# Platform Seeing: Image Ensembles and Their Invisualities

**Adrian MacKenzie**
Australian National University

**Anna Munster**
University of New South Wales

## Abstract

How can one 'see' the operationalization of contemporary visual culture, given the imperceptibility and apparent automation of so many processes and dimensions of visuality? Seeing – as a position from a singular mode of observation – has become problematic since many visual elements, techniques, and forms of observing are highly distributed through data practices of collection, analysis and prediction. Such practices are subtended by visual cultural techniques that are grounded in the development of image collections, image formatting and hardware design. In this article, we analyze recent transformations in forms of prediction and data analytics associated with spectacular performances of computation. We analyze how transformations in the collection and accumulation of images as ensembles by platforms have a qualitative and material effect on the emergent sociotechnicality of platform 'life' and 'perception'. Reconstructing the visual transformations that allow artificial intelligence assemblages to operate allows some sense of their heteronomous materiality and contingency.

## Keywords

artificial intelligence architectures, Bergson, image ensembles, invisuality, platform

Amid a flood of images enabled by the massification of 'black box' mechanisms, cheaper display options and, for the first time, portable storage solutions, in 1896, Henri Bergson described both the universe and matter as an aggregate (*ensemble* in French) of images: 'this aggregate [*ensemble*] of images I call the universe' (Bergson, 1991: 18) or 'I call

**Corresponding author:** Anna Marie Munster. Email: a.munster@unsw.edu.au
**Extra Material:** http://theoryculturesociety.org/

matter the aggregate of images' (p. 22). Bergson was, of course, writing in the midst of the popularization of photography driven in part by Kodak's 1888 release of its first Brownie box camera, and perhaps with the foresight that by 1900 there would already be more than 100,000 of these sold. But Bergson's assertion has turned out to be somewhat prescient in the claim that image aggregates or ensembles are what comprise the matter of the world.

In terms of sheer quantities of images, the trillions of photographs uploaded to platforms such as Google or Facebook exceed the limits of human visual imagining. But the notion – hinted at by Bergson in his concept of 'ensemble' – that images 'aggregate' through specific processes of assembling together is crucial for understanding the reorganization of contemporary forms of knowing and perceiving. Prior to the action of a perceiving subject (or the 'body'), Bergson proposed that 'present' images – by which we take him as meaning images directly registering *as and generating* material experience – acted upon each other to create 'pathways' or matter-flows:

> That which distinguishes it as a present image, as an objective reality, from a represented image is the necessity which obliges it to act through every one of its points upon all the points of all other images, to transmit the whole of what it receives, to oppose to every action an equal and contrary reaction, to be, in short, merely a road by which pass, in every direction, the modifications propagated throughout the immensity of the universe. (Bergson, 1991: 36)

Such pathways, considered as an ensemble, might begin to provide us with a sense of actual nonhuman, inorganic processes as generative technical forces of experience. If today images have not only become ubiquitous but, as we will suggest, are increasingly *mattering* for a new kind of (socio)technicity, we suggest it is useful to revisit Bergson as a touchstone for a different take on the ontogenesis of images. Here we reconsider Bergson via more contemporary approaches to the ontology and technics of images. As Gregory Flaxman has argued, Bergson's 'aggregate of images' is about simultaneously understanding matter as plastic *and* flux, the world's mode and manner of presenting itself (2000: 92). World and image are therefore inseparable even as they constantly change. Acts of perception are what, instead, cut into this plastic mass, producing certain kinds of 'framing' or ways of perceiving 'the' image. In reconsidering contemporary image aggregates as ensembles – for example, datasets made available for platforms – we take up a Bergsonian ontology in two different ways. First, continuing along the lines of approaches such as Flaxman's, we pursue a generative and presentational mode of thinking images as matter. In particular, we will argue that image datasets as ensembles have a generative force for organizing

all kinds of platform cultural actions and forms. We also take account of recent work by authors such as Luciana Parisi (2013), which argues for a specific generativity for technical entities, such as algorithms, that is nonsensuous. Hence for images to *matter* in contemporary computation, they must be understood as entities in a computational relationality. Second, we emphasize the importance of the formatting of image ensembles as datasets across contemporary data practices; the incorporation of platforms into hardware in devices; forms of parallel computation; and the computational architectures of contemporary artificial intelligence. These assemblages constitute the (nonhuman) activities of perception as mode of cutting into/selecting out of the entire flux of image-ensemble-world. In a different but not unrelated context, Orit Halpern suggests that first order cybernetics – especially Norbert Weiner's conception of both human and machine systems and Warren McCulloch's early work on neural networks – seized on a Bergsonian conflation of thought with action and perception, suggesting that 'viewing complex actions out of accumulated systematic behavior of networked neurons puts the emphasis on the process, or algorithmic pattern that facilitated processing, and viewed this processing as productive in itself' (2015: 56). Our approach here is not to detail the actual historical uptake of Bergson in technical developments, as Halpern so elegantly traces. Instead we ask: how can we think with the Bergsonian image ensemble at a time when images both lose their stability and uniqueness yet gather aggregated force?

For forms of artificial intelligence driven by contemporary neural network architectures (and for now these seem to outweigh any other algorithmic logic operating on platforms), an ensemble of images in the broadest sense matters greatly. At every level of platforms, ranging from chip architectures to display devices, from data centre architectures to mobile phone 'platforms-on-a-chip', signs of a realignment around image ensembles can be seen. Platforms are made up of levels, and changes in one level flow through to others. Bergson's original conception of image-matter was as relational entity – an *ensemble*. Today, we propose, it is the image *ensemble* – images, not simply quantified, but labelled, formatted and made 'platform-ready' – that enables the emergence of a new mode of perception, and indeed a reformulation of visuality itself. We call this *platform* seeing.[1]

These contemporary image ensembles are not simply quantitatively beyond our imagining but qualitatively not of the order of representation. Their operativity cannot be seen by an observing 'subject' but rather is enacted via observation events distributed throughout and across devices, hardware, human agents and artificial networked architectures such as deep learning networks.

Here we will also take up Bergson's indistinction at the level of subject-object determination between the image ensemble and perception. For him, a point we argue later in this article, perception amounts

only to a mode of acting upon the image ensemble in order to make a temporary cut or selection. This is an argument used by Louise Amoore and Volha Piotukh (2015) in their work on the ways in which certain kinds of analytical functions and processes come to count as modes of giving specific form to big data. Amoore and Piotukh, however, are interested in the ways in which Bergsonian perception can be deployed in relation to big data and its analytics, shifting attention to the qualitative transformations performed by functions, algorithms and data analysis on the seeming ontological homogeneity of 'data'.[2] Our purposes are aligned but differ in their focus. We will argue that image ensembles have a qualitative and material effect on the contouring of the emergent sociotechnicality of both platform 'life' and 'perception'; that is, at the conjunction of image ensembles and artificial intelligence architectures, devices and hardware, *platform seeing* transpires as a new mode of invisual perception. Such a mode suggests that while visual techniques and practices continue to proliferate – from data visualization through to LIDAR technologies for capturing nonoptical images – the visual itself as a paradigm for *how to see and observe* is being evacuated, and that space is now occupied by a different kind of perception. This is not simply 'machine vision', we argue, but a making operative of the visual by platforms themselves.

In the context of debates in visual cultural theory, software and science studies about the status of the image and the performance of algorithms as culture, we attend instead to ensembles; platforms; to the specificity of algorithmic processes; and to the conjunction of algorithm with device hardware. We want to suggest that the conjunctions between image ensembles and new modes of observing fundamentally challenge the claim for an 'autonomy' of computer vision associated primarily with artificial intelligence and its technical achievements. We suggest that the abundance of devices for producing images, together with the invisuality of platform seeing, create new opacities that even the most advanced seeing-devices – the machine learning-based predictive models used to organize and order image flows – cannot dispel. The imperative to make visible the invisible, so prominent within data science, data-inflected design and big data-driven commerce, may well be out of sync with the novel formations rendered by a radical platforming of visual culture.

## Ensembles

Glance through any contemporary data science article or developer's blog post on deep learning networks and you will quickly glean that images are discussed in quantitative terms that overwhelmingly obliterate any sense of their uniqueness, indexicality, or their perceptual or phenomenological experience. 'Millions of training examples' and 'an

endless stream of new impressions' (Mordvintsev et al., 2015), or 'we trained for a total of 50 million frames' (Mnih et al., 2015: 524) are typical phrases scattered across research reportage on contemporary forms of artificial intelligence (AI). These forms of AI are driven by deep learning architectures, in which images seem to function merely as a communicative baseload that might power new 'visions' for automated and autonomous decision-making and task performance. Ever since Google loaded 10 million YouTube thumbnails of cats into its neural network architecture in 2012, the conception that what an AI needs to learn to become more proficient is 'more (image) data' has taken hold.

We may indeed have become used to the mathematical sublime underpinning the term 'big data' through Google's cat-loaded artificial brain or by Facebook's facial recognition review and tagging features of its hundreds of billions of users' photos. But we have paid less attention to the imagistic nature of all this data. How has this collecting, formatting and processing of images – undertaken by emerging software, hardware techniques and processes that constitute a broader AI assemblage – affected what we understand of 'the' image and of observation today? We will suggest that attention to images as not just quantitative aggregates but rather as ensembles is needed in order to understand, first, how new forms of AI are assembling across technical, social and economic platforms; and, second, how indeed a new mode of nonrepresentational observation has become ascendant that we propose is *invisual*. Here, observation operates in and through the image but is not *of* the order of the visual.

Debates around contemporary images, especially scientific as well as media and entertainment ones, have often struggled with the problem of how to make sense of their abundance, as well as their increasingly opaque and obscure relations with referentiality. They have, however, rarely paid attention to the consistency and relationality of images as collections or sets. Critical theorists have long observed and commented on the increasing intimacy of machines or devices with vision. Whether in the Foucauldian-style approach of Jonathan Crary's *Techniques of the Observer* (Crary, 1992), the exponential acceleration of seeing described in Paul Virilio's *The Vision Machine* (Virilio, 1994), or the systemic ontological transformation proposed in Martin Heidegger's 'The Age of the World Picture' (1977), vision has been comprehensively described in terms of a machinic capture of seeing, and an increasing autonomy of vision techniques. Accounts of perception and visuality more attuned to practical transformations in seeing have tended to focus on shifts in the instrumentation of objectivity, and corresponding re-configuration of the position of observers and witnesses; for example, Lorraine Daston and Peter Galison's (2007) work on objectivity. But are these approaches still relevant for the transformations currently taking place via emergent

assemblages of and for vision predicated on image collections, the situational awareness being tasked for autonomous vehicles, or the Google DeepDream mission to evolve not simply intelligent but imaginative machines?

Drawing on science studies, visual cultural theory and software studies, this article explores some concrete settings and situations in which seeing (large collections) of images occurs today. We are especially interested in the aggregation of images around and through the modelling and predictive practices of deep learning since this field of technical practice comes into being in association with image ensembles. Within data science, it has been taken for granted that 'big data' must both be furnished for and be an artefactual consequence of the effective and robust training of deep learning AIs or 'models' (for example, Najafabadi et al., 2015). Yet the pervasiveness of image data throughout all domains in which models are being trained – even, for example, where the modality being trained for is aural and the domain being mined is music (for example, Jansson et al., 2017) – is startling, albeit often glossed over.

Our interest here in exploring the exchange between images and emerging platform assemblages is not to examine the onslaught of a monolithic 'algorithmic culture' (see Pasquale, 2015), figured either as algorithmic intelligence or dystopian vision of cultural expropriation (for example, the dystopianism of China's new social credit system that will be shaped and coordinated algorithmically). Instead, our approach will be to ask how large-scale collections of images, their persistent processing by everyday devices, their storage and analysis by media and hardware platforms from Google to NVidia and more, and their agency as (iterative) training grounds for deep learning networks, allow them to operationalize both a new mode of observing and new *agencement* for visuality that we are calling platform seeing. Here we deploy the term *agencement* in its original French following Deleuze and Guattari who understood the work done by an *agencement* as increasing the multiplicity of its conceptual and social (and aesthetic) dimensions as it expands and changes its agencies via connections with other machines both technical and social (Deleuze and Guattari, 2005: 8). We see this as important in the context of our arguments concerning platform seeing since both platforms and deep learning are enactive machinic agents that produce intensive and extensive sociotechnical multiplicities as they operate. The functions being performed by and on images as they are precisely formatted for inputting to models and in many cases labelled, as they are processed and used to configure small neural networks on board smartphones, and as they move from the devices of consumers to platforms and their data centres and back, transform them from bearers of indexical relations to elements within operational (image) collections. Along the way, images as unique bearers of content or expression finally give way to a multiplicative matrix of visual transformations. What

counts for platform seeing are images' capacities to yield and supply efficient edge detection, contrast ratios and recognizable compositional elements at immense scale and mobility.

The corollary to this reallocation of the image into particular kinds of ensembles that have become operative for deep learning architectures and the platforms they now enable is that it is impossible to see this operativity in any holistic or meta-observational display. The massive flows and iterations of images across and within devices, platforms and deep learning models are *plat-formatted* in operation. What do we mean by 'plat-formatted'? Take again Facebook's processing of its users' uploaded photos. These function simultaneously as content for users to share in social networks, data to train Facebook's image recognition models, a 'smart' collection able to serve selections via 'genre' (baby photos, pet action shots and so on) back to its users, and a portal into everyday life via its mobile app Moments. This networked 'image' collection is made possible by the intermediating agency – at once technical, cultural, economic and political – of the platform as ongoing operations that transform, order and circulate. Yet these multiple intermediations cut off the capacity to single-handedly or even collectively observe these, often, imperceptible movements, passes and processings of images. Seeing is performed by a multitude of human and computational agents whose 'vision' passes across and along platforms, eluding any singular coordinating position, and heterogeneously conjoining things and practices. Images as ensembles both heed and feed the technocultural logic of platforms; platforms are expanding into perception *agencements* via the contemporary operations of large-scale image collections and flows.

## Platforms

Ironically, the formatting of operations, as various visual processes and materials pass transversally through platforms, *cuts off* the ability to see across, look at, or step back and observe the vast array of contemporary distributed imaging operations. The platform itself clears visuality of such 'oversight'. Today, devices themselves perform many of the operations through which observation becomes (a) distributed event. Indeed, some devices specifically integrate distributed observation events by embedding the platform as their design matrix. The smart phone camera, for example, incorporates an entire raft of processes dedicated to and enabling (data) visualization that are distributed across its various sensors and chips. We might therefore begin to think of the smart camera as a device that supports an entire ecology of platform seeing practices, and we will unpack this a little later. But while now immanent to the camera, the platform nonetheless does not furnish a 'viewing' position from where, even at microcosmic scale, imaging operations can

themselves be 'seen' in total. By this, we mean that there is no position or place from which an 'observing subject' could view the ensemble of operations of image processing; either such operations are too small, since they take place on a microprocessor such as the image sensor of a camera; or they are too large, since the image's operativity only becomes clear by moving in a multi-scalar manner: across image databases, GPU arrays, server farms and data centres.

Platform 'seeing', we'd like to propose, is operative – only ever produced through the distributed events and technocultural processes performed by, on and as image collections are engaged by deep learning assemblages. Such distribution is both a logistical operation by platforms – Facebook, Apple via CoreML, NVidia, to name only a few – and an actual plat-formatting of observation. From now, seeing, as a continually mobilized set of perceptual and machinic operations, is re-configured via the generating, processing, and distribution of image data. The formatting and databasing of images, the re-formatting of image data as recognizable patterns through deep learning models and the re-assemblage of such patterns as predictive mechanisms for the very near future are all operations of platform seeing. Platforms, however, do not proffer a location for (a) single oversight.

We focus on three pivotal platform features: the platform as image collecting apparatus, machine learning algorithms as images of platform operations, and platforms as image-forms (what we will later term 'plat-formatting'). Platforms include devices (for example, games consoles or smart phones or satellites), instruments (medical scanners, for instance), and distributed systems (social media). The concept of platform has a complex relation to media regulation (Gillespie, 2010), to knowledge (imaging platforms associated with Keating and Cambrosio, 2003), and to an ever-expanding range of economic activities (Srnicek, 2016; McAfee and Brynjolfsson, 2017). Platforms constitute a privileged space of relationality between different groups and forms of belonging. Across all platforms, the problem of their relation to capitalism and, in particular, the ongoing structuring of things as assets capable of generating revenue streams (Doganovam and Muniesa, 2015), is writ large. Platform-specific image collections are primary assets. Many social media platforms amass vast collections of images generated by phones and cameras. Although these images were initially collected as part of the archival logic of social media platforms (Hogan, 2015b), those collections have now begun to take on a different value as future-oriented assets.

Although it is rarely mentioned as such, critical interest in problems of algorithmic accountability or agency (Kitchin, 2017; Aradau and Kaufmann, 2017; Neyland, 2016; Sandvig et al., 2014) follows in the wake of a significant re-configuration of platform algorithms. A diverse ecology of algorithms for sorting, counting, searching, finding, naming, ordering, and calculating things has accumulated around digital

platforms, but the algorithms that currently warrant critical attention are particularly concerned with predictive operations. One problem in understanding the work of 'algorithmic culture' (Hallinan and Striphas, 2014) or 'algorithms as culture' (Gillespie, 2017; Seaver, 2017) is that they largely overlook the image-centred logics of these changes. The polymorphous shifts in the algorithms should be framed, we are suggesting, by the large image collections on which platforms increasingly capitalize. Predictive work with images also has a close association with platforms.

## Images: 'Only the Pixels'

Demonstrations of machine seeing associated with science, internet media, and government rely on collections of images formatted for observation by neural networks. The existence of the collection, ensemble or aggregate of images is crucial to the model/AI. As we have already noted, Bergson assumes a plurality of images as world. In his pre-perceptual account (in the sense of perception understood as framing), images can be recast as a fluxing signal ensemble, an assembly of pathways that can be traversed in many different ways. Crucially for our purposes, acts of seeing or observation in the form of perceptions are provisionally stabilized derivatives of the image ensemble, and do not differ from the ensemble except in their connection to potential movements of observers (see also Flaxman, 2000: 94). Picture representations – images in the familiar sense – only differ, in turn, in their degree of connection or disconnection to potential action. An image in the form of a photograph, for Bergson (1991: 38–9), merely has a greater degree of virtualization of potential action than human visual perception.

What happens if we start from the primacy of image ensembles? Take, for example, DeepMind, the London-based deep learning startup bought by Google Corporation in 2015. DeepMind's activities – creating powerful game-playing artificial agents, reducing energy consumption in Google data centres, and working with Moorfields Eye Hospital on analysing OCT (Optical Coherence Tomography) eyescans – span media, industry and medicine. The contrast between game playing, data centre energy management, and eye disease already suggests a certain mobility or translatability of the techniques and devices that DeepMind has been developing; that is, deep learning neural networks combined with reinforcement learning. These settings, we suggest, share the configuration of the plat-formatted image ensemble. And, in spite of the different kinds of imaging processes – high resolution eyescans or small thumbnails – all kinds of images must be gathered together in a way that makes them both model readable *and* platform ready.

In 2016, DeepMind entered AlphaGo, the Go-playing system, in a match against the world's leading human Go player, Lee Sedol in

Seoul. AlphaGo won the Go tournament. This success has been widely seen as a sign of a historic step towards artificial intelligence. It has less often been recognized that AlphaGo's relation to the game of Go is not based on any abstract game intelligence, but on a vast archive of images of previous games of Go engineered imagistically to format moves in Go games. In a paper describing AlphaGo published in *Nature*, the designers and engineers directly point to an image ensemble:

> Deep convolutional neural networks have achieved unprecedented performance in visual domains.... They use many layers of neurons, each arranged in overlapping tiles, to construct increasingly abstract, localized representations of an image. We employ a similar architecture for the game of Go. We pass in the board position as a $19 \times 19$ image. (Silver et al., 2016: 484)

As Silver and co-authors describe it, AlphaGo deploys an architecture developed for 'visual domains' (face and object recognition) by treating Go board positions as 19x19 pixel image snapshots. When the DeepMind scientists write 'construct increasingly abstract, localized representations of an image', we might bear in mind that the neural networks represent by creating new images embodied as layers of different zones in the image ensembles. At core, deep neural networks are 'exploiting the local spatial correlations present in images, and building in robustness to natural transformations such as changes of viewpoint or scale' (Mnih, 2015: 529).

We draw attention here especially to the primacy of image ensembles; the model trained via the DeepMind platform 'learns' by observing many images. AlphaGo acts in the world to the extent that local spatial correlations can be associated with actions and rewards for those actions. The development of these systems centres on many cycles of observation followed by action. This cycling through observation and action constitutes the 'training' of the model; a training that seemingly requires very little 'prior knowledge' on the part of the model since it only receives pixels and game scores as input.

## Diagram

How should we think of the millions of images that AlphaGo processes? While the researchers at DeepMind refer to 'abstract, localized representations of images' (Silver et al., 2016: 484) or deriving 'efficient representations of the environment from high-dimensional sensory inputs and [using] these to generalize past experience to new situations' (Mnih et al., 2015: 529), are the models embodied as AlphaGo representations of images? The difficulty in understanding them as representations of images is that the representations 'act through' themselves in ways that cannot be represented. The representation of the images moving *en masse*

through AlphaGo Zero cannot itself be effectively presented as a representation, or at least not without another model, or a series of technical diagrams. Any talk of 'abstract representations' is misleading here. The model trained on an image collection is rather more like an image as described by Bergson, acting 'through every one of its points upon all the points of all other images, to transmit the whole of what it perceives' (Bergson, 1991: 36). It is possible to comprehend the complex and dynamic configuration of these devices (through 'training' techniques that optimize parameters to reduce error rates on subsequent predictions) as deriving 'efficient representations of the environment' where 'the environment' is something that can be rendered as an image collection. The combination of different techniques such as convolutional neural nets brought to bear on image collections creates a 'present image', but one that does not primarily work as representation of the world but 'produce[s] a new kind of reality, a new model of truth', as Gilles Deleuze and Felix Guattari put it (1994: 36), a model whose novel reality lies in the ensemble. In other words, machine learning systems such as AlphaGo operate diagrammatically, re-flowing relations in the image ensembles, generating materialities and experiences in their wake.

The analytical challenge for contemporary theory is to reframe the operations of such machines in ways that allow this primary and over-flowing diagrammatic dimension associated with image collections to appear in less abstract, less representational ways. The challenge is to understand them as situated, operative and as generative of new kinds of actualities. Amoore and Piotukh use Bergson to re-frame data analytic systems in terms of attenuated forms of perception (2015: 360). By contrast, we want to suggest that there is something more transversal happening in image aggregates. The transformations of images via a machine such as AlphaGo Zero are not so much technical (although as always there is a maze of technical details to navigate) but diagrammatic. The zigzag relations between a complex AI architecture such as AlphaGo and other 'machines' generate transversal platform changes, to which we now turn.

## Device

As platforms start to reorganize perception, our everyday vision machines – such as cameras – are also rapidly being plat-formatted (Plantin et al., 2018). Much theoretical and cultural work has been done on the actual images produced by smart phones – in particular recent critical media, photographic theory and software studies approaches to selfies. As Frosh (2015), Gómez Cruz and Thornham (2015) and Levin (2014) have all saliently argued, the emergence of the 'selfie' as an image stream rather than mere photographic practice is indebted to the place of the image within much larger networked

media flows. Scholarship on smart photography largely revolves around the techniques and technics enabled by the phone's front-facing camera. Nonetheless, all this research still emanates from conceiving both image as representational and human subject as the main perceiver within these flows. With the exception of Sarah Kember's (2014) work on the ways in which 'smart' photography emerges out of a re-assembling of imaging, informatics techniques and biopower, little work has yet been carried out on imaging *operations* in devices.

Yet a camera is no longer an image-taking device but rather an entire sensing 'platform' capable of carrying out the distribution and integration of different forms of processing. A device, then, becomes another location for the generation of image *ensembles*. But if the camera in general has moved from imaging to sensing device (Yoshida, 2015), it is the smartphone camera that has emerged as a conjunction for platform architectures and movements. The smartphone plus camera really discovered its operativity in 1997, when Philippe Kahn, a Silicon Valley technology developer, shared a photo of his baby by using his own wireless sharing software and a camera, both of which he integrated to function with his mobile phone. This was based upon a proto-platform web architecture: 'a web/notification system that was capable of uploading a picture and text annotations securely and reliably and sending linkbacks through email notifications to a stored list on a server and allowing list members to comment' (Kahn, 2017). Kahn's recollections underscore how mobile phone-generated image culture provides the pathways for the aggregation of web-servers and mobile phones into platforms.

As a number of technical and theoretical histories have noted, the in-phone camera developed primarily as a sensor of images rather than as an optical device for focusing light (Yoshida, 2015; Chesher, 2017). An image sensor is itself already a mini-workstation for image processing: in very recent high resolution digital cameras, sensors are panchromatic and arranged in an array. The position of sensors in the array conveys information about the amount of light being recorded. In slightly older digital cameras (2015 and before), sensors also occupy a mosaic grid but additionally use a 'Bayer' mask of 50 percent green, 25 percent red and 25 percent blue to filter the light being received. In both cases, it can be seen that a sensor does not merely receive light but processes light quantities alongside or in tandem with other information. Further processing then takes place, in the case of the co-operativity of sensor and Bayer filter, using 'demosaicing', a mode of interpolating the colour information of neighbour sensor elements to generate the digital image. As part of this new formation of plat-formatted seeing, in-camera 'observers' have become miniaturized to the level of individual sensor-pixel components. 'Higher' levels of observation might then be 'filled in' algorithmically (see also Rubinstein and Sluis, 2013).

The shift from recording to sensing is not confined to smart phone cameras. But image quantity and quality from camera phones accelerated early research on the camera's image sensor and fundamental processing capabilities. In order to resolve some ongoing smartphone photography issues such as motion blur and poor quality image capture in low light, image signal processors (ISPs) have become part of the computational architecture of smart phones. The ISP does not just process signal from the image sensor but also receives data from, for example, the gyroscope, which provides image stabilization and combines both signals into the one digital image. This assembling of different image data inputs through ISP chips, along with the capacity to generate digital images at standardized image resolution and size, creates a new kind of relationality among (smart phone) image ensembles that furnishes some of the conditions for machine learning-driven AI. Together with its networked connectivity, the smart phone offers a platform-ready form of imaging available for movement and distribution across and by the *agencement* that is platform seeing.

The reverse also holds: the smart phone image with its in-camera/in-phone processing capacities functions as an element in and for the distributed event of observation writ large. The multi-scalar pervasiveness of observers distributing and distributed across devices and image processing transforms the image from representational to purely processual. To return here to Bergson, processing is what acts throughout the image ensemble to actualize the ensemble's movement toward a form of action that is radically invisual. Here, it is image processing that conditions the image ensemble as platform-ready data (see Alpaydin, 2016: 99–100), becoming the 'input' for plat-formatted deep learning architectures that drive, for example, Google's Photos app.

The increased reliance on ISPs in smartphones to deliver image quality has become a hardware 'hook' for artificial intelligence to insert itself pervasively into everyday life, fostering a ubiquitous, platform-driven consumer-level deep learning. The A11 Bionic released in 2017, iPhone 8's chip, is optimized for image and video signal processing with a 64-bit and 6-core processor. But it is also optimized to work for machine learning using Apple's CoreML platform. This 'platform' (in a localized sense) enhances image and facial recognition among its raft of AI capabilities, which also include object detection and natural language processing (see Shaji, 2016). Hence the everyday contemporary image-ready device of the smart phone now forms part of the *agencement* that is platform seeing. Indeed, the platform has come home to roost in/on the phone with the entire device now becoming capable of 'conjoint operations': the processing of data with the delivery of data streams to installed apps, with these apps having been pre-trained using CoreML algorithms (see Newman, 2017). Both Apple and Samsung have invested considerably in developing proprietary ISPs, suggesting that plat-formatting moves both at the

level of and through the technical operation fundamental to contemporary data culture: image processing.

At the same time, the ascension of machine learning-driven AI also means that new kinds of platforms cut through the operations of imaging, as we will explore in the next section of this article in relation to the shift in computation from central (CPU) to graphics processing units (GPUs). In the context of smart phone image processing, one example that comes to mind here is the foray that the company ARM has made in to the smartphone market with the release of a new ISP in 2017 (Savov, 2017). ARM had been providing many of the CPUs for iPhones, iPads, Samsung Galaxy devices, and Google Pixels. But their 2017 ISP, aimed at the Chinese and south Asian market and embedded in Meizu, Huawei, and Xiaomi phones, is a downsized iteration of their image recognition processors for autonomous vehicles. This suggests that the plat-formatting of image processing also operates to generalize across both scales and domains in similar ways to how DeepMind performs conjunctions across gameplay and eye-scanning to generate the emergent *agencement* of platform seeing.

## Architecture

In order to grasp the structural texture of the changing position of images on and as platforms, we might turn to the other prominent strand of DeepMind's work that has not yet been discussed: its modelling of energy consumption in Google data centres. As in the case of platform and algorithm, a train of recent research has addressed the energy and environmental implications of data platforms and digital media. Whether in calculations of the energy costs of a Google search or in the $CO_2$ emissions of Facebook information architecture, critical attention to energy has been one way of grounding changes in platforms in their localized materialities (Hogan, 2015a). Platforms such as Facebook and Google, and even Amazon, which for many years showed no regard for $CO_2$ emissions, attend closely to energy consumption, partly for reasons of cost and partly for reasons of brand. DeepMind's modelling of Google's data centre operating efficiency using the same neural network techniques developed for AlphaGo and deep Q-networks attests to this interest (Evans and Gao, 2016). While it would not be convincing to attribute improvements in efficiency to either the critical literature or the machine learning models developed by DeepMind, the application of the image-based predictive models to data centre operations has an interesting recursivity to it. Platforms structurally change their own ensemble configurations today partly through such operational modelling.

Images animate the dynamism of platforms. Even if it is somewhat surprising that the same modelling techniques can be used to play Atari

console games, Go, analyse eyes, drive the 'smartness' of digital photo-
graphic practices or analyse data centre power consumption, it is perhaps
even more striking that all of these models depend on the graphic proces-
sing units – GPUs – first developed as hardware accelerators for compu-
ter game graphics in the early 1990s by the Taiwanese chip manufacturer
NVidia. The visual culture of real-time digital animation in gameplay has
an unlikely or non-local relation to machine learning through GPUs,
with their capacity to render calculation massively parallel. The architec-
ture of the GPU (barely mentioned in the scientific publications in *Nature*
but discernible in their methods sections and some of the tables of results)
is characterized by several thousand identical computing cores connected
in a grid on a single chip. GPU architectures initially addressed the often
subtle visual problems of texture and shading in computer graphics, but
in the context of deep learning neural nets, the same architecture is con-
cerned with a very different problem: the optimization of predictions
through the process of training a model. GPU architecture, the silicon
substrate of millions of first-person standpoint 3D action games, with
their pursuit of detailed and fluidly mobile game physics, has developed
to render images aggregately computable through massive calculative
parallelism. Graphics processing units – as distinct from central proces-
sing units that were the core of computation until only a few years ago –
specialize in vast numbers of discrete arithmetic operations carried out in
parallel lanes that generate images.

Whereas graphics applications used calculation to render high resolu-
tion shaded and textured-mapped images at speeds sufficient to support
'real-time' game play, the training of convolutional neural nets used in
AlphaGo, Q-networks and other similar models direct most of the cal-
culative capacity of the GPU to matrix multiplications of lower resolu-
tion input data (images) and model parameters (the so-called 'weights').
The architecture of deep learning neural nets is intricate, typically invol-
ving a dozen or more distinct layers, each of which has thousands of
elements with multiple parameters. While the images processed by
AlphaGo are quite small (19x19 pixels), training the model entails
many thousands of repeated adjustments of parameters over different
image aggregates (sometimes training sets; sometimes subsets of larger
aggregates; sometimes vast aggregates; sometimes combinations of all
these), to reduce prediction errors.

The techniques of training such models have intricate and heavily
mathematical underpinnings, but nearly everything that happens in the
construction of such models can be understood as reversing the flows of
image production that have defined visuality in recent decades. Instead of
generating images, these models observe images, they construct diagram-
matic abstractions of features common in images, and gather these loca-
lized abstractions into predictive statements that can be operationalized
as actions in the world: 'place a black stone at g9 (row 7, column 9)'. We

might think of the calculative observation of image collections as a generalized visuality since the structure of the eye (in the work on eye disease), the highly variable data power consumption of a data centre, edge detection for enhancement of digital photographs, or a series of moves in a game of Go or *Montezuma's Revenge* all operate according to the same logic: a large image collection allows a model to be trained given the computational capacity of GPUs. If computer vision now has 'superhuman' capacities to recognize things and faces, this in effect amounts to a generalized seeing taking shape in which operational situations are observed and potentially changed in a similar way.

### End: 'Image'; Begin: 'Images'

Engagement with the technical architectures of ISPs, GPUs and deep learning models offers a way to disinvest some of the numerical sublime that inflects responses to image aggregates, or indeed any 'big' data. Our central concern in this article has been image collections in their platform mode of existence as an index of an emergent platform seeing. Collections of images operate within and help form a field of distributed *invisuality* in which relations between images count more than any indexicality or iconicity of an image. This distributed invisuality constitutes the diagrammatic transversality that *moves* their mass. Bergson's account of images as ensembles asks us to turn from the perceived image to images, bathing in a flux of associations. He treats perceived images as forms temporarily cut out of the mass and held steady. Invisual image ensembles move across any clear contrasts between eye, lens, sensor, file, screen or database. The corollary of the processing of collections is that the platform itself as a raised, visible surface of intermediating cuts off the capacity to look from a position because it distributes seeing transversely: pixel, device, and hardware architecture are conjoined through the operations of this invisual diagram.

What are the implications of this engagement for accounts of platforms, media and contemporary experience? We have approached the image ensemble from several different platform-related angles. First, we have sought to show in the case of the deep learning models epitomizing platform data extraction – AlphaGo, Q-Network, etc. – that image collections are treated as large-scale patterns of associations between features. Second, in exploring the generalization of these architectures in environments that do not seem to be primarily image-based scenes, such as playing a board game, we have suggested that the platform-enabled modelling of associations enhances the importance of image collections for platform operativity. Third, in charting how the devices that produce images such as smart phones and cameras themselves undergo platformization, we have observed how image-sensing becomes the site of platform-derived predictive operations. These operations again work to generate baseload flows of

images that generalize the image collections into a new kind of ensemble. Taking into account these different perspectives we have brought to bear on contemporary 'image' operations, platform seeing strives for a generalized and multi-scalar yet precisely targeted vector destined to operationalize visuality wherever it lands.

## Notes

1. Why platform seeing? It is platforms such as Amazon, Google, Microsoft and many others who have been energetically marketing their image collections and the technical configurations needed to process them. But the recent re-emergence of AI owes much to platforms, and to their image ensembles. Machine learning, long pursued by researchers and engineers, suddenly gained traction and accelerated as 'AI'. We could track how the AI 'break-throughs' of recent years stem from the image collections associated with sharing photos on the internet combined with the hardware (GPUs) developed for computer gaming. This article begins this process of tracking how such breakthroughs occur by paying particular attention to the platform readiness of image datasets and the proliferation of the multi-levelling of all kinds of technical entities such as devices into platform.
2. Amoore has also examined the ways in which cloud computing's machinic vision acts on the world by rendering perceptible that which was previously imperceptible. Here she develops another aspect of Bergson's ontogenetic approach to seeing by suggesting that human and machine are 'organs' that function as a composite kind of perception (Amoore, 2018: 12).

## References

Alpaydin, Etham (2016) *Machine Learning*. Cambridge, MA: MIT Press.

Amoore, Louise (2018) Cloud geographies: Computing, data, sovereignty. *Progress in Human Geography* 42(1): 4–24.

Amoore, Louise and Piotukh, Volha (2015) Life beyond big data: Governing with little analytics. *Economy and Society* 44(3): 341–366.

Aradau, Aradau and Kaufmann, Mareile (2017) Politics of prediction. *European Journal of Social Theory* 20(3): 373–391.

Bergson, Henri (1991) *Matter and Memory*. New York: Zone Books.

Chesher, Chris (2017). Between Image and Information: The iPhone Camera in the History of Photography. In Jason Farman (eds), *Foundations of Mobile Media Studies: Essential Texts on the Formation of a Field*. Oxford and New York: Routledge.

Crary, Jonathan (1992) *Techniques of the Observer: On Vision and Modernity in the Nineteenth Century*. Cambridge, MA: MIT Press.

Daston, Lorraine and Galison, Peter (2007) *Objectivity*. New York: Zone Books.

Deleuze, Gilles and Guattari, Felix (1994) *What Is Philosophy?*, Tomlinson Hugh and Burchell. Graham (trans). New York: Columbia University Press.

Deleuze, Gilles and Guattari, Felix (2005) *A Thousand Plateaus*. Minneapolis: University of Minnesota Press.

Doganovam, Liliana and Muniesa, Fabian (2015) Capitalization devices: Business models and the renewal of markets. In: Kornberger, Martin,

Justesen, Lise, Koed Madsen, Anders and Mouritsen, Jan (eds) *Making Things Valuable*. Oxford: Oxford University Press.

Evans, Richard and Gao, Jim (2016) DeepMind AI reduces Google Data Centre cooling bill by 40%. *DeepMind*. Available at: https://deepmind.com/blog/deep mind-ai-reduces-google-data-centre-cooling-bill-40/ (accessed 21 March 2019).

Flaxman, Gregory (2000) Cinema Year Zero. In: Flaxman, Gregory (ed.) *The Brain Is the Screen: Deleuze and the Philosophy of Cinema*. Minneapolis: University of Minnesota Press .

Frosh, Paul (2015) The gestural image: The selfie, photography theory and kinesthetic sociability. *The International Journal of Communication* 9. Available at: http://ijoc.org/index.php/ijoc/article/viewFile/3146/1388 (accessed 21 March 2019).

Gillespie, Tarleton (2010) The politics of platforms. *New Media & Society* 12(3): 347–364.

Gillespie, Tarleton (2017) Algorithmically recognizable: Santorum's Google problem, and Google's Santorum problem. *Information, Communication & Society* 20(1): 63–80.

Gómez Cruz, Edgar and Thornham, Helen (2015) Selfies beyond self-representation: The (theoretical) f(r)ictions of a practice. *Journal of Aesthetics and Culture* 7. Available at: http://www.aestheticsandculture.net/index.php/jac/article/view/28073 (accessed 21 March 2019).

Hallinan, Blake and Striphas, Ted (2014) Recommended for you: The Netflix Prize and the production of algorithmic culture. *New Media & Society* 18(1): 117–137.

Halpern, Orit (2015) *Beautiful Data: A History of Vision and Reason since 1945*. Durham: Duke University Press.

Heidegger, Martin (1977) The age of the world picture. In: *The Question Concerning Technology, and Other Essays*. London: Harper and Row.

Hogan, Mél (2015a) Data flows and water woes: The Utah Data Center. *Big Data & Society* 2(2). Available at: https://journals.sagepub.com/doi/full/10.1177/2053951715592429 (accessed 21 March 2019).

Hogan, Mél (2015b) Facebook data storage centers as the archive's underbelly. *Television & New Media* 16(1): 3–18.

Jansson, Andreas, Humphrey, Eric, Montecchio, Nicola, et al. (2017) Singing voice separation with deep U-Net convolutional networks. Paper presented at the *18th International Society for Music Information Retrieval Conference, 23–27 October 2017, Suzhou, China*.

Kahn, Philippe (2017) The Creation of the Camera-Phone and Instant-Picture-Mail Available at: http://www.philippe.com/2017/06/09/the-creation-of-the-camera-phone-and-instant-picture-mail/ (accessed May 15 2019).

Keating, Peter and Cambrosio, Alberto (2003) *Biomedical Platforms: Realigning the Normal and the Pathological in Late-Twentieth-Century Medicine*. Cambridge, MA: MIT Press.

Kember, Sarah (2014) Face recognition and the emergence of smart photography. *Journal of Visual Culture* 13(2): 182–199.

Kitchin, Rob (2017) Thinking critically about and researching algorithms. *Information, Communication & Society* 20(1): 14–29.

Levin, Adam (2014) The selfie in the age of digital recursion. *InVisible Culture* 20. Available at: http://ivc.lib.rochester.edu/the-selfie-in-the-age-of-digital-recursion/ (accessed 21 March 2019).

McAfee, Andrew and Brynjolfsson, Erik (2017) *Machine, Platform, Crowd: Harnessing Our Digital Future*. New York: W.W. Norton & Co.

Mnih, Volodymyr, Kavukcuoglu, Koray, Silver, David, et al. (2015) Human-level control through deep reinforcement learning. *Nature* 518(7540): 529.

Mordvintsev, Alexander, Olah, Christopher and Tyka, Mike (2015) Inceptionism: Going Deeper into Neural Networks. Google AI Blog https://ai.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html (accessed May 15 2019).

Najafabadi, M, Villanustre, Flavio, Khoshgoftaar, Taghi, Naeem, M, et al. (2015) Deep learning applications and challenges in big data analytics. *Journal of Big Data* 2(1). Available at: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-014-0007-7 (accessed 21 March 2019).

Newman, Lily Hay (2017) Apple's machine learning engine could surface your iPhone's secrets. *Wired*. Available at: https://www.wired.com/story/core-ml-privacy-machine-learning-ios/ (accessed 18 January 2018).

Neyland, Daniel (2016) Bearing account-able witness to the ethical algorithmic system. *Science, Technology, & Human Values* 41(1): 50–76.

Parisi, Luciana (2013) *Contagious Architecture*. Cambridge, MA: MIT Press.

Pasquale, Frank (2015) *The Black Box: The Secret Algorithms that control Money and Information*. Havard: Harvard University Press.

Plantin, Jean-Christophe, Lagoze, Carl, Edwards, Paul N and Sandvig, Christian (2018) Infrastructure studies meet platform studies in the age of Google and Facebook. *New Media & Society* 20(1): 293–310.

Rubinstein, David and Sluis, Katrina (2013) The digital image in photographic culture: Algorithmic photography and the crisis of representation. In: Lister, Martin (ed.) *The Photographic Image in Digital Culture*. London: Routledge.

Sandvig, Christian, Hamilton, Kevin, Karahalios, Karrie and Langbort, Cedric (2014) An algorithm audit. In: Peña Gangadharan, Seeta (ed.) *Data and Discrimination: Collected Essays*. Washington, DC: New America. Available at: https://www.ftc.gov/system/files/documents/public_comments/2014/10/00078-92938.pdf (accessed 17 April 2019).

Savov, Vlad (2017) ARM designed an image signal processor, and that's a big deal. *The Verge*, 2017, https://www.theverge.com/2017/4/27/15449146/arm-mali-c71-isp-mobile-photography (accessed 18 May 2019).

Seaver, Nick (2017) Algorithms as culture: Some tactics for the ethnography of algorithmic systems. *Big Data & Society* 4(2). Available at: https://journals.sagepub.com/doi/full/10.1177/2053951717738104 (accessed 21 March 2019).

Shaji, Appu (2016) Understanding aesthetics with deep learning. *NVidia Developer Blog*, Available at: https://devblogs.nvidia.com/parallelforall/understanding-aesthetics-deep-learning/ (accessed 21 January 2018).

Silver, David, Huang, Aja, Maddison, Chris J, et al. (2016) Mastering the game of Go with deep neural networks and tree search. *Nature* 529(7587): 484–489.

Srnicek, Nick (2016) *Platform Capitalism*. Cambridge: Polity Press.

Virilio, Paul (1994) *The Vision Machine*. Indianapolis: Indiana University Press.

Yoshida, Junko (2015) How camera phones changed the world. *The EE Times*. Available at: https://www.eetimes.com/document.asp?doc_id=1328183&page_number=5 (accessed 19 January 2018).

**Adrian MacKenzie** is a Professor of Sociology at the Australian National University, Canberra. He is the author of four books, the latest being *Machine Learners: Archaeology of a Data Practice* (MIT Press, 2017).

**Anna Munster** is a Professor in Art and Design, University of New South Wales, Sydney, Australia. She is the author of two books, including *An Aesthesia of Networks* (MIT Press, 2013).