



Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Pfefferkorn, J

Title:

Montage, Memories, Machine: Seeing Photographically in Motion

Date:

2024

Citation:

Pfefferkorn, J. (2024). Montage, Memories, Machine: Seeing Photographically in Motion. *Media Theory*, 8 (1), pp.133-158. <https://doi.org/10.70064/mt.v8i1.1071>.

Persistent Link:

<https://hdl.handle.net/11343/355982>

License:

[CC BY](#)

Montage, Memories, Machine: Seeing Photographically in Motion

JASMIN PFEFFERKORN

The University of Melbourne, AUSTRALIA

Media Theory
Vol. 8 | No. 1 | 133-158

© The Author(s) 2024

CC-BY

<http://mediatheoryjournal.org/>

Abstract

This paper explores the entanglement of human perception and machine operations in contemporary photographic culture, using montage as a conceptual resource to interrogate our understanding of photography in the context of computational vision and algorithmic culture. It takes as a key example the ‘Memories’ feature of the Apple Photos application, in which ‘For You’ albums organize and (re)present images from a personal camera roll back to the viewer-user as a slideshow montage. Through a combination of technical analysis and auto-ethnography, it argues that the relation between stillness and motion in this instantiation of montage offers a productive way of understanding what it means to ‘see photographically’ within the contemporary socio-technical assemblages that comprise digital photography.

Keywords

Algorithmic curation, Computational photography, Montage, Archive

Stillness and motion are characteristics that have long held a special relationship in the history of photographic images. For Jean-Luc Nancy (2005), this inherent duality is a characteristic of the image. He writes:

The image contains the index of its frozenness (its form, its present, its representation) and at the same time the index of movement (force, appearing/disappearing). That is also why it engages both the indefinite proliferation of images as well as each image’s isolation and enframing, its being hung on the wall (98).

As Peter Osborne (2019) notes, much has already been written about the relationship between movement and stasis in relation to the photographic image. He asks: “How to say something new about the place of stasis within the temporalities of image and act?” (125). Yet in an era of machine vision and the algorithmic curation of images, the orientation of photography towards movement has arguably increased. For this reason, it is worth revisiting the relationship between motion and stillness in photographic images. In what follows, I mobilize the concept of montage as a way of exploring this relation.

Montage is a form of collecting (or selecting), organizing, and displaying images in a sequence, or flow, whereby “differences form configurations” and “dissimilarities, together, create unperceived orders of coherence” (Didi-Huberman in Bénichou, 2003: 179). By positioning images in relation to one another, montage contributes to our understanding of images as being ‘in motion’ both physically and conceptually. The practice of montage has been mobilized as a way of exploring and renegotiating narrative conventions. This occurs on a spectrum from linear narratives to fractured narratives. The term ‘montage’ is a deliberate invocation through this paper. It speaks to a body of literature that assesses the relationship between images and the *construction* of reality, as opposed to images as the *registration* of reality. “Montage,” Sam Rohdie writes, “rather than being unnatural and unrealistic, can be a means to help see and discover reality” (2015: 146). To ‘see photographically’ is to take into account the ways in which we perceive and construct the world through images. To ‘see photographically in motion’ is to place emphasis on the role of image sequencing in this perception and construction. Taking montage as a form, Rohdie’s sentiment speaks to how montage makes visible the contemporary experience of technological intensity. Taking montage as enacted through algorithmic processes, his sentiment underscores the role that computation now plays in shaping how we make meaning in the world through images.

Today, most photographic images are captured and stored on networked mobile devices that utilize both computer vision and machine learning. This has led to software like the Apple iPhone’s ‘For You’ albums within its ‘Memories’ feature, which serves as the key case study in this paper. Apple ‘Memories’ can be understood as a complex socio-technical ensemble that embeds a variety of technical processes

mobilizing personal image archives in distinct ways. Seeing photographically in motion, through the personal photographic archive, predates the computational. Batchen (2004) writes that when we “touch an album and turn its pages, we put the photograph in motion, literally in an arc through space and metaphorically in a sequential narrative” (49). What happens when the ‘we’ is not human, but predominantly an algorithmic process putting the photograph in motion? How does narrative – what John Berger (1982: 89) calls lending the photograph a ‘before’ and an ‘after’ – emerge in this paradigm? My aim is to explore ‘Memories’ to understand how the computational orients us towards seeing photographically in motion by drawing on montage as a useful lens for understanding motion through image sequencing and composition. This in turn offers insight into the role of both human and machine in creating narrative connections through personal photographs. At times, I utilize auto-ethnographic methods, exploring my own experiences of algorithmically curated albums to highlight some of the tensions that emerge between the algorithmic logic and affective logics of these sequenced images.

Introducing ‘Memories’

iPhone ‘Memories’ was released with iOS 10 in 2016 as part of the Apple Photos application. By scanning an iPhone user’s archive and applying a variety of algorithmic operations to categorize and select images, ‘Memories’ (re)presents photos to the user in the form of a slideshow arranged around a theme and set to music. These slideshows can be located within the Photos application by selecting the option titled ‘For You’.¹ They are automated montages combining images from a user’s camera roll, curated by an algorithm that utilizes various points of metadata (e.g. location) as well as machine vision (e.g. facial recognition). ‘Automation’ here is relative – while there is always a default selection offered, users can intervene in the creation of ‘Memories’ by making their own selections. Even with user intervention, ‘Memories’ remains underpinned by algorithmic processes that have key parameters to make connections between images within photographs in the Photos app. These technical operations can be contextualized through the concept of the ‘networked image’. “[T]he networked image is data,” Daniel Rubinstein and Katrina Sluis write, “visual information to be analysed and remapped to new contexts via algorithms” (2019: 359). The ‘Memories’ feature

algorithmically locates patterns, combining and presenting these as a slideshow to be experienced by the viewer as a moving sequence of images across a duration of one to two minutes. The serialization of images through algorithmically curated photo albums – understood here as automated narrative montages – orients various processes of selection, organization and display towards motion.

The technical operations underpinning ‘Memories’ and ‘For You’ albums have progressively increased alongside developments in machine learning and computer vision algorithms. Machine Learning (ML) models on the iPhone are integrated through Core ML. ML algorithms are applied to training data, and then used to make predictions when given new input data. Core ML was made possible by the development of the Apple Neural Engine, which enhanced processing power to run machine learning models on-device.² Gabriel Pereira (2019) takes up this development from a materialist perspective, looking at how a hardware accelerator, known as a Neural Processing Unit (NPU), enabled the speed, efficiency, and reduction in heat-generation necessary for the ‘matrix multiplications’ of machine learning applications like ‘Memories’. Matrix multiplications are a foundational element of machine learning, as they are the part of the algorithmic process that applies weight to features (assigning value to data) and calculates predictions. As Pereira notes, “Predictive modelling thus becomes embedded in the chip design itself” (2019, *Chip-engineering and the embedding of prediction*, para 3).

Core ML supports the functionality of Apple’s Vision framework, the domain-specific framework for analysing images – for instance, object recognition based on an image’s pixels and the subsequent classification of that object. Apple released their Vision framework as part of iOS 11 in 2017, a collection of computer vision algorithms that enabled more complex image recognition, image classification, and object detection.³ These algorithmic collections are exemplified by the Apple Neural Scene Analyzer (ANSA), which introduces the more detailed image analysis that forms the basis of the ‘Memories’ feature and thus ‘For You’ album slideshows.⁴ It includes scene tagging, semantic prints, utility content filters, sensitive filtering, saliency objectness, saliency attention, object detection, and perceptual prints.⁵ A series of processes occur in the tens of milliseconds it takes to perform automated image tagging in Apple Photos, which then contribute to the selection and organisation of ‘For You’ albums.

Taxonomic image tagging is a common component of training for machine learning and forms the baseline for the Vision framework. ANSA builds on this baseline. The signals produced by ANSA provide ‘Memories’ with its organizing logics and are seen as “critical to how users interface with the photos on their devices” (Apple, 2022). While it is impossible to detail the 16 million parameters that make up ANSA, I will draw out some of the key functionalities and technical operations to elucidate the algorithmic logics of computational montage in ‘For You’ albums and ‘Memories’.⁶

Predictive modelling is made manifest in two key ways within Apple ‘Memories’. First, through automating classification of detected content (e.g. object recognition). Second, through predicting affective connections that can be re-presented to the user in a presumably meaningful way (e.g. ‘For You’ album slideshows). iPhone ‘Memories’ oscillate between a mode of documentation (for instance, compilations of trips taken, determined by geolocate tags attached to images) and the formation of an affective reality (as intimated by the more abstract yet emotionally evocative album titles, like ‘Together’). The algorithm’s mathematical logic assesses my images, locating patterns to determine what depictions of people, places, and things might be significant to me. However, though algorithmic logic provides some key insights into how we see photographically today, it only provides part of the story. Regardless of the increasingly automated processes of making and viewing images, photography remains deeply affective and personal. Indeed, Apple’s designation of these features as ‘Memories’ and ‘For You’ indicate that they are meant to be perceived as affective and personalized. Apple refers to these technical operations not only as ‘features,’ but as ‘experiences’ (Apple, 2022). This description acknowledges a *socio-technical* process, in which the ‘technical’ and the ‘affective’ (which is both individual and collective) are deeply entwined.

In this paper, my focus is also on the socio-technical process, by showing how photographs are (re)presented through algorithmically curated photo albums. As such, what is important are the affective and aesthetic (sense-making) modes of interpretation produced by Apple ‘Memories’, rather than the complex relationship between photography and memory. It is, however, important to note that Apple deliberately and consistently invokes the concept of memory in relation to photographs, to the point of conflation. Sergio Martínez Luna (2018) writes that

“[w]ith the launch of Live Photos, Apple claimed that the photographic instant frozen in time becomes a permanently alive memory” (52).⁷ While memory has always been shaped and reshaped through our encounters with images, it is the sequencing of our photographs by machines that I want to attend to. iPhone ‘Memories’ are not memories, but an enframing. This enframing includes the software that remixes our personal photographic images. It also includes the device and interface, the hardware. The iPhone screen standardizes the image size and display, while the software organizes selection, sequence, timing, and music. Our experience of the artifact (Merleau-Ponty, 2005) becomes a cumulative part of our interactions with our own photographic archive. Engaging with our images through both iPhone and algorithm is an embodied act. For the purposes of this paper, my focus is on the software, rather than hardware. Affording machines increasing autonomy in turning our snapshots into narrative sequencing develops a newer human-machine entanglement of framing and personalization.

A brief history of motion and montage

In the early days of photography, transposing motion into stillness was a necessary component of the long exposure times required to capture images. We might think of the famous daguerreotype by Louis Daguerre, ‘Paris Boulevard’ (or ‘View of the Boulevard du Temple’), from 1838. At a brief glance, the Parisian streets appear otherwise emptied of people, though ghostly remnants of movement remain. A precursor to montage – the composite image – also emerges in the nineteenth century, both as a way of enhancing the image’s fidelity and stretching the bounds of possibility for image capture. We might think of Gustave Le Gray’s composites, like ‘The Great Wave’ from 1857 which used separate exposures to capture land and sky respectively. Another well-known example – also from 1857 – is Oscar Gustav Rejlander’s ‘Two Ways of Life,’ wherein he combined more than thirty separate negatives into a single large print.

While stillness was crucial in early image capture, by the end of the century it became possible to display images in motion. In the late 1800s, Thomas Edison and William Dickson developed the kinetoscope, a large box containing a mechanism to run sequential images over light, with a peephole for the viewer to look through. The

individual frames are made visible through jerky movements, but the subjects of the images were nonetheless presented as in motion. Indeed, a variety of ‘optical toys’ from the nineteenth century “depended on some kind of setting into motion of still images via a manual or mechanical apparatus to generate the impression that the images depict movement” (Hoelzl and Marie, 2015: 13).⁸ While the kinoscope was designed for personal viewing, the Lumière brothers developed a system for the public projection of moving images through the Cinématographe. These early films were composed of single, unedited shots. As Scott McQuire (2008) writes, it is not until the second decade of cinema that we see “the emergence of multi-shot narratives as films,” which “came to be composed by means of the fragmentation and re-assembly of the visual field” (65). Moving images foreground the juxtapositions that lead to montage being recognized as a live issue in new ways.

By the twentieth century, arguably driven by the increasing circulation and accessibility of photography and printed photographic images, the novel practice of ‘photomontage’ emerges. Scholarly writing tends to recognize some key differences in the approach and form of early photomontage through the work of Dada and Russian Constructivist artists. Varvara Stepanova (1989 [1928]) detailed two phases defining photomontage, the first which can be understood as combinatory (more reminiscent of collages, such as the works of Hannah Höch) and the second being a photographic series (for instance, the works of Alexander Rodchenko). Benjamin Buchloh (1998) wrote that the avant-garde photomontages of the 1910s were based on shock effects and discontinuity, while photomontage after 1925 spoke to the “logic of the archive” and continuity (43-56). He refers to these different types as heterogenous (a combination of unexpected visual elements which disrupts or blurs meaning) and homogenous (a formal organization of visual elements which creates clear meaning), respectively. Buchloh’s distinction speaks to a change in the perception of photographic images during the 1920s, a shift that hinges on the growing role of photography as an archive. Homogenous photomontage functions more like a photographic collection, with an “archival or mnemonic dimension” (Bénichou, 2003: 173). Montage thus invokes various practices: notably, the cutting, assembling, and serialization of images. It is both a technical operation and an aesthetic form. As a form it deals with narrative on a spectrum, from disjuncture to new forms of cohesion.

Montage also came to form a key term in film studies, where it also takes on a broad meaning. Ryan Conrath (2023) writes that filmic montage claims several roles; at times, it constructs spatiotemporal continuity, at others it creates disorder for intellectual engagement, and at others still it is used for expressions of duration for sensory experience. McQuire notes that “[t]he gradual development of recognizable narrative conventions enabled the construction of new forms of continuity from the formal discontinuity produced by montage” (2008: 67). This led to the prevalence of continuity editing as an “institutionalized mode of narration”, though also saw avant-garde practitioners breaking with film convention “to reconstruct the visible world” (McQuire, 2008: 67-68). It is through filmic montage that *relations between images* were seen to take precedence over the content of each shot, the sequencing of images mobilized to reshape and manipulate reality (Rohdie, 2015: 141).

According to Magda Dragu (2020), “[t]he film image and the photographic image share the same indexical nature, despite the different representations they propose, static in the case of the photograph and temporal and narrative in the moving image” (101). However, with digitization, the characterization of photography as ‘static’ starts to be further problematized as it became increasingly easy to circulate, edit, and remix images. With digitality, “images are no longer guarantors of visual truth, because they do not function as signifiers with a fixed meaning or value” (Martínez Luna, 2019: 44). As photography starts to be understood increasingly through (rather than in contradistinction to) motion, scholars (Hoelzl and Marie, 2015; Røssaak, 2011; Rubinstein and Sluis, 2019; Sutton, 2009) argue for the need to rethink the relations between photography and cinema. Damian Sutton (2009) goes so far as to claim that “perhaps the best way to understand photography and the photograph [is] through the lens of cinema” (ix). Drawing on the history of montage allows us to recognize how the sequencing of images creates narrative resources which not only register but help to construct the way we see photographically.

The contemporary moment cannot however be summed up in terms of a linear transition from composite images to photomontage to filmic montage. In fact, the composite image remains central to digital images and computational photography, albeit in a different way. “Technical images,” Flusser writes, “are not surfaces but mosaics assembled from particles” (2011 [1985]: 6). This is exemplified by photos

taken with iPhone cameras using Deep Fusion technology (from the iOS 13.2 release in 2019 onwards) to optimize detail, texture, and noise. Apple's promotion of this process often cites the technology's role in taking "better" images in low-level lighting (Apple Newsroom, 2022). This makes visible a key characteristic of the computational image, which distinguishes itself by renegotiating the technical operation of *photography* as the process of 'writing with light'.⁹ When a user taps the shutter in mid to low light, the iPhone camera takes nine images with different exposure levels (four short, four secondaries, one long exposure). It instantaneously begins post-processing, fusing the long exposure with the other images and using a pixel-by-pixel analysis to select the best pixels from each for the final composite image (Cervantes, 2023). This operation is performed automatically in the background, in the space of a second, and cannot be manually switched off. From 2022 (with all the iterations of iPhone 14 onwards), Deep Fusion was extended through the introduction of Apple's 'Photonic Engine' (Cervantes, 2023). Essentially, this enables Deep Fusion to be operationalized earlier in the process of image capture to work on uncompressed photos and further preserve detail. Computational photography, in assembling picture elements in such a way, produces a composite at an entirely different scale: at the level of the pixel.

Much of what Flusser (2011 [1985]) called the task of "envisioning", referring to the production of technical images and the task of gathering and ordering them, is performed through progressively automated processes. We are also gradually losing the capacity to adjust the apparatus. Apple hardware and software is notorious for promoting this shift, discouraging any "tinkering" (Gillespie, 2006) by voiding warranties if users attempt their own interventions. Forms of montage are now a default setting in Apple Photos, and this adds an additional element of complexity to the relationship between human perception and machine vision. Nicolas Malevé (2021) rightly asserts that computer vision has never been wholly automated (if we take automation to mean distinct from human labour). However, we can make the claim that we have increasingly shifted towards greater levels of automation in photograph production. In this way, the iPhone, its camera, and the default computations in the production and reception of our images are more akin to Flusser's television control panel, a "faulty key", those things "that permit me to choose but not to express myself" (2011 [1985]: 31). This notion aligns with what Jussi Parikka (2023)

refers to as “operational images”. The term originates with Harun Farocki’s early 2000s video installation trilogy *Eye/Machine 1-111* (2001–03) and denotes a shift away from representations and toward “the primacy of operations”. Parikka (2023) writes:

As instructions for life, operational images also imply a broader use of the term “algorithmic” as the training of bodies, the setting of institutional routines, and the rehearsing of automation in ways that tie machines to laboring human bodies. Imaging practices become operational in how they tie bodies into collective routines (9).

While Flusser locates a redistribution of agency between humans and technology, Parikka focuses on how this redistribution through operational images impacts human action. Extending this analysis to the computational underpinning of ‘For You’ montages implies that we are being trained to see and internalize algorithmically determined points of connection and significance between the objects, people, places, and events depicted in our photographs. The algorithm is optimized for “narrative montage”. Andrea Nelson (2006) defines narrative montage as “the careful sequencing of photographic images in order to convey visual arguments [...] a pedagogical model of visual literacy for mass audiences” (258). In the sections that follow, I illustrate how algorithms are operationalized through ‘Memories’ and ‘For You’ albums, and the implications of this for how we engage with and interpret our photographs.

Moving stills and scene analysis

Historically, montage has largely been the result of human intervention in the processes of image making, selecting, and sequencing. With the iPhone’s ‘For You’ album slideshows, this process is automated through computational photography and the algorithmic curation of personal archives. The contemporary turn to algorithmic curation is driven by a number of factors. One is the exponential growth in the scale and circulation of photography/imagery which has led to a reliance on machine computation and digital databases to enable searchability, sequencing, and narrative. Ingrid Hoelzl and Rumi Marie (2015) note that the “desire for movement” in contemporary photographic practice is further entrenched by a “desire for endlessness”. The “desire for endlessness”, they write, “can be found in digital image

editing and viewing software, in particular in navigable image databases that feature seemingly endless image spaces” (40).¹⁰

Hoelzl and Marie (2015) offer an example of the new relations of movement and stasis in photographs through their description of the ‘Ken Burns Effect’.¹¹ The Ken Burns Effect refers to the animation of an image using zooming and panning. Originally an analogue technique whereby a photograph would be re-filmed with a moving camera, the effect is now commonly seen in digital animation (Hoelzl and Marie, 2015: 17–20). The Ken Burns Effect frequently appears as a native feature or third-party plug-in for non-linear editing systems (Allegra, Stanco and Valenti, 2015: 93). Often this is explicit, for instance, “Openshot and iMovie software for Linux SO include a transition effect called “Ken Burns” [...] Final Cut Pro, Apple TV and Apple’s iMovie video editing programs have a photo slideshow option labeled “Ken Burns Effect”” (ibid). The effect is prevalent in the iPhone’s ‘For You’ slideshows, where “[m]ore often than not, it is used as a pure random feature to create animated digital slideshows, whose constant pans and zooms generate a liminal narration that is purely retinal – a kind of ambient motion” (Hoelzl and Marie, 2015: 20). Hoelzl and Marie refer to these images as “moving stills”, defining the moving still as “an image that displays a paradoxical distribution of stasis and movement via image motion techniques” (2015: 21). In terms of reception, “the digital screen functions as a ‘viewing window’ of a seemingly endless photographic space of which an ever-changing fragment appears within it” (ibid).

The automation of the Ken Burns Effect through Machine Vision is made possible through the addition of a ‘saliency’ algorithm within ANSA. Apple uses two categories to determine saliency: ‘attention based’ (identifying where viewers first look when they see an image – for photos of people, this is the face) and ‘objectness based’ (locating foreground content). As Brittany Weinert, a software engineer for Apple’s Vision Framework, has stated:

A lot of times, these photo-showing algorithms can be a little bit awkward. They zoom into seemingly random parts of the image, and it’s not always what you expect. But with saliency, you always know where the subjects are, so you can get a more documentary-like effect (Apple, 2019a).

The objectness saliency algorithm works in tandem with Apple’s image classification algorithm. To determine objectness saliency, one must first have a process of machine vision capable of recognizing objects. Apple developed a large-scale, on-device classification network which contains over one thousand different categories of objects. Rohan Chandra, a researcher in the Apple Vision team, notes that “this is a multi-label network capable of identifying multiple objects in a single image, in contrast to more typical mono-label networks that try to focus on identifying a single large central object in an image” (Apple, 2019a). Each category is conceived as part of a hierarchical taxonomy, beginning with the more general classification through to the more specific classification. For instance, a photo of your poodle will likely undergo the following series of classifications: the category of ‘animal’, then classified as mammal, then dog, before finally identifying the breed of dog. The taxonomy is mapped through semantic meaning, which creates a series of relationships between objects that can then be grouped together in a category. Each visible object within the image is given a probability weighting alongside the semantic tags, which denotes the algorithm’s confidence level for having accurately labelled the object. ‘Bounding boxes,’ whereby rectangular regions are positioned over objects and annotated, enable multiple objects to be isolated and analysed within the one image.¹²

In my own ‘Memories’, one ‘For You’ album has been titled ‘Pet Friends: Over the Years’ (fig. 1). The sequence is a montage of images of my dog in a variety of settings. Two images (re)presented to me stand out. The second image in the sequence is a close-up showing a patch of missing fur on my dog’s back, taken after he was injured in an altercation with another dog. The final image in the sequence is one in which I know my dog was featured, however, the image (re)presented is zoomed in on my face, effectively cropping any trace of my ‘pet friend’ from the photograph. These two images would seem to register the intersection of different algorithms within ANSA. For the first image, context aware object recognition enables the algorithm to trace images of my dog across multiple photographs, to begin to build a profile that includes various body poses, angles, distance, and clarity. This in turn enables machine vision to increase prediction accuracy – including recognizing an extreme close-up of fur. For the final image, we likely see attention-based saliency at work, mobilizing the Ken Burns effect as a narrative technique for this montage. It seems probable that an

attempt was made by the algorithm to select a point of significance – one of the ‘bounding boxes’ – in the photo and zoom in on it, using the saliency metric and motion to capture or retain my attention. The automated selection and display that is ‘Pet Friends’ aims towards, in Buchloch’s (1998) terms, a homogeneous narrative montage through image classification and pattern recognition. While it is largely successful in locating points of cohesion and connection between images based on a consistent feature, in this instance it produced a heterogeneous montage at the level of affect. The rupture created by the two images outlined above occur because the pattern does not (and cannot) fully account for my subjective, embodied experience of the photographic moments being (re)presented.

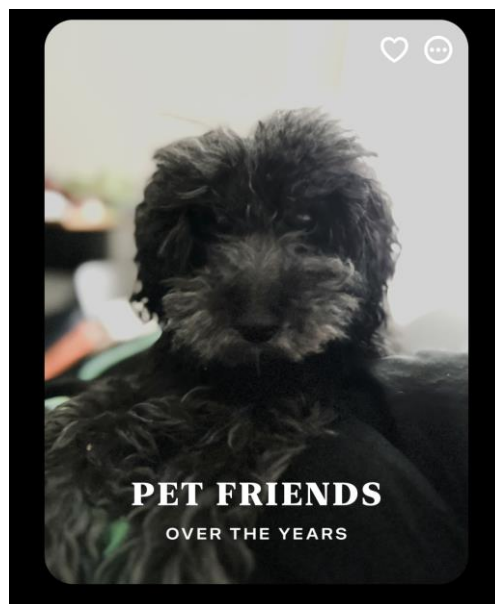


Fig. 1 Screenshot of key photo of ‘For You’ album ‘Pet Friends: Over the years,’ author’s own, March 10, 2024

Personalization and categorization: predicting affective connections

Nicolas Malevé and Katrina Sluis (2023) note the ‘onto-epistemic flattening’ of the photographic image within machine learning datasets used in image classification. Referencing the photo-sharing platform Flickr, and its significance for the canonical computer vision object recognition training dataset ImageNet, they highlight the importance of ‘amateur photography’ in the development of datasets used to train machine learning programs. In contrast to professional photography, ‘snapshots’ were seen by computer scientists as more representative of the world, more neutral,

authentic, all-encompassing – “objective ‘data’ for a machine vision pipeline,” or “a ground truth for machine vision” (Malevé and Sluis, 2023: n.p.). As such, popular photography from photo-sharing platforms (along with other images scraped from the Internet) were appropriated to train ML software. In the case of ImageNet, this was done with the aim of mapping the world through images. Apple ‘Memories’ operates as part of this wider context, also using supervised Machine Learning to classify images and recognize and categorize the objects depicted in them.¹³ However, the primary objective of ‘Memories’ is not to map the world, but to map which aspects of the world are significant to an individual user. An article on Apple’s Machine Learning Research website states:

Photos can also learn from identity information to build a private, on-device knowledge graph that identifies interesting patterns in a user’s library, such as important groups of people, frequent places, past trips, events, the last time a user took an image of a certain person, and more. The knowledge graph powers the beloved Memories feature in Photos, which creates engaging video vignettes centered around different themes in a user’s library. Memories uses popular themes based on important people in a user’s life, such as a memory for ‘Together’ (Apple, 2021).

The knowledge graph is key to the interplay between personalization and categorization in ‘Memories’. In ML, a knowledge graph aims to contextualize data by illustrating the relationship between nodes (people, objects, places) in a network. While object detection and image classification speak to a collective organizational logic through the systemization of semantic mapping, the knowledge graph adds the element of personalization required for predicting affective connections between the user and their images.¹⁴

Any one user’s knowledge graph consists of thousands of nodes and accompanying relationships.¹⁵ Some of the key components for the Apple Photos knowledge graph include events, user activity, people, places, and dates (Apple, 2019b: 9). Combined with ANSA, each component is further analyzed to attribute user significance.¹⁶ ‘Events’, for instance, utilizes scene analysis to determine moments of significance, such as weddings or concerts. ‘User activity’ measures interaction with a set of

photographs – whether they have been viewed multiple times, edited, shared, and so on. ‘People’ triangulates facial recognition with frequency of depiction (within photos) and frequency of communication (through Messages) to infer affinity between a user and their contacts. This can be extrapolated out to social groups by including people who are often in photos together within the knowledge graph. ‘Places’ utilizes Maps to determine the site of the user’s home and place of work, where the user frequently takes photographs, and whether photos are being taken at a new location or on a trip. In addition, object detection and scene analysis determines whether photographs are being taken at a landmark or cultural site, while scene level tags can also incorporate natural geographic features like beaches and mountains. ‘Dates’ uses data from a user’s Calendar and Contacts to incorporate birthdays and anniversaries, alongside user location to map holidays celebrated in the user’s country. The knowledge graph draws on multiple sources of data to locate patterns within the repository of images that make up a user’s Photo library, underpinning the organizational logic of affective predictions for the ‘Memories’ feature. The deployment of the knowledge graph returns us to Buchloh’s (1998) articulation of homogenous montage – that which has an ‘archival dimension,’ and aims to create cohesive meaning.

Unlike the automated curation of ‘For You’ albums, other forms of visually displaying personal photography have centred human choices. We might think of Berger’s (2008 [1972]) reference to the pinboard, a selection of paper paraphernalia – snapshots, notices, postcards, and so on, “chosen in a highly personal way to match and express the experience of the room’s inhabitant” (30). Don Slater (1995) picks up this reference, arguing that “the pinboard evokes a collage of affiliations, in which the representation of self is produced by and within the activities of the present” (in Lury, 1998: 84). In a similar vein, we might consider W. J. T. Mitchell’s (2017) invocation of the refrigerator door as an example of image display that speaks to a provisional assemblage (80). Each place importance on the selection and display of images, speaking to the way in which photography contributes to a greater degree of user organization of visual images.

While user selections are shaped by the culture they inhabit, users nevertheless make choices according to their own sensibility. The cohesive meaning of the pinboard comes from the particularity of the individual and the connections they have made

between images. The transition to computational photography automates the processes of image curation. Key to the ‘Memories’ feature is that it uses categories to build graphs of affinity. These overarching categories become more discriminating through the combination of signals generated via various data points. Connections within a collection of personal photography are co-constituted by the algorithm and user, as ‘For You’ albums (re)present patterns from a knowledge graph as a narrative montage. The algorithmic curation of our photographs is not based on ‘meaning’ as understood through the choices and juxtapositions forming the pinboard ensemble. Rather, it is a probabilistic assessment of potential meaning, based on predictions that particular selections and combinations will be significant and affective.

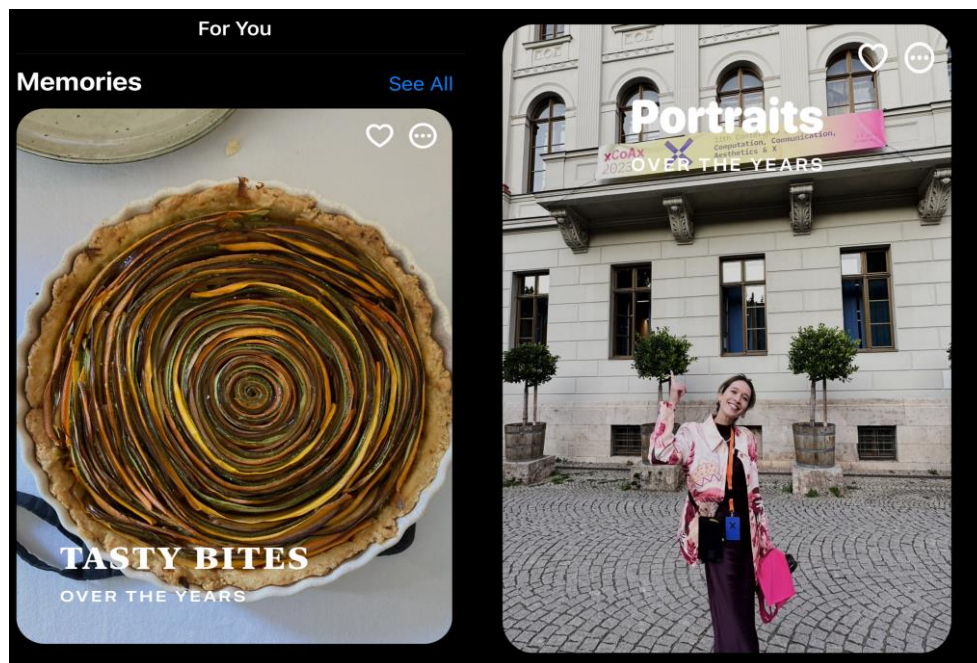


Fig. 2 (left) Screenshot of key photo of ‘For You’ album ‘Tasty Bites: Over the years,’ author’s own, March 10, 2024

Fig. 3 (right) Screenshot of key photo of ‘For You’ album ‘Portraits: Over the Years,’ author’s own, March 10, 2024

A distribution of agency between human and machine, and the tension between the algorithmic and the affective, is made apparent when the algorithm fails to produce a montage that is cohesive and significant to the user. An album slideshow titled ‘Tasty Bites’ (fig. 2) plays a sequence of meals I’ve eaten over the years, interspersed with images of food packaging I’ve taken to send to my partner in lieu of grocery lists. Another slideshow called ‘Portraits’ (fig. 3), with images taken over the course of many

years, oscillates between still images, Live Photos, and short clips extracted from my video footage. At times, two images are displayed together vertically in frame. These always seem to be photos of myself, but in entirely different contexts (for instance, a photograph where I am sitting in my yard in the sun is positioned together with a photograph of me sleeping next to my dog). Saccharine harp plays in tandem with the slideshow, and images subtly zoom in and out, creating additional movement on screen, the Ken Burns Effect. This experience does not offer a sense of cohesion between images. As Tara McLennan (2018) writes, “Despite efforts of narrativisation and sequencing, photographs often break out of the intended structures, practices and enunciations to which they are assigned” (33). When photographs are curated according to mathematical logic yet the rhetoric surrounding algorithms proclaim an affective logic, this break can seem even more marked.¹⁷

The above albums speak to Johanna Drucker’s ‘non-sequitur’ frame analysis. While we can still understand them in terms of narrative montage, the non-sequitur deviates from linear narrative, instead “stitching fragments of what are graphically related elements together in a narrative, or making our way through unrelated fragments until some chain of compelling connections captures our attention” (Drucker, 2011: 4). Drucker writes that non-sequitur connections are most evident in electronic space, where users have been predisposed to jump between points of attention and different embedded mediums. By doing the jumping for us, the algorithmically curated album retains our attention. As such, it aligns with a wider narrative of ‘attention economy’, found in the endless scroll of social media platforms, near-constant screen refresh, the interruptions of notification pings, the hyperlinked and hyperactive modes of consumption that are encouraged by our mediatized environs. The non-sequiturial aligns with what Mitchell (2017) writes is the central aim of montage: “making meaning from surprising cuts and juxtapositions” (83). While the algorithm performs the role of stitching fragments, the meaningfulness of the narrative remains the purview of the user, who meets the algorithm halfway to find the connections between images. What Apple is ostensibly attempting with ‘For You’ and ‘Memories’ is the provision of an imagined affordance of affective pattern recognition. In reality, the machine makes patterns of people, places, engagement. Its affective resonance emerges in the context of cultural and personal values intersecting with this technological ensemble.

There is more to these images and sequences than the narrative of hyper-consumption and fragmented concentration implies. Regardless of whether our images are more defined by the homogenizing effects of machine vision, and the narrowed optimization of algorithmic operations, they remain affectively personal. They speak to us in ways that they do not speak to others. As Seán Cubitt (2018) writes in relation to the mass image, “The space between significance and the *asignificant* is not a void but a sliding scale” (178). My ‘For You’ album titled ‘Golden Hour’ (fig. 4) – a slideshow of unimaginative images of sunsets I’ve captured – is likely indistinguishable from the millions of other sunset photos housed online and in other people’s camera rolls. While it could be argued that each sunset image is a trigger to remember a wider event or experience (perhaps a vacation or an evening with friends?), in all honesty, I could not say where or when most of these images were taken. What I can say is that they (re)present a version of my identity back to me – someone who many times over many years has stopped to think that a moment was beautiful. Rather than them being devalued by accumulation, it is precisely their presentation as a cumulative body that makes the value of these sunset photos visible. Even positioned within the logic of the mass image, ‘Golden Hour’ represents a personal attempt to enact the very common human quality of expressing wonder and gratitude for the natural world.



Fig. 4 Screenshot of key photo of ‘For You’ album ‘Golden Hour: Over the years,’ author’s own, March 10, 2024

The duality of the algorithmic album as simultaneously personal and common, as part of an affective logic and as part of a mathematical logic, echoes Lury (1998) who, drawing on Barthes (1981), writes:

... photography has contributed to the development of a self-identity constituted in a continuous, repetitive dis-internalisation of subjectivity and a simultaneous affirmation of new modes of intimacy, individual affect and self-representation (80).

This operates through a “simultaneous breaching and reaffirmation of the public and private” (Lury, 1998: 80). While the argument made by Barthes and taken up by Lury lends itself more easily to the practice of sharing photographs socially, in the context of the machine vision, the ‘publicness’ of images is extended. As data, the image is ‘public’ when it is interpreted by a machine in relation to all other images that have been tagged and used for training. As data, the image flows into the ocean of mass images, becoming a part of the moving swell. Through the algorithmically curated album our images return to us in the form of a wave, particles (or pixels) clustered together, no longer completely extractable from the current. The analysis and taxonomic categorization of images, ANSA, and the Photos knowledge graph organize our personal archive, positioning our photographs as networked images.

Though the ‘personal’ has never been completely autonomous from the social or the collective, contemporary capitalism is underpinned by an imperative for increasingly granular data. It is this granularity that predicates that which is conveyed as ‘personal’ data. The drive to collect and harness more representative and all-encompassing data is deeply embedded within the algorithmic curation of ‘For You’ albums. The stitching together of images through montage aims to (re)present the affective depth of photography, foregrounding personalization. The stance that has been taken here is that ‘personalization’ within algorithmic culture should not be conflated with ‘uniquely individual’. Our individuality is always constituted in our relation to others, rather than being self-contained. Apple ‘Memories,’ as data objects, undergo onto-epistemological flattening to produce sameness through classification. They are then re-inflated, contextualized by the thousands of nodes and edges produced by my data to predict

affective connections. Despite this, the logic of algorithmic curation never fully captures my subjective experience of that which is represented in ‘For You’ montages.

Conclusion

The popular montage imagery of the 1920s was arguably a response to the intensity of rapid urbanization and industrialization. We are again in a period of technological intensity, and observing changes in photographic culture such as new manifestations of montage within algorithmic curation offers a useful lens for unpacking our current socio-technical milieu. The predilection to visualize an overwhelming feeling of stimulus as montage is nothing new. Montage can provide both a reflection of disjunction, and a system for organizing and making sense of a vast repository of content. The photographic image has always been, to varying degrees, in motion. Today, it has become digitally circulated, networked and operationalized, subject to algorithms, automation, and machine vision pipelines, part of a database of mass images. The image has been catapulted into a flow of movement and sequencing. Our images, both in their creation and reception, register multiple tensions – of public and private, human and machine, disaffected and emotionally resonant. The serialization of images through algorithmically curated photo albums has shifted various processes of selection, organization, and display towards an automated narrative montage, while the combinatorial process of Deep Fusion brings the composite image to the level of the pixel.

Whether discernible to human visual perception or not, the photographs and other images on our networked mobile devices are parts of moving flows of data information. Narrative montages emerge through the mathematical logic of the algorithm – through the knowledge graph, metadata, image tagging, content recognition, scene analysis, and saliency. The algorithmic curation of our personal photographic collection adds a new dynamic to the relations between stillness and motion within photography. The machinic analysis of photographs as aggregate data points position images in relation to one another. The iterative nature of machine learning affords an automated and continual re-organization of our on-device archive. Through ‘Memories’ and ‘For You’ albums, patterns in how we take photographs, where we take them, and what we take them of are (re)presented to the user-viewer as

affective connections through the sequencing of images. While technology has always formed a crucial component in how we ‘see photographically’, the computational now shapes our engagement with photography not only at the point of image capture, but by increasingly automating the way photographs are selected, organized, and displayed.

References

- Allegra, D., F. Stanco and G. Valenti (2015) ‘A Semi-Automatic Algorithm for Applying the Ken Burns Effect’, in A. Giachetti, S. Biasotti and M. Tarini (eds.) *Smart Tools and Apps for Graphics*. Italian Chapter Conference: The Eurographics Association, pp.93-101. DOI:10.2312/stag.20151296.
- Apple (2019a) ‘Understanding Images in Vision Framework’. Available at: <https://developer.apple.com/videos/play/wwdc2019/222> (Accessed: 11 November 2023).
- Apple (2019b) ‘Photos: Private, on-device technologies to browse and edit photos and videos on iOS and iPadOS’, September. Available at: https://www.apple.com/id/privacy/docs/Photos_Tech_Brief_Sept_2019.pdf (Accessed: 18 March 2024).
- Apple (2021) ‘Recognizing People in Photos Through Private On-Device Machine Learning’, *Machine Learning Research*, July. Available at: <https://machinelearning.apple.com/research/recognizingpeople-photos> (Accessed 11 November 2023).
- Apple (2022) ‘A Multi-Task Neural Architecture for On-Device Scene Analysis’, *Machine Learning Research*, June. Available at: <https://machinelearning.apple.com/research/on-device-scene-analysis> (Accessed 11 November 2023).
- Apple Newsroom (2022) ‘Apple introduces iPhone 14 and iPhone 14 Plus’, press release, 8 September. Available at: apple.com/au/newsroom/2022/09/apple-introduces-iphone-14-and-iphone-14-plus/ (Accessed: 6 January 2024).
- Barthes, R. (1981) *Camera Lucida: Reflections on Photography*, trans. R. Howard. New York: Hill and Wang.
- Batchen, G. (2004) *Forget Me Not: Photography and Remembrance*. New York: Princeton Architectural Press.
- Bazin, A. (2009 [1967]) ‘The Evolution of the Language of Cinema’, trans. H. Gray, reprinted in L. Braudy and M. Cohen (eds.) *Film Theory and Criticism*. New York/Oxford: Oxford University Press, pp.41-53.

- Bénichou, A. (2003) 'Temporal Montage in the Artistic Practices of the Archive', trans. T. Barnard, in V. Lavole (ed.) *Now: Images of Present Time*. Montreal: McGill-Queen's University Press, pp.167-187.
- Berger, J. (2008 [1972]) *Ways of Seeing*. London/New York: Penguin Books.
- Berger, J. (1982) 'Appearances', in J. Berger and J. Mohr (eds.) *Another Way of Telling*. New York: Pantheon, pp.82-129.
- Buchloh, B. (1998) 'Warburg's Paragon? The End of Collage and Photomontage in Postwar Europe', in I. Schaffner and M. Winzen (eds.) *Deep Storage: Collecting, Storing and Archiving in Art*. Munich/New York: Prestel Verlag, pp.51-60.
- Cervantes, E. (2023) 'What is Apple's Photonic Engine all about?' *Android Authority*, 14 March. Available at: androidauthority.com/apple-photonic-engine-3208007/ (Accessed: 6 January 2024).
- Conrath, R. (2023) *Between Images: Montage and the Problem of Relation*. Oxford: Oxford University Press.
- Cubitt, S. (2018) 'Connectivity, Legibility and the Mass Image', in P. Hesselberth, J. Houwen, E. Peeren and R. de Vos (eds.) *Legibility in the Age of Signs and Machines*. Leiden/Boston: Brill/Rodopi, pp.166-179.
- Dragu, M. (2020) *Form and Meaning in Avant-Garde Collage and Montage*. New York/Oxon: Routledge.
- Drucker, J. (2011) 'Humanities approaches to interface theory', *Culture Machine* 12.
- Flusser, V. (2011 [1985]) *Into the Universe of Technical Images*, trans. N. A. Roth. Minneapolis/London: University of Minnesota Press.
- Gillespie, T. (2006) 'Designed to 'effectively frustrate': copyright, technology and the agency of users', *New Media & Society* 8(4): 651-669.
- Hoelzl, I. and R. Marie (2015) *Softimage: Towards a New Theory of the Digital Image*. Bristol/Chicago: Intellect Books.
- Lury, C. (1998) *Prosthetic Culture: Photography, Memory and Identity*. London/New York: Routledge.
- Malevé, N. (2021) 'On the data set's ruins', *AI & Society* 36: 1117-1131.
- Malevé, N. and K. Sluis (2023) 'The Photographic Pipeline of Machine Vision; or, Machine Vision's Latent Photographic Theory', *Critical AI* 1(1-2). <https://doi.org/10.1215/2834703X-10734066>.

-
- Martínez Luna, S. (2018) 'Still Images? Materiality and Mobility in Digital Visual Culture', *Third Text* 33(1): 43-57.
- McLennan, T. (2018) 'Memories in the networked assemblage: How algorithms shape personal photographs', *fusion journal* 14: 30-45.
- McQuire, S. (2008) *The Media City*. London: Sage Publications.
- Merleau-Ponty, M. (2005) *Phenomenology of Perception*, trans. C. Smith. London: Routledge.
- Mitchell, W. J. T. (2017) 'Method, madness and montage: assemblages of images and the production of knowledge', in J. Eder and C. Klonk (eds.) *Image operations: visual media and political conflict*. Manchester: Manchester University Press, pp.79-85.
- Nancy, J.-L. (2005) *The Ground of the Image*, trans. J. Fort. New York: Fordham University Press.
- Nelson, A. (2006) 'László Moholy-Nagy and *Painting Photography Film*: A Guide to Narrative Montage', *History of Photography* 30(3): 258-269.
- Osborne, P. (2019) 'The Image Is the Subject: Once More on the Temporalities of Image and Act', R. Göring, B. Gronau, and L. Schwarte (eds.) *Aesthetics of Standstill*. Berlin: Sternberg Press, pp.125-137.
- Parikka, J. (2023) 'Operational Images: Between Light and Data', *e-flux Journal* 133: 1-11.
- Pereira, G. (2019) 'Apple Memories and automated memory-making: Marketing speak, chip-engineering, and the politics of prediction'. Paper presented at *The 20th Annual Conference of the Association of Internet Researchers*, Brisbane, Australia, October 2-5. Retrieved from <http://spir.aoir.org>.
- Rohdie, S. (2015) *Film Modernism*. Manchester University Press.
- Røssaak, E. (2011) 'Algorithmic Culture: Beyond the Photo/ Film Divide', in E. Røssaak (ed.) *Between Stillness and Motion: Film, Photography, Algorithms*. Amsterdam: Amsterdam University Press, pp.187-206.
- Rubinstein, D. and K. Sluis. (2019) 'A Life More Photographic: Mapping the Networked Image', in L. Wells (ed.) *The Photography Cultures Reader: Representation, agency and identity*. New York/London: Routledge, pp.349-366
- Stepanova, V. (1989 [1928]) 'Photomontage', in Christopher Phillips (ed.) *Photography in the Modern Era: European Documents and Critical Writings 1913-1940*. New York: The Metropolitan Museum of Art/Aperture, p.236.

Sutton, D. (2009) *Photography, Cinema, Memory: The Crystal Image of Time*. Minneapolis: University of Minnesota Press.

Notes

- ¹ While these can be found as a separate tab within the camera roll, there are also occasions whereby unbidden notification alerts of 'You have a new memory' appear on-screen, linking the user to a slideshow. These are a default setting, though can be manually turned off.
- ² On-device operations are crucial for Apple. Without this capacity the company would not be able to adhere to its user privacy commitments.
- ³ Though Core ML and the Apple Vision framework are used within the users' Photo Application, they are also targeted to developers, so that they too can apply Apple's computer vision algorithms in building their applications.
- ⁴ ANSA also creates the titles for 'For You' albums and selects both the cover image and music to accompany the slideshow.
- ⁵ Scene tagging recognizes landscape features, such as beaches, mountains, and sunsets. Semantic prints classify images that are not suited for Memories, such as receipts and documents. Perceptual prints are what enable the recognition of image similarity and quality, to avoid (re)presenting duplicate or low-quality images. Saliency determines which aspects of an image are likely to be interesting to the viewer. Object detection recognizes the visual content of the image, including people, animals, flora, food, vehicles, and so on.
- ⁶ As of June 2022.
- ⁷ The Live Photo is a process introduced in 2015, whereby taking a photo with an iPhone captures one and a half seconds on either side of the 'photographic moment'. These can be later viewed as a three second video recording, effectively re-animating the image.
- ⁸ Hoelzl and Marie (2015: 13) list several: the phenakistoscope, zoetrope, daedalum, praxinoscope, and the mutoscope.
- ⁹ While computational photography automates the composite image and has raised questions around the role of the photographer's 'decisive moment', it must be noted that photography has always observed a distribution of agency between humans and technology.
- ¹⁰ Crucially, Hoelzl and Marie position this within a historical narrative that highlights the blurring of photographic and cinematic form, noting how shifts in the spatio-temporality of images predate the digital. As we have seen, the trajectory of montage forms has long been a way for exploring the construction and display images in sequences.
- ¹¹ Named after the award-winning documentary filmmaker Ken Burns, who used the technique to great effect in documentary series such as *The Civil War* (1990), which was inspired by Matthew Brady's famous war time photographs and made ground-breaking use of still images in a popular documentary film.
- ¹² This is first done through supervised learning, a subset of machine learning, whereby the algorithm is given a labelled dataset to begin to recognize patterns.
- ¹³ In fact, computer scientists at Apple evaluate ANSA for accuracy in relation to ImageNet (alongside a mixture of other historical 'landmark' datasets, as well as user-representative internal image libraries).
- ¹⁴ "The use of image-language embeddings has allowed the discovery of richer memory types that were previously inaccessible through fixed taxonomy tagging. Detecting human-object interactions like people seated around a table is one such use case" (Apple, 2022).
- ¹⁵ In ML, these relationships are referred to as 'edges'.
- ¹⁶ "Detecting food, mountains, beaches, birthdays, pets, or hikes helps build a personalized understanding of the user and their interests. This in turn allows for an experience that offers personalized and delightful curated content" (Apple, 2022).
- ¹⁷ Pereira's 2019 analysis of Apple's 2017 iPhone advert foregrounds the companies claims to affective resonance in relation to photo capture, storage, and display.

Jasmin Pfefferkorn is a Postdoctoral Research Fellow at the School of Culture and Communication at The University of Melbourne. Her research spans museum studies, digital and computational humanities, and visual culture. She is the co-director of the research group *CODED AESTHETICS* and is on the steering committee of the Centre for Artificial Intelligence and Digital Ethics' *Art, AI and Digital Ethics* research collective.

Email: jasmin.pfefferkorn@unimelb.edu.au

