

LIVING WITH IMAGES FROM LARGE-SCALE DATA SETS: A CRITICAL PEDAGOGY FOR SCALING DOWN

Gabriel Pereira & Bruno Moreschi

To cite this article: Gabriel Pereira & Bruno Moreschi (2023) LIVING WITH IMAGES FROM LARGE-SCALE DATA SETS: A CRITICAL PEDAGOGY FOR SCALING DOWN, *photographies*, 16:2, 235-261, DOI: [10.1080/17540763.2023.2189285](https://doi.org/10.1080/17540763.2023.2189285)

To link to this article: <https://doi.org/10.1080/17540763.2023.2189285>



© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 25 May 2023.



Submit your article to this journal [↗](#)



Article views: 358



View related articles [↗](#)



View Crossmark data [↗](#)

ARTICLE

Gabriel Pereira ***Bruno Moreschi** *LIVING WITH IMAGES FROM LARGE-SCALE
DATA SETS: A CRITICAL PEDAGOGY FOR
SCALING DOWN

The emergence of contemporary computer vision coincides with the growth and dissemination of large-scale image data sets. The grandeur of such image collections has raised fascination and concern. This article critically interrogates the assumption of scale in computer vision by asking: What can be gained by scaling down and living with images from large-scale data sets? We present results from a practice-based methodology: an ongoing exchange of individual images from data sets with selected participants. The results of this empirical inquiry help to consider how a durational engagement with such images elicits profound and variously situated meanings beyond the apparent visual content used by algorithms. We adopt the lens of critical pedagogy to untangle the role of data sets in teaching and learning, thus raising two discussion points: First, regarding how the focus on scale ignores the complexity and situatedness of images, and what it would mean for algorithms to embed more reflexive ways of seeing; Second, concerning how scaling down may support a critical literacy around data sets, raising critical consciousness around computer vision. To support the dissemination

CONTACT Gabriel Pereira g.pereira@lse.ac.uk Department of Media and Communications, London School of Economics and Political Science, London WC2A 2AE, UK

*Equal contributions

of this practice and the critical development of algorithms, we have produced a teaching plan and a tool for classroom use.

Introduction

In the 2020 exhibition *From Apple to Anomaly* by Trevor Paglen, 30 thousand images from the ImageNet dataset were shown as a large, unbroken mosaic at the Barbican's Curve Gallery. The exhibition showcases the normative way image data sets categorize visual elements for algorithmic systems. As described by Paglen himself: 'You're looking at these masses of images — they aren't actually for us to look at, but they are hardwired into technical systems that are looking at us.'¹ These are 'operational images', in the words of Harun Farocki: not images made for human consumption and reflection, but instead made for computers and their processes.² This exhibition and others, including *Training Humans* (by Kate Crawford and Trevor Paglen), offer a 'big picture', effectively problematizing the new ways of seeing of computer vision, as well as the general taxonomies through which these systems are constructed and work.³

Our question in this article is whether a 'big picture' approach, that of showing the large scale of data sets, is enough. Across projects that educate and expose the potentials and problems of computer vision, we rarely see data set images as singular and specific subjects of careful seeing. Our goal here is not a critique of endeavors that engage with data sets at a 'mosaic-scale', as these raise important questions. However, we believe that beyond showing a panoramic view of the grandeur of data sets, we can and should relate to images also through a different, more contextualized logic.

A guiding assumption of large-scale data sets such as ImageNet is, unsurprisingly, their scale.⁴ In fact, the emergence of contemporary computer vision coincides with the growth and dissemination of large-scale vision data sets, particularly through the development of machine learning as a 'sophisticated agent of pattern recognition.'⁵ Such collections are generated by scraping the open web to find images without copyright restrictions, a process which has been criticized for its lack of consent, as well as its wide reach into our everyday life.⁶ With very little regulation, we hear stories of companies such as Clearview AI indexing 20 billion images from the internet to build a facial recognition model.⁷ Many of these companies believe that more data would supposedly solve more problems, and allow for more calculations and generalizations.

A marker of the deeply-held belief of scale in computer vision was the consequential firing of computer scientist and researcher Timnit Gebru from Google in 2021. Gebru wrote an article with other researchers on the craze for ever-larger language models in machine learning. They asked: 'How big is too big? What are the possible risks associated with this technology and what paths are available for mitigating those risks?'⁸ They suggested developers should consider the 'environmental and financial costs first,' while also acting proactively to better care for dataset curation and documentation rather than continue the current belief of 'massive dataset size equals more diversity.'⁹ However, as widely reported, 'a group of product leaders and others inside the company [Google] . . . deemed the work unacceptable,' and fired Gebru — leading later to the ousting of other researchers

and the end of the Ethical AI group at Google.¹⁰ Though the company tried to argue Gebru and others were fired for violating company rules, the harsh response to this particular research direction shows the hegemonic belief in unrestrained, careless growth — and how undesirable it is for questions to be raised on such assumptions. The belief in scale is a core tenet of the techno-solutionist lens through which Silicon Valley sees the world, an approach that unabashedly assumes that ‘scaling up is better because it means more profit.’¹¹

This article emerges from the discussion around what scale means for computer vision data sets. We present and discuss the results of an empirical practice-based inquiry consisting of an ongoing exchange of individual images from large-scale data sets with selected participants. By stimulating people to have a long-term engagement with these images, we argue, it is possible to elicit new, profound, and variously situated new meanings beyond their apparent visual content used by machine learning algorithms.

The lens of critical pedagogy helps to approach the question of scaling down our relation with these images. Our focus is on teaching and learning as a political endeavor. First, a critical pedagogy approach means embracing how ‘machine learners’ are, in fact, ‘humans and machines involved in learning from data together’.¹² In this sense, computers are ‘taught’ to ‘see’ (more like observe patterns) through a relation with other humans, data, and machines, a process that seeks to educate them through training and testing — both pedagogical analogies which need to be critically interrogated. Second, a critical pedagogy frame means caring for how computer vision is explained to people, and what kinds of narratives we tell about what these systems do, can do, and should do.¹³ As McCosker and Wilken have proposed, ‘new kinds of literacy [are] needed to both comprehend and intervene in the automated visual systems’ emerging today.¹⁴ We are thus interested in what living with such image data may offer to a critical literacy of computer vision systems.

Large-scale image data sets and critical pedagogy

A key assumption driving large-scale data sets is their massive scale. Artist and researcher Nicolas Malevé argues that data sets are marked by the mobilization of the small ‘eye saccades’ of cheap microworker labor, used to glance at ‘billions of data items.’¹⁵ The paradigmatic large-scale data set ImageNet consists of ‘over 14 million images organized into about 20 thousand categories’, and its success has led in subsequent years to a ‘thorough commitment to ever growing datasets’.¹⁶ The belief in the ‘unreasonable effectiveness’ of such large collections occasioned in the production of data sets with billions of items, a move which necessarily trades the ‘subjective, situated, and contextualized nature of meaning-making’ for an unreflective, flat, uncritical mediation of the visual.¹⁷ As the scholar Nanna Bonde Thylstrup argues, ‘gathering, annotating, interpreting, governing, deploying and deleting data sets in machine learning models are all hermeneutic acts of governance that generate meanings and omissions.’¹⁸ That is, although industry hype says computers are being taught ‘to see like humans’,¹⁹ critical scholarship has explored how computer vision

models embed a particular lens for seeing the world, especially as data sets are formed through particular choices of categories and data.²⁰

From the outset, data sets for computer vision most often emerge from photographic representations of things.²¹ Such reliance on photography needs to be questioned considering its rationalization of vision, grounded as it is in a positivist history of employing visual techniques to ‘measure and collect facts about the world’.²² As Denton et al assess, computer vision data sets depart from ‘the assumed existence of an underlying and universal organization of the visual world into clearly demarcated concepts.’²³ Understanding a historical relationship between images and the desire for a mapping of the world relates to what sociologist Aníbal Quijano referred to as the ‘coloniality of power’: that the relations and discourses imposed during the colonial period continue well past colonization itself — even if in covert ways.²⁴

Situating computer vision within the wider history of visual culture reminds us that our relation with images at scale is not recent. First, even before the modern advent of digital photography, humans produced images at an ever-increasing scale. An explosion of images, often of art, occurred through museums, archives, catalogs, libraries, and other forms of image collections — particularly since the invention of the printing press, lithography, and other reproduction techniques. The invention of photography further led to an increased pace of visual culture production and the scale of collections. In a classic essay on the mechanical reproduction of images, Walter Benjamin wrote in 1935 that ‘since the eye perceives more swiftly than the hand can draw, the process of pictorial reproduction was enormously accelerated, so that it could now keep pace with speech.’²⁵ In fact, photographic equipment was already used at scale not long after its invention, among other contexts, by journalists and the police. Already in the 1880s, the collection of mugshots of criminalized individuals in Berlin was growing by the thousands every year.²⁶ Such large-scale collections of photographic records enabled new fields such as criminal anthropology, and even the racist pseudoscientific field of physiognomy.

Second, the longer history of visual culture shows that the decontextualized relationship with images we may associate with large-scale data sets is not entirely new. The mechanical reproduction of images through photography popularized the idea that images no longer need to be associated with their authors, and that they can exist as singular objects in other contexts — e.g. from advertising to library collections. When paintings were studied by art critics in the late 19th and early 20th century (such as Alois Riegl²⁷ and Heinrich Wölfflin²⁸), they remarked on the importance of looking at these images as part of a wider visual culture context. It was precisely the access to a broader scale of image production and distribution that allowed these scholars to detach their way of understanding art images from an artist’s individual effort, and towards seeing these images as they relate to an extensive network. More broadly speaking, Susan Sontag pointed to how ‘the [photographic] world becomes a series of unrelated, freestanding particles; and history, past and present, a set of anecdotes and *faits divers*.’²⁹

If large-scale image collections have long existed, and images have been circulating ever more quickly particularly since the 19th century, what is distinctively

problematic about the scale of current data sets? For one, *the scale of scale* has increased substantially, being oft ‘characterized in the apocalyptic terms of a deluge or avalanche, an explosion or eruption, a tsunami or storm.’³⁰ However, what most interests us is what this scale supports, which is the somewhat recent use for the ‘training’ of machines. Indeed, large-scale image collections have long been used to train *human eyes* – for example, in the case of 19th-century policing, where criminality was analyzed and classified through image archives of criminalized bodies³¹ –, but now images are used to train *computers* at an unforeseen scale. Teaching computers through image data sets is a largely invisible and opaque process, and its effects may be difficult to discern.

When teaching machines, data sets’ images work differently than previous forms of image collection because they exist beyond representation/sign: they primarily function as data to support algorithmic operations.³² The images in data sets are not merely teaching algorithms what the world looks like, they are rather determining what is recognizable, visible, and reproducible. As such, the scale of data sets is marked not only by its ever-increasing scale but also by what processes it supports and the consequences it generates.

Critical scholarship has discussed many of the consequences of the scale of data sets, particularly in terms of the biases that emerge from the collection, organization, and categorization of images.³³ Problems in data sets have direct consequences in their future algorithmic uses, for example as developers embody racial biases within data sets’ categories.³⁴ The sourcing of images through the extraction of ‘networked images’³⁵ from the web has also been widely criticized for its infringements on privacy, primarily raising questions concerning consent.³⁶ Several approaches have been suggested to deal with these ‘toxic traces’³⁷ within data sets, including through the creation of ‘datasheets for datasets’³⁸ or ‘data nutrition labels’.³⁹ These are ways for developers to reflect and report on what their data set contains, how it was generated, and its potential impacts.

Considering ‘teaching’ is central to understanding large-scale data sets, this article uses critical pedagogy as a lens, a conceptual move inspired by Malev’s work on ‘machine pedagogies’.⁴⁰ Critical pedagogy emerged in the 1960s, particularly through the work of Brazilian educator Paulo Freire. In his book *Pedagogy of the Oppressed*, Freire makes key arguments for applying critical theory to pedagogy, thus framing teaching and learning as a political proposition. As he argues, different forms of oppression are structurally built into educational systems, and the role of educators is to work ‘with, not for, the oppressed (whether individuals or peoples) in the incessant struggle to regain their humanity’.⁴¹ A key goal in the liberatory struggle against oppressions is using education to raise consciousness, unlearning the oppressive structures of power that surround us. Critical pedagogy necessarily translates into practices for social change (a ‘praxis’). Much of such practical engagement has found a focus on everyday life, as it is there that people can best relate to and understand complex social issues. This included, for example, how Freire carefully selected images that resonated with students’ lived experiences in his literacy methodologies.

Associated with Freire’s practice is the idea of ‘intellectual emancipation’ pointed out by the educator Joseph Jacotot in the 19th century.⁴² He suggests thinking of

teaching through building more critical autonomy, not as a hierarchical teacher-student relationship. The philosopher Jacques Rancière has used this concept to construct his idea of the ‘emancipated spectator’ – a spectator who not only contemplates the spectacle but also acts critically with such experience.⁴³ Our practice similarly hopes to empower people to relate to data set images not through contemplative spectatorship, but rather as reflexive and critical subjects.

The lens of critical pedagogy is useful for questioning scale in data sets for two key reasons. First, because the contemporary notion of training machines relies on traditional, top-down pedagogy. Machine learning continues a historical trajectory of humans ‘teaching’ computers.⁴⁴ The way machines are trained to see is based on the assumptions of a hierarchical and supposedly impartial form of teaching. For example, in the development of data sets, images are labeled according to strict categories by microworkers, which hides a level of cultural work that is crucial.⁴⁵ Besides the training data set, results are verified against a testing data set, which verifies the ‘correctness’ of the model. Across all of these emerges a training of algorithms that much resembles the ‘banking model of education’ criticized by Freire, one in which ‘learners are considered empty entities where their master make the “deposit” of fragments of knowledge’.⁴⁶ Here lies the potential, identified by Malevé, for a practice of reflexivity from critical pedagogy to question and unlearn the alienations of computer vision,⁴⁷ herein the problematic ideology of scale.

Second, critical pedagogy is useful because there is a pressing need for literacies for living with algorithms, including those of computer vision. Literacy is a multifaceted concept, with a long history that has spanned efforts to teach people of all ages how to read/write according to their social realities,⁴⁸ understand media messages,⁴⁹ or even understand social media metrics.⁵⁰ Across all these different uses of the term has been the consideration of literacy as an empowering process through which people can become more critically conscious of societal issues – ‘from passive to active, from recipient to participant, from consumer to citizen’.⁵¹ In this sense, we agree with the suggestion that the literacy we need is not just ‘learning to code’, but also learning ‘how forms of privilege are reproduced and naturalised through new ways of seeing’⁵² and ‘unveil what is behind the scenes when talking about data’.⁵³

We propose that a computer vision literacy must also consider the internal network of relations within these systems, attending to the complexity of the particularities of its operational images. This includes not just attending to the impacts of problematic data set construction, but also engaging with the particular images in large-scale datasets, their creation, operation, and use. Rather than only treating it as a large-scale collection of images, our suggestion is for an emancipated relation in the terms previously described by critical pedagogy. This finds resonance, for example, with the work of the artist Everest Pipkin, whose research involved watching all videos in MIT’s ‘Moments In Time’ data set in order to construct the art piece *Lacework*.⁵⁴ In the next two sections we discuss the results of our empirical practice, highlighting the potential of mobilizing these two different levels of critical pedagogy when relating data sets and scale.

Living with images from large-scale data sets

BRUNO MORESCHI: With millions of images to pick from, it may be difficult to understand why this is my favorite image of a large-scale dataset (Figure 1). I get it, the photo is actually very

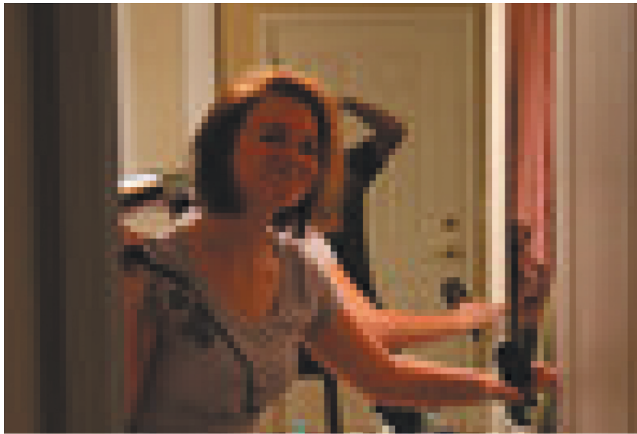


Fig. 1. Image taken from the ImageNet training dataset.

uninteresting at first glance: a woman opens a door and finds someone. But I had to look at this image more than once. And doing so, an ambiguous universe rich in possibilities soon appeared, worthy of comparison with the game of mirrors painted by Diego Velázquez in 1656 (*Las Meninas*) – yes, I really believe that. The woman who opens the door finds a person taking a picture of her. Behind her, a mirror reflects him, but we cannot see his reflection in its entirety because her head covers his body. We only know of this man through his reflection in the mirror that is inside the room behind her. The lack of image resolution makes it all the more dubious, just like thick brushstrokes: is she smiling at the camera? I could continue speculating on the door in the background (or its reflection?). . . It seems to be closed, but could it be open in a different image?

As the art critic John Berger so bluntly put it, ‘All photographs are ambiguous’.⁵⁵ Likewise, the images that train computer vision algorithms embody a vast field of possibilities. The vignette above reflects a moment of this realization: something can be gained by living with data set images and appreciating their complexity. At first

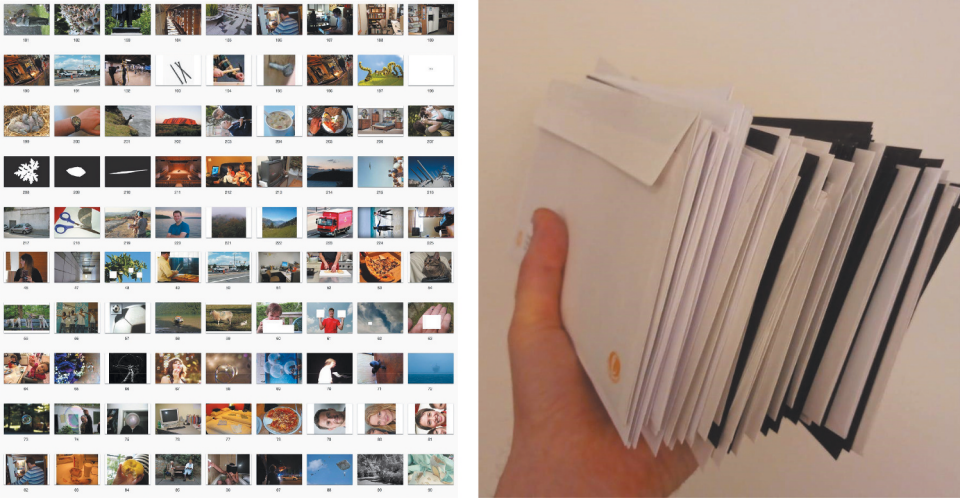


Fig. 2. Two photos documenting the process: on the left, organizing and preparing part of the data set images for printing as postcards; on the right, a pile of envelopes with postcards ready to be mailed to participants.

glance, one might characterize the above image (Figure 1) as just ‘a woman opening a door’. That is how Amazon Mechanical Turkers described and tagged such images for a data set. These microworkers received pennies for their labor, which is now reflected in the description the NeuralTalk2 model offers of the scene: ‘a woman opening a door’. There is a homogeneity in the way images are used to train models which aims to simplify and reduce them, instead of opening their ways of ambiguously representing the world.

Our practice-based inquiry for this article started in 2020, a period marked by the beginning of the COVID-19 pandemic. The first step was taking images from 11 machine learning training data sets out of their original setting by printing them into postcards (Figure 2). This size makes possible the easy distribution of such images through the post.⁵⁶ We sent sets of 3 postcard images to 111 previously-invited people around the world. Our goal with sending 3 images was that participants could relate not only to one image in itself, but also to the relation between images — which is, though at a different scale, a key goal of computer vision models. The three chosen images did not always come from the same data set, proposing a dynamic relation across these collections.

The people invited to participate were hand-picked because of their connection with images and/or algorithms. This was a form of resistance to the way such operational images are usually seen: by an unknown network of supposedly non-specialized people, the microworkers.⁵⁷ Participants include art historians, engineers, museum guards, artists, programmers, scholars, activists, mail workers, VR filmmakers, designers, and microworkers (particularly those who perform tasks related to tagging and image description).

The prompt sent to participants was to live with the three images, in whichever way they felt comfortable, for at least one week. They were asked to reflect on their

experience by creating a response. Rather than a standardized survey, we asked for responses in whichever format people felt comfortable with. Some participants sent written email responses and WhatsApp audio notes; others requested a video call to talk about their experience; and even others crafted a response through multimedia formats, such as video, painting, and sound. The project is still ongoing, and the writing of this article focused on the analysis of 28 responses. Written informed consent for the publication of details was obtained from the participants named in this article.

We understand this participatory and practice-based methodology as a form of critical technical practice. The objective of our inquiry is to question the assumptions of a sociotechnical system, computer vision data sets, through technical practice itself, with a reflexive disposition.⁵⁸ Critical technical practice as a form of research has been historically intertwined with artistic and pedagogical methods, thus unsettling ‘the binary split of theory and practice, thinking and doing, art and technology, humans and machines, and so on.’⁵⁹

‘The mail has arrived!’: a physical image that follows you

KAI YE: I mounted these images to the walls of my kitchen. There, my roommate and I encountered the images every time we cooked something or went to throw something away. I explained to her that these images were likely scraped from the Internet, such as the photos people posted on Flickr. In fact, they aren’t necessarily meant for only machines to see. Many of these images were taken by people, for an audience of one or two. What did we see? As I chop vegetables and wipe the counters, I can’t resist reading them as a narrative about labor, the dismal triple partition of daily life. The images show a hotel bed, a computer, a pizza. They may as well illustrate Marx’s famous saying: “8 hours for rest, 8 hours for work, 8 hours for what we will.”

Different reflections emerge from an analysis of the responses from participants. The first concerns what it means to live with such images once they are returned to the flow of life, and what happens when we engage once again with their original representational goal. Many of the participants decided to place the postcards somewhere at home where they can always be seen. As the computer scientist and artist Kai Ye mentions above, by living with such data set images we are reminded they are not just operational images, made for machines to see, but that they have emerged from people who created images for others (or themselves) to see.

LEONARDO IMPETT: They are funny images to look at so closely. So they ended up following me around for a long time, because I didn’t quite know what to do with them. So I was always kind of having them around . . . For the last two

weeks they have been on top of my [inaudible], which is also where I keep my postcards . . . So they've been with my postcards, looking at me . . .

Much is gained once these images are printed, as they affect and are affected by the physical environment. In printing these images — transforming them into a physical artifact — we are activating alternative ways of relating to images that relate to memories, postcards, and photo albums. By being singular and engaging people in relations, the postcard images turn into affective objects. Such affective charge significantly transforms our relationship with these images, further opening their potential meanings. In other words, these images that once seemed so far removed become now 'our' images.

These images are returned to human relations, where they spill into the world surrounding them. Julio Kraemer, a microworker from Amazon Mechanical Turk, received a banal picture of the sunset over a beach, which led to the following response:

JULIO KRAEMER: This postcard reminded me of a service available on Amazon Mechanical Turk: to assign a location to beaches based on images. There were always three alternatives to choose from, for example: Copacabana, Navagio, Maldives, Varadero, Maya Bay, and Tulum. However, due to what I believe was a bug in the system, there was always a recurring option: Acapulco. I now believe Acapulco can be the image of a beach that is Acapulco, but it can also be the image of a beach that maybe one day will be Acapulco (because of the recurrent bug), but that is not just yet Acapulco. I've never been to Acapulco. But I like it there.

Julio's response above is not a digression from the image itself, but rather points out how images are always complex, relational discourses, far beyond a watertight photographic truth. Images affect people's lives, always in relation to people's experiences and stories, and always in relation to other images.

Building narratives

RENATA MARQUEZ: Every event of photographic capture presupposes a second moment: that of a photographic interpretation.

A second reflection of living with data set images is that a sustained engagement enables the potential for these images to say something, at least for the person that is seeing them — as pointed out by the architect Renata Marquez above. The curator Cayo Honorato, for example, saw in his images not just the picture of two white men shaking hands:

CAYO HONORATO: They are dressed in business clothes and shaking hands in a way that is not protocol. They look each other in the eye. It makes me think that they have a relationship, maybe they're father and son. And more: maybe they are at an outdoor wedding, happy with the event.

These narratives interpret the images and assign them meaning, expanding them from just a data point to something much more complex. This resonates with the experience of the scholar William Uricchio, who received his image-postcards moments after he had explored the poetry of algorithmic image description with his lab:

WILLIAM URICCHIO: Based on experiments that my lab has been doing with alt-tech and AI-generated verbal descriptors of images, I have the sense that we have done a decent job in getting technological assemblages to simulate what we think we see (the photograph); but a pretty poor job in assigning verbal labels to what the system produces. Even stripping away the persistent tendency to narrativize; to situate in time, space, and causality; and to interpret – all of which combine to shape human descriptions of images, I would be surprised if image recognition software could identify the artifacts in image[-postcards] two and three.

Much as indicated by Uricchio, practically all responses we collected speculate, create stories, and remember or imagine things based on the images. The complexity coming from these localized interpretations of images is largely due to the human ability to create narratives, even when we are just trying to describe these images. This makes us wonder about how computer vision could be trained beyond the simple and mechanical exercise of describing images or tagging specific visual elements. It may be true that such complex understandings do not currently generate computable data, and that they don't fit well within the current understanding of accuracy in computer vision models. Our provocation here is whether recovering the 'persistent tendency to narrativize' may indeed make ways of machine seeing into something much more complex and situated, beyond the flat and supposedly neutral description of images that currently dominates computer vision.

Speculating with image trajectories

Beyond the stories of images' representations, our third reflection relates to participants' speculation on the trajectories of the images they received. That is, what they understand to be the use of such images, how they arrived in data sets, and finally how they have been used for training computer vision.

The photographer Paulo Avelino felt particularly compelled to find out the history of the images he received before they became training data. This led him to an investigative inquiry about the context from which such images were extracted, using a reverse image-search website. After much toil, and many dead ends, he was able to find the ‘origin’ of just one of the three images he received — precisely that of the woman opening the door that begins this section (Figure 1). It was initially posted on a social network, in high-resolution resolution, unlike how it appears now in ImageNet. With this better-quality image, we can finally ascertain that the woman, indeed, seems to be smiling. The investigation of Avelino tells us how these are not ‘comfortable images’, as they often ask us tough and (un-)answerable questions about their provenance, their history, their original goals and intentions.

Meanwhile, the scholar Alex Hanna reflected on how her images were used to train computers. Hanna, who is very familiar with how algorithmic systems work, described in an audio note how she imagines the images would be segmented by an algorithm: a building, a pond, some trees. When looking at an image of two men looking at each other (the same one Honorato received), though, she also highlights what ambiguities could emerge for computer seeing: are they in love? Related to these non-human ways of seeing also emerges a consideration about images as technical, posthuman artifacts, which is reflected in Uricchio’s response:

WILLIAM URICCHIO: But I will admit that ‘recognition’, so much a part of what photographic technologies have been designed to replicate, is but one way to read the image. Looking at the image as a list of color densities, ink requirements, and the technical data of exposure time and f-stops, and more, would all seem to offer equally legitimate descriptors of each card (and would all continue to be plagued by anthropocentric categories!).

Such explorations of the images as technical artifacts challenge us to think like the algorithms we know are trained by them, mimicking their ways of seeing. Images are data, they leave traces but are formed by traces, often in forms that are not completely understandable to us. There is a largely ignored potential, identified by Uricchio above, to see much more through the data and metadata of these singular images.

Generative images

The fourth point worth reflecting on is that a collection of three images is far too little for computer vision models, but is very generative for humans. Case in point: the scholar Richard Staley related the three images he received to wood (Figure 3). This stimulated Staley to remember two things. First, a recent train trip to Switzerland, and how logs might have been removed from the landscape he saw while there. Second, a walk in a forest off the coast of North Devon where he found a mysterious wooden crate open on a rainy day. In themselves, these two memories



Fig. 3. The three images Staley received.

already add external data, so to say, to the images he received. However, Staley went further: he felt motivated to respond to the images with other images, all photographed by him. One of them shows a pile of wood that was carefully organized by his wife Elisabeth to serve as shelter for hedgehogs that they regularly feed (Figure 4). This new image further generates unsettling reflections on practically all images that may be related to ‘wood’:

RICHARD STALEY: These trees don’t live. They have been cut, their history is now wood, it is in our future to burn them or shape them (maybe spiders and mosses, insects and birds make use of them in this pile, while we have not yet decided on that).

Similar to Staley, the scholar Junia Mortimer also generated new images to understand the images she received. Among other photos, one shows an image-postcard of dinosaur bones within an edited book on image theory, where they seem to serve as a bookmark. It is carefully placed in the chapter *Devolver uma Imagem* (‘To give back an image’), by the philosopher Georges Didi-Huberman. The placing seems to be intentional, as that is the precise page where the author discusses the etymology of the word ‘image’, and how its origins in ‘imago’ related to ideas of ‘possession’ and ‘refund’.



Fig. 4. Photograph taken by Staley in response to the images he received.

Beyond the production of new images, the artist and researcher Sheung Yiu produced a video experience. The postcards sent to him contained the same image in different sizes of an algorithmic face resulting from the superimposition of different faces from Flickr-Faces-HQ. This dataset was used to train the GAN algorithm network of the website *this Person Does Not Exist*, which generates deep fakes of faces. With the images in hand, Yiu reflected on the advances in machine learning, and how the plethora of facial image data gives rise to a multitude of new computer vision applications. He concluded that the three images were not enough as still images: they could only be understood if they were dynamically related to the face of the actress Cynthia Blanchette (Figure 5).

SHEUNG YIU: I flip your image back and forth between the face and the camera to evoke the perceptual phenomenon of persistence of vision. On the one hand, the persistence of vision is a unique feature of human vision that works wildly different from the statistical rendering of computer vision. It is an optical illusion, a fault, and a feature of the human visual system that preserves continuity. On the other hand, the persistence of vision makes moving images possible. Because of the lack of time resolution in our sight, continuous motions emerge from individual images. It signifies a switch in scales: between the individual static image – a discrete packet of time – and the continuous narrative.

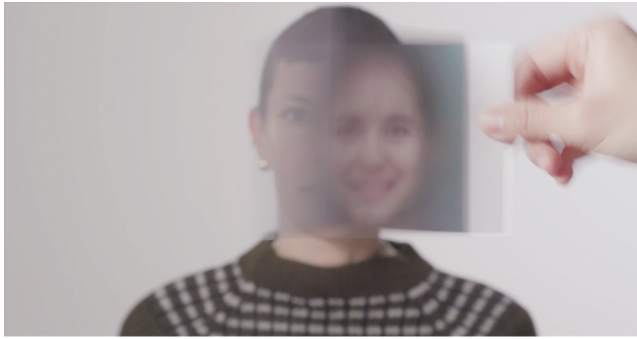


Fig. 5. A frame from the video created by Sheung Yiu.

Across these diverse responses is the realization that these images are generative, particularly when they are interpreted beyond as just an image. Humans look at these images not just within their representational quality or as technical artifacts, they become magnets for reflexivity, new images, and new data. Could a computer be taught to see beyond a passive ‘understanding’ of the visual? The provocation here is to consider what could be gained by seeing outside the frame and teaching/learning to be (creatively) affected by images.

Sometimes it doesn't mean much ... so burn it?

NICK COULDRY: I must be honest: I looked at the images and did not have any reaction. The images were, as it were, NOTHING. They did not communicate anything, in terms of meaning, though I could see (from a technical point of view) how they might be useful in training a machine. My wife similarly didn't have any reactions to the images. I was therefore ‘struck dumb’ by the images. I wonder if this is a common reaction?

Finally, we must accept that, quite often, these images may not actually have that much to say. As suggested above by scholar Nick Couldry, this is indeed a common reaction, and one which we must acknowledge as productive and generative within the pedagogical engagement of this research practice. In fact, many of the images in such data sets weren't meant to be seen by themselves, as they may have been produced exclusively for machine learning. Taking them off their context may make them seem empty, disconnected from their operational image roots. This is the case of the images from the Columbia University Image Library COIL-100, which brings together photographs of 100 consumer products with enough images to form a 360-degree view — useful data to increase the efficiency of machines that approach objects through shifting perspectives. One of



Fig. 6. A frame from the video of Julia Mortimer burning the image she received of the Louvre's pyramid.

the authors of this article received three such images (onion, eye drop, canned cat food):

GABRIEL PEREIRA: I didn't know what to do with these images. I tried many things, but still felt I had nothing meaningful to say about a low-res image of canned cat food. This is my response, then: People don't have to have something to say about these images, as they don't with so many other images. In computer vision data sets, there's always an answer, always a tag, always a description, always a way to operationalize these images. But, in this case, I choose not to have an answer, to refuse to produce more data, to enrich and extract meaning from these images. I want them to remain undefined.

This response of refusal was more radically put into practice by Junia Mortimer. Her second image postcard showed the glass pyramid of the Louvre, next to a wall with a poster written 'www'. Her response was recording a video in which one of her hands holds the postcard as, little by little, she burns the image with a candle (Figure 6). After that, the video shows her throwing what was left in a trash bin. Such action of burning and binning a data set image could be interpreted as a refusal to these large-scale image data sets. It recovers the capacity to choose for these images not to exist after living with them, to delete them and their traces like the creators of such images can't anymore.

What can be learned by scaling down?

Our goal here is not the exhaustive analysis of responses from participants, but to show the pedagogical potential held by thinking about data set images and scale. The methodology we have presented — that of living with images and looking closely at

them — may seem conservative, a return to much-critiqued normative ways of seeing art that ignore a current reality where we have learned to deal with an abundance of images. However, we believe this approach is useful in the context of large-scale data sets because these are ‘operative images’, images that are used to train machines and that are never taken seriously in their own right. As the responses show, much is gained from engaging these images from up-close and exploring the situated meanings they evoke — particularly when an open field of possibilities is encouraged, rather than a normative, right/wrong type of analysis. In other words, we are asking people to subvert a historically normative way of seeing art, but only as a way of productively breaking with the current normativity which trains machines to see the world as flat, clearly demarcated objects.

How do the results of this practice-based inquiry relate to the questions about critical pedagogy and scale we started with? First, our practice questioned the role large-scale image data sets play as they are used to train machine learning models. People’s responses show the potential held by relating closely to such ‘images that should not be seen,’ interpreting them through the many layers of human reflexivity. These specific ways of seeing transform these images, imbue them with meaning, and recover their inherent ambiguity. There are many (potential) stories in these images that computer vision models are not ‘taught’ to engage with or reflect upon. There is, in this sense, a deep potential held by these data sets. What would models look like if they dove deeper into already existing images, rather than collecting and building bigger and bigger data sets? Rather than accepting the depersonalization and decontextualization of images, what could be gained from understanding image data as personal and full of meaning?

Our provocation for computer vision is: do we need bigger data sets, or rather need new ways for teaching and learning about images? Contemporary models severely limit the way images are used and understood, trading all complexity away for teaching computers to see through a relatively flat, though actionable, understanding of the visual. Critical pedagogy reminds us that these teaching materials embody the power dynamics from which they originate, and remind us to center reflexivity in the process of teaching and learning. To make machines more reflexive, to go beyond the ‘banking model of education’, a shift of scales is needed, a break from current assumptions of bigger, larger data sets. This means we must look carefully and deeply at trivial data set images such as [Figure 1](#), not only at Velázquez’s painting. Important art critics have already taught us that to understand a complex painting, such as *Las Meninas*, it is necessary to look both from afar and closely, going back and forth, learning and reflecting from different perspectives. This means centering a practice of (computer) seeing no longer associated with tests of right and wrong, but with generative and complex situated perspectives.

Second, how can this experiment support efforts to consider a critical pedagogy around large-scale data sets? Our argument is that we need forms of critical literacy that engage with data set images beyond the ‘big picture’ of the grandeur of such collections. By getting people to engage with a few images, in a more close-up and personal way, we have shown how these data sets are much more complex than the panoramic view. Following critical pedagogy, a critical consciousness-raising effort

must enable people to recognize and understand not only that contemporary computer vision operates from these ‘images that should not be seen’, but that these images are complex and interesting in themselves. The different responses we have gathered show these images are not just data points, but embed many stories told and untold, working as generative artifacts — though also often as provocatively boring technical things we cannot fully understand. This practice thus invites participants to question data sets’ massive scale — something hard to relate to or understand — by acting as an ‘emancipated spectator’ that acts and produces critical, situated understandings around the images underlying systems of computer vision.

A final consideration is how a critical literacy of computer vision benefits from the act of returning images from data sets to a materiality (not necessarily physical). This means allowing people to have a durational experience of living with these images — regardless of the chosen medium. This engagement means images are returned to the flow of life, where they can find new and old meanings, affect and be affected by the world around them, as well as raise questions about their technical role within the wider computer vision assemblage they come from.

Taking it to the classroom

To spread the methodology of living with data set images beyond our practice, and inspired by other critical pedagogy projects,⁶⁰ we have produced a Teaching Plan (Attachment 1) to support classroom use. This Teaching Plan serves as an entry point to bring the debate on the role of scale in computer vision data sets to students. It centers an embodied, practice-based exercise which engages students in living with data set images. It also indicates pathways for leading a reflexive discussion around the results of such experience, linking it to broader problems and limitations of large-scale data sets. To support the use of the Teaching Plan, we have also created an online tool (Attachment 2: ‘Images to live with’) that generates printable sheets of images from data sets for teaching.

We piloted the Teaching Plan in an undergraduate classroom in Brazil, with students coming from a wide breadth of humanities and arts subjects. The exercise proved to be very generative for discussions, with a strong participation from students, who were interested in sharing their experiences with the images. Questions emerged, for example, regarding the context from which images were taken, what is *not* captured in data set images, and how value judgments are embedded in the making of data sets. Furthermore, students were empowered to produce their responses in diverse ways, such as cutting the physical images they received, writing on them, as well as taking them for a stroll.

Although we did not conduct an extensive test of this teaching methodology, our experience showed it was useful in getting students a hands-on experience with data sets and their visual elements, enabling them a more complex understanding of how computer vision is produced. Considering students have less prior expertise than the participants invited for the previous exchange, we understand this as a great indicator of the potential of this method for a broader public. As more broadly suggested by

critical pedagogy, the classroom is a crucial setting of consciousness-raising, as students are likely to impact the development and use of technological systems such as computer vision in the future.

Conclusion

This article suggests living with images from large-scale computer vision data sets as a way of learning and reflecting on them. This is a speculative experiment that helps to consider the limitations of the current reliance of algorithmic model creation on scale. Our practice questions the assumptions of scale in computer vision, particularly the widespread notion that more data is needed for further developing computer vision. We believe that a durational encounter with images proposes an approximation that we aren't otherwise offered in our relation to computer vision, which tends to always relate to the 'big picture' of data sets, ignoring the complexities of images. Rather than assuming a neutral perception of the visual that exists beyond the confine of each person, this experiment instead centered on situated perspectives. We propose this as a provocation for rethinking how we train/learn computer vision; and as a pedagogical intervention for considering what forms of literacy are needed for computer vision.

Our first contribution is to debates in visual studies and computer science. We suggest there is an untapped potential in data sets' images that is currently ignored. These images carry much more potential meaning than the way they are currently being used, and 'scaling down' can help us see through that. Our second contribution is to 'critical data set studies'⁶¹: a methodology for living with data sets as a pedagogical encounter. Relating to these images closely both confirms and complicates the scholarship on the problems of data sets. To disseminate these results beyond our practice, we offer a teaching plan and a teaching tool, which can be freely used to support critical teaching in classrooms.

Notes

1. Paglen, "Trevor Paglen: On 'From Apple to Anomaly.'"
2. Farocki, *Eye/Machine*.
3. Crawford and Paglen, "Excavating AI."
4. Denton et al., "On the Genealogy of Machine Learning Datasets"; Hanna et al., "Lines of Sight."
5. Zylinska, *AI Art*, 25.
6. Prabhu and Birhane, "Large Image Datasets."
7. ICO, "ICO Fines Facial Recognition Database."
8. Bender et al., "On the Dangers of," 610.
9. Ibid.
10. Simonite, "What Really Happened."
11. Ye, "Silicon Valley and the English language"; see also Liu, *Abolish Silicon Valley*.
12. Mackenzie, *Machine learners*.

13. See Zylinska, *AI Art*, chapter 2.
14. McCosker and Wilken, *Automating vision*, 114.
15. Malevé, “On the Data Set’s Ruins.”
16. Denton et al., “On the Genealogy of Machine Learning Datasets,” 6.
17. *Ibid.*, 9.
18. Thylstrup, “The Ethics and Politics of Data Sets,” 665.
19. Li, “How we’re Teaching Computers.”
20. Smits and Wevers, “The Agency of Computer Vision Models.”
21. See Offert and Bell, “Perceptual Bias.”
22. Lister, “Photography in the Age of Electronic Imaging,” 324; see also Malevé, “On the Data Set’s Ruins.”
23. Denton et al., “On the Genealogy of Machine Learning Datasets,” 6.
24. Quijano, “Coloniality and Modernity/Rationality”; for further reflection on data’s relation to colonialism, see Couldry and Mejias, *The costs of connection*.
25. Benjamin, Walter. *The Work of Art in the Age*.
26. Finn, Jonathan M. *Capturing the Criminal Image*, 10.
27. Riegl, Alois. “The Main Characteristics” and “The Place of the Vapheio Cups in the History of Art.”
28. Wölfflin, Heinrich. *Principles of Art History*.
29. Sontag, Susan. *On Photography*.
30. Dvořák and Parikka, *Photography off the Scale*, 4.
31. Sekula, Allan. “The Body and the Archive.”
32. See Hoelzl and Marie, *Softimage*; Zylinska, *Nonhuman photography*; and Dewdney, *Forget photography*.
33. See for example Silva et al., “APIs de Visão Computacional”; and Buolamwini and Gebru, “Gender Shades.”
34. Buolamwini and Gebru, “Gender Shades”; and Crawford and Paglen, “Excavating AI.”
35. Rubinstein and Sluis, “A life more photographic.”
36. Prabhu and Birhane, “Large Image Datasets”; see also Solon, “Facial Recognition’s ‘Dirty Little Secret.’”
37. Thylstrup, “Data out of Place.”
38. Gebru et al., “Datasheets for Datasets.”
39. Holland et al., “The Dataset Nutrition Label.”
40. See Malevé, “Machine Pedagogies.”
41. Freire, *Pedagogy of the oppressed*, 48.
42. Rancière, *The emancipated spectator*.
43. *Ibid.*; see also Boal, *Theatre of the oppressed*.
44. Turing, “Computing Machinery and Intelligence.” We must highlight there is much critique to the notion that machines are in fact learning, particularly because current AI ‘is not the kind of learning premised on understanding, but which rather involves being trained in making quicker and more accurate decisions on the basis of analysing extremely large data sets.’ (Zylinska, *AI Art*, 24). As suggested by Mackenzie (*Machine learners*), the learning in machine learning is better understood as the finding of a function with the least error.
45. Irani, “The Cultural Work of Microwork.”
46. Malevé, “Machine Pedagogies.”

47. Ibid.
48. Freire, *Pedagogy of the oppressed*.
49. Livingstone, "The Changing Nature and Uses of Media Literacy."
50. McCosker, "Data Literacies for the Postdemographic."
51. Livingstone, "The Changing Nature and Uses of Media Literacy," 2.
52. Cox, "Ways of Machine seeing as," 111.
53. Tygel and Kirsch, "Contributions of Paulo Freire for a Critical Data Literacy," 116; see also Markham and Pereira, "Analyzing Public Interventions."
54. Pipkin, "On Lacework."
55. Berger, "The Ambiguity of the Photograph," 91.
56. Data sets were chosen for their relative importance in the field: Visual Genome, Flickr-Faces-HQ Dataset, Tencent, Stanford Dogs Dataset, Leaf Dataset, Labelled Faces in the Wild Dataset, ImageNet, Columbia University Image Library COIL-100, TinyImage, the official Google Open Image and a second version of Open Image (with a decanonization experiment carried out by programmer Bernardo Fontes and which only showed the untagged areas of the images). The choice of these well-established data sets also occurs because they have mostly been audited as a result of critical research, which means they (as far as we could notice) no longer have too many upsetting images — which were not the focus of this project.
57. Tubaro et al., "The Trainer, the Verifier, the Imitator."
58. Agre, "Toward a Critical Technical Practice."
59. Soon and Cox, *Aesthetic Programming*, 246; see also Harwood, "Teaching Critical Technical Practice."
60. We were very inspired by previous approaches to teaching through practice-based engagements with technology, such as: Soon and Cox, *Aesthetic Programming*; and Harwood, "Teaching critical technical practice."
61. Thylstrup, "The Ethics and Politics of Data Sets."

Funding

The work was supported by the Danmarks Frie Forskningsfond [International Postdoc Grant]; Project Decay without mourning: Future thinking heritage practices, Riksbankens Jubileumsfond / Volkswagen Foundation.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Gabriel Pereira  <http://orcid.org/0000-0002-9267-4189>

Bruno Moreschi  <http://orcid.org/0000-0001-8938-5004>

Bibliography

- Agre, Phil. "Toward a Critical Technical Practice: Lessons Learned in Trying to Reform Ai." In *Bridging the Great Divide: Social Science, Technical Systems, and Cooperative Work*, edited by Geoffrey Bowker, Les Gasser, Leigh Star, and Bill Turner, 130–157. New York: Erlbaum Hillsdale, 1997.
- Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?." In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–23. FAccT '21. New York, NY, USA: Association for Computing Machinery, 2021. doi:[10.1145/3442188.3445922](https://doi.org/10.1145/3442188.3445922).
- Benjamin, Walter. *The Work of Art in the Age of Its Technological Reproducibility, and Other Writings on Media*. Cambridge: Harvard University Press, 2008.
- Benjamin, Garfield. "#fuckthealgorithm: Algorithmic Imaginaries and Political Resistance." In *2022 ACM Conference on Fairness, Accountability, and Transparency*, 46–57. Seoul Republic of Korea: ACM, 2022. doi:[10.1145/3531146.3533072](https://doi.org/10.1145/3531146.3533072).
- Berger, John. "The Ambiguity of the Photograph." In *Another Way of Telling: A Possible Theory of Photography*, edited by John Berger and Jean Mohr, e-book. London: Bloomsbury, 2016.
- Boal, Augusto. *Theatre of the Oppressed*. London: Pluto Press, 2008.
- Buolamwini, Joy, and Timnit Gebru. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." In *Conference on Fairness, Accountability and Transparency*, New York, 77–91, 2018.
- Crawford, Kate, and Trevor Paglen. "Excavating AI: The Politics of Training Sets for Machine Learning." 2019. <https://excavating.ai/>
- Denton, Emily, Alex Hanna, Razvan Amironesei, Andrew Smart, and Hilary Nicole. "On the Genealogy of Machine Learning Datasets: A Critical History of ImageNet." *Big Data & Society* 8, no. 2 (July 1, 2021). doi:[10.1177/20539517211035955](https://doi.org/10.1177/20539517211035955).
- Dewdney, Andrew. *Forget Photography*. Cambridge: MIT Press, 2021.
- Dvořák, Tomáš, and Jussi Parikka. *Photography off the Scale: Technologies and Theories of the Mass Image*. Edinburgh: Edinburgh University Press, 2021.
- Farocki, Harun. "Eye/machine." 2000. <https://www.harunfarocki.de/installations/2000s/2000/eye-machine.html>
- Finn, Jonathan M. *Capturing the Criminal Image: From Mug Shot to Surveillance Society*. Minneapolis, Minn: University of Minnesota Press, 2009.
- Freire, Paulo. *Pedagogy of the Oppressed*. 30th Anniversary Edition ed. New York: Bloomsbury Academic, 2000.
- Hanna, Alex, Emily Denton, Razvan Amironesei, Andrew Smart, and Hilary Nicole. "Lines of Sight." *Logic Magazine*, 2020. <https://logicmag.io/commons/lines-of-sight/>
- Harwood, Graham. "Teaching Critical Technical Practice." In *The Critical Makers Reader: (Un) Learning Technology*, edited by Loes Bogers and Letizia Chiappini, INC Reader, (12). Institute of Network Cultures, 29–37, 2019.
- Hoelzl, Ingrid, and Rémi Marie. *Softimage*. Bristol: Intellect Books, 2015.
- Holland, Sarah, Ahmed Hosny, Sarah Newman, Joshua Joseph, and Kasia Chmielinski. "The Dataset Nutrition Label: A Framework to Drive Higher Data Quality Standards." *arXiv* (May9, 2018). [10.48550/arXiv.1805.03677](https://arxiv.org/abs/10.48550/arXiv.1805.03677).

- ICO. "ICO Fines Facial Recognition Database Company Clearview AI Inc More Than £7.5m and Orders UK Data to Be Deleted." May 31, 2022. <https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2022/05/ico-fines-facial-recognition-database-company-clearview-ai-inc/>
- Irani, Lilly. "The Cultural Work of Microwork." *New Media & Society* 17, no. 5 (2015): 720–739. doi:10.1177/1461444813511926.
- Li, Fei-Fei. *How We're Teaching Computers to Understand Pictures*. TED Talks, 2015.
- Lister, Martin. "Photography in the Age of Electronic Imaging." In *Photography: A Critical Introduction*, edited by Liz Wells, 313–344, Routledge: London, 1997.
- Liu, Wendy. *Abolish Silicon Valley: How to Liberate Technology from Capitalism*. London: Repeater, 2020.
- Livingstone, Sonia. "The Changing Nature and Uses of Media Literacy." In *Media@LSE Electronic Working Papers 4*, edited by Rosalind Gill, Andy Pratt, Terhi Rantanen, and Nick Couldry, 1–35, London: Media@LSE, 2003.
- Mackenzie, Adrian. *Machine Learners: Archaeology of a Data Practice*. Cambridge: MIT Press, 2017.
- Malevé, Nicolas. "On the Data Set's Ruins." *Ai & Society*. doi:10.1007/s00146-020-01093-w.
- Malevé, Nicolas. "Machine Pedagogies." *Machine Research* (blog), September 26, 2016. <https://machineresearch.wordpress.com/2016/09/26/nicolas-maleve/>
- Markham, Annette N., and Gabriel Pereira. "Analyzing Public Interventions Through the Lens of Experimentalism: The Case of the Museum of Random Memory." *Digital Creativity* 30, no. 4 (October 2, 2019): 235–256. doi:10.1080/14626268.2019.1688838.
- McCosker, Anthony, and Rowan Wilken. *Automating Vision*. London: Routledge, 2020.
- Nick, Couldry, and Ulises A. Mejias. *The Costs of Connection: How Data is Colonizing Human Life and Appropriating It for Capitalism*. Stanford: Stanford University Press, 2019.
- Offert, Fabian, and Peter Bell. "Perceptual Bias and Technical Metapictures: Critical Machine Vision as a Humanities Challenge." *Ai & Society*. doi:10.1007/s00146-020-01058-z.
- Paglen, Trevor. "Trevor Paglen: From Apple to Anomaly." 2019. <https://sites.barbican.org.uk/trevorpaglen>
- Paola, Tubaro, Antonio A. Casilli, and Marion Coville. "The Trainer, the Verifier, the Imitator: Three Ways in Which Human Platform Workers Support Artificial Intelligence." *Big Data & Society* 7, no. 1 (2020): 205395172091977. doi:10.1177/2053951720919776.
- Pipkin, Everest. "On Lacework: Watching an Entire Machine-Learning Dataset." Accessed June 22, 2022. <https://unthinking.photography/articles/on-lacework>
- Prabhu, Vinay Uday, and Abeba Birhane. "Large Image Datasets: A Pyrrhic Win for Computer Vision." *ArXiv* (2020). <https://arxiv.org/abs/2006.16923>.
- Quijano, Anibal. "Coloniality and Modernity/Rationality." *Cultural Studies* 21, no. 2–3, (March 1, 2007): 168–178. doi:10.1080/09502380601164353.
- Rancière, Jacques. *The Emancipated Spectator*. London: Verso, 2009.
- Riegl, Alois. "The Main Characteristics of the Late Roman 'Kunstwollen'" and "The Place of the Vapheio Cups in the History of Art." In *The Vienna School Reader: Politics and Art Historical Method in the 1930s*, Christopher S. Wood, (org.), 87–132, NY: Zone Books, 2000.

- Rubinstein, Daniel, and Katrina Sluis. "A Life More Photographic: Mapping the Networked Image." *Photographies* 1, no. 1, (March 1, 2008): 9–28. doi:10.1080/17540760701785842.
- Sekula, Allan. "The Body and the Archive." *October* 39, no.3 (1986):3. doi:10.2307/778312.
- Silva, Tarcízio, André Mintz, Janna Joceli Omena, Beatrice Gobbo, Taís Oliveira, Helen Tatiana Takamitsu, Elena Pilipets, and Hamdan Azhar. "APIs de Visão Computacional: Investigando Mediações Algorítmicas a Partir de Estudo de Bancos de Imagens." *Logos: Comunicação e Universidade* 27, no. 1 (2020): 25–54
- Simonite, Tom. "What Really Happened When Google Ousted Timnit Gebru." *Wired*. Accessed June 20, 2022. <https://www.wired.com/story/google-timnit-gebru-ai-what-really-happened/>
- Smits, Thomas, and Melvin Wevers. "The Agency of Computer Vision Models as Optical Instruments." *Visual Communication*. doi:10.1177/1470357221992097.
- Solon, Olivia. "Facial Recognition's 'Dirty Little Secret': Social Media Photos Used Without Consent." *NBC News*. Accessed June 21, 2022. <https://www.nbcnews.com/tech/internet/facial-recognition-s-dirty-little-secret-millions-online-photos-scraped-n981921>
- Sontag, Susan. *On Photography*. S.I: Rosetta Books, 2020.
- Soon, Winnie, and Geoff Cox. *AESTHETIC PROGRAMMING: A Handbook of Software Studies*. London: Open Humanities Press, 2021.
- Thylstrup, Nanna Bonde. "Data Out of Place: Toxic Traces and the Politics of Recycling." *Big Data & Society* 6, no. 2 (2019): 205395171987547. doi:10.1177/2053951719875479.
- Thylstrup, Nanna Bonde. "The Ethics and Politics of Data Sets in the Age of Machine Learning: Deleting Traces and Encountering Remains." *Media, Culture & Society* 44, no. 4 (2022): 655–671. doi:10.1177/01634437211060226.
- Timnit, Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford. "Datasheets for Datasets." *Communications of the ACM* 64, no. 12 (December 2021): 86–92. doi:10.1145/3458723.
- Turing, A. M. "Computing Machinery and Intelligence." *Mind* LIX, no. 236 (October 1, 1950): 433–460. doi:10.1093/mind/LIX.236.433.
- Wölfflin, Heinrich. *Principles of Art History: The Problem of the Development of Style in Later Art*. New York: Dover, 1950.
- Ye, K. "Silicon Valley and the English Language." In *Affecting Technologies, Machining Intelligences*, edited by Dalida Maria Benfield, Bruno Moreschi, Gabriel Pereira, and K. Ye, CAD+SR, 2021. <https://book.affecting-technologies.org/for-more-diversity-in-technology/>
- Zylinska, Joanna. *Nonhuman Photography*, Cambridge, Massachusetts: The MIT Press, 2017.
- Zylinska, Joanna. *AI Art: Machine Visions and Warped Dreams*. 1sted. London: Open Humanities Press, 2020.

Pereira has a PhD degree from Aarhus University (Denmark). His research focuses on critical studies of data, algorithms, and digital infrastructures, particularly those of computer vision. <https://www.gabrielpereira.net/>

Bruno Moreschi is the co-coordinator of the Group on Artificial Intelligence and Art (GAIA)/C4AI at the University of São Paulo. He is a researcher of the project Decay without mourning: Future thinking heritage practices, Riksbankens Jubileumsfond/Volkswagen Foundation. Moreschi received his PhD in Arts at the State University of Campinas (Unicamp). Moreschi's investigations are related to the deconstruction of systems and the decoding of social practices in the fields of arts, museums, visual culture and technologies. He is a Senior Research Fellow at the Center for Arts, Design + Social Research.

Appendix

Attachment 1: Teaching guide for “Living with images from large-scale data sets”

(This exercise guide is just a skeleton. Please adapt it to your students' needs and possibilities, and complement it with discussion and exposition that fits your class' goals and framing.)

Learning goals: 1) Students will apply a practice-based research method to explore the formation of data sets through visual elements; 2) Students will analyze and reflect on their experience with these images, and discuss them in-class as they relate to issues of data set creation and use; 3) Through building responses to images and in-class discussions, students will build critical knowledge on issues around the scale of data sets.

Prior knowledge required: This exercise is well-suited for students at any level, particularly in courses dealing with critical approaches to data and algorithms in society. It fits well with other activities around the curation and use of data sets. If students are not at all familiar with data/algorithms, a brief explanation of the operation of computer vision is suggested.

Time required: This plan includes a take-home exercise. In the first class, students briefly discuss issues around data sets and their scale, as well as receive images to live with. Between the first and second class, at least one week passes where students live with their images and prepare any form of response. In the second class, students are asked to share their responses and discuss them. The second class concludes with the takeaway points from the exercise, as they relate to the development and use of data sets in society.

Materials required: 1 to 3 printed images per student. If it is not possible to print images, give each student a PDF file or ask them to pick their own images. See Attachment 2 for a tool that automates the process of generating images to live with. Please always double-check all images before giving them to students to avoid issues with sensitive images.

Suggested readings: Denton et al, “On the genealogy of machine learning datasets”; Malevè, “On the data set's ruins”; Thylstrup, “The ethics and politics of data sets”; Crawford and Paglen, “Excavating AI”; Prabhu and Birhane, “Large image datasets”; and Smits and Wevers, “The agency of computer vision models”.

First class: Introduce the exercise. (Around 30 minutes.)

- Ask students: Have you ever seen a data set? Do you think your images are part of a data set? What do you think happens when images are added to data sets?
- Use students' responses to set up for the assignment. You can engage students by showing the grandeur of data sets millions of images and/or with real-life cases and the questions that emerge from them, e.g.:
 - Clearview AI and their construction of a database indexing 20 billion images of people's faces, without their consent.
 - The case of ImageNet Roulette and the consequences of this intervention.
- Set up *why* students will be doing this assignment:
 - Some would call these “images that should not be seen”, as their goal is just the support of machine learning models. These are large-scale databases, with millions and millions of items. But what could happen if we engaged with them from up

close? What could be learned by looking at the images in themselves, not as a large-scale collection?

- Give students three images each.
- Ask students to write down what they see in this image in 20 seconds. Explain this is their initial impressions, which is what computer vision algorithms most often focus on (the tags created by microworkers). Our goal, however, is to get closer to such images and see what else is in them that we might not see at first glance. (Ask students to keep this first description for next class.)
- Explain the assignment:
 - You will be taking home three images. You should live with these images in whichever way you feel comfortable, and prepare a response. Try to create reflections on this process as you go, not only in the very end. These can take shape in any (multimedia) format: images, drawings, videos, etc.
 - Complete instructions will be published on student LMS (see section below).
- Set students' expectations that living with some of these images can be hard because they are pixelated and often may seem boring at first glance. If this difficulty comes up, treat it as part of the exercise and reflect on it!

Assignment instructions to be published in student LMS

This assignment asks that you live with three images from large-scale computer vision data sets. You have received (or will receive) three images from your instructor.

To live with an image means to have it around you, to spend time with these images by looking at them at least for a little bit of time every day. To remember to do this, you may want to place it somewhere they'll be seen, for example: on the table where you have breakfast, next to your computer screen, or even next to your toothbrush. You may also want to take your images on a hike, to get them out there to experience the world! Make it fun!

As you live with your image, you may want to take notes of your thoughts or any things that come up. You can write down your thoughts on your phone, computer, notepad, or even the back of the image or on the image itself.

Everything matters in getting closer to this image, so try to write things up! Here are some questions to get you going, but many others are possible:

- What do you think is the story of your image(s)? How would you tell them to your friends and family? Does this story change as you look more and more at this image?
- How do you think your three images relate? What images from your life would you like to put next to them? What relations emerge from the connection between these images and new images?
- How do you think these images have been used before? Where do you think they come from, and how do you think they are to be used by algorithms? How do you think an algorithm would understand it, and why?

For the next class, please come with your responses from living with your images. Your response can be the notes you've been taking, but if you feel creative, you can also bring something else. For example, you may write a poem about your image, create other images based on them, compose a song, or even film something with your phone. There are no limits to your creativity, as long as it relates to and reflects on your experience of these images.

(Depending on your preference, you may request students to publish their responses in the LMS before the second class.)

Second class: Post-assignment discussion. (From one to two hours.)

- Start the class with a prompt: In 30 seconds write down what you see in this image now that you've lived with it. How is it different than your initial perception?

- Open up the class for students to discuss their experiences and invite them to share their responses, e.g.:
 - Where did you place your images? What did it mean to live with them for you? What unexpected things happened by living with them?
 - How do you think getting close to them changed your experience of the images? Get them to show the images and share their responses with the class. (This can be done in small groups if class size allows.)
- Present key points around data sets and scale, by linking the experience of students to the suggested readings and/or exposition by the lecturer, e.g.:
 - Data sets are formed around simple descriptions or tags that are given to images. These descriptions are limiting and can be problematic.
 - There is an assumption of scale across the production of computer vision models. This leads to many potential problems, including in the provenance of data, microworker labor, and the right to deletion.
- Discuss take-away points with students, e.g.: How should data sets be developed and used? The project ‘Data sheets for data sets’ sought to engage data set developers in documenting data sets, including potential future use and how to maintain the data.

Attachment 2: ‘Images to live with’ tool

We have created an online tool to automatically generate PDFs containing images from large-scale image data sets.

The tool is available at

<https://learningfromlivingwithdata.herokuapp.com/>

or at

<https://www.gabrielpereira.net/imagestolivewith/>

Images to live with

This is a tool to create PDFs with data set images for printing.

To generate a PDF with images to live with, fill the form below with the number of images you want (max 50) and press OK.

The tool automatically picks from a selection of 14436 images randomly chosen from 8 different datasets: ("Tiny ImageNet", "COIL-100", "Labelled Faces in the Wild", "Leaf", "Open Image", "Stanford Dogs", "Tencent", and "VisualGenome").

For more details, read the article "Living with images from large-scale data sets: A critical pedagogy for scaling down", on the Photographies journal.

Full code can be found on [GitHub](#).



Created by [Gabriel Pereira](#) and [Bruno Moreschi](#).

