

Chiquita Room

MODELOS DE VER

UNA EXPOSICIÓN DE ESTAMPA

14.11.2024 - 21.12.2024

¿QUÉ SIGNIFICA VER A UNA PERSONA?

El problema de la visión ha atraído cada vez más la atención de la crítica en la era de la visión artificial. El algoritmo YOLO9000, explorado por Estampa con tintes lúdicos en esta exposición, ilustra perfectamente la naturaleza de esta cuestión. Yolo es un sistema de detección de objetos conocido por su precisión y velocidad. Su nombre le viene dado por una promesa que, en realidad, podría ser una amenaza: «You Only Look Once». En su célebre serie de televisión y libro de 1972, *Ways of Seeing*, el crítico de arte John Berger pone el foco en que ver nunca ha sido solo un proceso fisiológico, sino que también implica hábitos y convenciones culturales¹. Tanto en las máquinas como en los humanos —quienes cada vez dependen más de grandes y pequeñas máquinas para formarse una imagen del mundo— ver implica enmarcar. La fotografía ha perfeccionado la técnica de colocar rectángulos en el mundo móvil en un intento de estabilizarlo. A lo largo de los años, la teoría de la fotografía ha desarrollado un compromiso intencional con esta práctica. Escritoras y escritores como Susan Sontag, John Tagg, Ariella Aïsha Azoulay y Mark Sealy² han argumentado de diversas formas que, aunque las cámaras capturan objetos y los mantienen para ayudarnos a dar sentido al mundo, este proceso es solo temporal y parcial. Si bien hay cierta violencia en tales actos de captura, podemos minimizarla enfrentándonos a nuestro deseo de congelar, controlar y poseer el mundo.

Sin embargo, esta conciencia crítica en torno a las imágenes no parece haberse transferido a la informática convencional. Su subcampo estrella actual, la visión artificial, se basa en una serie de trucos conceptuales y técnicos en su implementación de la «visión»: la supresión de la profundidad dentro de la imagen, la fusión del primer plano con el fondo, y el aplanamiento de los objetos en patrones. De esta manera, los objetos tridimensionales del mundo real se representan como modelos bidimensionales mediante la selección de sus «características». La imagen se segmenta insertando múltiples marcos internos para aislar algunas entidades. Como resultado, una fotografía se reduce a «una colección de objetos para etiquetar» en lugar de ser vista como «un todo cohesionado o una

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

² Véase 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 Zylińska (2023) *The Perception Machine: Our Photographic Future Between the Eye and AI*. Cambridge: MIT Press; Joanna Zylińska (2024) «Diffused Seeing: The Epistemological Challenge of Generative AI», *Media Theory*, publicació especial a «Photographic Seeing», Vol. 8 N.º 1: 229–258.

composición con significado relacional»³. La relacionalidad en la fotografía se extiende al contexto histórico y social, la construcción deliberativa y, a menudo, conflictiva del significado a través de categorías y conceptos, que siempre son más que etiquetas. (Podemos tratar de asociar representaciones imaginativas promediadas a conceptos como «verdad», «democracia» y «justicia», por ejemplo, pero inevitablemente se producirán algunas generalizaciones burdas, por no mencionar la comisión de actos de violencia epistémica, pues surgirán preguntas con respecto a las decisiones tomadas sobre las imágenes, las formas de exclusión iniciadas y la imagen del mundo perpetuada a través de ese mismo acto de representar y etiquetar).

Por supuesto, los informáticos son conscientes de las acusaciones de «sesgo» dirigidas a sus resultados. Sin embargo, la mayoría de las respuestas a este problema han sido técnicas, basadas en la creencia de que un conjunto de datos más amplio permitiría acabar con cualquier sesgo evidente en los datos. Hace más de una década, el grupo de investigación liderado por el profesor Fei-Fei Li de la Universidad de Stanford dio un nuevo arranque a la visión artificial al expandir la base de datos de fotografías ImageNet de miles a millones de artefactos, extrayéndolos de diferentes partes de Internet. La clave de su éxito radicó en utilizar la fábrica clandestina de Amazon, MTurk, para que trabajadores anónimos etiquetaran esos millones de imágenes y los pares imagen-texto resultantes se utilizaron a gran escala para entrenar modelos en el reconocimiento de objetos; actualmente un proceso muy rápido que Estampa muestra de forma conmovedora. En los últimos años hemos sido testigos de varios intentos por ir más allá de la planitud de las imágenes en la visión artificial que elide el grosor fenomenológico del mundo físico. Y así, en 2024, Fei-Fei Li ha recaudado 230 millones de dólares para la *start-up* World Labs, cuyo objetivo es crear tecnología de IA que «pueda entender cómo funciona el mundo físico tridimensional»⁴. El proyecto promete lanzar la «inteligencia espacial» mediante el «desarrollo de [...] "grandes modelos mundiales" que utilizarán profesionales como artistas, diseñadores, desarrolladores e ingenieros»⁵.

Pero, esta última iniciativa plantea algunas preguntas: ¿qué entenderá *realmente* esa tecnología? ¿Qué tipo de inteligencia abarcará y promoverá; ¿qué tipo de espacios habitará, y qué tipo de mundos querrá construir? Es importante señalar que el hecho de plantear estas preguntas no supone volver al excepcionalismo humano, sugiriendo que el ser humano es la única especie que puede desarrollar

⁴ Anna Tong i Katie Paul (2024) «AI godmother' Fei-Fei Li raises \$230 million to launch AI startup». *Reuters*, 13 de septiembre, <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, 14 de septiembre, <https://techcrunch.com/2024/09/13/fei-fei-lis-world-labs-comes-out-of-stealth-with-230m-in-funding/>.

una comprensión verdadera y que las máquinas solo serán nuestra pura imitación. Los últimos avances en grandes modelos lingüísticos han demostrado que, en efecto, esto no es así. Sin embargo, sí supone retomar la pregunta que enmarca este ensayo: *¿qué significa ver a una persona?* A través del trabajo de Estampa, podríamos matizar aún más dicha pregunta: *¿qué significa ver a una persona que es un trabajador?* ¿Cómo la vemos? ¿Qué espacios ocupa? ¿Qué hace allí? ¿Con qué propósito y en qué condiciones? En su forma actual, la instrumentalidad de ver en visión artificial parece seguir siendo un punto ciego al tratar de examinar de verdad estas cuestiones. El megafuerzo por ver a través de todos los espacios y puntos de datos probablemente terminará promediando todos los valores atípicos en el conjunto de datos, con la consiguiente supresión total del desorden económico y ecológico de la imagen.

Joanna Zylinska

Artista, escritora y profesora de Filosofía de los Medios y Práctica Digital Crítica en el King's College de Londres. Es autora de varios libros, entre ellos, *The Perception Machine* (2023), *AI Art: Machine Visions and Warped Dreams* (2020) y *Nonhuman Photography* (2017). En su práctica artística experimenta con diferentes tipos de medios basados en imágenes.

MODELOS DE VER

Miramos y nos miran continuamente. El mundo se nos aparece a través de las pantallas y nos despierta el deseo de comunicarnos, de producir e intercambiar imágenes con nuestra red de relaciones. El interés económico para extraer datos estadísticos de todo este torrente de imágenes sociales ha requerido, en los últimos años, la construcción de nuevos artilugios de visión. Uno de los más extendidos en la red es el de la anotación automática de las imágenes. En cada nueva publicación en línea, se activan operaciones, invisibles en los ojos humanos, que identifican y etiquetan todo aquello que reconocen en las imágenes. Estos procesos, adheridos a la imagen digital, conforman el reverso de la cultura visual, una agencia de doble sentido en la cual quien mira las imágenes es también mirado.

El término «mirada» también lo usamos para referirnos a una cosmovisión o una cultura. Hablamos de teorías o autores que nos ofrecen una mirada particular sobre tal temática, por ejemplo. La mirada siempre implica esta perspectiva propia o un punto de vista específico sobre el mundo. Conseguir que los ordenadores «vean» también supone la construcción de estos sistemas a partir de una mirada particular, la cual coincide con los valores de quien los ha diseñado. Son herramientas de visión que codifican y ponen en circulación determinadas concepciones del mundo en forma de modelos matemáticos. Son modelos de mirar. Y son portadores de una mirada que actúa por delegación.

Estos modelos de mirar están basados en la asignación de palabras a fragmentos significativos de cada imagen. Los guía una pulsión ordenadora que opera por acumulación, formando una telaraña de etiquetas que pretenden explicar y domesticar las imágenes. Pero este proceso no resulta exento de alteraciones. Cuando imbricamos imagen y texto, tenemos dos opciones: bien emplear el texto para explicar la imagen, para cerrar el significado, o bien la posibilidad de multiplicar los sentidos. El caso es que los modelos de mirar, diseñados para funcionar según la primera opción, a menudo acaban concurriendo en la segunda. Este fenómeno, que para los profesionales de la ingeniería que trabajan en ello puede ser un quebradero de cabeza, desde el punto de vista creativo se vuelve muy interesante, porque ciertos descriptores influyen de manera inesperada en la interpretación de la imagen original, produciendo una intensificación y planteando un reto para la imaginación del espectador. Estas fugas de sentido, las fricciones provocadas por el intento de capturar aquello representado, nos remiten a la inacabable cuestión de superponer palabras e imágenes, a su tensión, a los límites y posibilidades poéticas de esta relación.

La gran mayoría de las imágenes aquí expuestas no son el producto de manipular la imagen ni de trastear con los modelos de inteligencia artificial, son el resultado de un proceso de comisariado de accidentes producidos por los automatismos de la visión artificial. Con curiosidad, nos hemos dedicado a procesar películas

de nuestro archivo y a poner la atención en estos equívocos y coleccionarlos con simples capturas de pantallas. En este gesto, los modelos son extraídos de su función operativa para mutar en una función poética que evidencia parte de su fracaso. En algunos casos, la salida gráfica de las etiquetas sí que ha sido manipulada, desbordando la imagen original hasta el punto de saturarla. La ubicuidad de las detecciones, habitualmente ocultas a la visión, desvela que las imágenes hoy son huéspedes de un lenguaje secreto que las habita y las transforma. Cada imagen digital se convierte, de esta forma, en una frontera entre la compulsión extractivista y todas las fugas fantasmales de la representación, entre nuestra mirada y la mirada de las imágenes.

En el contexto occidental, concebimos la mirada en términos receptivos. A partir del reflejo cultural con los dispositivos ópticos, entendemos que mirar equivale a recibir estímulos del exterior a través de la vista. Pero en otras culturas mirar también puede entenderse en términos proyectivos. Dice el antropólogo Roger Canals que en el ámbito digital nuestra mirada refuerza una dimensión más proyectiva, por el hecho de que es una mirada que deja traza en forma de datos. La simple observación de imágenes implica hoy un índice estadístico que será procesado en algún remoto centro de datos. Pero, ¿y en sentido contrario? ¿Cómo afecta a nuestra mirada que las imágenes nos estén mirando? ¿El estigma de las palabras deja traza en nuestras retinas? En este caso, haríamos bien al abrir las palabras y las imágenes, al deshacer su normatividad espectral, conjurando relaciones inesperadas.

Estampa

Colectivo artístico de programadores, realizadores e investigadores con residencia en Barcelona. Su práctica se basa en una aproximación crítica y arqueológica en las tecnologías audiovisuales y digitales, con un interés particular por el archivo y el audiovisual experimental. Desde 2017 una de sus principales líneas de trabajo ha sido la investigación de los usos e ideologías de la inteligencia artificial (IA).

ACTIVIDADES

Jueves 14 de noviembre, 19 h.

Inauguración de la exposición

Jueves 21 de noviembre, 19 h.

Presentación del libro *crater lover worker* y proyección de piezas audiovisuales de 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