

# NMA Bilderatlas: a practice-led model for critically examining machine vision in cultural collections

**Kieran Browne - Katrina Grant**

## Abstract

The transformation of machine vision brought about by deep learning has afforded new ways to make sense of digitised images at scale. However, before adopting machine vision into the digital humanistic toolkit we ought to interrogate how this technology mediates our connection to objects of study just as it appears to bring them closer. This article raises critical issues around the application of machine vision systems in museums and cultural institutions. We describe NMA Bilderatlas, an online interface to the National Museum of Australia's collection structured by the representational schemes of an off-the-shelf machine vision model. The interface serves as a surrogate for our critique of machine vision, which, for lack of interpretability, cannot be directly interpreted. Examining how a machine vision system structures the archive, we identify possible opportunities for cultural institutions as well as likely pitfalls that will arise if machine vision systems are applied uncritically.

## Introduction

Images have proved more resistant to numerical interpretation than text. While there are early examples of digital humanistic work on visual artefacts, these have generally been restricted to low-level statistical descriptions of pixel-values (Stork, 2009; Shen, 2009). Only in the last decade have methods emerged that allow us to quantitatively infer what is depicted in digital images. For museums and cultural institutions these methods are billed as a solution to swollen digital archives and patchy metadata. For art historians and scholars of visual culture, the technology could promise something akin to what "distant reading" promised literary studies: a glimpse of the whole archive, rather than a tiny fraction of it.

The purpose of this paper is to open up a critical engagement with this new wave of machine vision in humanistic inquiry. We argue that machine vision in its current form resists traditional modes of criticism. The computer science literature readily acknowledges that these methods produce behaviours that are obscure even to those who build them. Attempts to render machine learning "explainable" and "interpretable" have made little progress, frustrated by further increases in model complexity. As such, a close reading of these systems is unlikely. If Berry (2011) is right that the digital humanities (DH) has made a critical turn in its appraisal of computational methods, how is this complicated by systems which do not submit to interpretation?

This paper outlines the history of machine vision and the new "deep learning" paradigm. We argue that humanists and those working with cultural collections should consider how these systems mediate objects of study before adopting them wholesale. In the first part of this paper we describe existing critical work from new media studies and digital humanities with regard to computational systems and machine vision. We argue that a different approach is needed in this instance, one that engages in a longer history of critical engagement with art and images and the existing structures of knowledge that we use to describe and to understand this part of our culture. The second part of the paper is devoted to a practice-led exploration into reading the relations in machine vision systems. *NMA Bilderatlas* is an exploratory web interface to the National Museum of Australia's collection. The interface draws together images based on the representational schemes of a standard off-the-shelf machine vision model. The name deliberately cites Aby Warburg's experimental (almost wordless) project to create an encyclopaedic account of cultural memory and the image (in the Western tradition). *NMA Bilderatlas* borrows affordances from Warburg's original to reveal the obscured patterns of the machine vision system and operates as a prompt for critical and theoretical engagement with the results of these technologies set loose on mass digitised image collections from galleries and museums. In dialogue with the interface, we offer a critique of the use of machine vision in cultural institutions.

## AI vs Machine Vision

In this paper we generally refer to our object of study as "machine vision" rather than "artificial intelligence" (AI) or "machine learning" (ML). This is in order to better delineate the systems we are describing from related methods. AI is a particularly slippery term; it has referred to profoundly different technologies and methods over the decades. In general what is called AI today is ML, often specifically deep learning. Even ML entails a broad church of

methods. ML is also not new to DH; it has a long history in aiding humanistic inquiry and is already firmly placed as a DH methodology (Blanke, 2018). We choose to speak of "machine vision" as opposed to "machine learning" in order to foreground the specific challenges images present for quantitative interpretation.

## Machine vision in the humanities

As many scholars have noted, DH has inherited a strong textual focus (see e.g. Svensson (2009), Presner (2010), Hockey (2004), Champion (2017)). Drucker and Nowvickie (2004) even cautiously suggest that this amounts to *logocentrism*. In a sense, computers could always read. The mathematical models of computation that preceded electronic digital computers were themselves indebted to formalisations of language resulting from the linguistic turn in analytic philosophy. As such, foundational models of computation are described in terms of "alphabets" and "grammars."

The first wave of DH began as 'computing in the humanities', or 'humanities computing' and narrowly focussed on text analysis (Berry, 2011). The earliest example of the use of computation in aid of humanistic work is usually identified as Father Roberto Busa's *Index Thomisticus*; a lexical index of the writings of Thomas Aquinas initially conceived in the late 1940s (Hockey 2004). Renear (2004) notes that this project began so early as to be nearly coeval with digital computing itself. In the seven decades since Busa's pioneering work, a plethora of methods have been developed with which to analyse texts with computers. These methods have reached a level of pervasiveness and sophistication such that some scholars find it perfectly reasonable to say that machines can read (Hayles, 2010). Of course, this is a controversial position as exemplified by polemical debates over "distant reading" (Underwood, 2016). Nonetheless, it is clear that the machine's interpretive possibilities with regard to texts are well in advance of other digitally encoded media.

Against this rich history of literary digital humanities, the interpretation of visual artefacts by computational methods is only fledgling. Image-based digital analysis has not been absent, but like other applications of machine vision it has until recently fallen short of the semantic aspect of imagery. When Manovich (2012) wrote "How to compare one million images?" the comparisons described were statistical descriptions of pixel values e.g. the mean brightness, standard deviation etc. These offer fascinating proxies for aspects of form, but they can say nothing of content.

Machine vision has not been entirely barred from the content of digital images, but this is most advanced in service of text. Optical character recognition (OCR) has, for many years, made it possible to decompose digital images of text into constituent characters. The relative ease of identifying alphabetical signs with computers is most likely attributable to the "stabilised" and "conventionalised" visual form of writing systems (Drucker and McGann 2000). It is also possible that almost six centuries of moveable type in the West, and almost a millennium in China, has further stabilised written symbols. Art and object making arguably have done the opposite. Aside from certain persistent genres (portraiture, landscape) and examples of visual quoting and imitation (the Renaissance revival of antique style), art is characterised by a constant push for reinvention, disruption and very deliberate breaks with preceding styles.

## Machine vision as a way of seeing

Machine vision has a well-known origin story. In 1966, during the early days of MIT's AI lab, professors Seymour Papert and Marvin Minsky organised the "Summer Vision Project" and set some students to work on computer vision problems. Minsky, who was then the head of the lab, instructed a student to "connect a camera to a computer and have it describe what it sees" (Anand and Priya, 2020, chap. 1). The request seemed reasonable enough, but decades later no satisfactory solution existed, for this, or indeed for any other research problem in machine vision (Huang, 1996, sect. 3). The story may be apocryphal but its persistent retelling reveals something that those in the field know to be true: machine vision has proved much harder than anyone anticipated.

Exactly how hard depends on which sense of vision is meant. It is trivial, for example, for a computer to identify the brighter of two images or the darkest point in an image. Because digital images are stored as a list of numbers representing the brightness of pixels, numerical comparisons offer good proxies for these questions. Much more challenging are the gestalt and semantic aspects of vision. When a person looks at Magritte's pipe, they see it as a pipe and not as an arrangement of pigments or pixels; indeed, it is difficult *not* to see it as such. But how does one reproduce this feat in a computer program? The necessary numerical operations are far from evident. In Magritte, the human first sees the pipe then can work their way towards examining the structure (composition, style, pigment, brushstroke) but the machine sees the structure, the granular, first and then we have to coax it forward to the point where it sees a pipe.

The history of machine vision can be understood as a slow climb into higher abstractions. Lew et al. (2006) in a review of search in multimedia systems, offer a hierarchy of readability in images. Semantic aspects are some of the hardest to infer. Face detection is a good example of this. Where these approaches made some progress in "natural" (photographic) images, "artistic" images proved more resistant. Hurtut (2010) notes that artistic style perturbs many of the assumptions that one might attempt to leverage to interpret natural images. At least one cross-disciplinary art history and computer science project has investigated the potential to use facial recognition processes to identify anonymous portraits, on the face of it an ideal application of machine vision methods to art (Rudolph et al., 2017). However, this work found two key problems, the first being the lack of examples, a portrait of an unknown person may be the only portrait of that individual. The other is that the micro measurements that facial recognition depend on, which seem to remain relatively consistent across photographs of the same face, are not necessarily present in a painted representation of a face and the subjectivity of artistic representation. Both phenomena are easily explained by art history but serve as useful reminders that similarities and connections that can be made relatively easily by the human eye and traditional modes of humanistic research can be extremely challenging for machine vision. These more optimistic, if not entirely successful, projects are perhaps one reason that the humanities have exhibited a certain pessimism toward the use of computational methods to 'look' at culture, often accusing it of simplistic and uncritical methods that simply reinforce the canon while offering nothing much else of interest (Bishop, 2018). In other words, computational methods can offer nothing that isn't already being done more ably by human researchers. Art critic Horst Bredekamp in an interview published in *Art Bulletin* in 2012, claimed that it would be impossible "even in a thousand years" for a computer to recognise the chair in van Gogh's painting (Wood, 2012, p. 524-5). But already the methods to achieve this were being developed.

Progress in machine vision, like many areas of engineering, is measured in challenges. One of the most significant for machine vision is the annual ImageNet competition, based on the dataset of the same name. The competition requires competitors to build a system which identifies what is depicted in 150,000 photographs, from 1000 possible options. Examples include, "coffee mug", "pier", "neck brace", "sombbrero". In 2012, the third year of the competition, an approach based on the convolution neural network (ConvNet) achieved spectacular results, almost halving the error rates of the next best competitor (LeCun et al.,

2015). Further progress with ConvNets led to greater results in the ImageNet competition year after year, to the point that some now claim that we have reached human-level performance in this task (Geirhos, 2017). This approach has transformed machine vision; ConvNets are now the mainstream method used for almost all recognition and detection problems. The ConvNet has also proved more accommodating of artistic representations than previous methods, and more than capable of identifying van Gogh's chair, as pointed out by Leonardo Impett (Figure 1).

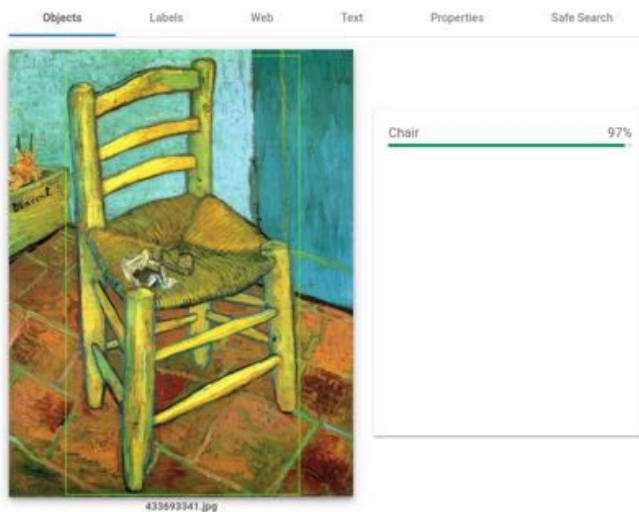


Figure 1: Van Gogh's chair identified as a chair by deep learning algorithm. Tweeted by Leonardo Impett. <https://twitter.com/Leolimpett/status/1225734199061159939/photo/1>

As discussed, the computability of content within photographic and especially artistic images emerged rather suddenly and even unexpectedly. This significantly expands the potential role of machine vision in the interpretive aspects of digital scholarship. Unlike texts, images are not readily searchable in their primary form. Metadata has long been the main way to identify the content of images for digital systems, yet reliance on metadata has known issues. Metadata is always imperfect and partial: it naturally adopts the terms and hierarchies of those who labelled the data, whose intentions might not be shared with your own (Baca et al. 2016; Palmer Albers, 2017). It also, even at its best, reduces images to how they can be described in text. The new wave of machine vision seems to promise a way of seeing more akin to human vision, that is, with an understanding of what is depicted.

These contemporary machine vision methods are already being applied in archives and cultural institutions. Rafdi et al.'s (2015) is the earliest project we have identified. They run a ConvNet on the British Library's collection to populate metadata about possible content.

Europeana has applied similar methods. The executive director of Europeana suggested in 2018 that these methods in audiovisual archives may "improve education, enrich tourist experiences and give meaning to the past" (Verwayen, 2018). In 2019, Europeana put together an AI taskforce to explore the possibilities and pitfalls of these methods (Europeana, 2019). The Metropolitan Museum of Art's digital department believes that AI will transform how audiences connect with art "within the next ten years" (Tallon, 2019). They have established collaborations with MIT and Microsoft to bring about this change. Google is pushing the use of AI as a way to develop novel applications that engage the broader public with the thousands of cultural institutions that have signed up their collections to the Google Arts and Culture program (Google Arts and Culture, 2020). The examples they choose to highlight tend more toward art as entertainment, such as the 'Art Selfie' or basic palette matching, with the exception of the MoMA exhibition archive project where a large scale (30 000 images) archive of exhibition installation photos was turned into a searchable archive of works of art found in the photographs (Google Arts and Culture, 2018).

The excitement is dizzying and not undeserved, but some of the claims made are reminiscent of those made of computational methods in literary studies before an equally passionate opposition emerged and attempted to debunk the whole project. Debates over computational literary studies were and remain highly polarised (Piper, 2018, Introduction). Da (2019) recently argued in *Critical Inquiry* that all work in computational literary studies is either non-significant or wrong. Piper responded that Da's claims rely on selective reading and make sweeping generalisations (2019). We would do well to avoid polemicising machine vision to the same degree, nonetheless, given the claims being made of machine vision, it seems likely that we are at the beginning of a similar dialectic.

## Critiquing computation

Machine vision is rapidly becoming part of the way that researchers, curators, and computer scientists engage with and use constantly expanding digital collections. In light of this, how can we support and extend the types of critical and theoretical engagement that are valued by the humanities? How does machine vision mediate our connection to the objects of study just as it appears to bring them closer? In 2011, Berry described a then-emerging critical move in DH which takes computation as its key issue and applies humanistic methods to this end (2011) . This is notably distinct from what Berry and others have called second wave DH, which applies computation to the usual humanistic objects of study. Posner (2015) calls for something similar but more direct, arguing that DH should rip apart and rebuild the

machinery of the archive "so that it doesn't reproduce the logic that got us here in the first place." This argument suggests that it is not (or should not) be possible to have a DH that simply applies computational methods to traditional disciplinary research questions without questioning the method itself. Many digital systems instead tend to achieve the opposite; Drucker (2013) has noted that reification runs rampant in digital media. The widely reported biases in popular apps from Google Arts and Culture, like the racism perceived in the selfie matching with historic portraits that went viral in 2017-18 and the ensuing discussion of 'the coded gaze', are amongst the most obvious examples (Chen, 2018).

Frameworks already exist, or are in development, that articulate a future for the critical examination of computation. Berry, for example, looks to software studies, platform studies and critical code studies (2011). These fields make up an unofficial coalition, each interrogating a different aspect of digital systems: software studies examines the interface and user-facing aspects; platform studies looks at the underlying hardware; and critical code studies examines the code itself (Marino quoted in Manovich, 2013 p. 42). It is notable that each of these adopts modes from traditional humanities: close reading, deconstruction and so on. Critical code studies most clearly embodies these traditional modes, as is made apparent by Marino's call to treat the code itself as a culturally-embedded text deserving of the deep analysis afforded to works of literature (Marino, 2006). Platform studies and software studies also employ humanistic modes inherited from media studies.

Though each has made a significant contribution to our understanding of the role of computation in shaping society and culture, it seems that none of the approaches taken are entirely applicable to the study of contemporary machine vision. New media scholars have argued that software erases medium specificity (Manovich 2013). Manovich goes so far as to ask whether it is still meaningful to talk about "different mediums" after software (2013). This seems to imply that there really is no difference between images and text in the digital context. All new media are, after all, written in the same, two-valued alphabet, regardless of medium they simulate. Intentionally or otherwise, this aligns with the poststructuralist tendency to construe all objects of study as texts; which is to say as a system of signs.

A key problem is that none of these critical disciplines seem to account for the comparative difficulty of interpreting digital images with computers. Because deep neural networks are universal approximators and their behaviours and representations are extracted from training data, the source code reveals little about how these systems structure the world. Taken as a



software artefact, little can be made of these systems beyond the obvious critique that they are hopelessly obscured. In computer science literature, deep neural networks, like the ConvNets that underwrite machine vision, are commonly described with the metaphor of the "black box" (Breiman, 2001). These methods rely on extraordinarily complex statistical models that tend to perform well when combined with large datasets and significant computational power, and as such do not submit easily to interpretation. This creates a strange dynamic between programmer and program. In a number of cases when cultural biases have been discovered in ML systems, the engineers of those systems have blamed the training data, pleading ignorance (see e.g. Cresci, 2017; Quach, 2020).

We are left with the conundrum that as images become machine visible, the systems that achieved the feat have become obscure. For humanists, the understanding of machine vision as a 'black box' appears to disallow the normal approaches to criticism. If we wish to understand how machine vision mediates images, it seems that we cannot proceed on the basis of close reading of either the software or its source code. Perhaps progress in explainable/interpretable AI will make this possible in the time. However, critical evaluation is still needed. Without it, we allow machine vision to reinforce dominant perspectives. There is no neutral use of machine vision in cultural collections. Our proposed response, in the absence of a readable system, is to attempt a critique by *showing* how machine vision functions in an archive. To do so we have applied a machine vision system to a cultural collection and attempted to visualise this with an exploratory interface that provides an unfiltered visualisation of the manner in which a machine vision system mediates a digitised cultural collection. The aim is to reveal the semantics of the system to give us a position from which to open up a critique.


## NMA Bilderatlas

NMA Bilderatlas is an experimental online interface to the National Museum of Australia's (NMA's) collection. The project builds on the NMA's digital collection explorer and public API to order the museum's collection around the classifications of an off-the-shelf machine vision model published by Microsoft (He et al., 2015). This practice of repurposing pretrained models is common amongst other scholars who have applied machine vision systems to cultural archives (Wang et al., 2016), and represents a possible approach other cultural institutions might take in introducing these technologies to the archive. Unlike similar applications which use machine vision to expand metadata or further describe an image in

text (see e.g. Wang et al., 2016; Daley, 2019), NMA Bilderatlas instead makes the classification central and draws together images from the archive, which then serve as a typology of the class. This juxtaposes the classification with the interpreted images from the archive, serving comparison between like images. The aim of the interface is not as a tool that assists public discovery of a collection (though it could be used like this), but rather one that sets out to provoke a more critical engagement with classification, visual pattern matching and discovery. The interface is intended to support the same interpretive modes afforded by the project's namesake, art historian Aby Warburg's *Bilderatlas Mnemosyne*. Warburg's *Bilderatlas* arranged clusters of images (everything from photographs and postcards to diagrams, adverts and newspaper clippings) across a series of wooden panels. The images ranged from antique sculptures, to medieval painting, engravings of Renaissance ephemeral architecture to a contemporary photo of a woman playing golf (Warburg, 2016; Johnson, 2012). The original *Bilderatlas* was never completed and never really spurred any direct imitators, but it has remained an object of particular interest for many art historians and other researchers focused on visual culture, particularly those trained in the Western tradition. This is in part because Warburg left behind his library of images with its unique classification system ordered by subjects like 'Ritual', 'Gestures' and 'Magic' instead of the more usual artist, period, nationality (Forster, 1976). More recently, however, researchers have been looking at Warburg's methods of work and their resonance with the way that we experience imagery on the web and through digital applications, and how new digital methods could extend (or even fully realise) Warburg's original idea (Patti and Quiviger, 2014; Du Preez, 2020).

The original *Bilderatlas Mnemosyne* project began in 1924 and occupied Warburg until his death in 1929 (Weigel, 2013). It was intended as a three-volume publication but was never completed and is preserved only through photographic records of the more than 60 panels completed by Warburg before his death. Each table included a collection of reproduced images united under a common heading. Images are arranged in tightly packed constellations, revealing commonalities across cultures and eras. Warburg used the notion of "wandering" [*Wanderung*] to refer to the recurring presence of symbols and images across cultures and throughout time, but also as the approach to reading images afforded by his *Bilderatlas* (Weigel, 2013). Presenting imagery in spatial constellations reveals previously unseen relations and resonances as the eye travels freely between the constituent images. Warburg claimed to have traced an afterlife of patterns from antiquity through history and

into contemporary visual culture, many of which had so far resisted attempts at interpretation (Weigel, 2013).



e.g. whiskey jug...

*Figure 2: Screenshot from NMA Bilderatlas entry page.*

NMA Bilderatlas attempts to apply this practice of "wandering" as a way of interpreting the obscured relationships that structure machine vision. The interface appears at first glance to be a standard search driven interface. On arrival a visitor to the page is greeted by a single grey search box in the centre of an otherwise blank white screen (Figure 2). The box prompts the visitor with a suggestion of a query understood by the machine vision model (it is not an open keyword search). The possible queries come from the one thousand classes understood by the pretrained model, though not all of these were present within the NMA's collection, e.g. no panpipes or ostriches were identified amongst the objects in the collection. When the visitor enters a relevant query the machine vision system will identify images in that class and begin to populate the screen, packing the images tightly into available space around the search box (see Figure 3). When the available space is exhausted the process pauses. If the user then moves their cursor around the screen, the page scrolls in that direction creating further space where images continue to fill. This allows the user to wander through constellations of images from tens to potentially hundreds.



involved Indigenous consultation, management and curation (Morphy, 2006), Andrews has noted the sense of unease in parts of the museum amongst Indigenous staffers, especially while these places maintain the “typological conventions and aura of ethnographic storerooms of old” (Andrews, 2017). She notes that through the act of collection, the material is introduced into a system of classification that struggles to contain it (Andrews, 2017). Part of our interest in applying these technologies in the NMA is in how the technology might meet or fall short of the museums stated commitment to plurality.

NMA Bilderatlas aims to promote the kind of interpretive process made possible by its namesake. Humanistic engagements with interface design have developed a language for how interface mediates content. Proximity, order, continuity, etc, change how we make sense of images. An interface, then, performs a “quasi-semantic function” giving structure to the way we read and view ‘content’ (Drucker, 2011). Many of the choices made in designing NMA Bilderatlas serve to impede the ingrained left-to-right, top-to-bottom ordered mode of reading. Humans have an incredible ability to draw connections between pieces of information. How one interprets these patterns depends as much on what one already knows and believes as it does on what is presented to the eyes. Because of this, Drucker argues, our interpretations of the interface must understand the user/viewer “as a situated and embedded subject” (2011). Graphical organisation only provides “provocations to cognition”, the interface therefore is conceived as a collection of affordances which mediate cognitive activities. Whitelaw has previously critiqued standard search interfaces of many cultural institutions for failing to match the richness of our digital collections (2015). “Generous interfaces”, Whitelaw suggests, would support exploration and offer multiple ways into a collection. At its best, NMA Bilderatlas’ arrangements, and the machine vision semantics which they represent, offer just that; new ways into the collection. However, as we will address in the next section, due to a prevailing Western gaze in the machine vision model, these representations fall short of the full richness of the collection.

## Wanderings

We will now discuss a number of cases that reveal qualities of the machine vision system in the NMA’s collection pointing to possible opportunities and pitfalls, which may emerge as machine vision systems are applied in cultural institutions.

## Pier

Below, Figure 4 shows the constellation of images collated for the query: *pier*. It is clear on close inspection that few of the images drawn together actually depict a pier by the dictionary definition of the term. It is difficult to identify a single way in which all the images are alike, and yet there are clearly resonances amongst them. The items appear to have been drawn predominantly from the NMA's collection of postcards and photographs. Many images depict bridges of various architectural styles. There are several images of the Sydney harbour bridge, in its various states of construction. Another image with similar industrial steel structures depicts Broken Hill's silver mines, far from the coast, while another appears to show some kind of quarry. Other images show no sign of bridges or piers but contain bodies of water, some include sailboats. One image appears to depict a partial view of a circus ring, while another shows a section of horse-racing track. In both of these, the long curve of a boundary appears similar to the curve of a bay depicted in another image. Or perhaps the mottled form floor, clearly dirt or sand from context, has a similar pattern to the surface of a body of water. An engineer might group these into positive and negative cases, but this would be to refuse the clear visual accord amongst the images. Their relations clearly go beyond categorical concerns - there are distinct visual similarities across the constellation of images. The constellation of images suggests a broader visual relation is captured than the English language label, "pier", ascribed to it. Rather than reading this as a collection of successes and failings of the machine vision system, if we interpret the images which emerge for a given query as a typology, it is possible to make sense of the actual (fuzzy) semantics captured by the machine vision system.



Figure 4: Screenshot from NMA Bilderatlas with "pier" query.

Scabbard, Bonnet, Goblet

For many queries, NMA Bilderatlas generates a constellation of images which more or less align with objects we would expect to find under that categorical label. Such is the case for *scabbard*, *bonnet* and *goblet*. These are perhaps not the kind of objects one might expect a state-of-the-art machine vision system to have been trained for. However, given the influence of the ImageNet competition, almost all pretrained machine vision systems for object detection developed after 2010, will come off-the-shelf with some capacity to identify these and numerous other antiquated European artefacts. The opportunity for museums and archives to take advantage of this is clear. Indeed, it is unclear who else would benefit from a system that can automatically identify bonnets.

## Slug, Matchstick

In stark contrast to the success with which this off-the-shelf machine vision system drew together Western artefacts and thereby rendered accessible the material culture of Australia's European settlers, the Indigenous objects in the NMA's collection are sporadically and inconsistently identified. The most obvious cause for this is the complete absence of Indigenous artefacts from the machine vision system's ontology. '*May pole*' and '*croquet ball*' are part of the pretrained objects while even iconic and internationally recognisable objects like 'boomerang' are absent. In light of this, we ought to be sceptical of broad claims that machine vision will make archives more accessible. What will be made accessible, and to whom? It is also not just absence that is the problem here - we could perhaps excuse the absence of Australian Indigenous categories in a free piece of software provided by an American technology company. However, also at issue is that this system doesn't simply omit Indigenous material culture, it actually misclassifies it in a way that many would consider offensive. Indigenous objects show up incorrectly in many of the constellations, and while in some cases this can be dismissed as part of the level of misclassifications one must expect of a probabilistic system of this kind, many are nonetheless problematic. A number of examples of boomerangs (the NMA collection includes a significant number and also highlights these as a key Defining Moment on its website) are drawn together by the system under the category *slug*. The *slug* constellation is particularly egregious because there is no apparent visual connection between the boomerangs and the other images identified as similar; some kind of animal preserved in a jar of formaldehyde. Compare this to the clear resonances in the constellation for *pier*. In its present off-the-shelf form, these technologies embody an unmistakable Western-centrism in the categories present, but also in the sense of visual similarity captured. These cases offer a clear warning to cultural institutions looking to apply machine vision in their digital systems. Investment will be required if these systems are to make sense of non-Western artefacts to an acceptable degree.

## Platypus, Wallaby

One might be surprised, given the complete absence of Indigenous categories from the machine vision system's ontology, to find that a number of Australian native animals are in fact included. Platypus, wallabies, koalas and dingos are amongst the animals the ImageNet competitors must train their systems to interpret. While it is encouraging to see these species included in the list, the machine vision model did not appear to function well in the NMA's collection. The *platypus* and *wallaby* queries each identified only five items in the



collection. This seems highly suspect given that the NMA has a large natural history collection as well as a great deal of print culture.

## Maze

In the constellation for *maze* (Figure 5) we see a collection of artworks and images drawn from the collection regardless of culture. Though the text label given to this relation comes from a Western ontology, it is notable that none of the images in the constellation are, categorically, a maze. Instead the actual relation used by the machine vision system seems to be a transcultural one, characterised by concentric lines. We see amongst these images artworks, earthworks, gardens etc. and again we see this clear resonance amongst these images. Conceived as but one way into the collection, these kinds of relations might offer a way of discovering genuinely new cross-cultural connections between objects in the archive given that categorical metadata and even human curators might not have thought to unite these objects. Used appropriately, these relations have the potential to generate what Gurrumuruwuy and Deger (2019) call “unexpected moments of mutuality” and to stimulate intercultural understanding.



Figure 5: Screenshot from NMA Bilderatlas with “maze” query.

## Conclusions

Machine vision has emerged suddenly and unexpectedly with new capabilities that extend the potential horizons of digital scholarship and cultural heritage of images. Though the humanities have a growing literature and methodological toolkit for making sense of digital systems, aspects of this new wave of machine vision render interpretive methods ineffective. We should consider how these systems mediate objects of study before adopting them wholesale. Applying these technologies should not be seen as a goal in itself. We should be wary, as always, of computational solutions to cultural problems. If we accept machine vision as a way of seeing, what are its properties? Which perspectives might it privilege, and which might it restrict? Revealing these representations to human eyes is key to understanding and critically engaging with machine vision. NMA Bilderatlas reverses the standard application of machine vision to make sense of the archive, instead using the archive to reveal the representations and properties of the machine vision system. The interface serves as a surrogate for our critique of machine vision, which at present cannot be directly interpreted. It is clear from our explorations that, perhaps predictably, the machine vision system embodies a Western gaze. For museums attempting to decolonise and pluralise, machine learning may appear to offer an impartial point of view from which to reorganise and represent collections. This is a grave error. Our critique has been framed by the specifics of the NMA, and its collection, however, given the connected experiences of colonialism in many parts of the world, as well significant non-Western material culture held in European cultural institutions, we believe these considerations are not unique to the Australian context. Indeed, the NMA Bilderatlas model has the potential to be deployed within other collections to see what those collections may further reveal about the machine vision system. It could also be used to examine the effects of alternative training sets on the machine vision system, though no alternatives currently come near to the scale and influence of ImageNet. Before accepting machine vision into the digital humanities toolbox, it ought to be brought into dialog with existing (and emerging) modes of critical and theoretical study of images, whether from art history, museum or cultural studies. Such engagement can avoid obvious missteps where those trained in machine vision misunderstand basic precepts of art and object making, and can also offer a more meaningful way to challenge existing hegemonies of knowledge contained both within the institutions and the machine vision systems.

*NMA Bilderatlas is hosted at <https://kieranbrowne.com/bilderatlas/>*

## Bibliography

- Anand, S. and Priya, L., 2020. *A Guide for Machine Vision in Quality Control*. CRC Press.
- Andrews, J., 2017. Indigenous perspectives on museum collections. *Artlink*, 37(2), pp.88-91.
- Baca, M. "Introduction to Metadata." InteractiveResource. Getty Research Institute, Los Angeles, July 20, 2016. <http://www.getty.edu/publications/intrometadata>.
- Berry, D., 2011. The computational turn: Thinking about the digital humanities. *Culture machine*, 12.
- Bishop, Claire. "Against Digital Art History." *International Journal for Digital Art History*, no. 3 (July 27, 2018). <https://doi.org/10.11588/dah.2018.3.49915>.
- Blanke, T., 2018. Predicting the Past. *DHQ: Digital Humanities Quarterly*, 12(2).
- Breiman, L., 2001. Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical science*, 16(3), pp.199-231.
- Caffey, S., 2009. Privileging the Text, Subordinating the Image. *Reviews in American History* 37(4).
- Chen, A., 2018, 'The Google Arts & Culture App and the Rise of the "Coded Gaze"', *The New Yorker*, January 26, 2018,
- Crawford, K. and Paglen, T., 2019. Excavating AI: The Politics of Training Sets for Machine Learning <https://excavating.ai>
- Champion, E.M., 2017. Digital humanities is text heavy, visualization light, and simulation poor. *Digital Scholarship in the Humanities*, 32(suppl\_1), pp.i25-i32.
- Chun, W. H. K. *Control and freedom: Power and paranoia in the age of fiber optics*. MIT Press, 2008.

Cresci, E., 2017. FaceApp apologises for 'racist' filter that lightens users' skintone. *The Guardian* April 25, 2017.

<https://www.theguardian.com/technology/2017/apr/25/faceapp-apologises-for-racist-filter-which-lightens-users-skintone>

Da, N.Z., 2019. The computational case against computational literary studies. *Critical inquiry*, 45(3), pp.601-639.

Daley, B., 2019. How we're using smart tech to create richer cultural experiences. *Europeana*. Accessed July 14 2020.

<https://pro.europeana.eu/post/how-we-re-using-smart-tech-to-create-richer-cultural-experiences>

Drucker, J., 2011. Humanities approaches to interface theory. *Culture machine*, 12.

Drucker, J., 2013. Performative Materiality and Theoretical Approaches to Interface. *DHQ: Digital Humanities Quarterly*, 7(1).

Drucker, J. and McGann, J., 2000. Images as the Text: Pictographs and pictographic rhetoric. *Information design journal*, 10(2), pp.95-106.

Drucker, J. and Nowviskie, B., 2004. *Speculative computing: Aesthetic provocations in humanities computing* (pp. 431-447). Oxford: Blackwell.

Du Preez, A. "Approaching Aby Warburg and Digital Art History: Thinking Through Images." In *The Routledge Companion to Digital Humanities and Art History*, edited by K. Brown. New York: Routledge, 2020. <https://doi.org/10.4324/9780429505188>.

Europeana, 2019. AI in relation to GLAMs. *Europeana*.  
<https://pro.europeana.eu/project/ai-in-relation-to-glams>

Forster, K. W. "Aby Warburg's History of Art: Collective Memory and the Social Mediation of Images." *Daedalus* 105, no. 1 (1976): 169–76.

Fuller, M., 2006 'Software Studies Workshop'

<http://web.archive.org/web/20100327185154/http://pzwart.wdka.hro.nl/mdr/Seminars2/softstudworkshop>

Geirhos, R., Janssen, D.H., Schütt, H.H., Rauber, J., Bethge, M. and Wichmann, F.A., 2017. Comparing deep neural networks against humans: object recognition when the signal gets weaker. *arXiv preprint arXiv:1706.06969*.

Google Arts and Culture, "MoMA & Machine Learning by Google Arts & Culture | Experiments with Google." 2018. Accessed July 15, 2020.

<https://experiments.withgoogle.com/moma>.

Google Arts & Culture. "Unlock Culture at Home With Machine Learning." Accessed July 14, 2020.

<https://artsandculture.google.com/story/unlock-culture-at-home-with-machine-learning/kwKSLHCd3edAlg>.

Gurumuruwuy, P. and Deger, J., 2019. The law of Feeling: Experiments in a Yolngu museology. *The Routledge International Handbook of New Digital Practices in Galleries, Libraries, Archives, Museums and Heritage Sites*. Routledge.

Hayles, N.K., 2010. How we read: Close, hyper, machine. *ADE bulletin*, 150(18), pp.62-79.

He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

Hinrichs, U., Forlini, S. and Moynihan, B., 2019. In defense of sandcastles: Research thinking through visualization in digital humanities. *Digital Scholarship in the Humanities*, 34(Supplement\_1), pp. i80-i99.

Hockey, S., 2004. The history of humanities computing. *A companion to digital humanities*, pp.3-19.

Huang, T., 1996. Computer vision: Evolution and promise.

<https://cds.cern.ch/record/400313/files/p21.pdf>

Hurtut, T., 2010. 2D artistic images analysis, a content-based survey.

Johnson, C.D., 2012. *Memory, Metaphor, and Aby Warburg's Atlas of Images*, Ithaca, Cornell University Press.

LeCun, Y., Bengio, Y. and Hinton, G., 2015. Deep learning. *nature*, 521(7553), pp. 436-444.

Lew, M.S., Sebe, N., Djeraba, C. and Jain, R., 2006. Content-based multimedia information retrieval: State of the art and challenges. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 2(1), pp.1-19.

Liu, A.Y., 2012. Where is cultural criticism in the digital humanities? *Debates in the digital humanities* (pp. 490-509). University of Minnesota Press.

Manovich, L., 2012. How to compare one million images?. In *Understanding digital humanities* (pp. 249-278). Palgrave Macmillan, London.

Manovich, L., 2013. *Software takes command*. Bloomsbury.

Marino, M.C., 2006. Critical Code Studies. *electronic book review*, 11(27).  
<http://electronicbookreview.com/essay/critical-code-studies/>.

Morphy, H., 2006. Sites of persuasion: Yingapungapu at the National Museum of Australia. In *Museum frictions: Public cultures/global transformations*. Duke University Press.

Palmer Albers, K. "The Pig and the Algorithm." *Plot* 16 (March 4, 2017), Accessed July 14, 2020. <https://plot.online/plot/points/the-pig-and-the-algorithm/>.

Patti, E., and F. Quiviger. "'Linking Venus'. New Technologies of Memory and the Reconfiguration of Space at the Warburg Library." *Between* 4 (December 1, 2014).  
<https://doi.org/10.13125/2039-6597/1349>.

Pieris, A., 2012. Occupying the centre: Indigenous presence in the Australian capital city. *Postcolonial Studies*, 15(2), pp.221-248.

Piper, A., 2018. *Enumerations: data and literary study*. University of Chicago Press.

Piper, A., 2019. Do we know what we are doing?. *Journal of Cultural Analytics*. April 1, 2019.

Posner, M., 2015. What's next: The radical, unrealized potential of digital humanities. In *Keynote lecture, Keystone Digital Humanities Conference* (Vol. 22).

Presner, T., 2010. Digital Humanities 2.0: a report on knowledge.

Quach, K., 2020. Once again, racial biases show up in AI image databases, this time turning Barack Obama white. *The Register* June 24 2020.

[https://www.theregister.com/2020/06/24/ai\\_image\\_tool/](https://www.theregister.com/2020/06/24/ai_image_tool/)

Rafdi, M., Sarraf, A., Durrant, J., & Baker, J, 2015. British Library Machine Learning Experiment. *Zenodo*. <http://doi.org/10.5281/zenodo.17168>

Renear, A.H., 2004. Text encoding. *A companion to digital humanities*, pp.218-239.

Rudolph, C., Roy-Chowdhury, A., Srinivasan, R. and Kohl, J., 2017. Faces: Faces, art, and computerized evaluation systems—a feasibility study of the application of face recognition technology to works of portrait. *Artibus et historiae: an art anthology*, (75), pp.265-291.

Shen, J.. "Stochastic Modeling Western Paintings for Effective Classification." *Pattern Recognition* 42, no. 2 (February 2009): 293–301.

<https://doi.org/10.1016/j.patcog.2008.04.016>.

Stork, D. G., 2009. Computer Vision and Computer Graphics Analysis of Paintings and Drawings: An Introduction to the Literature. *Computer Analysis of Images and Patterns*, edited by Xiaoyi Jiang and Nicolai Petkov, 5702:9–24. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009.

[https://doi.org/10.1007/978-3-642-03767-2\\_2](https://doi.org/10.1007/978-3-642-03767-2_2).

Svensson, P., 2009. Humanities Computing as Digital Humanities. *Digital Humanities Quarterly*, 3(3).

Tallon, L., 2019. Sparking Global Connections to Art through Open Data and Artificial Intelligence. *The Metropolitan Museum of Art*. Accessed April 15, 2020.

<https://www.metmuseum.org/blogs/now-at-the-met/2019/met-microsoft-mit-art-open-data-artificial-intelligence>.

Underwood, T., Distant Reading and Recent Intellectual History. *Debates in the Digital Humanities*, edited by Matthew K. Gold and Lauren F. Klein, 2016.

<https://dhdebates.gc.cuny.edu/read/untitled/section/3b96956c-aab2-4037-9894-dc4ff9aa1ec5>.

Verwayen, H. 2018. Could AI and data mining technologies overcome issues in cultural heritage?

<https://pro.europeana.eu/post/could-ai-and-data-mining-technologies-overcome-copyright-issues-in-cultural-heritage>

Warburg, 2016, "Online BilderAtlas Mnemosyne," July 19, 2016 [Accessed July 14, 2020],

<https://warburg.sas.ac.uk/library-collections/warburg-institute-archive/online-bilderatlas-mnemosyne>.

Wang, K., Zhao, L. and Do, B., 2016. SherlockNet: tagging and captioning the British Library's Flickr images *British Library Digital Scholarship Blog*

<https://blogs.bl.uk/digital-scholarship/2016/08/sherlocknet-tagging-and-captioning-the-british-librarys-flickr-images.html>

Weigel, S., 2013. Epistemology of wandering, tree and taxonomy. The system figuré in Warburg's Mnemosyne project within the history of cartographic and encyclopaedic knowledge. *Images Re-vues. Histoire, anthropologie et théorie de l'art*, (Hors-série 4).

Whitelaw, M., 2015. Generous Interfaces for Digital Cultural Collections. *Digital Humanities Quarterly*, 9(1), pp.1-16.

Wood, C.S., 2012. Iconoclasts and Iconophiles: Horst Bredekamp in Conversation with Christopher S. Wood. *The Art Bulletin*, 94(4), pp.515-527.