

Access & Discovery of Documentary Images (ADDI): A Platform for the Exploration and Critique of Computer Vision Algorithms

Arnold, Taylor

tarnold2@richmond.edu

University of Richmond, United States of America

Tilton, Lauren

ltilton@richmond.edu

University of Richmond, United States of America

An important current topic in the digital and public humanities is the role that artificial intelligence can serve in the access and discovery of large, digitized collections of cultural heritage materials. There is considerable excitement about the possibilities for automated techniques to assist in the exploration archives from new angles and to engage with lesser-known aspects of large collections. At the same time, there is hesitation about the potential for algorithmic biases and the potential threat to the important knowledge conveyed through human interventions through librarians, archivists, and subject-matter experts. More work is needed to understand the possibilities and limits.

In this presentation, we present the project *Access & Discovery of Documentary Images (ADDI)*, collaborative work that with the U.S. Library of Congress Lab's Computing Cultural Heritage in the Cloud (CCHC) Initiative. The initiative "pilot[s] ways to combine cutting edge technology and the collections of the largest library in the world, to support digital research at scale." The ADDI project adapts and applies computer vision algorithms to aid in discovering and using digital collections, specifically photography. Rather than treating cultural heritage images as a monolith, whereby computational approaches are often developed and applied without attention to the form of cultural heritage in technical scholarship, ADDI pursues technical research with computer vision that considers the specificity of photography as a medium, social practice, and source of evidence for humanistic inquiry. Our project focuses on five photography collections from the early 20th century, totaling over a quarter of a million documentary photographs, all held by the Library of Congress. The project produced an open digital platform for exploring these collections using both archival metadata and automatically generated annotations from the application of computer vision algorithms.

The construction of the ADDI digital interface is guided by engagement with existing scholarship in the digital and public humanities. ADDI aims to provide what Whitelaw termed a *generous interface*, a "rich, browsable interfaces that reveal the scale and complexity of digital heritage collections" (Whitelaw 2012, 2015). This was achieved through the construction of a platform that allows for search through both automatically generated metadata and archival metadata. At the same time, engagement with D'Ignazio and Klein's concept of *Data Feminism* led us to be cautious about the power dynamics of giving too much weight to annotations that were produced by algorithms originally designed by

for-profit companies for specific industry applications (D'Ignazio and Klein, 2020). These cautions led us to make an interface that makes a clear separation between archival data and algorithmic data. Further, every single data element in the interface is followed by a question mark icon that provides detailed information about the source of the information. For algorithmically generated features, we provide full citations of the algorithms used to surface both possible sources of biases as well as to credit the underlying labor that went into each model.

Our project also attends to the conclusions of the whitepaper "Computer-Aided Metadata generation for Photo archives" (CAMPI). As the paper notes, the best use of computer algorithms for metadata generation is that which is "tightly integrated with a digital archival collection management system [to] ... strategically leverage machine learning models without making the collection beholden to them" (Corrin et al., 2020). Our collaboration with the Library of Congress included working with to better integrate with the Library of Congress' new API for accessing archival metadata and digitized resources. Our work is also indebted to the model carried out by the SNCC Digital Gateway Project. While most of the individuals involved in the digital archives under consideration in the ADDI project are no longer living, our project seeks to engage in a similar "transfer of informational wealth" through close collaboration with archivists at the Library of Congress (Cox et al., 2020).

Through the engagement with the existing scholarship mentioned above, the ADDI interface introduces three novel features. First, the interface allows visitors to explore the probabilistic nature of the automatically generated computer vision features through a slider that selectively turns on and off annotations based on the algorithm's confidence in their correctness. Secondly, ADDI allows visitors to selectively fade in and out the documentary image to provide more-or-less focus on the automatically generated features, such as detected faces, poses, and objects. Finally, a set of automatically generated recommendations based on visual similarity, we also provide a set of (carefully labelled) randomly generated recommendations. These three novel features offer a paradigm shift that flips the traditional relationship between computer vision and digitized collections of culture heritage.

While typically we see computer vision as a tool for exploring digitized collection, ADDI additionally shows how cultural heritage collections can also offer a critical lens back onto computer vision algorithms themselves. Using our interface, in close collaboration with our partners at the Library of Congress Labs, we have investigated the performance of state-of-the-art computer vision algorithms for the detection of faces (Cao et al., 2018), poses (Cao et al., 2019), object detection (Woo et al, 2022), and image similarity (Douze et al, 2021). We found that the algorithms tend to have high precision but low recall; features that are detected by modern computer vision algorithms tend to be valid however there are many objects, poses, and faces that the algorithms fail to detect. Interestingly, the failure to detect all the features present in the image seems to be a result not of problems with black-and-white images or technological difference between film and born-digital materials. Rather, challenges result from differences between the relatively simple images—typically consisting of only a small number of people and a focused subject—computer vision datasets are built around and the complexity often seen in documentary photography.

We will conclude by exploring future avenues of research for our collaboration with the Library Congress, with particular focus on how the critical lens provided by the ADDI interface can offer new areas of research at the intersection of DH, public humanities, and computer science.

Bibliography

- Arnold, T., Ayers, N., Madron, J., Nelson, R., and Tilton, L.** (2020) "Visualizing a Large Spatiotemporal Collection of Historic Photography with a Generous Interface." Presented at 5th Workshop on Visualization for the Digital Humanities (Vis4DH Workshop at IEEE VIS2020).
- Arnold, T. and Tilton, L.** (2020) "Distant Viewing Toolkit: A Python Package for the Analysis of Visual Culture." *Journal of Open Source Software* , 5(45), 1800.
- Barz, B., & Denzler, J.** (2019). Hierarchy-based image embeddings for semantic image retrieval. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)* (pp. 638-647). IEEE.
- Burke, R., Sonboli, N., & Ordonez-Gauger, A.** (2018). Balanced neighborhoods for multi-sided fairness in recommendation. In *Conference on Fairness, Accountability and Transparency* (pp. 202-214), PMLR.
- Cao, Q., et al.** (2018). VGGFace2: A dataset for recognising faces across pose and age. In 2018 13th IEEE international conference on automatic face & gesture recognition (pp. 67-74).
- Cao, Z., et al.** (2019) "OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields." *IEEE transactions on pattern analysis and machine intelligence* 43.1: 172-186.
- Cox, C., Forner, K., Gartrell, J., Hogan, W., Lawson, J., Moore, I., & Nelson, N.** (2020). Building and Transferring Movement Informational Wealth: The SNCC Digital Gateway. *The Journal of African American History* , 105 (4), 626-647.
- Douze, M., Tolias, G., Pizzi, E., Papakipos, Z., Chanussot, L., Radenovic, F., Jenicek, T., Maximov, M., Leal-Taixé, L., Elezi, I. and Chum, O.,** (2021). "The 2021 image similarity dataset and challenge." *Computer Vision and Pattern Recognition*.
- D'Ignazio, C., & Klein, L. F.** (2020). *Data Feminism* . MIT Press.
- Hinton, S. & Whitelaw, M.** (2010) "Exploring the digital commons: an approach to the visualization of large heritage datasets." In *Proceedings of Electronic Visualization and the Arts* (2010).
- Corrin, J., Davis, E., Lincoln, M., Weingart, S.** (2020). *CAMPI: Computer-Aided Metadata Generation for Photo archives Initiative* . Carnegie Mellon University. Preprint. <https://doi.org/10.1184/R1/12791807.v1>
- Marchionini, G.** "Exploratory search: from finding to understanding." *Communications of the ACM*, 49(4) (2006): 41-46;
- Rogers, Y.** (2012). HCI theory: classical, modern, and contemporary. *Synthesis lectures on human-centered informatics* , 5 (2), 1-129.
- Sherratt, T. and Bagnal, K.** "The People Inside." *Seeing the Past: Experiments with Computer Vision and Augmented Reality in History* , eds Kevin Kee and Timothy Compeau. University of Michigan Press, 2019: 11-31.
- Whitelaw, M.** "Towards generous interfaces for archival collections." In *Proceedings of International Council on Archives Congress* . 2012.
- Whitelaw, M.** "Generous interfaces for digital cultural collections." *Digital Humanities Quarterly* , 9 no 1. 2015.
- Woo, S., Park, K., Oh, S. W., Kweon, I. S., & Lee, J. Y.** (2022). Bridging Images and Videos: A Simple Learning Framework for Large Vocabulary Video Object Detection. In *European Conference on Computer Vision* (pp. 238-258). Springer, Cham.