Chiquita Room

MODELS OF SEEING

AN EXHIBITION BY ESTAMPA

14.11.2024 - 21.12.2024

WHAT DOES IT MEAN TO SEE A PERSON?

The problem of seeing has attracted increased critical attention in the age of computer vision. The YOLO9000 algorithm that has been playfully explored by Estampa in this exhibition perfectly illustrates the nature of this problem. YOLO is an object detection system known for its accuracy and speed. Its name stands for a promise that may actually be a threat: 'You Only Look Once'. In his celebrated 1972 TV series and book, *Ways of Seeing*, art critic John Berger points out that seeing has never been just a physiological process; it also involves cultural habits and conventions.[1] In both machines and humans (who increasingly rely on machines big and small to form a picture of the world) seeing entails framing. Photography has perfected the technique of placing rectangles within the mobile world in an attempt to stabilise it. Over the years photography theory has developed a knowing engagement with this practice. Writers such as Susan Sontag, John Tagg, Ariella Aïsha Azoulay and Mark Sealy[2] have variously argued that, even though cameras capture objects and hold them to help us make sense of the world, this process is only ever temporary and partial. Even though there is a violence to such acts of capture, we can minimise it by confronting our desire to freeze, control and possess the world.

This critical awareness around images does not seem to have transferred to mainstream computer science though. Its current star subfield, computer vision, relies on several conceptual and technical tricks in its enactment of 'seeing': the elimination of depth within the image, the merging of foreground and background, and the flattening of objects into patterns. 3D objects from the real world are thus rendered as 2D models through the selection of their 'features'. The image becomes segmented through the insertion of multiple internal frames into it to isolate some entities. As a result, a photograph is reduced to 'a collection of objects to label' rather than being seen as 'a cohesive whole or a composition with relational meaning'.[3] Relationality in photography extends to the historical and social context, the deliberative and often conflicting construction of meaning through categories and concepts – that are always more than labels. (You could

---

[1] See John Berger (1972) *Ways of Seeing*. Londres: British Broadcasting Corporation and Penguin Books.

[2] See Susan Sontag (1977) *On Photography*. Nova York: Farrar, Straus and Giroux; John Tagg (1988) *The Burden of Representation: Essays on Photographies and Histories*. Amherst: University of Massachusetts Press; Ariella Aïsha Azoulay (2019) *Potential History: Unlearning Imperialism*. Londres: Verso, 2019; Mark Sealy (2019) *Decolonising the Camera: Photography in Racial Time*. Londres: Lawrence & Wishart.

[3] Amanda Wasielewski (2023) «Authenticity and the Poor Image in the Age of Deep Learning», *photographies* 16(2): 191-210, 195-196. Per a més informació sobre aquest tema, vegeu Joanna Zylinska (2023) *The Perception Machine: Our Photographic Future Between the Eye and AI*. Cambridge: MIT Press; Joanna Zylinska (2024) «Diffused Seeing: The Epistemological Challenge of Generative AI», *Media Theory*, publicació especial a «Photographic Seeing», Vol. 8 N.º 1: 229-258.

try attaching averaged imagistic representations to 'truth', 'democracy' and 'in/justice', for instance, but you will inevitably produce some crass generalisations, not to mention commit acts of epistemic violence, because questions will arise with regard to the decisions made about the images, the forms of exclusion initiated and the picture of the world perpetuated through that very act of picturing and labelling.)

Computer scientists have of course become aware of the accusations of 'bias' aimed at their outputs. Yet most answers to this issue have been technical, premised as they have been on the belief that a larger data set will make any bias evident in the data simply disappear. Over a decade ago Professor Fei-Fei Li's research group at Stanford University rebooted computer vision by expanding the database of photographs called ImageNet from thousands to millions of artefacts by scraping them from different parts of the Internet. The crux to their success lay in using the Amazon labour sweatshop, MTurk, to get anonymous workers to label those millions of images, with the resultant image-text pairs being used at scale to train models in object recognition – now a very fast process which the Estampa show poignantly visualises. In recent years we have witnessed various attempts to move beyond the flatness of images in computer vision that elides the phenomenological thickness of the physical world. And thus in 2024 Fei-Fei Li raised $230 million for a startup called World Labs whose goal is to create AI technology that 'can understand how the three-dimensional physical world works'.[4] The project promises to launch 'spatial intelligence' by 'developing … "large world models" that will be used by professionals such as artists, designers, developers, and engineers'.[5]

But this latest initiative does raise some questions. What will such technology really understand? What kinds of intelligence will it embrace and promote; what kinds of spaces will this intelligence inhabit – and what kinds of worlds will it want to build? Importantly, to pose such questions is not return to human exceptionalism, suggesting that humans are the only species that can develop true understanding and that machines will only ever be our pure imitation. Recent developments with large language models have shown that this is indeed not the case. But it is to return to the framing question of this essay: *What does it mean to see a person?* Through Estampa's work, we could qualify that question further: *What does it mean to see a person who is a worker?* How do we see them? What spaces do they occupy? What are they doing there? To what purpose and under what conditions?

---

[4] Anna Tong i Katie Paul (2024) «AI godmother' Fei-Fei Li raises $230 million to launch AI startup», *Reuters*, September 13, https://www.reuters.com/technology/artificial-intelligence/ai-godmother-fei-fei-li-raises-230-million-launch-ai-startup-2024-09-13/.

[5] Marina Temkin (2024) «Fei-Fei Li's World Labs comes out of stealth with $230M in funding», *TechCrunch* In Brief, September 14, https://techcrunch.com/2024/09/13/fei-fei-lis-world-labs-comes-out-of-stealth-with-230m-in-funding/.

As it currently stands, the instrumentality of seeing in computer vision seems to remain a blind spot when it comes to truly examining these questions. Its mega-effort to see across all spaces and data points will most likely end up averaging all the outliers in the data set, with the economic and ecological messiness completely erased from the picture.

Joanna Zylinska

Artist, writer and Professor of Media Philosophy + Critical Digital Practice at King's College London. She is an author of a number of books, including *The Perception Machine* (2023), *AI Art: Machine Visions and Warped Dreams* (2020) and *Nonhuman Photography* (2017). Her art practice involves experimenting with different kinds of image-based media.

MODELS OF SEEING

We look and are constantly being looked-at. The world appears to us through screens and awakens a desire to communicate with it, to produce and exchange images with our network of relationships. An economic interest in extracting statistical data from all the barrage of social images has led—in recent years—to the assembly of new contraptions of vision. One of the most widespread of these on the web is automatic image tagging. Each new online post sets in motion operations that, while invisible to the human eye, identify and label everything they pick out in these images. These processes, affixed to the digital image, constitute the flip-side of visual culture, a form of agency that runs two ways in which the person looking is also being looked at.

We also use the term "vision" to refer to a worldview or a culture. For instance, we speak of theories or authors that offer us a specific vision of a given topic. Vision always entails an inherent perspective or specific point-of-view on the world. Getting computers to "see" also involves the assembly of these systems based on a specific vision, in tune with the values of the person who designed them. They are instruments of vision that codify certain conceptions of the world and put them into circulation as mathematical models. They are models of seeing, and they are the vectors of a vision that acts by proxy.

These models of seeing are based on words being mapped onto meaningful fragments of each image. They are guided by an organising drive that operates by accumulation, weaving out a web of tags intended to tame and explain away images. However, this process is not exempt from disturbances. When we interweave image and text, we have two options: either to use the text to explain the image, to preclude its meaning, or to use it to multiply its meanings. These models of seeing, designed to act according to the first option, often end up falling in line with the latter. This phenomenon—which can be a real headache for the engineers working on them—becomes rather interesting from a creative perspective since certain descriptors bear an unexpected influence on the interpretation of the original image, leading to an intensification and posing a challenge to the viewer's imagination. These flights of meaning, the frictions brought about in trying to capture that which is represented lead us to the interminable matter of overlapping words and images, to the tension therein, to the boundaries and poetic possibilities of this relationship.

The vast majority of images on display here are not the product of image manipulations or of tinkering with artificial intelligence (AI) models: they are the result of a process of curating accidents generated by automated systems of artificial vision. Out of curiosity, we set about processing films from our archive and setting our sights on these misperceptions and collecting them with simple screenshots. In this gesture, the models are drawn out of their operational functions to shift into a poetic function that exposes part of their

own failure. In these cases, the graphic output of the tags has been tampered with, overwhelming the original image until it is swamped. The ubiquitousness of recognition systems, usually hidden from our sign, lays bare the fact that images now harbour a secret language that inhabits them and transforms them. Thus, each digital image becomes a borderline between the extractivist drive and all the phantom seepages of representation, between our vision and the vision of the images.

In the Western world, we understand the gaze in receptive terms. Based on a cultural mirroring with optical mechanisms, we understand that looking is equivalent to receiving external stimuli through sight. Yet in other cultures, looking can also understood in projective terms. Anthropologist Roger Canals argues that in digital space, our gaze emphasises a more projective dimension since it is a gaze that leaves a trace in the form of data. Today, even the mere viewing of images now leads to a statistical index that will be processed in some far-flung data centre. But what if we turn this around? How is our gaze affected by the images looking at us? Does the stigma of the words leave a trace on our retina? In this case, it would behove us to open up words and images, to disentangle their spectral normativity, conjuring forth unforeseen relationships.

Estampa

Barcelona-based artistic collective of programmers, filmmakers and researchers. Their practice is based on a critical and archaeological approach to audiovisual technologies, with a specific eye to archives and experimental video. Since 2017, one of their main lines of action has been researching the uses and ideologies of AI.

ACTIVITIES

Thursday 14 November, 7 p.m.

Opening of the exhibition

Thursday 21 November, 7 p.m.

crater lover worker book launch and screening of films by Estampa:

- ¿Qué es lo que ves, YOLO9000? (2019) – 13:43
- Espècies marcianes (2021) – 02:47
- The Vertigo of the Ways of Seeing (2022) – 17:02

Chiquita Room

Villarroel 25, Barcelona Tel. (+34) 93 146 68 75 hola@chiquitaroom.com www.chiquitaroom.com