

Tracy Stuber

Present or Absent? Reframing Photographic Discovery and Interpretation with Computer Vision

Computer vision, a branch of artificial intelligence aimed at understanding digital images, promises to help backlogged archives make their digitized collections more discoverable. However, resonances between this new technology and the history of photography make its application in photographic archives practically and methodologically fraught, especially with images of people. This article describes a critical, pared-back approach to person detection, a common computer vision task, designed in response to the shared physiognomic underpinnings of computer vision and photography. Reframing person detection around the question of presence and absence, it demonstrates how computer vision can enable archival discovery beyond its historically textual and classificatory limits.

Keywords: photography; archives; computer vision; artificial intelligence

When you picture a photo archive, what do you picture? Thirty or forty years ago, before the widespread availability of the internet and the prominent investment in the “archival turn,” it might have been a large space filled with shelves of boxes, not unlike that captured in the image gracing the Wikipedia entry for the term.¹ While those spaces have not gone away, the visual idea of the archive has, for many, become digital. The archive now occupies the space of the web browser; shelves have morphed into rows or grids of thumbnail boxes, each now containing only a single photograph. The visual similarity between interfaces for digital collections of LAMs—libraries, archives, and museums—and interfaces for online shopping websites reflects a parallel shift in their operations. Where shopping used to require going to a store, Amazon

delivers packages to your home; where an archivist once delivered requested boxes and folders to a reading room, you now click, search, and filter until the images you want show up on your home computer.

Looking at these interfaces, it is easy to identify mass digitization as a culprit of archival absence. The ubiquity and instantaneity of internet access in the Global North make it difficult to escape the feeling that if something isn’t online, it doesn’t exist, even for people well aware of the factual fallacy of this articulation. At any rate, for users of online archives, digitized materials have a different presence than non-digitized materials, as often captured in the difference between a thumbnail image and a “No image” placeholder.

But digitization, at least as a shorthand for the creation of a digital surrogate (often a digital image) of an archival photograph, is not the only mechanism of absence in digital archives. As information science professional Murtha Baca succinctly puts it, “Digitization does not equal access. The mere act of creating digital copies of collection materials does not make those materials findable,

Corresponding author:

Tracy Stuber, Harvard University, Cambridge, MA, USA
email: tstuber@fas.harvard.edu

<https://doi.org/10.1515/zkg-2025-3010>

understandable, or utilizable to our ever-expanding audience of online users.”² What is needed for access is metadata—data about data—that describes the digitized materials.

Faced with enormous backlogs of materials that have yet to be cataloged, and therefore lack this necessary metadata, cultural heritage institutions are joining almost every other sector in the first half of the 2020s in experimenting with artificial intelligence (AI).³ For archives, as well as museums, libraries, and other collections of historical photographs, the branch of AI known as computer vision, which deals with interpreting images, stands out for its potential to intervene in the long-standing relationships between images and text that shape photographic meaning. I begin this article by grounding the opportunities and concerns associated with computer vision in the specific disciplinary context of archives as they have evolved in dialogue with libraries and museums as related but distinct institutions. Examining how archival history and practice manifest in contemporary digital image collections, I triangulate photography, archives, and computer vision as interrelated technologies of image management and control that perpetuate similar gaps and omissions. Drawing on research by the collaborative project *Photography Unbound*, I propose how computer vision, and specifically the task of person detection, can disrupt the paradigms underlying these absences.

Computer Vision for Photographic Archives

The ubiquity of the generalized collections interface further blurs a long-standing haziness around the distinctions between libraries, archives, and museums, especially from the perspective of external, non-scholarly users.⁴ Today, the abstract equivalence of data, together with the abstract equivalence of the photograph, defines the horizons of expectation for access to

the historical photographic record. But while interfaces for photographic archives, broadly defined as institutionally maintained photographic collections, are visually structured in similar ways, the structure of metadata behind them often varies according to a given institution’s historical practices, values, and resources, including funding, staffing, and equipment. Metadata as an information concept long predates digital ecosystems of information management, and it is typically created by transforming existing information from its predecessors, such as card catalogs and other analog, paper-based systems, into digital form. This transformation accords with guidelines, known as metadata standards, appropriate to the discipline or institution.⁵

Compare, for example, the collections of the Frick Art Reference Library and the New York Public Library’s Picture Collection. As an art-historical image collection, the Frick’s Library is structured like a museum; photographs are primarily attributed to the artist of the depicted artwork (rather than the photographer), and the subject of the photograph is equivalent to the subject of the artwork, be it “Still Life with Lemons” or “Madonna and Child.” Meanwhile, the New York Public Library’s Picture Collection groups clipped-out images according to content to serve its constituency of artists, illustrators, and educators seeking visual reference points.⁶ Where the Frick would locate a photographic reproduction of British painter Henry Herbert La Thangue’s 1903 *A Provençal Farm* under “Paintings,” “British School,” and “Henry Herbert La Thangue,” the NYPL places La Thangue’s image of goats in a pasture under “Animals—Goat.”

Despite these differences, museums and libraries have typically treated photographs as individual objects in a manner that corresponds well with the structure of contemporary digital databases. Most databases used by contemporary cultural heritage institutions to store their metadata are essentially flat file systems in which digital surrogates exist as individual items.⁷

These databases assume a one-to-one relationship between object and surrogate, such that “object-level” or “item-level” records represent individual artworks (in museums) or books (in libraries).

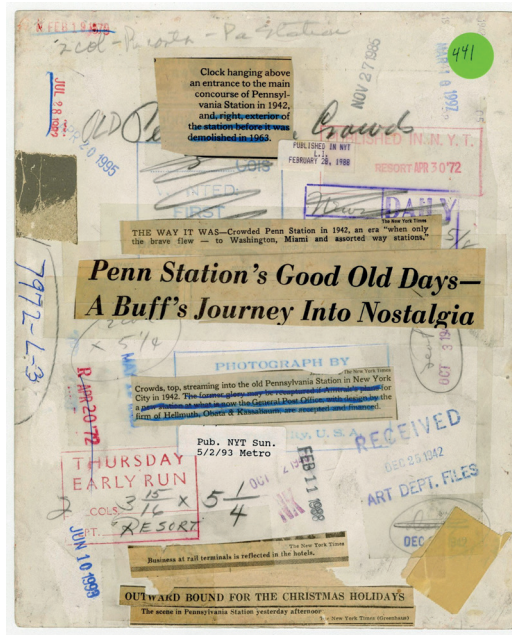
Archives as specific types of institutions, however, face further stumbling blocks in the path to digital access. Where museums are selective in their accessions, archives often acquire large collections of heterogeneous material. Where libraries benefit from the consistent information that accompanies published works, archival collections are unique. For these reasons, archival processing has historically been particularly time-consuming and labor-intensive, and the outpacing of processing by acquisition means that the majority of archives are besieged by backlogs of materials that are functionally invisible.⁸ Although decisions about what materials to process, and how to process them, have impacted what is visible to researchers since before the digital era, the advent of mass digitization has further compounded the problem. Unlike library and museum records, archival records are not typically at the item level but instead are structured as hierarchical documents called finding aids.⁹ First employed in the 1930s, a finding aid describes the physical organization of a collection or fonds, meaning an organically-created collection of materials related to a single person, family, or organization.¹⁰ Starting from the entire collection, the finding aid progresses from general to specific, descending through series (groups of related items) to files, such as folders and boxes.

While both finding aids and digital databases collect descriptive information about archival records, they are therefore structurally distinct.¹¹ Encoded Archival Description (EAD), the standard used to transform finding aids into digital documents, also maintains this hierarchy through tags in the markup tag language XML. As a result, digitized archival documents and digital archival systems are not automati-

cally interoperable, even when finding aids do descend to the item level. Transforming archival information from one information structure to the other requires additional time and labor that many backlogged archives do not have.

It is primarily in the context of archival backlogs that archives, along with other cultural heritage institutions, are currently exploring the potential value of artificial intelligence (AI). An umbrella term for technologies wherein computers perform tasks that have historically depended on human intelligence or intervention, artificial intelligence as discussed in archival contexts often centers on machine learning (ML) and deep learning. These disciplines utilize massive amounts of existing data to train algorithms to make predictions or classifications about previously unseen inputs. They are thus well-suited, at least ostensibly, to speeding up cataloging as an essentially classificatory activity. Thomas Paddilla usefully outlines some potential uses of ML in archives and similar institutions: “semantic metadata can be generated from video materials using computer vision, text material description can be enhanced via genre determination of full-text summarization using machine learning, and audio material description can be enhanced using speech-to-text transcription.”¹² These activities apply well-established machine learning technologies like image processing, natural language processing, and speech recognition to the data already captured in archival records.

Of these machine learning technologies, computer vision is especially relevant for photographic archives because it pertains to the analysis and interpretation of digital images. Computer vision (CV) is fundamentally concerned with recovering information about the three-dimensional world, including space and objects, from its two-dimensional representation in images.¹³ Research in this field dates to the 1960s, but a coincidence of technological developments in the early twenty-first century—including the availability of large datasets and



1 Back of a print of Alfred Eisenstadt, *Pennsylvania Station in 1943* from the *New York Times* Archival Library

more powerful computers—has precipitated significant advances. In 2012, widely considered a landmark year in computer vision research, researchers from the University of Toronto successfully used deep convolutional neural networks, commonly called CNNs, to classify images. This popular form of deep learning is inspired by the human visual cortex and consists of many (“deep”) interconnected layers that, when exposed to training data such as images, “learn” to detect significant patterns across the corpus.¹⁴ Sorting 1.2 million images into 1000 different classes, the Toronto team achieved an error rate of around sixteen percent, about ten percentage points better than prior methods.¹⁵

The development of deep CNNs precipitated a boom in computer vision research that, a little over a decade later, powers contemporary technologies that are close to becoming commonplace. The ability to “search by image” uses computer-identified patterns among images so

that users can find other images or information based on pixel data alone, without entering textual keywords. Photo management apps like iPhoto and commerce-oriented image platforms like Pinterest often combine this numerical, pixel-based comparison of images, known as unsupervised learning, with supervised learning tasks that connect data (often language) to pixel patterns. In these image-oriented applications, existing data is used to train algorithms to identify pre-selected classes or objects, like people or furniture, and generate data for keyword searches. But as art historian Amanda Wasielewski succinctly points out, “with most computer vision methods, the assumed area of investigation is not primarily the digital image itself but the digital photographic image that contains what is of interest.”¹⁶ In the example of self-driving cars, digital images in the form of video feeds provide data that, when analyzed by computer vision detection algorithms trained to see people and traffic, literally and figuratively drive decisions about the vehicle’s movement.¹⁷

Archivists and archive administrators are particularly interested in a branch of computer vision known as content-based image retrieval. When used to find digitized images in databases, computer vision functions as a kind of content-based *object* retrieval that provides access to materials via their digital surrogates. Some photo archives are employing the common computer vision task of text detection to capture existing cataloging information as structured metadata. In a 2018–2020 collaboration, the *New York Times* and Google targeted the millions of photographic prints and clippings in the newspaper’s archive, also known as its “photo morgue.” Using computer vision, they computationally analyzed and digitally extracted the typed and hand-written text captured in digital images of photograph backs, mounts, and folders, as well as the archive’s card catalog (fig. 1).¹⁸ They then used additional text-based machine learning methods like natural language processing to clean up the text

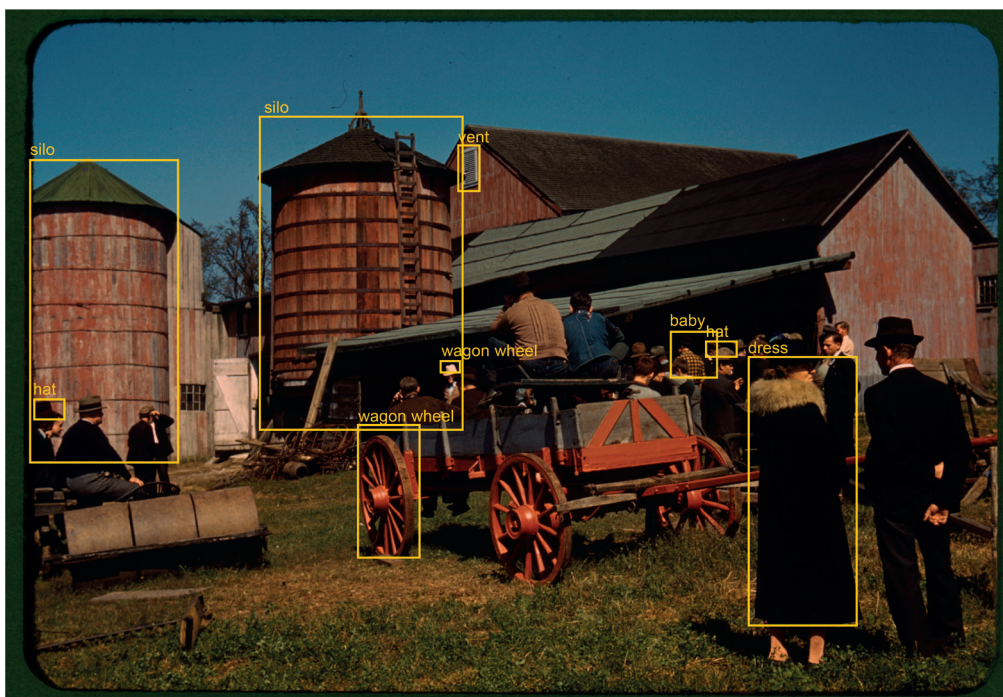
output and parse it into categories. Since then, other large-scale archival digitization initiatives have taken on a similar approach. PHAROS, an international consortium of art-historical photo archives, is using computer vision to analyze 1.5 million images of works of art collected by five different institutions, and a group of researchers at Northeastern University is currently tackling the approximately 1.65 million photographs the *Boston Globe* produced across the twentieth century.¹⁹

These projects recognize that many archival photographs essentially wear their metadata on their backs, mounts, and plastic sleeves. Photographs are often materially bound to archival cataloging information such as titles, photographers, and provenance that participated in and/or precipitated from the physical organization and use of the archive. Digitizing the back of a photograph, as well as the front, values the photograph not just as an image, but as an object, and more specifically as an archival object. It draws attention, as Costanza Caraffa has deftly argued, “to the operations conducted in the archive and the persons involved in them.”²⁰ These operations and people thereby persist as parts of a photograph’s history, of which the digitized record is only a further episode. On a practical level, this approach captures accrued textual information as essentially visual metadata that digital technology (including computer vision) can extract, structure, and make searchable. This workflow positions computer vision as internal, rather than external, to the existing lineage of archival practice.²¹

Other uses of computer vision in photo archives are intended to generate new metadata about archival objects. Carnegie Mellon University’s Computer-Aided Metadata Generation for Photoarchives Initiative (CAMPI) generated feature vectors of its digitized images—essentially lists of numbers that describe the pixel content of images—to help archivists organize uncataloged digitized photographs.²² Using these feature vec-

tors, archivists could quickly find and tag images depicting visually similar scenes. The Museum of Modern Art worked with Google to use a related workflow to match artworks pictured in its exhibition installation photos to images in its online collection, thereby linking not only two databases, but also two moments in the history of art documentation.²³

Image classification and object detection algorithms offer two different methods of generating metadata about an image’s subject matter. Object detection algorithms identify and locate things, including people, *within* images. For example, the EyCon (Early Conflict Photography 1890–1918 and Visual AI) Project is using object detection to identify and locate weapons, vehicles, and other entities relevant to conflict within its photographic corpus.²⁴ Depending on how an algorithm is trained, the objects available to be detected can vary, as demonstrated by the experimental interface created by the Access and Discovery of Documentary Images (ADDI) Project for selected photographic collections from the Library of Congress.²⁵ When applied to Jack Delano’s 1940 photograph of a farm auction in Derby, Connecticut, an algorithm trained on the popular COCO (Common Objects in Context) dataset primarily identifies people, while the algorithm trained on the LVIS (Large Vocabulary Instance Segmentation) dataset points out silos, hats, a baby, a dress, and a wagon wheel (fig. 2).²⁶ Image classification algorithms, meanwhile, aim to assign a label or category to an image as a whole. Incorporating image classifications generated by Google’s Vision API, the Getty Research Institute’s *12 Sunsets* interactive uses this categorical data to facilitate browsing of over 60,000 photographs of Los Angeles by the artist Ed Ruscha. Many of the classifications, such as airplane or fence, function similarly to object detection outputs in describing what is *in* an image. Other labels, as Nathaniel Dienes has pointed out, manifest an interesting if not always accurate relationship to form, such as when



2a – b Object detection results with COCO-trained algorithm (above) and LVIS-trained algorithm (below). Photograph: Jack Delano, *Farm Auction, Derby, Conn.*, 1939, color slide. Washington D.C., Library of Congress

sunlit curbs, blobs of masking fluid, and other swaths of white appear to produce an unexpectedly high occurrence of the label “snow” in the Ruscha’s images of Southern California.²⁷

Like object detection algorithms, different image classification algorithms can label images differently, depending both on the algorithms’ internal vocabularies and the visual qualities of images to which they are attuned. The Harvard Art Museum’s AI Explorer project exhibits this phenomenon by collecting classifications generated by several different commercial image analysis services—including Amazon’s AWS Rekognition, Google’s Vision API, Microsoft’s Azure AI Vision, and two image analysis companies, Clarifai and Imagga—for images of artworks in its collections.²⁸ These are multi-label classifications, meaning a service can assign more than one label to an image, and the classification lists considered in full often have significant conceptual overlap across services. Nonetheless, the precise terminology of each service and the labels it deems most probable often differ. In the output generated for a 1903 photograph (fig. 3) of a New York City immigration station, made by J. H. Adams for Harvard’s Social Museum Collection, the top Clarifai labels focus on the people in the image (“group,” “child”); the top Amazon labels mention what those people are wearing (“clothing,” “apparel”); and the top Google labels classify the image itself (“photograph,” “snapshot”).²⁹

Although the data generated by these computer vision processes may be new, it fits into and therefore easily augments existing archival practices. The concept of object detection could also be used to describe the act of identifying the “human beings, animals, plants, houses, tools, and so forth” that comprise Erwin Panofsky’s primary or “pre-iconographic” subject matter, a foundational reference point for image description guidelines in libraries and information science (LIS).³⁰ Image classification aligns with the sorting of photographs into descriptive cat-



3 J. H. Adams, *Races, Immigration: United States. New York. New York City. Immigrant Station: Regulation of Immigration at the Port of Entry. United States Immigrant Station, New York City: To Be Deported*, c. 1900, glossy collodion silver print, 21.1 × 18.4 cm. Cambridge, MA, Harvard Art Museums (transfer from the Carpenter Center for the Visual Arts, Social Museum Collection)

egories by assigning standardized vocabularies such as the Library of Congress Subject Headings. Both types of metadata creation involve interpretation, whether it is an algorithm or a human cataloger performing the task of decoding an image’s meaning and translating it into text.

Furthermore, the data generated by these computer vision algorithms maps onto the two main ways contemporary users browse digitized archival collections: searching and filtering. The ubiquitous search bars into which users type text-based queries, known as free text search, add objects to a list of results by matching the entered values across many fields of metadata. The ability to search images by content requires textual descriptions of that content that can be searched. Filtering, by contrast, removes or narrows results based on existing lists, keeping only objects that match the given criteria. Filters

operate on structured metadata that has been organized into categories. While the presence or absence of metadata affects whether archival objects are accessible, the presence or absence of specific terms through which that metadata is queried affect *how* those archival objects are accessed. The words used to describe an image are the words by which it can be found.

The value placed on subject matter in photographic archives gives credence to the value of computer vision as an aid to archival cataloging, and the established relevance of CV-generated data is part of its appeal. In a 2021 study by Anna Dahlgren and Karin Hansson, seventy-four percent of surveyed museum, archive, and library information specialists reported creating metadata about what an image depicts.³¹ Research to this point has at least provided proof of concept of computer vision workflows that can aid this research, whether by using text detection to capture existing cataloging information or using object detection and image classification to generate new, more specific data about subject matter. For digitized but unprocessed archival collections of archival material where everything—including photographs, drawings, but also digitized letters and other text documents—is technically a “photograph,” these CV tasks could also be employed to sort digitized images into materially-oriented categories that, with little effort, can drastically cut down on the number of records users need to wade through to get to the type of material they are seeking. Furthermore, unlike in physical archives, in which objects can only occupy one place in a classificatory schema, data about photographers can co-exist with data about image content or provenance. Given archival backlogs and the labor of transforming physical cataloging information into digital metadata, there is still a lot of opportunity in this regard that computer vision may help fulfill. Recalling the earlier comparison of the Frick and the New York Public Library and the capacity of different metadata schema to serve

different audiences, the variability and malleability of computer vision is potentially a feature rather than a bug.

At the very least, the sorts of classifications in the example of Adams’s photo may be sufficient to reach the baseline of image description achieved by some existing collections metadata. For example, in her 2022 analysis of metadata in a dataset of 30,000 digitized stereograph cards from the Library of Congress, digital humanist Zoe LeBlanc noted the prevalence of general subject terms about photographic format (“stereograph”), location (“United States”), and time (“history”) that did not actually convey much information about the image content.³² The recognition that human-created metadata is not always perfect or complete offers one response to concerns about the general accuracy of computer vision algorithms among LAM professionals as well as users.³³ Any implementation of this technology on a public-facing, institutional level will require the development of so-called “human-in-the-loop” workflows in which subject matter experts and archivists review generated data for accuracy and relevance.³⁴

While there are broad concerns about accuracy that apply to almost any use of computer vision as a technology of automation, there are other practical and historical considerations that apply to its use in photographic archives specifically. The example of Adams’s photograph also points in this direction: the generated description of this image as “a photograph of people wearing clothing” remains paltry in light of the photograph’s caption, which records that the people in the photograph are immigrants waiting to be deported. The next section addresses the use of computer vision in photographic archives, and particularly in photographic archives depicting people, in more detail.

Photographic Archives Against Computer Vision

Even if we recognize the foregoing opportunities, there are still good reasons to be hesitant about using computer vision in almost all contexts. Resonances between the present of computer vision and the history of photography, however, make its utilization in photographic archives conceptually as well as practically fraught.

In her summary of concerns about the use of computer vision in visual arts collections, Jessica Craig highlights three interrelated issues of algorithmic labor, transparency, and bias.³⁵ Regarding labor in archives, computer vision as a tool for automating archival processing can easily be construed as a threat to the livelihoods of archivists, even amid assurances that AI is meant to enhance rather than replace their efforts. But labor concerns are also inextricable from the technological development of CV, as the large datasets that have enabled recent research have largely been created by employing poorly-paid microworkers, often via Amazon's Mechanical Turk platform, to annotate images. It is not just the low wages and repetitive actions associated with this work that are dehumanizing; with their contributions to CV technology broadly unacknowledged in technical literature, these workers are abstracted away under the premises of "mechanical" objectivity.³⁶

Furthermore, there is a wealth of evidence that computer vision algorithms demonstrate significant biases, especially when directed at detecting and analyzing images of people.³⁷ In an influential 2018 paper, Joy Buolamwini and Timnit Gebru found that the over-representation of lighter-skinned people in facial recognition datasets resulted in facial analysis algorithms that performed significantly worse on images of non-white females.³⁸ This performance gap owes to a lack of diversity in the data used to train algorithms, which itself reflects a lack of diversity among the computer scientists who select the

training data. The phenomenon whereby "algorithms trained with biased data have resulted in algorithmic discrimination" is not isolated to computer vision, and it affects everything from search engine results to bank loan approvals to healthcare.³⁹ The specific shortcomings of computer vision with regard to race are especially dire when this technology is employed to help identify suspects or determine sentences within the criminal justice system.⁴⁰

In photographic archives, the high likelihood of people being depicted in processed images means there is a great potential for harm to depicted subjects and community users. As Craig points out, "The visual collections held by LAMs often have cultural, social, religious, and political significance," such that "the responsibility of LAMs to care for such objects extends beyond their physical or material state, to their digital representation and access."⁴¹ Much as existing societal biases have shaped datasets and the algorithms they train, applying these algorithms in photo archives could easily further perpetuate the oversights, occlusions, and legacies of erasure these collections already manifest.⁴² One example of an effort to mitigate these potential negative outcomes comes from the Teenie Harris Archive Project, an initiative to process and organize approximately 80,000 of the photographer's images of Black life in Pittsburgh. This project consciously combined commercial facial recognition software with input from representatives from the local Black community to facilitate the identification of Harris's sitters and subjects.⁴³

If, for photo historians, these issues around computer vision may sound vaguely familiar, one need only look at tendencies in photo archive metadata. Like microworking annotators, photographers have often been historically treated as mere operators. As existent metadata for historical photographs tends to favor their subject matter, it also tends to privilege the perspectives of the institutions that commissioned and col-

lected them. Substantial critiques by archivists and librarians have demonstrated the Western biases and widespread problematic terminology of commonly used metadata standards and vocabularies such as the Library of Congress's subject headings that make these standards inadequate for cataloging minority archives.⁴⁴

These resonances between computer vision and photo history are part of a larger, shared logic between these two technologies, especially where it concerns images of people. Computer vision research in the last decade has demonstrated an alarming tendency toward physiognomy, meaning the study of facial and bodily features to determine human characteristics, often along group lines. Despite being thoroughly discredited as a biased and baseless pseudoscience, physiognomy has regularly resurfaced in CV research that claims to predict not only race, gender, and sexuality, but also criminality, trustworthiness, and a host of other human characteristics by analyzing human faces.⁴⁵ In their thorough assessment of such "physiognomic artificial intelligence," Luke Stark and Jevan Hutson attribute its prevalence to the technology's conceptual underpinnings. Working from two-dimensional images, computer vision breaks the observable world into comparable units (binary code) and identifies patterns (salient features) within the data.⁴⁶ These patterns inform statistical inferences about the three-dimensional world. Stark and Hutson argue that this inferential aspect of CV, necessary to move from two to three dimensions, means that applications of the technology to the human body are inherently physiognomic. The basic task of facial recognition involves "the crudest possible type of physiognomic judgment": when a computer vision system infers the presence (or absence) of a face, it is "making a judgment about the 'face' as a characteristic of an image."⁴⁷

Although Stark and Hutson briefly mention the role of photography and photographers in establishing physiognomy as an organized disci-

pline in the nineteenth century, the significance of photography in their discussion of physiognomic AI remains largely implicit.⁴⁸ In photo history, however, the centrality of photography in earlier moments of physiognomy's flourishing is well-established. Allan Sekula's influential 1982 essay "The Body and the Archive" excavates the role of photography in advancing the paradigms of visual representation and interpretation that pervaded nineteenth-century culture.⁴⁹ Sekula singles out the anthropometric techniques of French police officer Alphonse Bertillon, inventor of the mug shot, and English anthropologist and eugenicist Francis Galton, inventor of composite portraiture. For Sekula, Bertillon's facial measurements and Galton's facial averages exemplify realist hermeneutic paradigms in inverted but complementary ways, with both using photographic representation as a repressive tool to define and delimit the normal from the abnormal, criminal, or otherwise "other."

As Christy Lange has observed, Sekula's text "eerily presages contemporary applications of facial recognition."⁵⁰ Already in the early 1980s, Sekula saw the continued presence of Bertillon and Galton in the national security state and discourse of biological determinism; today, they live on with renewed vigor in physiognomic AI. Artist Nancy Burson's computer-generated composite faces, which Sekula singles out for critique, are essentially "eigenfaces," representations of average faces essential to the development of facial recognition. Now, these faces can be found easily in computer vision tutorials and, on a larger scale, in the technology behind the augmented reality filters that populate social media apps like TikTok and Snapchat.

Looking at computer vision through Sekula's text frames this technology, like photography before, as inextricable from the archive. In the last few pages of "The Body and the Archive," Sekula links his account of the history of physiognomic photography to the emergence of the archive as "the dominant institutional basis for

photographic meaning” at the turn of the twentieth century.⁵¹ Between them, Bertillon and Galton mapped out the structuring premises and practices of photographic archives. Photography and text “[regulate] the semantic traffic in photographs,” working together to stabilize archives as seats of power by keeping each other in check.⁵² The paradigm of the photograph as a tool of reproduction enables the library, the archive, and the computer vision algorithm to function. At the same time, the paradigms of the library, needed to tame these “expansive and unruly collections of photographs,” end up inscribing and circumscribing photographic meaning.⁵³ In particular, Bertillon’s strategy of embedding photographs within numerical and textual records presaged the adoption of classificatory models from the concurrently emerging field of bibliographic science. The application of the numerical paradigm of the Dewey Decimal System to literature on photography cataloged prints by topic and classified the medium itself under the graphic arts.

Joan Schwartz, one of the most important figures in archival scholarship on photography, has argued that this uncritical use of bibliographic models for photographic materials extends to an ontological misunderstanding of photographs among archivists. The treatment of photographs as visual information, as demonstrated by the standards of image description discussed above, has precipitated a focus “on the factual content rather than the functional origins of visual images” and a prioritization of the single image over and outside of its context of creation and circulation.⁵⁴ This tendency stubbornly persists despite the work of archivists like Schwartz, Tim Schlak, and Terry Cook in pushing the field to absorb and adopt the postmodernist awareness of the malleability and instability of photographic meaning articulated by theorists like Sekula.⁵⁵ “The archivist thus finds herself in a particularly difficult position,” as Schlak puts it, not because of simple obstinance, but because

photography and archives have been historically constructed to be tautologically intertwined.⁵⁶ Digitization and the subsequent need for semantically structured metadata—driven by growing user demands for fully indexed, thickly cross-referenced item-level records—has only further tightened this knot.

What for Sekula is the “fundamental problem of archives, the problem of volume,” is the main problem that computer vision has been proposed to address.⁵⁷ But as Sekula’s text triangulates the related histories of photography, archives, and computer vision, it becomes clear that the issues around this technology are even bigger than massive backlogs. Using computer vision on photographic archives emphasizes the deep, historical continuity between these two technologies of understanding the world through images. For Sekula, photography is a “double system” in which its repressive function is the overlooked yet unavoidable pendant to its more lauded “honorific” function.⁵⁸ Every snapshot, in short, is shadowed by a mugshot. Does the “honorific” work of cataloging, of making photographs digitally discoverable, adequately balance the repressive function that computer vision has already taken on in contemporary society? In the final section, a case study on photographic research with person detection demonstrates how computer vision may be used to address archival absences not simply by filling them, but by rethinking the structures and practices behind them.

Person Detection in Photographic Archives

Based on the foregoing practical and conceptual concerns about computer vision, person detection may seem like an especially unlikely, even ill-advised application of this technology in photographic archives. In this section, my participation in the research project *Photography Unbound* frames a reevaluation of this technology’s

relevance for probing, mitigating, or bridging archival absences. When used to generate metadata about photographs, person detection enables archival discovery beyond its historically textual and classificatory limits.

Photography Unbound is an ongoing collaborative initiative, sponsored by the Getty Research Institute, that is designed to explore how computer vision can be used to advance research on early photography. The project's team—including myself, Emily Pugh, David Ogawa, Alexander Supartono, Zakaria Abou El Houda, and Alexandros Louizos—combines expertise in art history, photo history, digital humanities, and computer vision.⁵⁹ Focusing on nineteenth-century photography, loosely defined, the project responds to the relative paucity of research on computer vision as applied to photographs from this era, as compared to the twentieth or twenty-first century.⁶⁰ It works with a dataset of approximately 30,000 digitized photographs assembled from a range of institutions, including archives (Getty Research Institute, American Academy in Rome, Istituto Centrale per il Catalogo e la Documentazione); museums (J. Paul Getty Museum, National Gallery of Singapore); and libraries (Universitaire Biblioteken Leiden). This breadth of sources ensures our dataset reflects the realities of art- and photo-historical research as spread across different collections with differing metadata coverage, metadata standards, and digitization protocols.

An early step in the project was to brainstorm potential applications of computer vision that could aid and enhance existing research questions. As a team, we identified well-established tasks, like detecting objects; experimental tasks, like identifying vanishing points; and a range of ideas in between. Given the size of the initial dataset, we were conscious not to try too many different tasks at once because the sheer volume of data generated would be difficult to analyze. Instead, we wanted to focus on one or two tasks that could be expected to generate useful infor-

mation, based both on what is *possible* with contemporary CV and what is *relevant* to contemporary photo-historical research.

Person detection fulfilled both requirements. First, the status of person detection as a “solved” computer vision problem enabled us to focus on evaluating how standard or baseline algorithms, in our case YOLOv5, performed on digitized nineteenth-century photographs.⁶¹ Second, this task aligned with identified research questions in our group. In his research on photography in colonial Java, Supartono was studying how images of local religious sites by the indigenous photographer Kassian Cephas compared to images of the same places made by Europeans, such as the Dutch photographer Isidore van Kinsbergen and the British duo Woodbury and Page.⁶² Supartono hypothesized that where European photographers' archaeological impulses led them to avoid capturing people in their pictures, Cephas was more likely to include people, as places like the Buddhist temple at Borobodur were, for him, active rather than ancient sites of religious activity. This research led us to identify other photo-historical arguments about the inclusion of people, for example the common trope of using human figures for scale in architectural or archaeological images, that the generated data could be used to test and examine. Starting from the standpoint of photo-historical research, we did not orient our work toward archival search and discovery. The process of generating and working with the person detection data, however, suggested broader applications for this approach as a method of metadata generation.

In working with person detection, we consciously employ a pared-back approach designed to circumvent some of computer vision's most problematic applications. Facial recognition or gender detection systems take person detection to its endpoint: they involve detecting people, locating them in the image, and making predictions about them based on their isolated pixels. While Amalia Foka suggests these applications



4 Bounding box visualization for detected people. Photograph: Esteban Gonnet, *[Rancho]*, approximately 1866, albumen silver print. Los Angeles, Getty Research Institute

might be used with good intentions, for example to determine national identity to facilitate the analysis of colonialism, this practice accepts, rather than critiques, computer vision's physiognomic underpinnings.⁶³ In contrast to research on facial recognition in photographic archives, then, our pared-back method stops at the point of merely detecting people and does not try to identify them.⁶⁴ This approach avoids making predictions about people's character or identity and instead stresses the basic function of a person detection algorithm: returning a Boolean (true or false) value about whether people are present in an image. Starting from this point, data about the number and location of people in the image forges new pathways through photographic archives.

Foremost, the ability to search for images that contain any number of people, including zero, makes it possible to literally find photo-

graphs of absence. While this may sound simple, the premise of browsing photographs based on what they do *not* show runs completely counter to norms of image discovery. It challenges the foundational standards of archival image description, as discussed in section one, and their concern with what is present in an image. At the same time, it alters the way researchers find, and potentially interpret, historical photographs. Filtering a digitized collection to only show images without people makes absence a determining framework. Conversely, browsing photographs based on presence motivates attention to the people that appear, whether or not they are the intended subject of the image. As results of a search for people, even seemingly incidental figures take on new significance. A circa 1866 photograph taken in Buenos Aires by French photographer Esteban Gonnet, described only by its archive-created title of "Rancho," becomes a photograph



5 Bounding box visualization for detected people. Photograph: Photographer unknown, *Retiro Railway Station*, 1860s, albumen silver print. Los Angeles, Getty Research Institute

of the child and two women seated outside the thatched roof building (fig. 4). In the context of the album to which the photograph belongs, this photograph, and these people, take on a new relation to the sitters of the more classic portraits on other pages. In an era when the historical record is populated primarily by those in power, especially in colonial contexts, computer vision can contribute to historical research by highlighting people who, while present in the *photographic* record, have not registered in the *archival* record.

This use of computer vision as a tool for seeing people arguably works even when the algorithm fails to recognize all the people in an image, especially if one approaches the technology with skepticism rather than blind faith. Reviewing the algorithm's output for accuracy, either as an archivist correcting the data or as a researcher browsing the results, requires looking for people the algorithm may have missed. Seeing the

one person detected in an 1860s image from the same album captioned only as "Retiro Railway Station" calls the reviewer's attention to the presence of the photograph's other figures—the men standing outside the building's entrance, as well as laborers elsewhere around the station grounds—and to their absence from both the generated data and the existing institutional metadata (fig. 5).

In addition to impacting how we read individual photographs and albums, person detection as a source of metadata also carries historiographic weight as a response to past and current debates around photographic cataloging. Postmodern critics in the 1970s and 1980s took issue with art history's absorption of photography into its institutions and norms. For example, at the New York Public Library, librarian Julia van Haften collected nineteenth-century photographic books from across the institution

into a single division with art and prints. Critic and theorist Douglas Crimp lamented how van Haaften removed photographs from their subject-based classifications and reorganized them according to an art-historical model of authorship: “What was Egypt will become Beato, or du Camp, or Frith; Pre-Columbian Middle America will be Désiré Charnay; the American Civil War, Alexander Gardner and Timothy O’Sullivan; the cathedrals of France will be Henri LeSecq; the Swiss Alps, the Bisson Frères; the horse in motion is now Muybridge; the flight of birds, Marey; and the expression of emotions forgets Darwin to become Guillaume Duchêne de Boulogne [sic].”⁶⁵

In Crimp’s pre-digital moment of writing, these sorts of reclassifications mattered because, by literally taking photographs out of their previous contexts, the Library changed both the information pathways through which photographs could be found and the discursive frameworks through which photographic meaning was presented. Similarly, Rosalind Krauss, Christopher Phillips, and Joel Snyder critiqued the assimilation of nineteenth-century photographs into art-historical genres.⁶⁶ In particular, the modernist celebration of survey photographs reconceived images by figures like Timothy O’Sullivan as landscapes—a term inextricable from the history of painting—rather than views, which Krauss argues was the preferred photographic term. This reclassification also shifted the terms of photographic authorship by repositioning the previously overlooked camera operator into the primary determinant of photographic value. Krauss explicitly recognizes the implications of these changes when she connects her argument to “the possibility of storing and cross-referencing bits of information and of collating them through the particular grid of a system of knowledge” evidenced by stereo cabinets and, on a larger scale, public libraries.⁶⁷

When we consider that the mere concepts of “portrait” and “landscape” hinge in some basic

way on whether images do or do not focus on people, respectively, person detection suggests a compelling alternative. Filtering on either genre would separate studio portraits, on the one hand, and natural vistas with people in the distance, on the other. By contrast, filtering for images with two people presents the results as continuous and interrelated. Such a search therefore destabilizes genre classifications as inherited from art history, and in particular Western art history. In colonial Southeast Asia, for example, “landscapes” and “portraits” served the aims of European colonizers. As curator Charmaine Toh explains, the picturesque quality of Charles McWhirter Mercer’s views communicates “the civilizing benefits of colonialism,” while studio portraits of people of various races by G. R. Lambert & Co. cast their sitters as representative types instead of named individuals.⁶⁸ Being context-dependent, the labeling of a photograph as a “portrait” or a “type” may be best left to subject matter experts. Meanwhile, by surfacing images that merely contain people, person detection makes photographs accessible outside of these categories.

By circumventing genres, person detection favors the “shadow archive” Sekula posits as existing across, and binding together, discrete disciplinary archives. Although their subjects occupy very different positions within the terrain of nineteenth-century society, a sitting portrait of an aristocrat and a front-facing mugshot of an accused criminal are equivalent when judged on the number of people they depict. Surfacing this equivalence is especially valuable in those portions of the photographic record now “archived” by art museums. In the collection of the J. Paul Getty Museum, an 1886 photograph of “Horse Thief” George E. Swain is absorbed into the realm of art when it is described as a “portrait of a man with short hair and stubble on his face” (fig. 6).⁶⁹ Cataloging this object as a photograph of one person avoids this art-historical framing while still demonstrating, in the form of search



6 Unknown maker, American, [George E. Swain alias Sargeant, Horse Thief], June 1886, albumen silver print, 9,5 × 6,7 cm. Los Angeles, J. Paul Getty Museum

results, the inseparability of this image from the thousands of cabinet cards and other bourgeois portraits also owned by the museum.

Along with these past debates, person detection as a source of metadata also bears on current movements in archives, museums, and libraries toward reparative archival description and inclusive cataloging.⁷⁰ Reparative description involves identifying outdated, offensive, or otherwise problematic language in metadata. By contextualizing or substituting this language with more accurate and inclusive terms, reparative description aims to prevent harm to marginalized people and communities, as well as promote access to archival materials. Inclusive cataloging also enhances discoverability by introducing more specific terms and alternative

information schema that incorporate the voices and perspectives of underrepresented groups. The Homosaurus, for example, is a linked data vocabulary of LGBTQ+ terms that responds to the paucity of relevant Library of Congress subject headings. As applied to the Digital Transgender Archive, a centralized resource for pre-2000 digital archival material about transgender history, this vocabulary organizes photographs, articles, and ephemera according to the values and identities of the people and communities they document.⁷¹

The important work behind this and other inclusive and reparative cataloging efforts is not a simple matter of find-and-replace. Extensive and often collaborative labor is required to construct vocabularies, research archival objects, and ensure that the proper terms are applied. As part of a larger movement to decolonize archives, reparative description is often restorative description, in that it adds back into the archive information about the culture and contributions of colonized peoples that was previously, often purposely, left out or erased.⁷² Person detection should by no means be used as a substitute for inclusive cataloging, but it can complement these efforts as a generalizing counterpoint that enhances discoverability. On a practical level, the tendency toward specificity in inclusive cataloging necessarily entails the expansion of how many categories users need to navigate. For users specifically interested in finding photographs of people, using person detection data as a first step in browsing may expedite searches by reducing subject terms to shorter, more legible lists, potentially making more specific terms easier to find. Moreover, as institutions pursue the work of inclusive cataloging, person detection data that is comparably easy to generate can ensure that depicted people described as “natives” or “slaves” are also described as people.

Especially when viewed through the calls for social justice and decolonization that have motivated practices like reparative and inclu-

sive cataloging, what I have presented here as a relatively simple mechanism—determining the absence or presence of people—might look like a dangerous concentration of power in both archives and technology to bestow “humanity” upon photographed figures. In cases where pictured people’s humanity has already been threatened by the use of photography, the systems of archives, or both, the stakes of false negatives in person detection are arguably quite high. As with all uses of computer vision, this task should be pursued with transparency and accountability.⁷³ The decision to implement this technology needs to be accompanied by concrete plans for how machine-generated data will be reviewed and evaluated and how it will be identified in collection interfaces, among other considerations.

Yet even with these strategies, does person detection and its assertions of an encompassing “humanity” not support the kind of universalizing humanism Sekula critiqued in propagandistic endeavors like Edward Steichen’s *Family of Man*? Is archival discoverability, with all its colonial connotations, not of a piece with the liberal values of transparency and accessibility to which photography has also been yoked? Without entirely discounting this suggestion, we may look as a counterpoint at *The Real Face of White Australia*, a project by Tim Sherratt and Kate Bagnall using facial recognition on immigration records held in the National Archives of Australia.⁷⁴ These late nineteenth- and early twentieth-century records date to a period when the country enacted discriminatory “White Australia” policies in an effort to limit non-European immigrants. After extracting faces from the Bertillon-like documents, Sherratt and Bagnall created a website that displays the faces at different sizes in a masonry-style layout. Disrupting the regularized grid and the homogenous image of a “White Australia” that the country’s government sought to project, the accumulation

of people of color on the website provides a more accurate visualization of the Australian population in the first half of the twentieth century. Linking each image to the larger digitized document to which it belongs, the site links a history of discrimination to the individuals it affected. This project can be read as “counter-forensics,” a concept Sekula articulated elsewhere to describe a practice of political resistance that seeks to restore names and identities to the people of “a buried archive.”⁷⁵

The approach to person detection I describe in this article does not, and purposely so, attempt to restore names and identities to the people it detects. Rather, it seeks to give presences to absences in the historical photographic record by bringing people to the fore. Person detection may not necessarily fulfill Schwartz’s call to focus on photographs’ “functional origins,” but it also does not occlude them in the ways that practices adopted from library science and art history have.⁷⁶ Existing alongside extant descriptions and categories, this data also cuts across them, attacking their non-coherence and discontinuity from within. This use of computer vision as a tool of metadata generation facilitates search that can support research, but it purposely leaves that research to the scholars, archivists, and other people in the loop.

TRACY STUBER is the Digital Humanities Specialist for Arts and Humanities Research Computing at Harvard University. Previously, she was Research Specialist in Digital Art History at the Getty Research Institute, where she focused on applications of emerging technologies such as computer vision in photographic archives. Her research addresses the institutionalization of photographic history in US-American arts institutions in the 1970s and its ramifications for contemporary digital collections. She received her Ph.D. in Visual and Cultural Studies from the University of Rochester.

- 1 Archive, in: Wikipedia, URL: <https://en.wikipedia.org/wiki/Archive> (last accessed 15 June 2024).
- 2 Murtha Baca, Introduction, in: eadem (ed.), *Introduction to Metadata*, Los Angeles 2016, URL: <http://www.getty.edu/publications/intrometadata/introduction> (last accessed 15 June 2024).
- 3 Thomas Padilla, *Responsible Operations: Data Science, Machine Learning, and AI in Libraries* (OCLC Research), Dublin 2019, 6–22, DOI: <https://doi.org/10.25333/xk7z-9g97> (last accessed 15 June 2024).
- 4 For an overview of the archival turn, see Marlene Manhoff, Theories of the Archive from Across the Disciplines, in: *Portal: Libraries and the Academy* 4, 2004, no. 1, 9–25.
- 5 Libraries, archives, and museums use different—and multiple—metadata standards. Some common ones include MARC (Machine-Readable Cataloging) Format, in libraries; Encoded Archival Description (EAD), in archives; and Dublin Core and Visual Resources Association (VRA) Core, among others, in museums. Cf. Anne J. Gilliland, Setting the Stage, in: Baca 2016 (as in note 2), URL: <http://www.getty.edu/publications/intrometadata/setting-the-stage> (last accessed 15 June 2024).
- 6 Diana Kamin, *Picture-Work: How Libraries, Museums, and Stock Agencies Launched a New Image Economy*, Cambridge, MA 2023, DOI: <https://doi.org/10.7551/mitpress/14086.001.0001> (last accessed 27 February 2025).
- 7 The exception to this structure is linked data, which is built as a web or network. For an overview of linked data, see Edie Davis and Bahareh Heravi, Linked Data and Cultural Heritage: A Systematic Review of Participation, Collaboration, and Motivation, in: *Journal of Computing and Cultural Heritage* 14, 2021, no. 2, 1–18, DOI: <https://doi.org/10.1145/3429458> (last accessed 27 February 2025).
- 8 Mark Greene and Dennis Meissner, More Product, Less Process: Revamping Traditional Archival Processing, in: *American Archivist* 68, 2005, no. 2, 208–263.
- 9 Gregory Wiedeman, The Historical Hazards of Finding Aids, in: *The American Archivist* 82, 2019, no. 2, 381–420.
- 10 Fonds, in: *Dictionary of Archives Terminology*, URL: <http://dictionary.archivists.org/entry/fonds.html> (last accessed 15 June 2024).
- 11 Jane Zhang and Dayne Mauney, When Archival Description Meets Digital Object Metadata: A Typological Study of Digital Archival Representation, in: *The American Archivist* 76, 2013, no. 1, 174–195.
- 12 Padilla 2019 (as in note 3), 12.
- 13 Richard Szeliski, *Computer Vision: Algorithms and Applications*, New York 2022.
- 14 Laith Aluzbaidi et al., Review of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, Future Directions, in: *Journal of Big Data* 8, 2021, DOI: <https://doi.org/10.1186/s40537-021-00444-8> (last accessed 15 June 2024).
- 15 The Revolutionary Technique That Quietly Changed Machine Vision Forever, in: *MIT Technology Review*, 9 September 2014, URL: <https://www.technologyreview.com/2014/09/09/171446/the-revolutionary-technique-that-quietly-changed-machine-vision-forever> (last accessed 15 June 2024).
- 16 Amanda Wasielewski, *Computational Formalism: Art History and Machine Learning*, Cambridge, MA 2023, 5.
- 17 What is Computer Vision?, IBM, 27 July 2021, URL: <https://www.ibm.com/topics/computer-vision> (last accessed 15 June 2024).
- 18 The NYT Open Team, From Print to Digital: Making over a Million Archived Photos Searchable, in: *NYT Open*, 2020, URL: <https://open.nytimes.com/from-print-to-digital-making-over-a-million-archived-photos-searchable-b146037705d2> (last accessed 15 June 2024).
- 19 PHAROS: The International Consortium of Photo Archives, URL: <http://pharosartresearch.org> (last accessed 15 June 2024); The Boston Globe Collection Description, Northeastern University, 2024, URL: <http://globe.library.northeastern.edu/collection-description> (last accessed 15 June 2024).
- 20 Costanza Caraffa, From “Photo Libraries” to “Photo Archives”: On the Epistemological Potential of Art-Historical Photo Collections, in: eadem (ed.), *Photo Archives and the Photographic Memory of Art History*, Berlin and Munich 2011, 11–46, here 12.
- 21 Eadem et al., PHAROS: A Digital Research Space for Photo Archives, in: *Art Libraries Journal* 45, 2020, 2–11.
- 22 Matthew Lincoln et al., CAMPI: Computer-Aided Metadata Generation for Photo Archives Initiative, Carnegie Mellon University, preprint, 2020, DOI: <https://doi.org/10.1184/R1/12791807.v2> (last accessed 15 June 2024).
- 23 Identifying Art through Machine Learning, Museum of Modern Art, URL: <https://www.moma.org/calendar/exhibitions/history/identifying-art> (last accessed 15 June 2024).
- 24 Katherine Aske and Marina Giardinetti, (Mis)Matching Metadata: Improving Accessibility in Digital Visual Archives through the EyCon Project, in: *ACM Journal of Computing in Cultural Heritage* 16, 2023, DOI: <https://doi.org/10.1145/359472> (last accessed 15 June 2024).
- 25 Taylor Arnold and Lauren Tilton, Method Paper: Access & Discovery of Documentary Images (ADDI), white paper, 2022, URL: https://raw.githubusercontent.com/distant-viewing/addi/main/o3_methods_paper/methods_paper.pdf (last accessed 15 June 2024).

- 26 *Farm Auction, Derby, Conn.*, Access and Discovery of Documentary Images, URL: https://dis-tant-viewing.github.io/addi/o6_interactive_viz/build/?id=2017877382 (last accessed 15 June 2024).
- 27 Nathaniel Deines, Does it Snow in L.A., in: *Getty News & Stories*, 7 October 2020, URL: <http://www.getty.edu/news/does-it-snow-in-la> (last accessed 15 June 2024).
- 28 AI Explorer, Harvard Art Museums, URL: <http://ai.harvardartmuseums.org> (last accessed 15 June 2024).
- 29 The differences are also not rigid: Google, for example, does not label every photograph in Harvard's collections as a "photograph."
- 30 Erwin Panofsky, *Studies in Iconology: Humanistic Themes in the Art of the Renaissance*, New York 1939, 5; Megan Winget, Describing Art: An Alternative Approach to Subject Access and Interpretation, in: *Journal of Documentation* 65, 2009, no. 6, 958–976, here 961.
- 31 Anna Dahlgren and Karin Hansson, What an Image Is: The Ontological Gap between Researchers and Information Professionals, in: *Art Documentation* 40, 2021, no. 1, 21–32, here 25.
- 32 Zoe Leblanc, *Subject Unknown: Stereograph Cards, Historical Metadata, and Feedback for Data Jam 2022* (Computing Cultural Heritage in the Cloud Data Jam), Library of Congress Labs, 25 October 2022, URL: <https://www.loc.gov/item/2024698320>, 40:54–41:30 (last accessed 15 June 2024).
- 33 Padilla 2019 (as in note 3); Tracy Stuber et al., Interfacing in the Archive: Making Online Collections Work with and for Digital Humanities Research, in: Isabel Galina Russell and Glen Layne-Worthey (eds.), *The Routledge Companion to Libraries, Archives, and the Digital Humanities*, New York 2025 (forthcoming).
- 34 For one example of such a workflow, see Vikram Mohanty et al., Second Opinion: Supporting Last-Mile Person Identification with Crowdsourcing and Face Recognition, in: *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing* 7, 2019, 86–96, DOI: <https://doi.org/10.1609/hcomp.v7i1.5272> (last accessed 15 June 2024).
- 35 Jessica Craig, Computer Vision for Visual Arts Collections: Looking at Algorithmic Bias, Transparency, and Labor, in: *Art Documentation* 40, 2021, no. 2, 198–208.
- 36 Nicolas Malevé, On the Data Set's Ruins, in: *AI & Society* 36, 2021, 1117–1131, here 1127, DOI: <https://doi.org/10.1007/s00146-020-01093-w> (last accessed 27 February 2025).
- 37 Alexander Monea, Race and Computer Vision, in: Andreas Sudmann (ed.), *The Democratization of Artificial Intelligence: Net Politics in the Era of Learning Algorithms*, Bielefeld 2019, 189–208.
- 38 Joy Buolamwini and Timnit Gebru, Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification, in: *Proceedings of Machine Learning Research* 81, 2018, 1–15, URL: <https://proceedings.mlr.press/v81/buolamwini18a.html> (last accessed 27 February 2025).
- 39 Ibid., 1; Safiya Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism*, New York 2018.
- 40 Sidney Perkowitz, The Bias in the Machine: Facial Recognition Technology and Racial Disparities, in: *MIT Case Studies in Social and Ethical Responsibilities of Computing*, 6 February 2021, DOI: <https://doi.org/10.21428/2c646de5.62272586> (last accessed 27 February 2025).
- 41 Craig 2021 (as in note 35), 202–203.
- 42 On "legacies of erasure," see Safiya Noble, Decolonizing Standards, in: *Information Services and Use* 43, 2023, 327–333, DOI: <https://doi.org/10.3233/ISU-230214> (last accessed 15 June 2024).
- 43 Andy Ptashinski, CMU Wins NEH Grant for Advanced Computer Analysis of Teenie Harris Archive, in: Carnegie Mellon University News, Pittsburgh 2017, URL: <https://www.cmu.edu/news/stories/archives/2017/september/neh-grant.html> (last accessed 15 June 2024).
- 44 For a useful summary, see Kate Holterhoff, From Disclaimer to Critique: Race and the Digital Image Archivist, in: *Digital Humanities Quarterly* 11, 2017, no. 3, URL: <http://www.digitalhumanities.org/dhq/vol/11/3/000324/000324.html> (last accessed 15 June 2024).
- 45 Luke Stark and Jevan Hutson, Physiognomic Artificial Intelligence, in: *Fordham Intellectual Property, Media, and Entertainment Law Journal* 32, 2022, no. 4, 922–978.
- 46 Ibid., 941.
- 47 Ibid., 942–943.
- 48 Ibid., 938.
- 49 Allan Sekula, The Body and the Archive, in: *October* 39, 1986, 3–64.
- 50 Christy Lange, Surveillance, Bias and Control in the Age of Facial Recognition Software, in: *Frieze* 2018, no. 196, URL: <http://www.frieze.com/article/surveillance-bias-and-control-age-facial-recognition-software> (last accessed 15 June 2024).
- 51 Sekula 1986 (as in note 49), 56.
- 52 Ibid., 55.
- 53 Ibid., 56.
- 54 Joan Schwartz, Coming to Terms with Photographs: Descriptive Standards, Linguistic "Othering," and the Margins of Archivry, in: *Archivaria* 54, 2002, 142–171, here 144.
- 55 Tim Schlak, Framing Photographs, Denying Archives: The Difficulty of Focusing on Archival Photographs, in: *Archival Science* 8, 2008, no. 2, 85–101, here 95.

- 56 Ibid., 91.
- 57 Sekula 1986 (as in note 49), 46.
- 58 Ibid., 3.
- 59 Emily Pugh is Principal Research Specialist in Digital Art History at the Getty Research Institute; David Ogawa is Associate Professor of Art History and Department Chair at Union College; Alexander Supar-tono is Lecturer in the School of Arts and Creative Industries at Edinburgh-Napier University; Alexan-dros Louizos is Deep Learning Engineer at Quantiva; and Zakaria Abou El Houda is Chief Technology Of-fice at Quantiva.
- 60 This imbalance may be because twentieth-century gelatin silver and chromogenic prints are more simi-lar in both form and content to the digital, full-color 21st-century photographs that comprise most CV training.
- 61 Haiying Liu et al., SF-YOLOv5: A Lightweight Small Object Detection Algorithm Based on Improved Fea-ture Fusion Mode, in: *Sensors* 22, 2022, DOI: <https://doi.org/10.3390/s22155817> (last accessed 15 June 2024).
- 62 Charmaine Toh, Living Photographs, in: eadem (ed.), *Living Pictures: Photography in Southeast Asia*, Singa-pore 2022, 8–13, here 10.
- 63 Amalia Foka, Computer Vision Applications for Art History, in: *Proceedings of EVA*, London 2021, 73–80, here 76, DOI: <http://dx.doi.org/10.14236/ewic/EVA2021.12> (last accessed 15 June 2024).
- 64 Kenzo Milleville et al., Enriching Image Archives via Facial Recognition, in: *Journal on Computing and Cultural Heritage* 16, 2024, no. 4, 1–18, DOI: <http://doi.org/10.1145/3606704> (last accessed 15 June 2024).
- 65 Douglas Crimp, The Museum's Old/The Library's New Subject, in: Richard Bolton (ed.), *The Contest of Meaning: Critical Histories of Photography*, Cam-bridge, MA 1989, 3–14, here 7.
- 66 Joel Snyder, *American Frontiers: The Photographs of Timothy H. O'Sullivan, 1867–1874* (exh. cat. Philadel-phia, PA, Philadelphia Museum of Art), NY 1981; Rosa-lind Krauss, Photography's Discursive Spaces: Land-scape/View, in: *Art Journal* 42, 1982, no. 4, 311–319; Christopher Phillips, The Judgment Seat of Photogra-phy, in: *October* 22, 1982, 27–63.
- 67 Krauss 1982 (as in note 66), 314.
- 68 Toh 2022 (as in note 62), 8–13.
- 69 Object description, [*George E. Swain alias Sargeant, Horse Thief*], Los Angeles, J. Paul Getty Museum, URL: <https://www.getty.edu/art/collection/object/107No9> (last accessed 15 June 2024).
- 70 For example, Samantha Ceja, Sara McGillivray, and Helen Kim, How Getty Archivists Support Racial Jus-tice, in: *Getty News & Stories*, 16 June 2021, URL: <https://www.getty.edu/news/getty-archivists-support-racial-justice> (last accessed 15 June 2024); List of Statements on Bias in Library and Archives Description, in: *The Cataloging Lab*, last updated December 2024, URL: <http://cataloginglab.org/list-of-statements-on-bias-in-library-and-archives-description> (last accessed 15 June 2024).
- 71 Marika Cifor and K. J. Rawson, Mediating Queer and Trans Pasts: The Homosaurus as Queer Information Activism, in: *Information, Communication & Soci-ety* 26, 2023, no. 11, 2168–2185, DOI: <https://doi.org/10.1080/1369118X.2022.2072753/> (last accessed 15 June 2024).
- 72 Abeba Birhane and Zeerak Talat, It's Incomprehen-sible: On Machine Learning and Decoloniality, in: Simon Lindgren (ed.), *Handbook of Critical Studies of Artificial Intelligence*, Northampton 2023, 128–140, here 131.
- 73 Craig 2021 (as in note 35), 203.
- 74 Tim Sherratt and Kate Bagnall, The People Inside, in: Kevin Kee and Timothy Compeau (eds.), *Seeing the Past with Computers: Experiments with Augmented Reality and Computer Vision for History*, Ann Ar-bor 2019, 11–31, DOI: <https://dx.doi.org/10.3998/mpub.9964786> (last accessed 28 February 2025).
- 75 Allan Sekula, Photography and the Limits of Na-tional Identity, in: *Grey Room* 55, 2004, 28–33, here 30, 32 (originally published in *Culturefront* 2, 1993, 54–55).
- 76 Schwartz 2002 (as in note 54), 144.

Photo Credits: 1 Courtesy of *The New York Times*, New York. — 2 Library of Congress, Access & Discovery of Docu-mentary Images (ADDI), Washington, D.C. — 3 © President and Fellows of Harvard College Harvard Art Museums, Cambridge, MA. — 4, 5 Getty Research Institute, Los Angeles. — 6 J. Paul Getty Museum, Los Angeles.