Solving AI's last-mile problem with crowd-augmented expert work

Kurt Luther, Department of Computer Science, Virginia Tech

Abstract

Visual search tasks, such as identifying an unknown person or location in a photo, are a crucial element of many forms of investigative work, from academic research, to journalism, to law enforcement. While AI techniques like computer vision can often quickly and accurately narrow down a large search space of thousands of possibilities to a shortlist of promising candidates, they usually cannot select the correct match(es) among those, a challenge known as the *last-mile problem*. We have developed an approach called *crowd-augmented expert work* to leverage the complementary strengths of human intelligence to solve the last-mile problem. We report on case studies developing and deploying two visual search tools, GroundTruth and Photo Sleuth, to illustrate this approach.

Introduction: AI and the last-mile problem

Visual search tasks are a crucial element of many forms of investigative work, from academic research, to journalism, to law enforcement. One common type of visual search task requires the investigator to identify unknown content in a photo or video, such as a person, location, or object. For example, museums seek to identify portraits of unknown subjects in their collections [10], human rights investigators seek to determine the location where a mass murder was video-recorded [1], and law enforcement seeks to identify toys and clothing in photos depicting sexual exploitation of minors [14].

Performing these searches often resembles finding a needle in a haystack, as the investigator must compare an unknown image to hundreds or thousands of identified candidates. First, a large quantity of candidate images must be narrowed down to a shortlist of high-similarity candidates. Second, each shortlist candidate must be visually inspected in detail and compared to the unidentified image to select the match(es). Third, the potential match(es) must be considered in the broader context of the investigation.

Artificial intelligence (AI) techniques, such as computer vision, have been developed to support these visual search tasks, with varying results with respect to performance and uptake by investigators. For example, a growing body of computer vision research in academia and industry focuses on image geolocation, i.e., identifying the precise location on Earth where a photo or video was taken [12,13]. While these efforts have demonstrated success in highly constrained settings, they cannot yet provide reliable and accurate results for global-scale searches. Consequently, practitioners rarely use these tools in their investigations, instead relying on a largely manual process [6].

Another example of AI techniques supporting visual search, face recognition, has seen greater uptake by investigators, especially in law enforcement contexts. A growing number of police departments and federal agencies leverage face recognition to identify unknown suspects in surveillance footage and other imagery [5]. As this technology is more widely adopted, however, its limitations have attracted greater scrutiny. Researchers have identified substantial accuracy issues in major service providers, including disproportionate error rates for darker-skinned faces [2], and civil rights advocates have raised serious objections to the (mis)use of ubiquitous biometric surveillance [3].

Thus, while computer vision techniques like image geolocation and face recognition offer great potential to support visual search tasks, investigators must first overcome major challenges to their effective and

ethical use. First, these technologies encounter widespread opposition when deployed for investigative purposes, especially without expert oversight. Second, these technologies remain highly constrained and context-sensitive in their ability to produce accurate results. While both face recognition and image geolocation software can often quickly and accurately narrow down a large search space of thousands of possibilities to a shortlist of promising candidates, they usually cannot select the correct match(es) among those. Drawing parallels to similar challenges in the transportation and communication industries, i.e., the complexities of getting from the nearest airport to the front door of one's home, we term this the *last-mile problem*.

Our approach: Crowd-augmented expert work

Our research in the Crowd Intelligence Lab seeks to address both challenges (i.e., oversight and last-mile accuracy) by creating visual search tools that combine the complementary strengths of AI, crowdsourcing, and experts. As argued above, computer vision is effective for rapidly searching through thousands (or more) possibilities to generate a shortlist, making it well-suited for the first subtask. However, it cannot reliably select among them. In contrast, human intelligence is poorly suited to massive searches, due to fatigue and other issues. However, humans are well-suited for the other two subtasks: performing fine-grained comparative analysis, and synthesizing diverse information sources and broader context.

Furthermore, two forms of human intelligence, novice crowdsourcing and individual experts, offer complementary strengths in performing these two subtasks. Individual experts are highly skilled and experienced, but fundamentally limited in the time and attention they can give to any one investigation. Novice crowds lack this skill and experience, but instead offer highly scalable, parallelizable general intelligence. Thus, crowds can rapidly perform many comparative analysis microtasks without fatigue, while experts can draw on their skill and experience to verify the results and weigh the evidence. Together, crowds and experts can strengthen each other while providing combined human oversight of AI-generated results.

We have developed an approach called *crowd-augmented expert work* that leverages these complementary strengths to narrow down a candidate dataset in three phases. First, AI narrows down a huge set of candidates to a promising shortlist. Second, novice crowdsourcing performs detailed comparative analysis of the shortlist in near real-time. The comparisons are facilitated by an expert who highlights salient areas of the unknown image to guide crowd analysis. Third, the expert reviews the aggregated and visualized crowd analyses while conducting their search, taking responsibility for the final decisions about potential matches. Below, we describe two case studies where we employed this approach in building visual search tools.

Case Study 1: Geolocating images with GroundTruth

This case study addresses the challenge of verifying or debunking photos and videos shared on social media through the process of image geolocation described above. We took inspiration from computer vision-based systems like PlaNet [13] and Im2GPS [12], which attempt to automatically geolocate unknown photos using deep learning networks trained on millions of reference photos collected from the web. A recent comparison found that Im2GPS achieved 47.7% accuracy vs. 37.6% for PlaNet at a region-level (200km) localization [12]. Expert investigators such as open source intelligence analysts require

point-level (<1km) localization with perfect accuracy [6]. Therefore, we explored whether crowds and experts could work together to solve this last-mile problem.

Prior work [6] showed that when experts perform image geolocation, they first examine visual clues in the image content and the image's metadata and context. If these details are insufficient, they often resort to a brute-force process of manually reviewing large areas of satellite imagery in the suspected region. This work is facilitated by an aerial diagram of the unknown (typically ground-level) photo drawn by the expert, enabling an easier comparison to satellite imagery.

In one study [7], we explored whether providing this diagram to novice crowd workers could allow crowds to search the satellite imagery. In two experiments, we found that an aerial diagram with a medium level of detail enabled crowds to narrow down the search area by 50% in about 10 minutes while including the correct location 98% of the time. In contrast, providing crowds with the ground-level photo (rather than the expert's diagram) reduced the true positive rate to an unacceptably low 78%. Additionally, because multiple workers search each area, high intra-crowd agreement provides a reliable signal of where to prioritize the search.

In a second study, we built a system, GroundTruth [11], to enable experts and crowds to work together on image geolocation tasks. Experts define an initial search area which the system divides into a grid of cells and allocates to crowd workers. As workers review cells, their judgements are aggregated and visualized for the expert to consider while they perform their own parallel search. An evaluation with 11 real geolocation experts and 562 crowd workers found that GroundTruth enabled the combination of crowds and experts to pinpoint the correct location in all but one case.

Case Study 2: Identifying historical portraits with Photo Sleuth

This case study focuses on the visual search task of identifying unknown people in historical photos. The American Civil War (1861-1865) was the first major conflict to be extensively documented with photography, and by one estimate, over 4 million portraits from that era survive today. However, only 10-20% of these photos are identified. We explored whether a combination of AI, crowds, and experts could help rediscover these lost identities.

In our first study, we developed Photo Sleuth [9], a web-based software platform that combines crowdsourced human expertise and AI-based face recognition to identify unknown soldiers in American Civil War-era portraits. Our approach employed a needle-in-the-haystack metaphor with three phases. First, users build the haystack by uploading photos to the reference database--either identified photos to enrich the database, or unknown photos with the goal of identifying them. We seeded the website with 21,000 identified soldier portraits from public collections, and users have contributed over 8,000 more. Second, users narrow the haystack by visually inspecting the unknown photos for clues (e.g., uniform insignia) that are linked to search filters and military records. For example, tagging a soldier's rank chevrons will exclude from search results any soldiers whose military records indicate they never held that rank. Additionally, the haystack is further narrowed by face recognition, which eliminates any candidates with facial similarity below a certain threshold. Third, users find the needle by inspecting the shortlist of candidates with matching military records, sorted by facial similarity. A comparison interface

provides zoom/pan controls and displays biographical details to help users perform detailed facial analysis and consider broader context.

We publicly launched Photo Sleuth in August 2018 (www.civilwarphotosleuth.com) and evaluated its first month of usage. During this period, users uploaded over 1000 photos and proposed new identifications for over 100 previously unknown photos. An expert review found that over 80% of these identifications were probably or definitely correct. Today, the site has over 10,000 registered users and nearly 30,000 soldier photos. As users seek to identify unknown photos, we discovered they would frequently turn to friends and followers on social media to seek feedback on potential matches returned by face recognition.

To support this emergent practice, we built a follow-up system, Second Opinion [8], to help experts and crowds collaborate on last-mile person identification tasks. Once the software's AI-based face recognition returns a shortlist of high-similarity candidates, Second Opinion draws on theories of similarity from cognitive psychology to help experts and crowds collaborate on picking the correct match(es). First, experts highlight unique "high-diagnostic" facial features to focus crowd attention. Second, crowds perform fine-grained facial analysis of each candidate in near-real time. Third, the results are aggregated and visualized for the expert to review in making their final decisions. We evaluated Second Opinion with 10 Civil War photo experts and 300 crowd workers. We found that crowds can eliminate 75% of high-similarity false positives, and that experts were enthusiastic about using Second Opinion in their work.

Final thoughts: Embracing crowd-AI collaboration for the future of work

In *Ghost Work* [4], Mary Gray and Siddharth Suri argue that crowdsourcing and AI are fundamentally intertwined. Crowds generate training data for AI systems, they handle edge cases where AI fails, and they prototype user experiences for future AI systems. While software developers often downplay the role of human workers in their systems and hope to eliminate them, their critical role persists as AI technologies advance and raise the bar for even more automation. Gray and Suri call this phenomenon, which they trace back to the Industrial Revolution, "the paradox of automation's last mile."

Rather than seeking to downplay or eliminate crowds, our work embraces and foregrounds the enduring value of human intelligence in AI-infused systems. We seek to build systems that harness the complementary strengths of crowds, experts, and AI. For visual search tasks, a type of needle-in-the-haystack challenge common in many types of investigative work, our GroundTruth and Photo Sleuth case studies show how crowds and experts can collaborate to solve the last-mile problem raised by computer vision techniques. In future work, we are exploring new models of crowd-AI collaboration in tasks beyond visual search, from visualizing biological networks to analyzing historical social networks.

Acknowledgements

I thank the members of the Crowd Intelligence Lab, especially Vikram Mohanty, Jacob Thebault-Spieker, and Sukrit Venkatagiri. This research was supported by NSF IIS-1651969 and IIS-1527453.

References

1. Bellingcat Investigation Team. 2017. How a Werfalli Execution Site Was Geolocated. *bellingcat*. Retrieved June 10, 2018 from https://www.bellingcat.com/news/mena/2017/10/03/how-an-execution-site-was-geolocated/
2. Joy Buolamwini and Timnit Gebru. 2018. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In *Conference on Fairness, Accountability and Transparency*, 77–91. Retrieved October 10, 2019 from http://proceedings.mlr.press/v81/buolamwini18a.html
3. Clare Garvie, Alvaro Bedoya, and Jonathan Frankle. 2016. *The Perpetual Line-Up*. Georgetown Law Center on Privacy & Technology. Retrieved January 24, 2019 from https://www.perpetuallineup.org/
4. Mary L. Gray and Siddharth Suri. 2019. *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*. Houghton Mifflin Harcourt, Boston.
5. Drew Harwell. 2019. Oregon became a testing ground for Amazon's facial-recognition policing. But what if Rekognition gets it wrong? *Washington Post*. Retrieved October 10, 2019 from https://www.washingtonpost.com/technology/2019/04/30/amazons-facial-recognition-technology-is-supercharging-local-police/
6. Rachel Kohler and Kurt Luther. 2017. Crowdsourced Image Geolocation as Collective Intelligence. In *Collective Intelligence 2017*.
7. Rachel Kohler, John Purviance, and Kurt Luther. 2017. Supporting Image Geolocation with Diagramming and Crowdsourcing. In *Proceedings of the Fifth AAAI Conference on Human Computation and Crowdsourcing* (HCOMP '17), 98–107. Retrieved January 18, 2018 from https://aaai.org/ocs/index.php/HCOMP/HCOMP17/paper/view/15812
8. Vikram Mohanty, Kareem Abdol-Hamid, Courtney Ebersohl, and Kurt Luther. 2019. Second Opinion: Supporting last-mile person identification with crowdsourcing and face recognition. In *Proceedings of AAAI HCOMP 2019*.
9. Vikram Mohanty, David Thames, Sneha Mehta, and Kurt Luther. 2019. Photo Sleuth: Combining Human Expertise and Face Recognition to Identify Historical Portraits. In *Proceedings of the 24th International Conference on Intelligent User Interfaces* (IUI '19), 547–557. https://doi.org/10.1145/3301275.3302301
10. Steve Szkotak. 2012. Civil War photos: Help sought to solve old mystery. *Yahoo! News*. Retrieved May 14, 2018 from https://www.yahoo.com/news/civil-war-photos-help-sought-solve-old-mystery-092732336.html
11. 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. *Proceedings of the ACM on Human-Computer interaction* 3, CSCW: 107:1-107:30.
12. N. Vo, N. Jacobs, and J. Hays. 2017. Revisiting IM2GPS in the Deep Learning Era. In *2017 IEEE International Conference on Computer Vision (ICCV)*, 2640–2649. https://doi.org/10.1109/ICCV.2017.286
13. Tobias Weyand, Ilya Kostrikov, and James Philbin. 2016. PlaNet - Photo Geolocation with Convolutional Neural Networks. *arXiv:1602.05314 [cs]*. Retrieved July 6, 2016 from http://arxiv.org/abs/1602.05314
14. Stop Child Abuse – Trace an Object. *Europol*. Retrieved October 10, 2019 from https://www.europol.europa.eu/stopchildabuse