MATTHIAS SCHÄFER
University of Art & Design Linz

This Person Does Exist

KEYWORDS Computer Vision, Creative Commons, Dataset Art, Data Mining, Machine Learning, Photography.

HOW TO CITE Schäfer, Matthias. 2021. "This Person Does Exist". *Temes de Disseny* 37: 214-225. https://doi.org/10.46467/TdD37.2021.214-225

ABSTRACT This Person Does Exist is an artistic approach to exploring a large dataset of photographic portraits in a randomised manner. The dataset was originally created by Nvidia Research Lab, which has scraped and analysed creative commons images from the popular image hosting platform Flickr. These pictures were then used to train a machine learning model which can create new stochastic images of faces. In contrast to a popular website that showcases the computer generated images, I am displaying random faces from the dataset with their corresponding metadata.
This essay looks into extractivist mechanisms in current machine learning techniques, using the internet to populate and refine databases, while focusing on artistic approaches that expose them. I make the case for Dataset Art as an emerging field which reframes scientific corpora by placing them into galleries and exhibiting them as found objects online.
Finally, I argue that this artistic practice is a legitimate way of opening up a larger public discourse, although artists working with human data must be aware of ethical issues and responsibilities regarding privacy and consent.

MAKING ART WITH HUMAN DATASETS

As machine learning techniques in computer vision advance, they demand larger training sets. Scientists filled the gap by algorithmically scraping and categorising images from sites like Flickr where they could exploit creative commons licenses to copy and create derivatives of people's photographs. This practice has been widely unknown to the general public, as shown by interviews of affected individuals by the New York Times (Hill and Krolik 2019) and NBC News (Solon 2019).
Training sets were not made for human consumption, but rather created for the statistical analysis and modelling of patterns between individual data points. By shifting our gaze onto these vast collections, we can find hidden biases that are transported into the model. We can also find a great number of individual efforts, from taking pictures to labelling and categorising them. In this paper, I discuss my approach of looking into one specific dataset and works of art that have dealt with similar issues in the recent past.

INTRODUCTION TO FLICKR FACES

In 2018 the research lab of Nvidia, one of the leading companies for visual computing, published a paper introducing a machine learning architecture called StyleGAN (Karras, Laine, and Aila 2019). They improved on generative adversarial networks (Goodfellow et al. 2014) in such a way that it was possible to create controllable synthetic high-resolution images. In simple terms, scientists are able to abstract large amounts of images with a model that in turn outputs similar-looking pictures. In this case, they were able to generate realistic-looking images of human faces. As previous face datasets were too low in resolution,[1] a new corpus was created by scraping images with Creative Commons, Public Domain or U.S. Government Works licenses from the social photo sharing platform Flickr. The new dataset was named Flickr-Faces-HQ (FFHQ).[2]
Since 2004 Flickr endorsed the use of CC licenses and their open access made it legally possible for researchers to scrape and analyse the material. In fact, the largest open media dataset existing is shared by the research lab of the parent company Yahoo, with 100 million media objects (Thomee et al. 2016).
In response to scrutiny about the failures of IBM's face detection algorithm, which affected black women in particular (Buolamwini and Gebru 2018), the research department created a public database of 1 million faces from the Yahoo dataset. They called it Diversity in Faces (DiF) (Merler et al. 2019) and added questionable metadata of estimated age, gender and craniofacial measurements with the aim of making face detection fairer.[3]
In comparison, FFHQ, which was used to train the above-mentioned generative model, is more than 10 times smaller, as it only contains 70,000 images. The dataset itself is published under a CC-BY-NC-SA license and the instructions for use and download are very clear, making it manageable in terms of size, license and ease-of-use to discover the underlying characteristics.
The crawled images were automatically aligned and cropped around a face to a square ratio with a dimension of 1024^2px with the open-source library dlib. The library provides fast facial recognition and identifies 68 points around the face outlining the chin, eyebrows, eyes, nose and mouth. The resulting dataset was finally checked by Amazon Mechanical Turk workers "to remove the occasional statues, paintings, or photos of photos."[5]
In contrast to other scraped datasets, the Nvidia researchers provide a tool to see if an image is part of the collection and allow the removal of the photograph from the FFHQ training set. As an investigative article from NBC News (Solon 2019) suggests, most photographers on Flickr are oblivious of the fact that they are part of the larger IBM DiF dataset and they are divided in their opinions on being used for facial recognition research. We can assume that this is the case in the FFHQ corpus too, as the researchers did not get consent before creating the collection.

Fig. 1. Cropped 1024 x 1024 px image from the FFHQ dataset, overlaid with face landmarks from the dataset by the author.
Fig. 2. "In The Wild" image scraped from Flickr, overlaid with face landmarks and cropped from the dataset by the author. License: CC-BY-2.0. Crossing Europe Filmfestival Linz.

THIS PERSON DOES EXIST

After Nvidia released its StyleGAN paper, the software developer Phillip Wang published a website[7] that showcases generated faces of people that do not exist. The site quickly took off and alarmed people about the capabilities and potential impact of AI systems generating synthetic media.

As a counter-narrative to the AI image creator, I wanted to showcase the people who were used to train this generative system. My first look into FFHQ was in the spring of 2020 and I faced technical roadblocks such as downloading and disseminating the images from the dataset. I moved the cropped and aligned face images to my server and built a website[8] that displays the faces from people that do exist.[9]

Looking at the individual images facilitates an interpersonal connection with the unknown person and evokes a feeling for the images that were used to train the generative model. As Flickr is used mostly by hobbyists and professional photographers, one can find portraits of children and families, speakers at conferences or people on holiday. The authors claim that FFHQ includes more variety than other face image sets in terms "of age, ethnicity and image background, and also has much better coverage of accessories such as eyeglasses, sun-glasses, hats, etc." (Karras, Laine, and Aila 2019). They admit to biases inherited from the Flickr platform but fail to mention them. One example (Fig. 1) shows such a cropped image from the dataset with face landmarks. The landmarks are part of the metadata and were used to align images for training. The corresponding original (in the wild) image reveals more context and the title clarifies that the image was taken at a film festival brunch (Fig. 2).

To get a sense of scale of the dataset, I compiled all face images into a grid, reducing the size of each image to 16 by 16px (Fig. 3). This simple montage makes it possible to get a feeling for the vast amount of normalised image data.

Another technique to visually find biases in image sets is to average their pixel values. This suppresses outliers, but it allows us to see an overall trend of the dataset. The resulting composite image The Flickr Face (Fig. 4) might reveal a trend towards smiling and light-skinned people in the data set. This correlates with my subjective experience, but further analysis is needed to be able to measure biases in this image set.

The following section presents other artistic projects that used similar techniques to make large image sets experienceable and critique the unsolicited collection of human data.

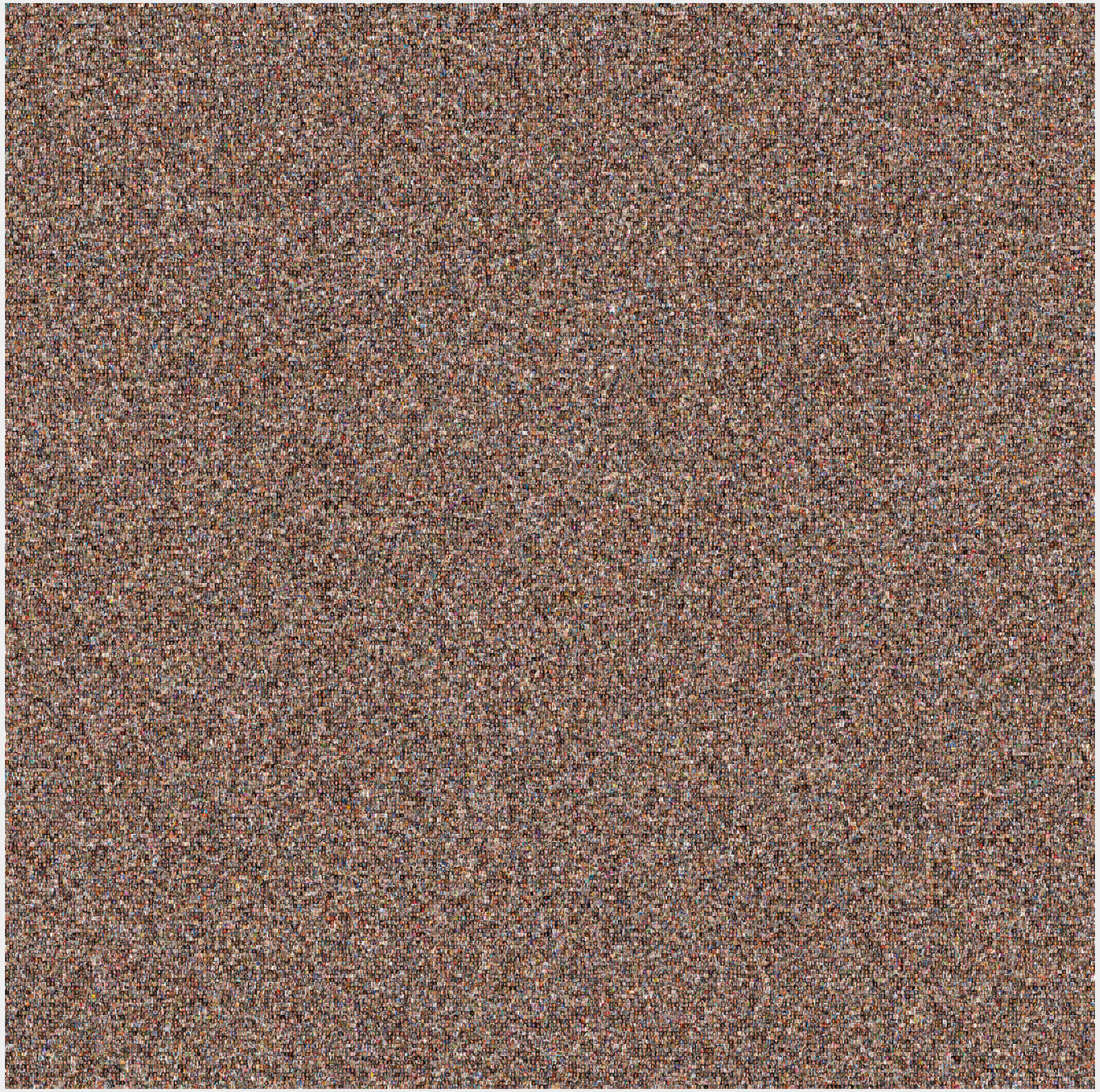
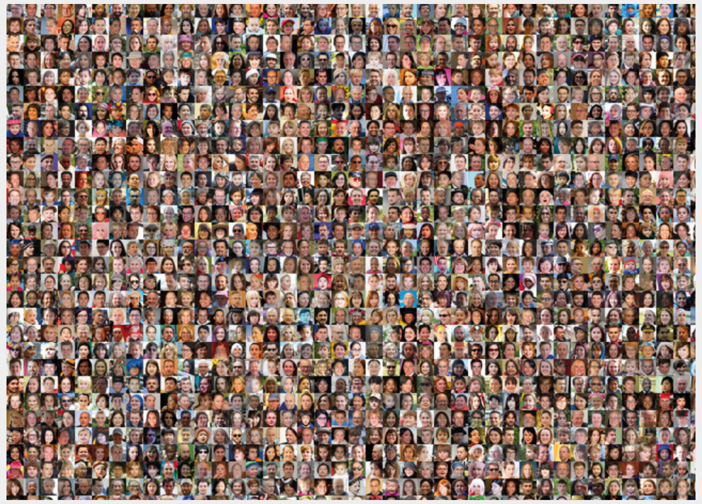


Fig. 3. 69,960 faces from the FFHQ dataset compiled into a grid by the author. On the right, a zoomed area of the grid. License: CC-BY-NC. Original Image Credits can be found at:
https://github.com/NVlabs/ffhq-dataset
https://this-person-does-exist.com/credit.html



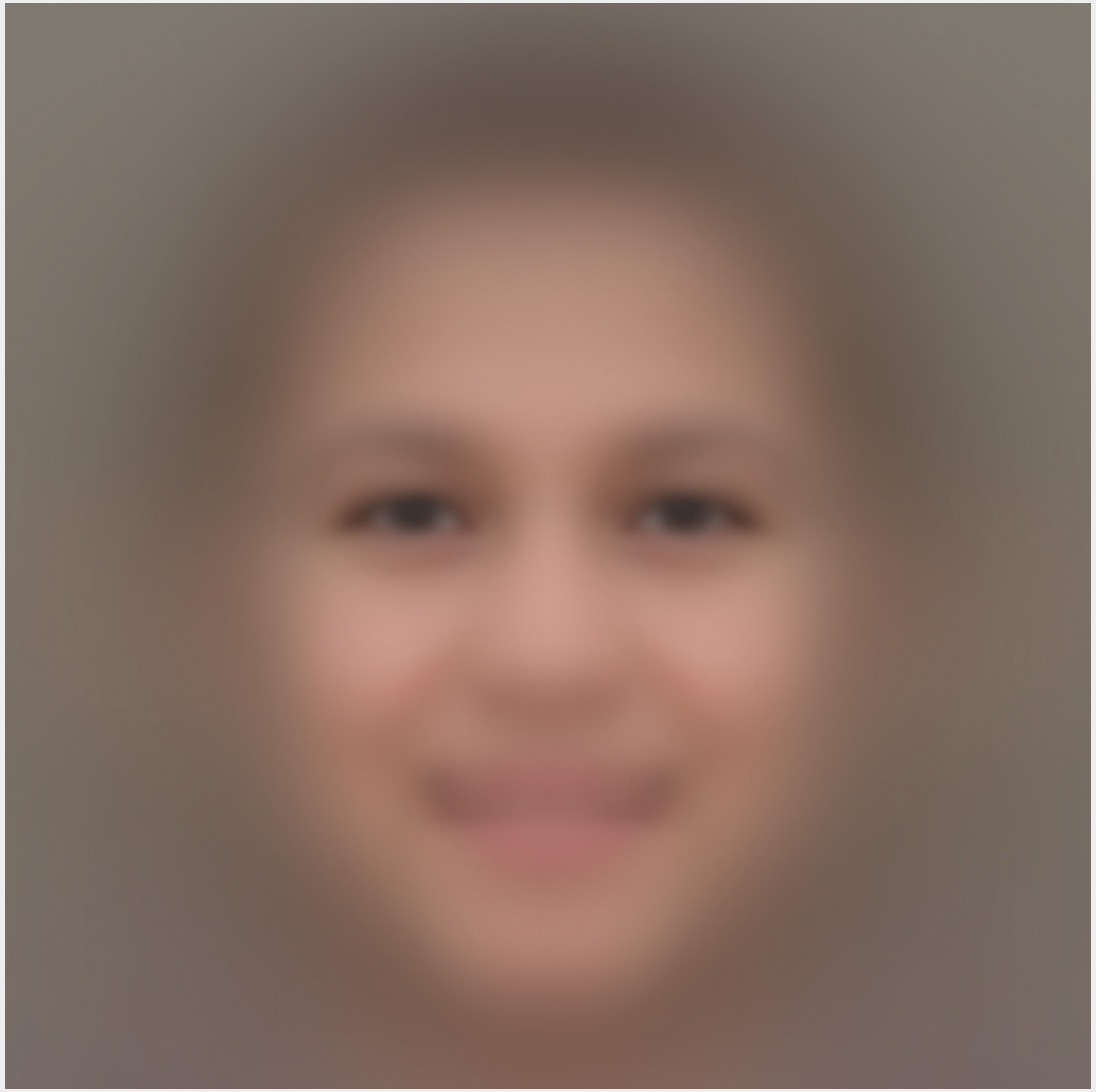


Fig. 4. Averaged Flickr Face. Generated by the author by calculating the average pixel values of all 70k aligned face images.

MEGAPIXELS, TRAINING HUMANS & HUMANS OF AI

Since 2017 Adam Harvey and Jules LaPlace have been working on the art and research publication *MegaPixels*,[10] where they analyse facial recognition and detection datasets. As a result of their research and media attention, multiple datasets have been revoked and removed, for example, the MegaFace dataset (Kemelmacher-Shlizerman et al. 2016) with roughly 4.7 million faces from Flickr (Thomee et al. 2016). The corpus was used by private companies for commercial purposes and the image authors were not sufficiently credited. Furthermore, the dataset contains images of people in Illinois, where the collection, capturing and purchase of biometric information is prohibited (Hill and Krolik 2019).

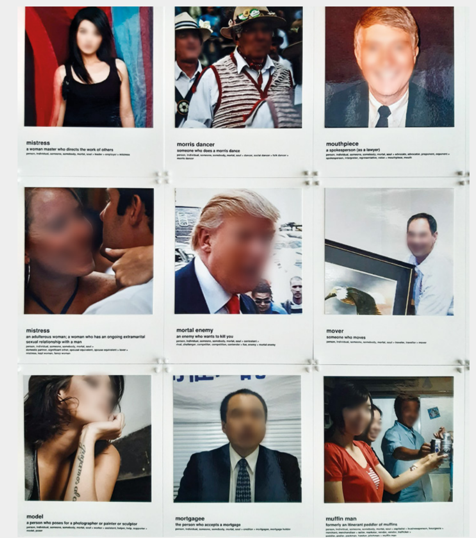
While Harvey and LaPlace have been applying traditional forms of data visualisation and academic publishing, they also found ways of making scraped media in machine learning much more personal.

In 2017 Adam Harvey showed[11] an interactive installation for the Glass Room exhibition in which visitors' face landmarks are compared with images from the MegaFace dataset and the closest matches were displayed.

In 2019 the artist Trevor Paglen and researcher Kate Crawford collaborated on an exhibition titled Training Humans, dedicated to human image databases[12] (Fig. 5). One of the main exhibits was a vast collection of human images from the ImageNet dataset (Fig. 6), initiated by Stanford University's AI professor Fei Fei Li (Deng et al. 2009). It was created to tackle object recognition tasks and consists of 14 million images organised and labelled by Amazon Mechanical Turk workers. Some images of people fall into categories such as "Bad Person, Call Girl, Drug Addict, Closet Queen, Convict" and so on (Crawford and Paglen 2019a). The artists used these absurd, racist and misogynistic labels to train "ImageNet Roulette,"[13] a recognition algorithm that was accessible online and in an interactive installation. A result of the media attention that followed was that 600,000 images were removed from ImageNet and, due to maintenance,[14] the website and dataset have not been properly accessible since (Yang et al. 2019).

But Training Humans comes with its own issues, as Michael J. Lyons (2020) recently wrote in a critique of the work of Crawford and Paglen. Exhibiting some of the image sets violates the terms of use, which only allows for non-commercial scientific research. He criticises the double standards of Crawford and Paglen using human data without prior informed consent. Knowing fully about the ethical dilemma, the artists argue against privacy "to have the real conversation about what's going on with these [AI] systems" (Crawford and Paglen 2019b). Even though visitors were able to remove their images from the exhibition, it was sometimes too late for them. This was the case of the photographed volunteers from the exhibited JAFFE dataset by Lyons, Kamachi, and Gyoba (1998), which was never intended for use as a "training set." Only after the exhibition closed did they realise that the images intended for scientific use were distributed all over the internet as advertisement and a bad example for emotion recognition research.

Two other artists focused on the labelling and annotation process of Microsoft's COCO dataset (Lin et al. 2014), which was scraped from Flickr and annotated by clickworkers for object recognition models.



↓ **Fig. 5.** A photograph from the exhibition Training Humans.
→ **Fig. 6.** A photograph of the exhibited ImageNet dataset and their labels. Faces blurred by the author.

Fig. 7. Screenshot of a Certificate of Appreciation by Philipp Schmitt.

Noticing the missing image attributions, the artist Philipp Schmitt published Humans of AI.[15] He created certificates of appreciation for all 34,248 contributors (Fig. 7) and made the list into a new dataset coCOCO (context for Common Objects in Context). Using YOLO (Redmon and Farhadi 2018), an object detection system that was trained on COCO, he created the Declassifier (Fig. 8), which superimposes images from the dataset onto another picture. Finally, he produced a slideshow that fades between all images from the collection, showcasing an endless amount of mundane amateur photographs. Another artist, Sebastian Schmieg, has acknowledged the work of the unknown clickworkers outlining objects for the COCO set by creating segmentation.network,[16] a website that endlessly redraws and overlays the 600,000 hand drawn outlines (Fig. 9).

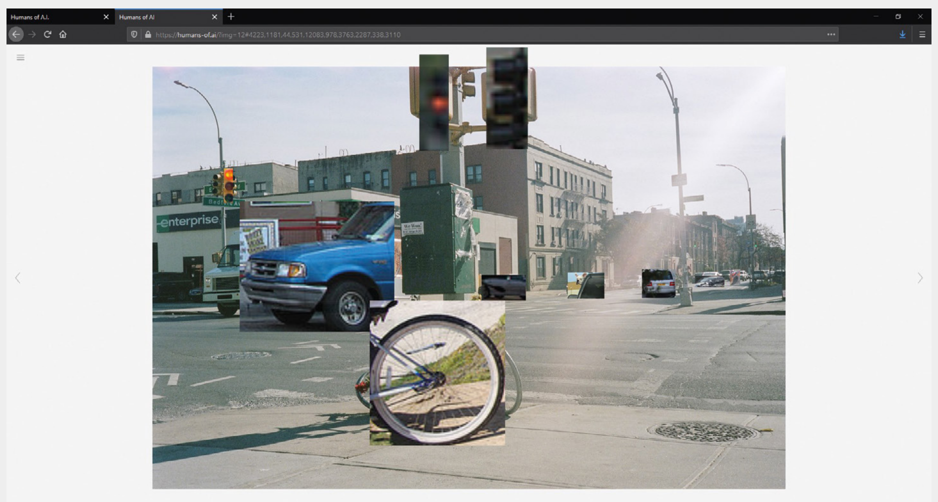
CONCLUSION

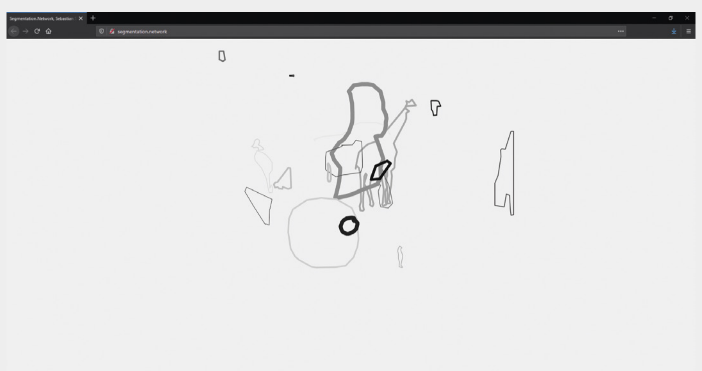
While big datasets have enabled new image technologies like generative adversarial networks and detection algorithms, the research community exploits open licenses to do so. In most cases, the contributors to scraped datasets do not receive fair recognition and are unaware of their material being used.

Artists have responded to the creation of these datasets by analysing and exhibiting them in galleries and online. Their works make large datasets comprehensible and engaging. In some cases they have created enough media attention and scientific discourse for institutions to remove questionable parts or entire collections.

With my research into the FFHQ dataset, I found common issues that tie it to the other presented works, which include the use of human data without prior informed consent, particularly the use of images with CC licenses from Flickr and the use of clickworkers to refine the dataset. Other issues were not present: the authors attributed the creators accordingly, created a tool to find and delete images from the dataset and released the collection under a CC non-commercial license itself.

Still, looking through the automatically cropped images of human faces creates a certain unease of human identities reduced to decontextualised statistical occurrences. My work shows the person that does exist behind the AI model, but it does violate personality rights as much as the original dataset. This ethical dilemma needs to be explored further, but so far I think Dataset Art, as shown in this paper, helps to broaden the discussion and create awareness of practices in ML research.[17]





↗ Fig. 8. Screenshot of the Declassifier from Humans of AI by Philipp Schmitt.
→ Fig. 9. Screenshot of the website segmentation.network by Sebastian Schmieg.

BIOGRAPHY

Matthias Schäfer
University of Art & Design Linz

Matthias Schäfer is an artist who explores algorithmic cultures. As pitscher, he creates performances, installations and web based works with which he engages with centralised online platforms in a deconstructive and humorous way. His works have been exhibited at transmediale, Ars Electronica Festival, Xie Zilong photography museum, Roehrs & Boetsch and multiple other exhibitions and festivals.
Schäfer is currently enrolled in the Interface Cultures Master programme at the University of Art & Design in Linz, where he is focusing on the field of machine learning, automation and its core mechanism of data collecting.

ENDNOTES

1. For the preceding Paper 'Progressive Growing of GANs for Improved Quality, Stability, and Variation' (Karras et al. 2017) the authors upscaled images from the popular CelebA (30,000 images) dataset to meet their demands.
2. See https://github.com/NVlabs/ffhq-dataset
3. The inclusion of computer-generated face measurements, age and gender assumptions were highly criticized, because of their roots in phrenology. According to a discussion on GitHub (https://github.com/tensorflow/datasets/issues/299) the dataset is not available from IBM, though it is still being publicised by IBM.
4. See http://dlib.net/
5. Quote from the readme file in the GitHub repository of the FFHQ dataset (endnote 2). Accessed 25 January 2021.
6. See https://nvlabs.github.io/ffhq-dataset/search/
7. See https://thispersondoesnotexist.com/
8. See https://this-person-does-exist.com/
9. Six days before my website went live, Vincent Woo published a page with the same concept, but that gets the images directly from Flickr and crops them in the browser.
   See http://thispersonexists.net/ and the corresponding tweet https://twitter.com/fulligin/status/1335030372187312128
10. See https://ahprojects.com/megapixels/ and https://exposing.ai/
11. See *MegaPixels*: Faces, https://ahprojects.com/megapixels-glassroom/
12. See *Fondazione Prada, "KATE CRAWFORD | TREVOR PAGLEN: TRAINING HUMANS"*, http://www.fondazioneprada.org/project/training-humans/?lang=en
13. See *Trevor Paglen*, ImageNet Roulette https://paglen.studio/2020/04/29/imagenet-roulette/
14. On 11 March 2021 the team updated https://image-net.org/. They addressed the criticism by removing 2,702 synsets and experiment with blurring faces to preserve people's privacy.
15. See *Philipp Schmitt*, Humans of AI, https://humans-of.ai/
16. See *Sebastian Schmieg*, Segmentation.Network http://segmentation.network/ further information on: http://sebastianschmieg.com/segmentation-network/
17. The publication http://plottingd.at/a/ explores datasets through artistic interfaces and critical reflections much further.

REFERENCES

Buolamwini, Joy, and Timnit Gebru. 2018. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency, PMLR 81*, 77-91. Cambridge, Mass.: Proceedings of Machine Learning Research. http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf

Crawford, Kate, and Trevor Paglen. 2019a. "Excavating AI". The AI Now Institute, NYU. Accessed January 24, 2021. https://www.excavating.ai/

Crawford, Kate, and Trevor Paglen. 2019b. "What are the ethics of exhibiting datasets of faces? An interview with Kate Crawford & Trevor Paglen". Interview by Gaia Tedone. *The Photographers' Gallery*, https://www.youtube.com/watch?v=BSaYROvz29E

Deng, Jia, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. "Imagenet: A large-scale hierarchical image database." In *IEEE Conference on Computer Vision and Pattern Recognition*, Miami, Florida, 2009, 248-255. https://doi.org/10.1109/CVPR.2009.5206848

Goodfellow, Ian J., Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. "Generative Adversarial Networks". https://arxiv.org/abs/1406.2661

Hill, Kashmir, and Aaron Krolik. 2019. "How Photos of Your Kids Are Powering Surveillance Technology" New York Times, October 11, 2019. https://www.nytimes.com/interactive/2019/10/11/technology/flickr-facial-recognition.html

Karras, Tero, Samuli Laine, and Timo Aila. 2019. "A Style-Based Generator Architecture for Generative Adversarial Networks." In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, California, 4396-4405. https://doi.org/10.1109/CVPR.2019.00453

Karras, Tero, Timo Aila, Samuli Laine, and Jaakko Lehtinen. 2017. "Progressive Growing of GANs for Improved Quality, Stability, and Variation." In *6th International Conference on Learning Representations, ICLR 2018*, Vancouver, BC, Canada, April 30-May 3, 2018. https://arxiv.org/abs/1710.10196

Kemelmacher-Shlizerman, Ira, Steven M. Seitz, Daniel Miller, and Evan Brossard. 2016. "The megaface benchmark: 1 million faces for recognition at scale". In *IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, Nevada, 2016, 4873-4882. https://doi.org/10.1109/CVPR.2016.527

Lin, Tsung-Yi, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. 2014. "Microsoft COCO: Common Objects in Context." In *Computer Vision - ECCV 2014*, edited by David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars, 740-755. Cham: Springer. https://doi.org/10.1007/978-3-319-10602-1_48

Lyons, Michael. 2020. "Excavating 'Excavating AI': The Elephant in the Gallery." *Zenodo*. https://doi.org/10.5281/zenodo.4037538

Lyons, Michael, Miyuki Kamachi, and Jiro Gyoba. 1998. "The Japanese Female Facial Expression (JAFFE) Dataset". *Zenodo*. https://doi.org/10.5281/zenodo.3451524

Merler, Michele, Nalini Ratha, Rogerio S. Feris and John R. Smith. 2019. "Diversity in Faces". https://arxiv.org/abs/1901.10436

Redmon, Joseph, and Ali Farhadi. 2018. "YOLOv3: An Incremental Improvement". Preprint, submitted 8 April. https://arxiv.org/abs/1804.02767

Solon, Olivia. 2019. "Facial recognition's 'dirty little secret': Millions of online photos scraped without consent" *NBC News*, March 17, 2019. https://www.nbcnews.com/tech/internet/facial-recognition-s-dirty-little-secret-millions-online-photos-scraped-n981921

Thomee, Bart, David A. Shamma, Gerald Friedland, Benjamin Elizalde, Karl Ni, Douglas Poland, Damian Borth, and Li-Jia Li. 2016. "YFCC100M: the new data in multimedia research." *Communications of the ACM 59* (January 2016): 64-73. https://doi.org/10.1145/2812802

Yang, Kaiyu, Klint Qinami, Li Fei-Fei, Jia Deng, and Olga Russakovsky. 2019. "Towards Fairer Datasets: Filtering and Balancing the Distribution of the People Subtree in the ImageNet Hierarchy". ImageNet, September 17, 2019. http://image-net.org/update-sep-17-2019

Matthias Schäfer

This Person Does Exist

Traducció al Català

PARAULES CLAU Visió per Computador, Creative Commons, Dataset Art, Mineria de Dades, Aprenentatge Automàtic, Fotografia.

RESUM *This Person Does Exist* (Aquesta persona sí que existeix) és una aproximació artística per investigar de manera aleatoritzada un gran conjunt de dades de retrats fotogràfics. El laboratori d'investigació d'Nvidia va crear el conjunt de dades a partir d'una extracció i anàlisi d'imatges amb llicència Creative Commons de la popular plataforma d'allotjament d'imatges Flickr. Aquestes imatges es van usar posteriorment per entrenar un model d'aprenentatge automàtic que és capaç de crear noves imatges estocàstiques de rostres.

A diferència d'un popular lloc web que mostra les imatges generades per ordinador, en aquest article es mostren rostres aleatoris que provenen del conjunt de dades, així com les seves metadades corresponents.

En el present article estudia els mecanismes extractivistes de les tècniques actuals d'aprenentatge automàtic, que utilitzen internet per completar i perfeccionar les bases de dades, i se centra en els enfocaments artístics que les exposen. Es presenta l'argument segons el qual l'art de conjunts de dades és un camp emergent que replanteja els corpus científics en situar-los a galeries i exposar-los com a objectes que s'han trobat a la xarxa.

Per acabar, s'argumenta que aquesta pràctica artística és un model legítim d'obrir un debat que arribi a un públic més ampli, tot i que els artistes que treballin amb dades humanes hagin de ser conscients dels problemes ètics i de les responsabilitats que tenen pel que fa a la privacitat i el consentiment.

CREAR ART A PARTIR DE CONJUNTS DE DADES HUMANES

A mesura que avancen les tècniques d'aprenentatge automàtic de la visió per computador, aquestes necessiten conjunts de dades més amplis per al seu entrenament. Els investigadors van omplir aquest buit a través d'algoritmes que recollien i categoritzaven imatges de llocs com Flickr, on podien valer-se de les llicències Creative Commons (CC) que tenien les imatges per copiar les fotografies d'altres persones i crear imatges derivades a partir d'aquestes. Tal com demostren diverses entrevistes a persones afectades, com les realitzades pel The New York Times (Hill i Krolik 2019) i NBC News (Solon 2019), el públic en general no sap que es produeixen aquestes pràctiques.

Els conjunts d'entrenament no es van dissenyar per al consum humà, sinó que es van crear per dur a terme anàlisis estadístiques i modelar patrons entre punts de dades individuals. En examinar aquestes àmplies col·leccions, s'hi poden trobar prejudicis ocults que es transporten al model. També s'hi pot veure una gran quantitat de treball individual, que va des de fer les fotografies fins a etiquetar-les i categoritzar-les.

En el present article s'explica l'enfocament que s'ha seguit, que es basa en un conjunt de dades concret, així com les obres d'art que han tractat problemes similars durant els últims anys.

INTRODUCCIÓ ALS ROSTRES DE FLICKR

El laboratori d'investigació d'Nvidia, una de les principals empreses del sector de la visió per computador, va publicar el 2018 un article que presentava una arquitectura d'aprenentatge automàtic anomenada StyleGAN (Karras, Laine i Aila 2019). Nvidia va millorar les xarxes generatives antagòniques (Goodfellow et al. 2014) de tal manera que va ser possible crear imatges controlables i sintètiques d'alta resolució. Dit en termes més simples, els investigadors van ser capaços de recollir una gran quantitat d'imatges a través d'un model que, alhora, produïa imatges similars. En aquest cas, van poder generar imatges versemblants de rostres de persones. Com que els conjunts de dades de rostres anteriors tenien molt poca resolució[1], es va crear un corpus nou a partir d'una extracció de fotografies de la plataforma d'allotjament d'imatges Flickr. Per a això, es van utilitzar fotografies que tinguessin llicències Creative Commons, de domini públic o del govern dels Estats Units. A aquest nou conjunt de dades se'l va anomenar Flickr-Faces-HQ (FFHQ).

Des del 2004, Flickr dona suport a l'ús de llicències CC, i el seu accés obert fa que sigui legal que els investigadors n'extreguin i analitzin el material. De fet, el conjunt de dades de mitjans visuals d'accés obert més gran que existeix, que disposa de 100 milions d'objectes, es comparteix amb el laboratori d'investigació de Yahoo, l'empresa matriu de Flickr (Thomee et al. 2016)[2].

En resposta a l'escrutini sobre els errors de l'algoritme de detecció facial d'IBM, que va afectar les dones negres en particular (Buolamwini i Gebru 2018), el seu departament d'investigació va crear, a partir del conjunt de dades de Yahoo, una base de dades pública que contenia un milió de rostres. La van anomenar Diversity in Faces (la diversitat en els rostres, DiF per les sigles en anglès) (Merler et al. 2019) i li van afegir unes metadades qüestionables, com una estimació de l'edat, el gènere i les mesures craniofacials, amb la intenció que la detecció facial fos més justa[3].

En comparació, FFHQ, que es va utilitzar per entrenar el model generatiu mencionat anteriorment, conté tan sols 70 000 imatges (menys d'una desena part). Aquest conjunt es va publicar sota una llicència CC-BY-NC-SA i les instruccions per al seu ús i descàrrega són molt clares, la qual cosa fa que sigui manejable en termes de mida, llicència i treball necessari per descobrir-ne les característiques subjacents.

Es va utilitzar la biblioteca d'accés obert dlib[4] per alinear i retallar les imatges extretes de manera automàtica al voltant del rostre i aconseguir que tinguessin un format d'imatge quadrat d'una mida de 1.024^2 px. Aquesta biblioteca proporciona un reconeixement facial ràpid i identifica 68 punts al voltant del rostre que delineen la barbeta, les celles, els ulls, el nas i la boca. Per acabar, els treballadors d'Amazon Mechanical Turk van comprovar el conjunt de dades resultant per "eliminar les possibles estàtues, quadres i fotografies d'altres fotografies"[5].

A diferència d'altres conjunts de dades que es basen en imatges extretes de llocs web, els investigadors d'Nvidia proporcionen una eina[6] que permet comprovar si una imatge determinada forma part de la col·lecció i eliminar la fotografia esmentada del conjunt d'entrenament FFHQ. Com suggereix un article d'investigació d'NBC News (Solon 2019), la majoria de fotògrafs de Flickr no sap que forma part del conjunt de dades DiF d'IBM, i hi ha opinions diverses sobre el fet que s'utilitzin les seves fotografies per a la investigació del reconeixement facial. Es pot assumir que això també succeeix amb el corpus FFHQ, ja que els investigadors no van obtenir cap tipus de consentiment abans de crear la col·lecció.

AQUESTA PERSONA SÍ QUE EXISTEIX

Després que Nvidia publiqués el seu article sobre la StyleGAN, el desenvolupador de software Phillip Wang va publicar un lloc web[7] en el qual presenta els rostres generats de persones que no existeixen. Aquest lloc es va fer popular en molt poc temps i va generar inquietuds sobre les capacitats i l'impacte potencial dels sistemes d'intel·ligència artificial quan s'usen per produir imatges sintètiques.

Com a contrapunt al creador d'imatges d'intel·ligència artificial, vaig voler mostrar les persones que es van utilitzar per entrenar el sistema generatiu esmentat. La primera investigació sobre el corpus FFHQ es va dur a terme a la primavera del 2020, i s'hi van trobar esculls tècnics com la descàrrega i la disseminació d'imatges del conjunt de dades. Es van moure les imatges de rostres, ja retallades i alineades, al meu servidor i es va crear un lloc web[8] que mostra els rostres de persones que sí que existeixen[9]. El fet de veure les imatges individuals facilita que es creï una connexió interpersonal amb la persona desconeguda i proporciona una idea de les imatges que es van utilitzar per entrenar el model generatiu.

Atès que els usuaris de Flickr solen ser fotògrafs, tant aficionats com professionals, s'hi poden trobar retrats de nens i familiars, ponents en conferències o persones de vacances. Els autors afirmen que el FFHQ inclou més varietat d'imatges que altres conjunts de retrats facials en termes d'"edat, etnicitat i fons d'imatges, i mostra una major presència d'accessoris com ulleres, ulleres de sol i barrets, entre d'altres" (Karras, Laine y Aila 2019).

Així mateix, si bé admeten que hi ha prejudicis que es van heretar de la plataforma de Flickr, no especifiquen quins són. Un exemple (Fig. 1) mostra una imatge retallada del conjunt de dades amb els punts de referència del rostre. Aquests punts de referència formen part de les metadades i s'utilitzen per alinear les imatges i poder entrenar el sistema. La imatge original corresponent (en estat natural) afegeix un context major i el títol clarifica que es va fer en un *brunch* d'un festival de cinema (Fig. 2).

Per entendre millor l'escala del conjunt de dades, es van compilar les imatges dels rostres en una quadrícula i es va reduir la mida de cada imatge a 16 px × 16 px (Fig. 3). Aquest muntatge tan senzill permet fer-se una idea de la gran quantitat de dades d'imatge normalitzades que conté el conjunt.

Una altra tècnica per trobar biaixos de forma visual en els conjunts de dades és fer una mitjana dels seus valors de píxel, ja que permet veure'n la tendència general en suprimir els casos atípics. La imatge composta resultant, anomenada The Flickr Face (El rostre de Flickr) (Fig. 4), revela una tendència cap al somriure i cap a les persones de pell clara en el conjunt. Aquest fet es correlaciona amb l'experiència subjectiva prèvia, però cal una anàlisi més exhaustiva per poder mesurar els prejudicis del conjunt de dades esmentat.

En la secció següent es presenten altres projectes artístics que utilitzen tècniques similars perquè es puguin experimentar els grans conjunts de dades i criticar la recollida de dades humanes sense consentiment previ.

ELS PROJECTES I EXPOSICIONS *MEGAPIXELS*, "TRAINING HUMANS" I "HUMANS OF AI"

Des del 2017, Adam Harvey i Jules LaPlace han treballat en el projecte d'art i investigació *MegaPixels*[10], en el qual analitzen el reconeixement facial i els conjunts de dades de detecció. Com a resultat de la seva investigació i de la cobertura mediàtica que aquesta ha rebut, ja s'han retirat i eliminat diversos conjunts de dades, com el MegaFace (Kemelmacher-Shlizerman et al. 2016), que contenia uns 4,7 milions de rostres extrets de Flickr (Thomee et al. 2016). Algunes empreses privades utilitzaven aquest corpus amb finalitats comercials i no es reconeixia del tot l'autoria de les imatges. A més, entre les imatges que contenia el conjunt, n'hi havia algunes de persones d'Illinois, on està prohibit recollir, obtenir i comprar informació biomètrica (Hill i Krolik 2019).

Tot i que Harvey i LaPlace han fet servir mètodes tradicionals de visualització de dades i de publicació acadèmica, també van idear maneres perquè les imatges extretes per utilitzar-les en aprenentatge automàtic siguin quelcom molt més personal.

El 2017, Adam Harvey va organitzar[11] una instal·lació interactiva per a l'exposició "Glass Room", en la qual es comparaven els punts de referència facials dels visitants amb les imatges del conjunt de dades MegaFace i es mostraven les persones que més s'hi assemblaven.

El 2019, l'artista Trevor Paglen i la investigadora Kate Crawford van col·laborar en una exposició sobre les bases de dades d'imatges humanes[12] (Fig. 5). Una de les obres principals era una àmplia col·lecció d'imatges humanes procedents del conjunt de dades ImageNet (Fig. 6), que va ser creat per Fei Fei Li, professor d'Intel·ligència Artificial de la Stanford University (Deng et al. 2009). Aquest conjunt es va crear per abordar les tasques de reconeixement d'objectes i està format per 14 milions d'imatges etiquetades i organitzades per treballadors d'Amazon Mechanical Turk. Algunes de les imatges estaven categoritzades com a "mala persona, dona de companyia, drogoaddicte, amanerat, presidiari", entre d'altres (Crawford i Paglen 2019a). Els artistes van utilitzar aquestes etiquetes absurdes, racistes i misògines per entrenar ImageNet Roulette,[13] un algoritme de reconeixement que estava disponible a internet i en una instal·lació interactiva. Com a resultat de l'atenció mediàtica que va provocar, es van eliminar 600 000 imatges del conjunt de dades i tant el lloc web com el conjunt en si mateix no han estat disponibles des de llavors per motius de manteniment (Yang et al. 2019)[14].

D'altra banda, l'exposició "Training Humans" també comporta els seus propis problemes, tal com comenta Michael J. Lyons (2020) en la seva crítica a l'obra de Crawford i Paglen. El fet d'exposar algunes de les imatges va en contra de les seves condicions d'ús, ja que només es permet que les imatges s'utilitzin per a la investigació científica, no per a finalitats comercials. Lyons critica la doble moral de Crawford i Paglen, els quals fan servir dades humanes sense haver obtingut un consentiment informat previ. Malgrat que coneixen molt bé el dilema ètic, els artistes es posicionen en contra de la privacitat "per poder parlar de veritat sobre què succeeix en aquests sistemes [d'intel·ligència artificial]" (Crawford i Paglen 2019b). Si bé els visitants van poder eliminar les seves imatges de l'exposició, en el cas dels voluntaris fotografiats per al conjunt de dades JAFFE exposat per Lyons, Kamachi i Gyoba (1998), que mai va estar destinat per a utilitzar-se com a "conjunt d'entrenament". Va ser tan sols un cop l'exposició ja havia acabat quan es van adonar que les imatges destinades a l'ús científic s'havien distribuït sense cap limitació per internet.

Dos artistes més es van centrar en el procés d'etiquetatge i anotació del conjunt de dades COCO de Microsoft (Lin et al. 2014), que utilitzava imatge extretes de Flickr i anotades per *clickworkers* per entrenar els seus models de reconeixement d'objectes.

En adonar-se que no es reconeixia l'autoria de les imatges, l'artista Philipp Schmitt va organitzar l'exposició "Humans of AI"[15]. A més, va elaborar certificats d'agraïment per als 34 248 contribuents (Fig. 7) i els va incloure a la llista d'un nou conjunt de dades, el coCOCO (context d'Objectes Comuns en Context). A partir de YOLO (Redmon i Farhadi 2018), un sistema de detecció d'objectes que va entrenar amb COCO, Schmitt va crear el Declassifier (Fig. 8), un programa que superposa imatges del conjunt de dades sobre altres imatges. Finalment, va elaborar una presentació de diapositives que s'esvaïen entre totes les imatges de la col·lecció per poder mostrar una infinitat de fotografies quotidianes d'aficionats.

Un altre artista, Sebastian Schmieg, va reconèixer el treball dels *clickworkers* anònims que van delinear els objectes del conjunt COCO a través de segmentation.network[16], un lloc web que redibuixa i superposa de manera constant els 600 000 contorns dibuixats a mà (Fig. 9).

CONCLUSIONS

Si bé els grans conjunts de dades han permès desenvolupar noves tecnologies de la imatge com les xarxes generatives antagòniques i els algoritmes de detecció, la comunitat d'investigació es val de les llicències d'accés obert per aconseguir-ho. En la majoria de casos no es reconeix com correspondria les persones que van contribuir als conjunts de dades, o aquestes ni tan sols són conscients que s'utilitza el seu material.

Els artistes han respost a la creació d'aquests conjunts de dades a través d'anàlisis i d'exposicions tant en galeries com *online*. Les seves obres fan que els grans conjunts de dades siguin comprensibles i dignes d'atenció. En alguns casos han aconseguit atreure l'atenció dels mitjans de comunicació, i això ha fet que diverses institucions eliminin les seves col·leccions més qüestionables, o algunes parts d'aquestes.

A través de la investigació del conjunt de dades FFHQ, s'han trobat problemes comuns que la relacionen amb les altres obres ja mencionades. Alguns d'aquests problemes són l'ús de les dades humanes sense tenir el consentiment informat previ, com les imatges amb llicències CC de Flickr, i l'ús de *clickworkers* per refinar el conjunt. No obstant, el FFHQ no comptava amb alguns dels problemes que sí que tenien altres conjunts, ja que els autors van reconèixer l'autoria de les imatges de manera correcta, van crear una eina per trobar i eliminar imatges del conjunt i van publicar la pròpia col·lecció sota una llicència CC no comercial. Tot i així, les imatges de rostres retallades de manera automàtica creen certa sensació d'inquietud, en veure que les identitats humanes es poden reduir a aquests esdeveniments estadístics descontextualitzats. Aquest article mostra la persona que sí que existeix darrere d'aquest model d'intel·ligència artificial, però infringeix els drets personals de la mateixa manera que ho fa el conjunt de dades original. Encara que s'ha d'estudiar més aquest dilema ètic, fins ara es pot afirmar que l'art de conjunts de dades, tal com es demostra en aquest article, ajuda a ampliar el debat i a conscienciar el món sobre les pràctiques que es duen a terme en la investigació de l'aprenentatge automàtic[17].

BIOGRAFIA

Matthias Schäfer
University of Art & Design Linz

Matthias Schäfer és un artista que explora les cultures dels algoritmes. Des de Pitscher, crea representacions, instal·lacions i obres basades en llocs web amb les quals tracta el tema de les plataformes *online* centralitzades d'una manera desconstructiva i graciosa. Algunes de les seves obres s'han exposat al Transmediale, l'Ars Electronica Festival, el Xie Zilong Photography Museum, el Roehrs & Boetsch i en moltes altres exposicions i festivals. Actualment, Schäfer està inscrit en el màster en Interfaces Culturals de la University of Art & Design de Linz, on se centra en el camp de l'aprenentatge automàtic, l'automatització i el seu mecanisme central: la recollida de dades.

NOTES FINALS

1. Per a l'article anterior, "Augment progressiu de les xarxes generatives antagòniques per millorar la qualitat, estabilitat i variació" (Karras et al. 2017), els autors van augmentar la definició de les imatges del popular conjunt de dades CelebA (30 000 imatges) perquè complissin amb els seus requisits.
2. Vegeu https://github.com/NVlabs/ffhq-dataset
3. La inclusió de mesures facials i suposicions sobre edad i gènere generades per ordinador van ser molt criticades per les seves bases frenològiques. Segons un debat a GitHub (https://github.com/tensorflow/datasets/issues/299), el conjunt de dades no està disponible per part d'IBM, tot i que se segueix publicant.
4. Vegeu http://dlib.net/
5. Cita de l'arxiu "llegiu-me" del repositori de GitHub del conjunt de dades FFHQ (nota final 2). Obtingut el 25 de gener de 2021.
6. Vegeu https://nvlabs.github.io/ffhq-dataset/search/
7. Vegeu https://thispersondoesnotexist.com/
8. Vegeu https://this-person-does-exist.com/
9. Tan sols sis dies abans de publicar la meva pàgina, Vincent Woo va publicar un lloc web que seguia el mateix concepte, però que extreia les imatges directament des de Flickr i les retallava en el navegador. Vegeu http://thispersonexists.net/ i el tweet corresponent https://twitter.com/fulligin/status/1335030372187312128
10. Vegeu https://ahprojects.com/megapixels/ i https://exposing.ai/
11. Vegeu *MegaPixels*: Faces, https://ahprojects.com/megapixels-glassroom/
12. Vegeu *Fondazione Prada*, "*KATE CRAWFORD | TREVOR PAGLEN: TRAINING HUMANS*", http://www.fondazioneprada.org/project/training-humans/?lang=en
13. Vegeu *Trevor Paglen*, ImageNet Roulette, https://paglen.studio/2020/04/29/imagenet-roulette/
14. L'11 de març de 2021 l'equip va actualitzar https://image-net.org/. Van abordar les crítiques eliminant 2702 synsets i van experimentar desenfocant les cares per preservar la privacitat de les persones.
15. Vegeu *Philipp Schmitt*, Humans of AI, https://humans-of.ai/
16. Vegeu *Sebastian Schmieg*, Segmentation. Network http://segmentation.network/ Més informació a: http://sebastianschmieg.com/segmentation-network/
17. La publicació http://plottingd.at/a/ explora els conjunts de dades molt més a fons a partir d'interfícies artístiques y reflexions crítiques.

FIGURES

Fig. 1. Imatge retallada de 1.024 px × 1.024 px del conjunt de dades FFHQ, amb els punts de referència del rostre que provenen del mateix conjunt superposats per l'autor.
Fig. 2. Imatge "en estat natural" extreta de Flickr, amb els punts de referència del rostre del conjunt de dades superposades per l'autor. Llicència: CC-BY-2.0. Crossing Europe Filmfestival Linz.
Fig. 3. 69.960 rostres del conjunt de dades FFHQ compilades en una quadrículaper l'autor. A la dreta, una zona de la quadrícula ampliada. Llicència: CC-BY-NC. Els crèdits de les imatges originals estan disponibles a https://github.com/NVlabs/ffhq-dataset o a https://this-person-does-exist.com/credit.html.
Fig. 4. El rostre mitjà de Flickr, generat per l'autor a partir del càlcul de la mitjana de valors de píxel de les 70 000 imatges de rostres.
Fig. 5. Fotografia de l'exposició "Training Humans".
Fig. 6. Fotografia de l'exposició del conjunt de dades ImageNet i les seves etiquetes. Rostres difuminats per l'autor.
Fig. 7. Captura de pantalla d'un certificat d'agraïment de Philipp Schmitt.
Fig. 8. Captura de pantalla de Declassifier, del projecte *Humans of AI* de Philipp Schmitt.
Fig. 9. Captura de pantalla del lloc web segmentation.network, de Sebastian Schmieg.

REFERÈNCIES

Veure llistat complet de referències a la pàgina 221.

Matthias Schäfer

This Person Does Exist

Traducción al Castellano

PALABRAS CLAVE Visión por Computador, Creative Commons, Dataset Art, Minería de Datos, Aprendizaje Automático, Fotografía.

RESUMEN *This Person Does Exist* (Esta persona sí existe) es una aproximación artística para investigar de un modo aleatorizado un gran conjunto de datos de retratos fotográficos. El laboratorio de investigación de Nvidia creó el conjunto de datos a partir de una extracción y análisis de imágenes con licencia Creative Commons de la popular plataforma de alojamiento de imágenes Flickr. Dichas imágenes se usaron posteriormente para entrenar un modelo de aprendizaje automático que es capaz de crear nuevas imágenes estocásticas de rostros.
A diferencia de un popular sitio web que muestra las imágenes generadas por ordenador, en el presente artículo se muestran rostros aleatorios que provienen del conjunto de datos, así como sus metadatos correspondientes.
El presente artículo estudia los mecanismos extractivistas de las técnicas actuales de aprendizaje automático, que utilizan internet para completar y perfeccionar las bases de datos, y se centra en los enfoques artísticos que las exponen. Se presenta el argumento de que el arte de conjuntos de datos es un campo emergente que replantea los corpus científicos al situarlos en galerías y exponerlos como objetos que se han encontrado en la red.
Por último, se argumenta que dicha práctica artística es un modo legítimo de abrir un debate que llegue a un público más amplio, aunque los artistas que trabajen con datos humanos deban ser conscientes de los problemas éticos y de las responsabilidades que tienen en cuanto a la privacidad y el consentimiento.

CREAR ARTE A PARTIR DE CONJUNTOS DE DATOS HUMANOS

Según avanzan las técnicas de aprendizaje automático de la visión por computador, estas necesitan conjuntos de datos más amplios para su entrenamiento. Los investigadores llenaron este hueco a través de algoritmos que recogían y categorizaban imágenes de sitios como Flickr, donde podían valerse de las licencias Creative Commons (CC) que tenían las imágenes para copiar las fotografías de otras personas y crear imágenes derivadas a partir de estas. Tal como demuestran varias entrevistas a personas afectadas, como las realizadas por el New York Times (Hill y Krolik 2019) y NBC News (Solon 2019), el público en general no sabe que se producen dichas prácticas.
Los conjuntos de entrenamiento no se diseñaron para el consumo humano, sino que se crearon para llevar a cabo análisis estadísticos y modelar patrones entre puntos de datos individuales. Al examinar estas amplias colecciones, se pueden encontrar prejuicios ocultos que se transportan al modelo. También se puede ver una gran cantidad de trabajo individual, que va desde tomar las fotografías hasta etiquetarlas y categorizarlas.
En el presente artículo, se explica el enfoque que se ha seguido, que se basa en un conjunto de datos concreto, así como las obras de arte que han tratado problemas similares durante los últimos años.

INTRODUCCIÓN A LOS ROSTROS DE FLICKR

El laboratorio de investigación de Nvidia, una de las mayores empresas del sector de la visión por computador, publicó en 2018 un artículo que presentaba una arquitectura de aprendizaje automático llamada StyleGAN (Karras, Laine y Aila 2019). Nvidia mejoró las redes generativas antagónicas (Goodfellow et al. 2014) de tal modo que fue posible crear imágenes controlables y sintéticas de alta resolución. Dicho en términos más simples, los investigadores fueron capaces de recopilar una gran cantidad de imágenes a través de un modelo que, al mismo tiempo, producía imágenes similares. En este caso, pudieron generar imágenes verosímiles de rostros de personas. Debido a que los conjuntos de datos de rostros anteriores tenían muy poca resolución,[1] se creó un corpus nuevo a partir de una extracción de fotografías de la plataforma de alojamiento de imágenes Flickr. Para ello, se utilizaron fotografías que tuvieran licencias Creative Commons, de dominio público o del gobierno de los Estados Unidos. A este nuevo conjunto de datos se le llamó Flickr-Faces-HQ (FFHQ)[2].
Desde 2004, Flickr apoya el uso de licencias CC, y su acceso abierto hace que sea legal que los investigadores extraigan y analicen su material. De hecho, el conjunto de datos de medios visuales de acceso abierto más grande que existe, que cuenta con 100 millones de objetos, se comparte con el laboratorio de investigación de Yahoo, la empresa matriz de Flickr (Thomee et al. 2016).
En respuesta al escrutinio sobre los fallos del algoritmo de detección facial de IBM, que afectó a las mujeres negras en particular (Buolamwini y Gebru 2018), su departamento de investigación creó, a partir del conjunto de datos de Yahoo, una base de datos pública que contenía un millón de rostros. La llamaron "Diversity in Faces" (la diversidad en los rostros, DiF por sus siglas en inglés) (Merler et al. 2019) y le añadieron unos metadatos cuestionables, como una estimación de la edad, el género y las medidas craneofaciales, con la intención de que la detección facial fuera más justa[3].
En comparación, FFHQ, que se utilizó para entrenar el modelo generativo mencionado anteriormente, contiene tan solo 70 000 imágenes (menos de una décima parte). Dicho conjunto se publicó bajo una licencia CC-BY-NC-SA y las instrucciones para su uso y descarga son muy claras, lo que lo convierte en algo manejable en términos de tamaño, licencia y trabajo necesario para descubrir sus características subyacentes.
Se utilizó la biblioteca de acceso abierto dlib[4] para alinear y recortar las imágenes extraídas de forma automática alrededor del rostro y conseguir que tuvieran un formato de imagen cuadrado de un tamaño de 1024^2 px. Dicha biblioteca proporciona un reconocimiento facial rápido e identifica 68 puntos alrededor del rostro que delinean la barbilla, las cejas, los ojos, la nariz y la boca. Por último, los trabajadores de Amazon Mechanical Turk comprobaron el conjunto de datos resultante para "eliminar las posibles estatuas, cuadros y fotografías de otras fotografías"[5].
A diferencia de otros conjuntos de datos que se basan en imágenes extraídas de sitios web, los investigadores de Nvidia proporcionan una herramienta[6] que permite comprobar si una imagen determinada forma parte de la colección y eliminar dicha fotografía del conjunto de entrenamiento FFHQ. Como sugiere un artículo de investigación de NBC News (Solon 2019), la mayoría de fotógrafos de Flickr no sabe que forma parte del conjunto de datos DiF de IBM, y existen opiniones diversas sobre que se utilicen sus fotografías para la investigación del reconocimiento facial. Se puede asumir que esto también ocurre con el corpus FFHQ, pues los investigadores no obtuvieron ningún tipo de consentimiento antes de crear la colección.

ESTA PERSONA SÍ EXISTE

Después de que Nvidia publicara su artículo sobre la StyleGAN, el desarrollador de *software* Phillip Wang publicó un sitio web[7] en el que presenta los rostros generados de personas que no existen. Dicho sitio se hizo popular en muy poco tiempo y generó inquietudes sobre las capacidades y el impacto potencial de los sistemas de inteligencia artificial cuando estos se usan para producir imágenes sintéticas.
Como contrapunto al creador de imágenes de inteligencia artificial, quise mostrar las personas que se utilizaron para entrenar a dicho sistema generativo. La primera investigación sobre el corpus FFHQ se llevó a cabo en la primavera de 2020, y se encontraron escollos técnicos como la descarga y la diseminación de imágenes del conjunto de datos. Se movieron las imágenes de rostros, ya recortadas y alineadas, a mi servidor y se creó un sitio web[8] que muestra los rostros de personas que sí existen[9].
El hecho de ver las imágenes individuales facilita que se cree una conexión interpersonal con la persona desconocida y proporciona una idea de las imágenes que se utilizaron para entrenar al modelo generativo. Dado que los usuarios de Flickr suelen ser fotógrafos, tanto aficionados como profesionales, se pueden encontrar retratos de niños y familias, ponentes en conferencias o personas de vacaciones. Los autores afirman que el FFHQ incluye más variedad de imágenes que otros conjuntos de retratos faciales en términos de "edad, etnicidad y fondo de imágenes, y muestra una mayor presencia de accesorios como gafas, gafas de sol y sombreros, entre otros" (Karras, Laine y Aila 2019). Asimismo, si bien admiten que existen prejuicios que se heredaron de la plataforma de Flickr, no especifican cuáles son. Un ejemplo (Fig. 1) muestra una imagen recortada del conjunto de datos con los puntos de referencia del rostro. Dichos puntos de referencia forman parte de los metadatos y se utilizaron para alinear las imágenes y poder entrenar al sistema. La imagen original correspondiente (en estado natural) añade un contexto mayor y el título clarifica que se tomó en un *brunch* de un festival de cine (Fig. 2).
Para entender mejor la escala del conjunto de datos, se compilaron las imágenes de los rostros en una cuadrícula y se redujo el tamaño de cada imagen a 16 px × 16 px (Fig. 3). Este montaje tan sencillo permite hacerse una idea de la gran cantidad de datos de imagen normalizados que contiene el conjunto.
Otra técnica para encontrar sesgos de forma visual en los conjuntos de datos es hacer una media de sus valores de píxel, pues esto permite ver su tendencia general al suprimir los casos atípicos. La imagen compuesta resultante, llamada The Flickr Face (el rostro de Flickr) (Fig. 4), revela una tendencia hacia la sonrisa y hacia las personas de piel clara en el conjunto. Este hecho se correlaciona con la experiencia subjetiva previa, pero se necesita un análisis más exhaustivo para poder medir los prejuicios de dicho conjunto de datos.
En la siguiente sección se presentan otros proyectos artísticos que utilizan técnicas similares para que se puedan experimentar los grandes conjuntos de datos y criticar la recolección de datos humanos sin consentimiento previo.

LOS PROYECTOS Y EXPOSICIONES *MEGAPIXELS*, "TRAINING HUMANS" Y "HUMANS OF AI"

Desde 2017, Adam Harvey y Jules LaPlace han trabajado en el proyecto de arte e investigación *MegaPixels*[10], donde analizan el reconocimiento facial y los conjuntos de datos de detección. Como resultado de su investigación y de la cobertura mediática que esta ha recibido, ya se han retirado y eliminado varios conjuntos de datos, como el MegaFace (Kemelmacher-Shlizerman et al. 2016), que contenía unos 4,7 millones de rostros extraídos de Flickr (Thomee et al. 2016). Algunas empresas privadas utilizaban dicho corpus con fines comerciales, y no se reconocía del todo la autoría de las imágenes. Además, entre las imágenes que contenía el conjunto había algunas de personas de Illinois, donde está prohibido recoger, tomar y comprar información biométrica (Hill y Krolik 2019).
Aunque Harvey y LaPlace han utilizado métodos tradicionales de visualización de datos y de publicación académica, también idearon modos de hacer que las imágenes extraídas para usarse en aprendizaje automático sean algo mucho más personal.
En 2017, Adam Harvey organizó[11] una instalación interactiva para la exposición "Glass Room" en la que se comparaban los puntos de referencia faciales de los visitantes con las imágenes del conjunto de datos MegaFace y se mostraban las personas que más se parecían a ellos.
En 2019, el artista Trevor Paglen y la investigadora Kate Crawford colaboraron en una exposición sobre las bases de datos de imágenes humanas[12] (Fig. 5). Una de las obras principales era una amplia colección de imágenes humanas procedentes del conjunto de datos ImageNet (Fig. 6), que fue creado por Fei Fei Li, profesor de Inteligencia Artificial de la Stanford University (Deng et al. 2009). Dicho conjunto se creó para abordar las tareas de reconocimiento de objetos y está formada por 14 millones de imágenes etiquetadas y organizadas por trabajadores de Amazon Mechanical Turk. Algunas de las imágenes estaban categorizadas como "mala persona, mujer de compañía, drogadicto, amanerado, presidiario", entre otras (Crawford y Paglen 2019a). Los artistas utilizaron dichas etiquetas absurdas, racistas y misóginas para entrenar a "ImageNet Roulette"[13], un algoritmo de reconocimiento que estaba disponible en internet y en una instalación interactiva. Como resultado de la atención mediática que provocó, se eliminaron 600 000 imágenes del conjunto de datos y tanto el sitio web como el conjunto en sí no han estado disponibles desde entonces por motivos de mantenimiento (Yang et al. 2019)[14].
Sin embargo, la exposición "Training Humans" también conlleva sus propios problemas, tal como comenta Michael J. Lyons (2020) en su crítica a la obra de Crawford y Paglen. Exponer algunas de las imágenes va en contra de sus condiciones de uso, pues solo permiten que las imágenes se utilicen para la investigación científica, no para fines comerciales. Lyons critica el doble rasero de Crawford y Paglen, quienes utilizan datos humanos sin haber obtenido un consentimiento informado previo. A pesar de que conocen muy bien el dilema ético, los artistas se posicionan en contra de la privacidad "para poder hablar de verdad sobre qué ocurre en esos sistemas [de inteligencia artificial]" (Crawford y Paglen 2019b). Si bien los visitantes pudieron eliminar sus imágenes de la exposición, en el caso de los voluntarios fotografiados para el conjunto de datos JAFFE expuesto por Lyons, Kamachi y Gyoba (1998), que nunca estuvo destinado para utilizarse como "conjunto de entrenamiento". Fue solo después de la finalización de la exposición cuando se enteraron de que las imágenes destinadas al uso científico se habían distribuido sin ninguna limitación por internet.
Otros dos artistas se centraron en el proceso de etiquetado y anotación del conjunto de datos COCO de Microsoft (Lin et al. 2014), que utilizaba imágenes extraídas de Flickr y anotadas por *clickworkers* para entrenar a sus modelos de reconocimiento de objetos.
Al percatarse de que no se reconocía la autoría de las imágenes, el artista Philipp Schmitt organizó la exposición "Humans of AI"[15]. Además, elaboró certificados de agradecimiento para el total de 34 248 contribuyentes (Fig. 7) y los incluyó en la lista de un nuevo conjunto de datos, el coCOCO (contexto de Objetos Comunes en Contexto). A partir de YOLO (Redmon and Farhadi 2018), un sistema de detección de objetos que entrenó con COCO, Schmitt creó el Declassifier (Fig. 8), un programa que superpone imágenes del conjunto de datos sobre otras imágenes. Por último, elaboró una presentación de diapositivas que se desvanecían entre todas las imágenes de la colección para poder mostrar un sinfín de fotografías cotidianas de aficionados.
Otro artista, Sebastian Schmieg, reconoció el trabajo de los *clickworkers* anónimos que delinearon los objetos del conjunto COCO a través de segmentation.network[16], un sitio web que redibuja y superpone de forma constante los 600 000 contornos dibujados a mano (Fig. 9).

CONCLUSIONES

Si bien los grandes conjuntos de datos han permitido desarrollar nuevas tecnologías de la imagen como las redes generativas antagónicas y los algoritmos de detección, la comunidad de investigación se vale de las licencias de acceso abierto para conseguirlo. En la mayoría de casos, no se reconoce como corresponde a las personas que contribuyeron a los conjuntos de datos, o estas ni siquiera son conscientes de que se utiliza su material.
Los artistas han respondido a la creación de dichos conjuntos de datos a través de análisis y de exposiciones tanto en galerías como *online*. Sus obras hacen que los grandes conjuntos de datos sean comprensibles y dignos de atención.. En algunos casos, han conseguido atraer la atención de los medios de comunicación, lo que ha provocado que varias instituciones eliminen sus colecciones más cuestionables, o algunas partes de ellas.
A través de la investigación del conjunto de datos FFHQ, se han encontrado problemas comunes que lo relacionan con las otras obras ya mencionadas. Algunos de estos problemas son el uso de los datos humanos sin contar con un consentimiento informado previo, como las imágenes con licencias CC de Flickr, y el uso de *clickworkers* para refinar el conjunto.
Sin embargo, el FFHQ no contaba con algunos de los problemas que *sí tenían* otros conjuntos, pues los autores reconocieron la autoría de las imágenes de forma correcta, crearon una herramienta para encontrar y eliminar imágenes del conjunto y publicaron la propia colección bajo una licencia CC no comercial.
Aun así, las imágenes de rostros recortadas de forma automática crean una cierta sensación de inquietud, al ver que las identidades humanas se pueden reducir a estos sucesos estadísticos descontextualizados. El presente artículo muestra la persona que sí existe detrás de ese modelo de inteligencia artificial, pero infringe los derechos personales del mismo modo que lo hace el conjunto de datos original. Aunque debe estudiarse más ese dilema ético, hasta el momento se puede afirmar que el arte de conjuntos de datos, tal como se demuestra en el presente artículo, ayuda a ampliar el debate y a concienciar al mundo sobre las prácticas que se llevan a cabo en la investigación del aprendizaje automático[17].

BIOGRAFÍA

Matthias Schäfer
University of Art & Design Linz

Matthias Schäfer es un artista que explora las culturas de los algoritmos. Desde Pitscher, crea representaciones, instalaciones y obras basadas en sitios web con los que trata el tema de las plataformas *online* centralizadas de un modo deconstructivo y gracioso. Algunas de sus obras se han expuesto en el transmediale, el Ars Electronica Festival, el Xie Zilong photography museum, el Roehrs & Boetsch y en muchas otras exposiciones y festivales.
En la actualidad, Schäfer está inscrito en el máster en Interfaces Culturales de la University of Art & Design de Linz, donde se centra en el campo del aprendizaje automático, la automatización y su mecanismo central: la recolección de datos.

NOTAS FINALES

1. Para el artículo anterior, "Aumento progresivo de las redes generativas antagónicas para mejorar la calidad, estabilidad y variación" (Karras et al. 2017), los autores aumentaron la definición de las imágenes del popular conjunto de datos CelebA (30 000 imágenes) para que cumplieran con sus requisitos.
2. Véase https://github.com/NVlabs/ffhq-dataset
3. La inclusión de medidas faciales y suposiciones sobre edad y género generadas por ordenador fueron muy criticadas por sus bases frenológicas. Según un debate en GitHub (https://github.com/tensorflow/datasets/issues/299), el conjunto de datos no está disponible por parte de IBM, aunque se sigue publicitando.
4. Véase http://dlib.net/
5. Cita del archivo "léeme" del repositorio de GitHub del conjunto de datos FFHQ (nota final 2). Obtenido el 25 de enero de 2021.
6. Véase https://nvlabs.github.io/ffhq-dataset/search/
7. Véase https://thispersondoesnotexist.com/
8. Véase https://this-person-does-exist.com/
9. Tan solo seis días antes de publicar mi página web, Vicent Woo publicó otra con el mismo concepto, pero que obtiene las imágenes directamente de Flickr y las recorta en el navegador. Véase http://thispersonexists.net/ y el correspondiente tweet https://twitter.com/fulligin/status/1335030372187312128
10. Véase https://ahprojects.com/megapixels/ y https://exposing.ai/
11. Véase *MegaPixels*: Faces, https://ahprojects.com/megapixels-glassroom/
12. Véase *Fondazione Prada, "KATE CRAWFORD | TREVOR PAGLEN: TRAINING HUMANS"*, http://www.fondazioneprada.org/project/training-humans/?lang=en
13. Véase *Trevor Paglen*, ImageNet Roulette, https://paglen.studio/2020/04/29/imagenet-roulette/
14. El 11 de marzo de 2021 el equipo actualizó https://image-net.org/. Abordaron las críticas eliminando 2702 synsets y experimentaron desenfocando los rostros para preservar la privacidad de las personas.
15. Véase *Philipp Schmitt*, Humans of AI, https://humans-of.ai/
16. Véase *Sebastian Schmieg*, Segmentation.Network http://segmentation.network/ más información en: http://sebastianschmieg.com/segmentation-network/
17. La publicación http://plottingd.at/a/ explora los conjuntos de datos mucho más a fondo a través de interfícies artísticas y reflexiones críticas.

FIGURAS

Fig. 1. Imagen recortada de 1024 px × 1024 px del conjunto de datos FFHQ, con los puntos de referencia del rostro que provienen del mismo conjunto superpuestos por el autor.
Fig. 2. Imagen "en estado natural" extraída de Flickr, con los puntos de referencia del rostro del conjunto de datos superpuestos por el autor. Licencia: CC-BY-2.0. Crossing Europe Filmfestival Linz.
Fig. 3. 69 960 rostros del conjunto de datos FFHQ compiladas en una cuadrícula por el autor. A la derecha, una zona de la cuadrícula ampliada. Licencia: CC-BY-NC. Los créditos de las imágenes originales están disponibles en https://github.com/NVlabs/ffhq-dataset o en https://this-person-does-exist.com/credit.html.
Fig. 4. El rostro medio de Flickr, generado por el autor a partir del cálculo de la media de valores de píxel de las 70 000 imágenes de rostros.
Fig. 5. Fotografía de la exposición "Training Humans".
Fig. 6. Fotografía de la exposición del conjunto de datos ImageNet y sus etiquetas. Rostros difuminados por el autor.
Fig. 7. Captura de pantalla de un certificado de agradecimiento de Philipp Schmitt.
Fig. 8. Captura de pantalla de Declassifier, del proyecto *Humans of AI* de Philipp Schmitt.
Fig. 9. Captura de pantalla del sitio web segmentation.network de Sebastian Schmieg.

REFERENCIAS

Ver listado completo de referencias en la página 221.