 4–23. https://doi.org/10.1016/j.ijhcs.2012.07.007

[93] Elizabeth Yardley, Adam George Thomas Lynes, David Wilson, and Emma Kelly. 2018. What's the deal with 'web-sleuthing'? News media representations of amateur detectives in networked spaces. *Crime, Media, Culture* 14, 1 (2018), 81–109. https://doi.org/10.1177/1741659016674045 arXiv:https://doi.org/10.1177/1741659016674045

[94] Ling Yu, Sheryl Ball, Christine Blinn, Klaus Moeltner, Seth Peery, Valerie Thomas, Randolph Wynne, Ling Yu, Sheryl B. Ball, Christine E. Blinn, Klaus Moeltner, Seth Peery, Valerie A. Thomas, and Randolph H. Wynne. 2015. Cloud-Sourcing: Using an Online Labor Force to Detect Clouds and Cloud Shadows in Landsat Images. 7, 3 (2015), 2334–2351. https://doi.org/10.3390/rs70302334

Received April 2019; revised June 2019; accepted August 2019