

DE FONTS D'ARXIU A CONJUNT DE DADES. EXPERIÈNCIES CREATIVES AMB EINES D'INTEL·LIGÈNCIA ARTIFICIAL I ELS FONTS DEL CRDI

Roc Albalat, Pau Artigas, Marc Padró, Marcel Pié i Daniel Pitarch
Estampa

RESUMEN

Las herramientas de inteligencia artificial (IA) actuales se basan en el aprendizaje automático (machine learning) y en las llamadas redes neuronales de aprendizaje profundo. Éstas requieren de gran cantidad de datos para realizar su proceso de entrenamiento. En este sentido, los fondos de archivo constituyen un recurso interesante para ser explorado y trabajado con estas herramientas, convirtiéndose así en conjuntos de datos (datasets). La comunicación explicará una experiencia concreta de trabajo creativo con los fondos del CRDI, que ha implicado el uso y entrenamiento de redes generativas de imágenes (en concreto, redes GAN). Así como mostrará otros trabajos artísticos del colectivo Estampa con herramientas de IA, que exploran usos alternativos de esta tecnología.

SUMMARY

Today's artificial intelligence (AI) tools are based on machine learning and so-called deep learning neural networks. These require many data to complete its training process. In this sense, archives collections are an interesting resource to be explored and worked with these tools, thus becoming datasets. The communication will explain a specific experience of creative work with CRDI funds, which has involved the use and training of image generating networks (specifically, GAN networks). It will also show other works of art by the Estampa collective with AI tools, which explore alternative uses of this technology.

Els darrers anys han vist un important creixement del camp de la intel·ligència artificial (IA), tant a un nivell de desenvolupament tecnològic com de discurs públic sobre les seves possibilitats. A nivell tècnic es pot dir que aquest increment és conseqüència d'explorar la via de l'aprenentatge automàtic (*machine learning* en anglès) i, dins d'aquest, la de la tecnologia de les xarxes neuronals d'aprenentatge profund. És interessant remarcar que el que és nou no és tant la tecnologia en si, sinó el fet que aquestes xarxes donen bons resultats quan treballen amb grans quantitats de dades. Si la IA ha crescut tant en els darrers anys és per la combinació d'aquesta tècnica amb la possibilitat d'acumular dades de manera massiva (principalment a través d'internet i, per tant, en part gràcies als nostres dispositius personals), conjuntament amb la potència de hardware requerida per poder-les processar. Aquest desenvolupament tècnic s'ha vist acompanyat per un discurs públic al voltant de la IA, sobretot a partir de notícies sobre les seves possibilitats i novetats, per exemple aplicada al camp de la creació (notícies sobre IA que escriuen, pinten, componen música, etc.). Aquest discurs també constitueix la IA, ja que crea un horitzó d'expectatives, legitima els seus usos i li atorga unes capacitats.

Explorar aquestes eines i els discursos que les envolten ha estat una de les línies de treball del col·lectiu Estampa els darrers cinc anys. Vam començar a treballar-hi l'any 2017 amb un projecte que vam titular programàticament *El mal alumne. Pedagogia crítica per a intel·ligències artificials*. Aquest títol agafava una de les metàfores d'aquesta tecnologia –la de l'aprenentatge– per girar-la i reivindicar-ne un ús al marge del marc normatiu. El projecte també suposava treballar i apropiat-se d'aquestes tecnologies, encara que fos fora dels estàndards quantitius industrials. I posava l'accent en desmistificar aquestes eines; és a dir, en entendre què fan, com ho fan i com en parlem i utilitzar-les de primera mà.

Elements bàsics de les xarxes neuronals d'aprenentatge profund

Un element important per evitar mistificar aquestes eines és utilitzar un vocabulari concret i no genèric i intentar entendre els elements que en formen part. En general intentem referir-nos no a IA sinó a la tec-

nologia concreta que s'empra avui, que és l'aprenentatge automàtic i les xarxes neuronals d'aprenentatge profund. Si bé aquests noms també estan plens de metàfores –neuronal, aprenentatge, etc.– la seva concreció i aparent complicació esquiva les del terme intel·ligència artificial, que, en certs contextos, sembla legitimar de per si els seus resultats.

Una definició simple de què és «aprenentatge automàtic» la dona l'artista Kyle McDonald que diu: «aprenentatge automàtic és automatitzar tasques per mitjà d'exemples (dades d'entrenament) i no d'instruccions escrites (codi)».¹ D'aquesta definició, volgudament simple, ja és interessant que comença afirmant l'objectiu d'aquestes eines que és «automatitzar»; un verb que no ens porta al mateix camp semàntic que «intel·ligència» o «creació» (encara que tingui les seves interseccions). Una xarxa neuronal d'aprenentatge profund automatitzarà una tasca, és a dir, donarà una sortida quan li donem una entrada; això pot ser crear una imatge, etiquetar el contingut d'una fotografia, escriure un text, etc. Com automatitza aquesta eina la tasca no es construeix per mitjà d'instruccions concretes (per exemple, si volguéssim que identifiqui ulls en les imatges podríem intentar pensar instruccions concretes com podrien ser: buscar píxels contigus amb molt contrast per identificar línies en les imatges, verificar si són línies corbades, mirar si hi ha una certa simetria entre aquesta i una altra que seria l'altre ull, etc.), sinó proporcionant molts exemples d'imatges on està identificat què és un ull (s'indica quin part de la imatge és un ull, per mitjà del que s'anomena *bounding box*). Aquests exemples són les dades d'entrenament i es podria dir que una xarxa neuronal és una eina que és capaç d'autoconfigurar-se a partir d'aquestes (és a dir, el que s'ha programat és aquesta eina que s'autoconfigura i no com configurar cada tasca que es vol automatitzar).

Amb l'explicació d'aquesta definició surten tres paraules que és important tenir clares: xarxa neuronal, conjunt de dades (*dataset* en anglès) i entrenament. La xarxa neuronal és l'eina capaç d'autoconfigurar-se; existeixen diferents tipus de xarxes en funció d'amb quin tipus de dades volem que treballi (amb imatges, amb text...) o de quin procés volem automatitzar (són diferents les de visió artificial –etiquetar el contingut de les imatges– que les de generació d'imatges,

1 - Aquesta definició la va utilitzar en la conferència "Limits of Learning" impartida a Sonar+D 2019

per exemple). El conjunt de dades són els exemples que li proporcionem i amb els que s'autoconfigurarà. Aquest element és, per tant, molt important en aquestes eines. Pensem en una eina de visió artificial, que li proporcionem com exemples –tant a nivell de les imatges com de l'etiqueta verbal– serà el que després identificarà. Com es construeixen els conjunts de dades –i això implica també d'on es treuen els exemples i quina força de treball ho fa– és bàsic i una decisió fonamental al treballar amb aquesta tecnologia. Per últim, el procés d'autoconfiguració de la xarxa amb el conjunt de dades és el que s'anomena entrenament. El resultat d'aquest és la mateixa xarxa inicial, però modificada per realitzar aquella tasca concreta que vulguem que realitzi (que generi un tipus d'imatge, que identifiqui certes coses, etc.).

Conjunts de dades i fons d'arxiu

Com s'acaba de veure, en aquest camp els conjunts de dades són essencials, ja que suposen la base per a l'entrenament de les xarxes. Hi ha un gran nombre de qüestions interessants al voltant d'aquests *datasets*. D'una banda, se'n pot fer una història; per exemple, a la dècada dels 80 i 90 del segle passat es van construir un conjunt de dades de dígitos escrits a mà per a eines de visió per computadora que es poguessin aplicar a correus o a la banca.² D'una altra, els processos per aconseguir grans quantitats de dades poden ser discutibles en molts casos. Per exemple, la recopilació d'imatges per internet pot implicar des d'imatges penjades en xarxes socials o repositoris fotogràfics, fins a l'ús de webcams obertes i això pot entrar en conflicte amb les llicències d'ús de les imatges (que no accepten ni prohibeixen aquesta possibilitat perquè no existia) o pot ser discutible per qüestions ètiques. També com viatgen i s'utilitzen els conjunts de dades mereix una atenció detallada. Els artistes Adam Harvey i Ju-

les LaPlace han realitzat un excel·lent projecte sobre aquestes dues qüestions –*Exposing AI* (<https://exposing.ai/>)– investigant la construcció i ús de *datasets* concrets. En la creació dels conjunts de dades d'entrenament és quan apareixen també biaixos i altres elements que ens porten als límits d'aquestes eines. Segurament un dels casos més cèlebres és el del *dataset* ImageNet que és l'estàndard amb el que es testen moltes eines de visió per computadora. Aquest conjunt de dades està format per paraules i exemples visuals d'aquestes. En la seva creació van decidir partir d'una taxonomia de la llengua anglesa preexistent anomenada WordNet. No va haver-hi cap filtratge d'aquesta i això va portar a que totes les paraules associades amb persones s'incloguessin en aquest conjunt de dades i es busquessin imatges d'exemple. Això incloïa nombroses paraules que no poden tenir una definició visual (per exemple, «mala persona») o que comportaven, en la seva associació entre paraules i imatges, plantejaments racistes o misògins. El cas va passar a la discussió pública en part gràcies al projecte *ImageNet Roulette* de l'artista Trevor Paglen i la investigadora Kate Crawford.³ Un altre aspecte a estudiar és com es construeixen aquests *datasets* des del punt de vista més material i quina força de treball impliquen; per exemple, ImageNet havia realitzat gran part del seu etiquetat a través d'una plataforma com Amazon Mechanical Turk que ofereix feines concretes i breus en línia, buscant minimitzar-ne al màxim els costos. Aquesta plataforma es publicita explícitament com important per a processos d'aprenentatge automàtic⁴ i el seu nom fa referència a un conegut cas en la història dels autòmats (el jugador d'escacs conegut com «el turc», del segle XVIII)⁵, en el qual el suposat autòmat es movia en realitat per un humà que s'hi amagava dins. Per tant, el seu nom implica, irònicament o no, aquest fet de la força de treball humana que s'amaga rere la tecnologia. En el món glo-

2 - LeCun, Yann, et al. "Backpropagation applied to handwritten zip code recognition." *Neural computation* 1.4 (1989): 541-551

3 - Veure el seu article «Excavating AI. The Politics of Images in Machine Learning Training Sets» a <https://excavating.ai/>

4 - Així ho descriuen a la web: «MTurk can be a great way to minimize the costs and time required for each stage of ML development. It is easy to collect and annotate the massive amounts of data required for training machine learning (ML) models with MTurk. Building an efficient machine learning model also requires continuous iterations and corrections. Another usage of MTurk for ML development is human-in-the-loop (HITL), where human feedback is used to help validate and retrain your model. An example is drawing bounding boxes to build high-quality datasets for computer vision models, where the task might be too ambiguous for a purely mechanical solution and too vast for even a large team of human experts.» <https://www.mturk.com/>

5 - Sobre aquest i altres casos d'autòmats veure el llibre de Marta Peirano i Sonia Bueno: *El rival de Prometeo* (Impedimenta editorial, 2018)

balitzat actual això participa de dinàmiques econòmiques i polítiques de deslocalització i precarització de la força de treball.

Del que s'ha explicat abans i d'aquests exemples que ara s'esmenten es veu que el conjunt de dades és un element important i complex d'aquestes tecnologies i, per extensió, del nostre món contemporani. Crear o trobar *datasets*, quins es converteixen en estàndard, qui els ha creat, qui pot crear-ne de massius, qui els utilitza o podria utilitzar-los, etc. són qüestions travessades per elements econòmics, polítics, socials i culturals. Els conjunts de dades es poden veure com un element proper als fons d'arxiu, però no són elements equivalents. Els conjunts de dades són una forma nova amb les seves pròpies característiques. Per exemple, construir un *dataset* implica pensar en termes d'homogeneïtat i heterogeneïtat de les dades; havent d'avaluar tots dos aspectes per obtenir resultats satisfactoris. Alhora la importància dels conjunts de dades forma part d'una preocupació genèrica per treballar amb grans quantitats de dades (que és també la que porta al desenvolupament de les tecnologies d'aprenentatge automàtic) i que es mou en el que podríem anomenar «ideologia de les dades massives» i la seva necessitat de monitorar i conservar tot. Tot i que això també es pot desenvolupar en moltes direccions, ho diem ara principalment per destacar-ho com una altra connexió parcial amb els arxius i els seus fons.

En els nostres projectes hem treballat amb xarxes neuronals d'aprenentatge profund i fons d'arxiu d'imatges de dues maneres: d'una banda, com a conjunts de dades d'entrenament i de l'altre com a dades per a ser visualitzades. El primer cas l'hem aplicat a xarxes generatives d'imatges, és a dir, a xarxes que automatitzen la producció d'imatges similars a les que els hi ensenyem com exemples; en el segon, treballem amb eines de visió per computadora que poden servir, per exemple, per ordenar un conjunt d'imatges a partir de la seva similitud gràfica. En la resta de l'article parlarem d'uns tests realitzats amb fons d'arxiu del CRDI i després d'altres casos que hem formalitzat en forma d'instal·lació.

Tests en curs amb fons del CRDI

Dels fons del CRDI hem treballat amb un fons de retrats fotogràfics d'estudi i amb una part del fons del fotògraf Martí Massafont. Amb el primer fons hem fet diverses proves (principalment orientades a la generació d'imatges) i amb el segon només alguns tests.

Del nombrós fons de retrats fotogràfics hem treballat amb una col·lecció acotada cronològicament entre 1900 i 1927, provinent de diversos estudis de fotografia de Girona (Fotografia Unal, Foto Lux i Josep Jou) i que resulta en un total de 3591 fotografies. Es va seleccionar aquest fons com a punt de partida perquè suposa un tipus d'imatge alhora heterogènia i homogènia. És a dir, hi ha una sèrie de convencions del retrat (tipus de pla, il·luminació, etc.) i una varietat dins aquesta tipologia de fotografia que són les persones retratades.⁶

Aquestes imatges van constituir el conjunt de dades d'entrenament d'una xarxa generativa (en concret la xarxa StyleGAN 2); és a dir, es buscava tenir una eina que generés nous retrats del mateix estil que aquests. Els fons d'arxiu es va estructurar en dos conjunts de dades diferents –i aquí trobem una altra diferència entre totes dues formes, un fons d'arxiu pot donar lloc a diferents *datasets*, no tenen perquè ser equivalents numèricament. El primer conjunt de dades era la fotografia sencera i el segon conjunt de dades era només els rostres de les fotografies reenquadrats en un primer pla (aquest procés es va automatitzar per mitjà d'instruccions de codi que identificaven les cares –amb eines de visió per computadora–, sumaven un marge d'un número de píxels concrets al voltant i retallaven la imatge). El primer conjunt de dades té el mateix nombre d'imatges que les fotos originals (de fet, algunes menys perquè s'han utilitzat només les verticals, que són 3356), però en el segon s'arriben a obtenir més imatges –3807 en total–, ja que diverses fotografies retraten a més d'una persona alhora.

L'entrenament de la xarxa generativa amb les imatges originals va donar com a resultat unes imatges majoritàriament no realistes, però que poden tenir un interès gràfic i també de veure certes convencions de les imatges originals. No desenvoluparem ara aquesta qüestió, sobre la que tornarem a partir d'un altre projecte després.

6 - En el nostre projecte *Espais latents. Imaginacions màquines* vam fer diversos experiments amb com es podia definir un *dataset* d'entrenament de xarxa GAN, tenint en compte unitats conceptuals a més de gràfiques: <https://tallerestampa.com/estampa/espais-latents/>

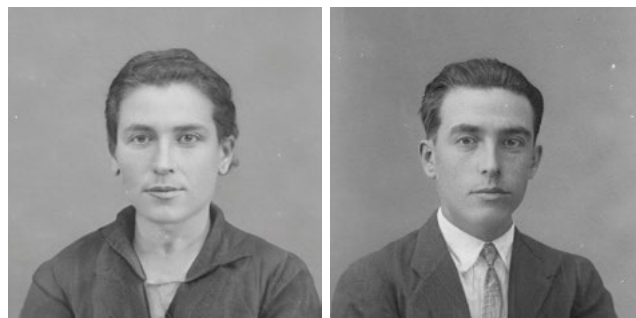


Imatge 1. Resultat de l'entrenament amb les fotografies senceres.

Per contra el resultat de l'entrenament amb els rostres reenquadrats sí que va obtenir resultats aparentment fotorealistes. En aquest cas el conjunt de dades d'entrenament és molt més homogeni, ja que els elements del decorat, etc. s'obvien per concentrar-se en el rostre.

Els resultats de les xarxes generatives poden ser en format d'imatges fixes o de vídeo. En forma de vídeo el que s'obté és una mena de transformació contínua entre les imatges que la xarxa sap generar. És interessant entendre, més o menys, com es fa això, perquè permet comprendre millor les xarxes generatives. Una xarxa entrenada treballa amb un sistema de coordenades multidimensional; en concret la xarxa StyleGAN 2 treballa amb 512 dimensions.

Això vol dir que cada imatge que genera és una posició en aquest sistema de coordenades (un valor per cada dimensió). El sistema de coordenades complet s'anomena «espai latent»; és a dir, s'entenen totes les imatges que pot generar aquesta xarxa com una mena d'espai a transitar o descobrir. Els resultats en vídeo de les xarxes generatives consisteixen en explorar aquest espai multidimensional generant una imatge i alguna de les seves contigües successivament (és a dir, qualsevol que impliqui variar poc algun o diversos d'aquests 512 valors). D'aquesta manera s'obté un recorregut per aquest espai latent que, gràficament, es tradueix en un *morphing* continu.



Imatge 2. Resultats del segon entrenament, només amb els rostres.

Part de l'interès de treballar amb aquestes eines és explorar aquest espai latent, buscant i trobant imatges que considerem interessants. L'interès pot ser molt divers. D'una banda, és evident que hi ha una atracció en la creació d'aquestes imatges fotorealistes que provenen d'originals fotogràfics i que no en són una còpia sinó una mena de versió (segurament el cas més conegut d'això és la pàgina web <https://www.thispersondoesnotexist.com>). De l'altra, qualsevol entrenament generatiu suposa una mena d'anàlisi del conjunt de dades original que identifica motius importants. En aquest cas elements d'estil de l'època com els pentinats o la vestimenta es fan particularment evidents –sobretot en els recorreguts en vídeo que en mostren exemples consecutius. Però també es fan evidents aspectes formals de la pròpia fotografia com, per exemple, els tipus d'il·luminació emprats en aquests retrats –a les exploracions en vídeo és particularment maco veure com la llum s'anima i canvia de posició. També es fan evidents motius temàtics, com, per exemple, la importància de les fotos de primera comunitat i els seus elements característics.



Imatge 3. Resultats del segon entrenament, només amb els rostres.

Si és significatiu què apareix de manera més o menys fotorealista a les fotografies, també ho pot ser què no sap fer amb la mateixa qualitat. Per exemple, les ulleres són un element que apareixen poques vegades i sense gaire consistència o no és capaç de generar dents (cosa que només intenta amb alguns rostres femenins):



Imatge 4. Resultats del segon entrenament, només amb els rostres

Abans s'ha parlat dels resultats en vídeo que són *morphings* que s'obtenen generant imatges consecutives en l'espai latent (passar d'una coordenada a una propera). Aquesta metàfora espacial permet entendre que aquestes eines generen també el que ens poden semblar síntesis d'elements disperss del conjunt de dades, com si proposessin variacions respecte les fotografies originals que combinen elements de diverses d'aquestes. En aquest cas, això es pot veure, per exemple, en alguns elements de vestuari que barregen elements i semblen proposar un estil diferent:



Imatge 5. Resultats del segon entrenament, només amb els rostres

Seguint aquesta metàfora de l'espai latent, existeixen eines que permeten identificar vectors en aquest espai. És a dir, permeten anar cap a un cert tipus d'imatge o d'element de les imatges. Això vol dir, per exemple, que podríem identificar en quina zona hi ha els retrats de persones amb ulleres i no només generar imatges d'aquest tipus sinó també apropar qualsevol imatge

generada a aquella zona (hipotèticament i si hi hagués prou exemples, que en no és el cas aquí, serviria per posar ulleres a qualsevol de les imatges anteriors).



Imatge 6. Resultats del segon entrenament, només amb els rostres. Mostra de l'increment del psi

Les xarxes generatives GAN tenen un valor amb el que es pot treballar (és a dir, es pot modificar quan es demana que generi una imatge) que s'anomena psi (ψ). Aquest valor indica com de semblants volem que siguin els resultats al conjunt de dades d'entrenament. Si li posem un valor baix trauríem imatges més semblants, és a dir més probables (podent arribar a treure la imatge més probable de totes, que és, de fet, la primera que s'ha mostrat abans). Com més augmentem el valor menys probables són i, per tant, més elements poc comuns apareixen (per exemple, en valors baixos no hi ha ulleres, ni altres elements de roba com gorres o tampoc bigotis), més barreges es fan i menys realistes són els resultats. Les imatges següents mostren, explicat de manera senzilla, un mateix valor de coordenades (una mateixa imatge, podríem dir) però amb diferents psi en ordre creixent (0,2, 0,5, 0,7 i 1):

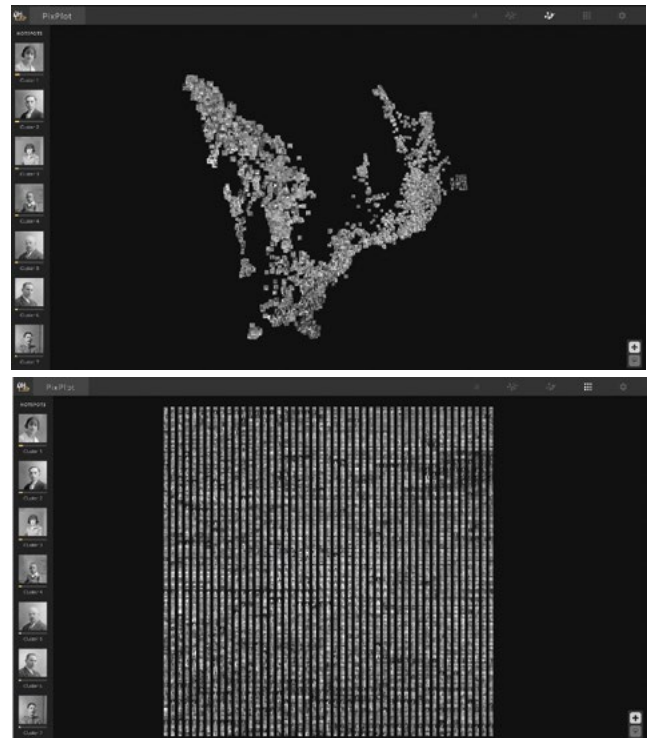
Com es veu, la metàfora de l'espai latent és útil per entendre aquestes eines i ens porta a entendre-les com un espai a explorar. En aquest sentit, hi ha, com s'ha vist, diferents recursos per fer aquesta exploració. Alguns que s'han desenvolupat més recentment impliquen poder donar una instrucció de cerca dins aquestes imatges hipotètiques. Per exemple, el que s'anomena «projecció» permet cercar dins aquest espai latent una imatge que s'assembli a una altra que li donem com a entrada. Aquest és per exemple el resultat de buscar dins la xarxa entrenada un conegut retrat de l'escriptor Franz Kafka (la de l'esquerra és el resultat que proposa l'eina i la de la dreta una imatge intermèdia; ja que és un procés de «cerca» que es pot visualitzar en forma de vídeo):



Imatge 7. Projecció d'un retrat de Franz Kafka dins l'entrenament de rostres

A més d'aquestes eines generatives, com s'ha comentat, es poden utilitzar eines de visió per computadora, que també utilitzen xarxes neuronals d'aprenentatge profund, per analitzar col·leccions d'imatges. Donarem alguns breus exemples de les seves possibilitats en relació a la visualització de col·leccions d'imatges.

Aquestes eines es poden utilitzar per organitzar un conjunt d'imatges per criteris de similitud gràfica; són recursos que es coneixen amb el nom d'UMAP o t-SNE. Les imatges es situen sigui en un espai 2D o en un espai 3D, aproximant entre elles les més semblants gràficament, en base a unes característiques extretes per una xarxa neuronal. Aquests serien un exemple de cada cas amb les fotografies originals re-enquadrades a primer pla:



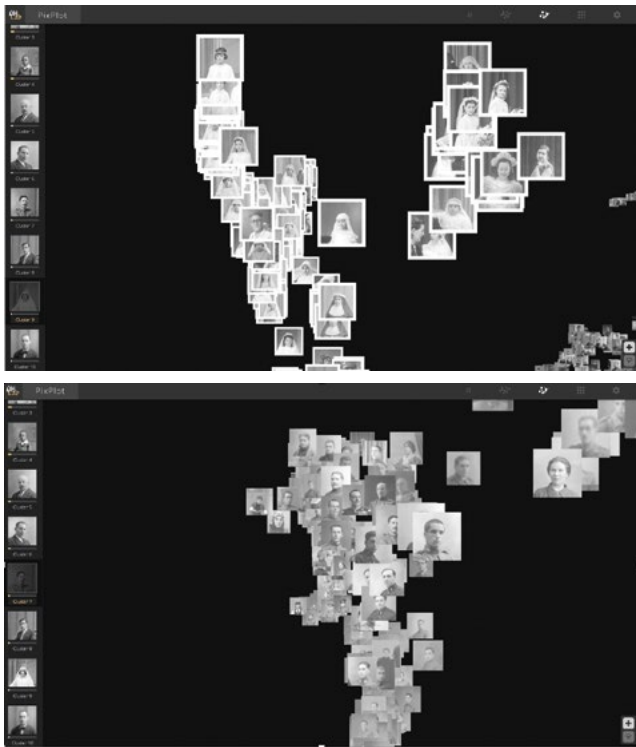
Imatge 8 i 9. Ordenació del fons d'arxiu en un espai 3D i un 2D

Si en la primera ordenació ens apropem veiem les imatges en un espai virtual 3D de la següent manera:



Imatge 10. Visualització més propera de l'espai 3D

El que es veu a l'esquerra són clústers d'imatges que l'eina identifica. Aquestes són dues captures de zones on principalment hi ha imatges de primera comunió i imatges de militars, respectivament (i en totes dues hi ha altres imatges que podem entendre per què hi són en tant pot haver-hi una similitud formal entre elles):



Imatge 11 i 12. Visualització més propera de l'espai 3D

Una altra possibilitat en relació a la visió per computadora és poder fer cerques a les imatges en funció del seu contingut. Això implica utilitzar alguna xarxa entrenada amb conjunt d'imatges com dels que es parlava a la primera part d'aquest article (per exemple, ImageNet). Aquest camp, però, ha evolucionat recentment amb el desenvolupament de la xarxa CLIP (Contrastive Language-Image Pre-training). Aquesta xarxa no parteix d'una taxonomia/vocabulari per a la qual s'han buscat exemples visuals, sinó que ha estat entrenada amb parells d'imatge amb un text descriptiu. Això vol dir, resumint i simplificant molt, que s'intenta que l'eina no reconegui un conjunt de paraules tancat sinó que s'apropi més al llenguatge natural.

També resumint i simplificant es pot destacar que el que fa CLIP és situar en un únic espai de coordenades els dos elements amb els que treballa: imatges i paraules (úniques o frases). Això vol dir que pot jutjar com de pròximes són entre ells una imatge i un text.

Si s'explica això és perquè CLIP es pot utilitzar amb aquesta intenció de buscar dins un conjunt d'imatges, escrivint paraules o frases. Per exemple, aquest són alguns dels resultats que ens ha donat quan hem buscat «dancing» o «sunglasses» dins el fons d'imatges Massafont (com es veu en el segon no sempre identifica bé, però podem preguntar-nos i entendre el perquè d'aquest resultat):



Imatge 13. Resultats de cercar «sunglasses» dins el fons d'imatges Massafont del CRDI

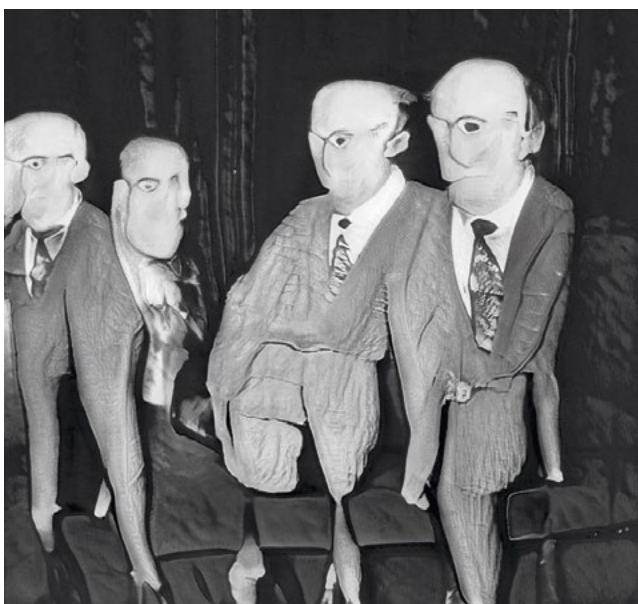
Altres projectes amb fons d'arxiu o xarxes GAN

El col·lectiu Estampa hem realitzat molts projectes amb eines d'IA; no només aplicades a les imatges sinó també a la generació de text. Els nostres projectes es poden veure i consultar dins el repositori que constitueix la

nostra pàgina web: tallerestampa.com Entre els realitzats, parlarem molt breument de dos que tenen relació amb els tests que s'han explicat en la secció anterior.

BCN 1965–1975. Una mirada des de la intel·ligència artificial (2021) va ser una exposició que vam realitzar al Castell de Montjuïc de Barcelona, organitzada per l'ICUB. En aquesta exposició vam treballar amb fons de l'Arxiu Fotogràfic de Barcelona, així com amb algunes filmacions de l'època. Es van aplicar els dos tipus de tècniques mencionades: la generació d'imatges (xarxes GAN) i la ordenació/visualització (UMAP). La visualització ordenava els fons originals en un núvol d'imatges que reordenava la memòria d'aquells anys. Un aspecte que ens sembla interessant és com aquestes eines poden servir per suggerir connexions; és a dir, que els seus resultats ens poden semblar interessants o suggerents segons la nostra mirada. Per exemple, l'eina va ajuntar en aquest cas les imatges amb vehicles i això creava una relació entre la indústria automobilística de l'època (la Seat, un saló de l'automòbil, etc.) i les imatges de vehicles militars.

També es van realitzar dos entrenaments de xarxes generatives GAN. Un va ser amb els fons de l'arxiu fotogràfic, amb un conjunt de dades que era petit quantitativament (514 imatges) i alhora molt heterogeni. El resultat de l'entrenament no és fotorealista, però no per això no és interessant. La funció d'anàlisi de l'imaginari amb la que s'ha entrenat la complex i, de vegades, sembla suggerir-ne paròdies.



Imatge 14. BCN 1965–1975. Una mirada des de la intel·ligència artificial (2021), Estampa

Com en aquest tipus d'imatge (que genera sovint) que sembla sintetitzar la imatge del poder que es conserva a l'arxiu (mostrant el seu gènere, la seva vestimenta i fins i tot metaforitzant la seva tasca en aquest ull únic):



Imatge 15. BCN 1965–1975. Una mirada des de la intel·ligència artificial (2021), Estampa

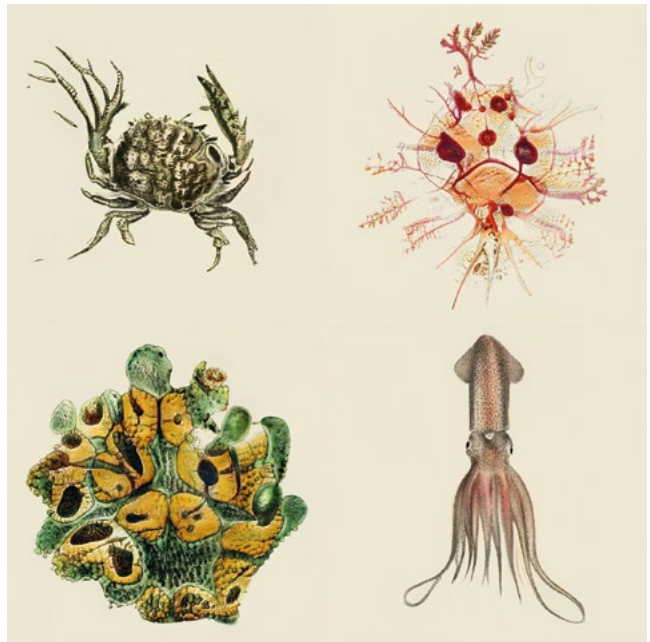
Si bé el resultat seria incorrecte des d'un punt de vista industrial (el *dataset* d'entrenament no compleix els requisits industrials i la imatge final no és realista), no ens sembla però que la imatge no tingui interès. En

certa manera en els nostres projectes volem entendre elements com aquests, no com errors que descartar, sinó com possibles proposicions.

Un altre entrenament va ser amb un conjunt de dades més elevat i homogeni: 1500 fotografies aèries de l'època. En aquest cas el resultat és més realista i sembla proposar una altra planta de la ciutat o una síntesi impossible de perspectives:

El projecte més gran amb xarxes GAN que hem realitzat és la instal·lació *Espècies marcianes* per a l'exposició *Mart. El mirall vermell* del CCCB comissariada per Juan Insúa. Per aquesta instal·lació es va construir un conjunt de dades de 24.000 imatges.

Aquestes eren il·lustracions científiques –principalment del s. XIX i principis del s. XX– d'espècies animals, vegetals, insectes o microorganismes relacionades amb l'imaginari marcianà, sigui per la seva capacitat de sobreviure en condicions extremes (com els fongs) o per la seva vinculació en la cultura popular (per exemple, els pops o els cactus). Aquest conjunt de dades va servir per entrenar una xarxa GAN i també es visualitzava en un UMAP de dues dimensions. La proposta s'integrava dins una part de l'exposició dedicada als desantropocentrismes i proposava utilitzar les xarxes GAN com una mena d'imaginació artificial que permetia imaginar altres espècies, éssers híbrids combinats fins a cert punt– al marge de les nostres convencions o ordenacions.



Imatge 16. *Espècies marcianes* (2021), Estampa

En aquesta presentació s'ha fet un recorregut senzill per alguns principis de les xarxes neuronals d'aprenentatge profund i alguns casos pràctics de la nostra experiència amb aquestes eines. Ens trobem en un moment de desenvolupament constant d'aquesta tecnologia i alhora en el que podríem anomenar el moment de la seva primera recepció cultural (com s'entén en l'imaginari públic). És per això que ens sembla particularment important poder fer ús d'aquestes tecnologies, per entendre-les, proposar usos i també per intervenir en el seu discurs.