# Aplicació de Tècniques d'Intel·ligència Artificial en la Postproducció

Adrià Mateos Martínez

**Resum**— Aquest article analitza i estudia diferents mètodes centrats en l'envelliment i rejoveniment de rostres mitjançant tècniques d'Intel·ligència Artificial i com aquestes eines s'estan implantant en la producció de productes audiovisuals. En concret s'analitzen tres models: Lifepan Age Transformation Synthesis (LATS), High Resolution Face Age Editing (HRFAE) i Style-based Age Manipulation (SAM). L'estudi compara els resultats de cada mètode tenint dos conceptes en el punt de mira, l'edat aparent dels rostres modificats i si es manté o no la identitat del rostre. Per tal de corroborar l'objectivitat de l'anàlisi, s'empra l'eina "deepFace" i diverses funcionalitats de la llibreria "face_recognition".

**Paraules clau**— Intel·ligència Artificial, Postproducció, LATS, HRFAE, SAM, GAN, deepFace, xarxes neuronals, Detecció facial, Reconeixement facial.

**Abstract**— This article analyzes and studies different methods focused on aging and rejuvenating faces through Artificial Intelligence techniques and how these tools are being implemented in the production of audiovisual products. Specifically, three models are analyzed: Lifepan Age Transformation Synthesis (LATS), High Resolution Face Age Editing (HRFAE), and Style-based Age Manipulation (SAM). The study compares the results of each method with two concepts in mind: the apparent age of the modified faces and whether the identity of the face is maintained or not. To corroborate the objectivity of the analysis, the "deepFace" tool and various functionalities of the "face_recognition" library are employed.

**Index Terms**— Artificial Intelligence, Post-production, LATS, HRFAE, SAM, GAN, deepFace, neural networks, Facial Detection, Facial Recognition.

## 1 Introducció

La postproducció és una pesa clau en la realització audiovisual, profundament vinculada a altres etapes com la preproducció i la producció. Es tracta de la fase final on s'aconsegueix una estructura coherent que permet reflectir l'idea final d'un projecte dotant-lo de la perfecció tècnica i artística necessària. La postproducció es podria definir com la manipulació del material audiovisual on es realitzen tots els processos necessaris per finalitzar el producte i que aquest sigui de la màxima qualitat possible.

Els professionals que intervenen en aquest procés són els muntadors (encarregats del muntatge o edició de les imatges), els muntadors de so, els músics, els mescladors (unió de tots els sons), els artistes VFX i els etalonadors (encarregats que igualen les diferents lluminositats i donen una continuïtat a l'obra).

En els últims anys, l'aplicació de tècniques d'intel·ligència artificial (IA) en la postproducció ha experimentat un ràpid creixement. La IA s'ha consolidat com una eina poderosa per millorar l'eficiència i la qualitat dels processos dins d'aquest àmbit. Aquest fet és a causa de la capacitat que tenen les IA per analitzar grans conjunts de dades i aplicar algoritmes avançats de processament d'imatge i so.

Entre les diverses tècniques emprades en la postproducció, cal destacar l'ús del "Deepfake". És un acrònim de "fake" i "Deep Learning". Tal com indica la paraula, du a terme un aprenentatge profund per tal de crear imatges "falses". Un ús concret d'aquesta tecnologia és la millora dels doblatges [1, 2]. En aquest cas, es sincronitza els moviments de la boca per tal que concordi amb l'àudio d'un doblatge.

Una altra tècnica és el rejoveniment o envelliment d'actors, permetent-los interpretar personatges més joves o més vells [3]. En conseqüència, fa possible la creació de pel·lícules com la recent d'Indiana Jones, on Harrison Ford interpreta un Indiana Jones més jove, o "Captain Marvel", on Samuel L. Jackson interpreta un Nick Fury amb uns quaranta anys.

En aquests casos, en una primera observació i amb l'experiència personal d'haver vist una gran quantitat de seqüències animades en videojocs. Es pot apreciar certs aspectes als rostres que et fan adonar-te'n de l'engany. Fet que demostra que queda bastant per millorar si s'utilitza aquesta tècnica per modificar cares que es veuen amb claredat i durant una estona prolongada.

Un altre exemple és l'ús d'aquestes tècniques per suplantar als dobles d'acció en les escenes de risc que realitzen i posar la cara dels actors principals.

Amb això present, l'ús de tècniques d'IA en la postproducció ofereix una sèrie d'avantatges significatius. En primer lloc, millora la qualitat visual i auditiva mitjançant correccions d'imatge i so precises i automatitzades, assegurant un producte final de primera qualitat. A més, possibilita la identificació i segmentació precisa d'objectes per aplicar


- *E-mail de contacte: adriamm99@gmail.com*
- *Menció realitzada: Computació*
- *Treball tutoritzat per: Coen Antens (CVC)*
- *Curs 2023/24*

efectes visuals específics. En segon lloc, l'ús d'IA pot conduir a una reducció substancial en els costos de producció. Aquest fet es deu a la capacitat de la IA per minimitzar la necessitat de mà d'obra per a tasques repetitives i manuals, optimitzant així els recursos i l'eficiència del procés de postproducció.

## 2 Objectius

Per tal de comprendre el funcionament d'aquestes tècniques, l'objectiu és estudiar les diferents tecnologies que utilitzen intel·ligència artificial en l'àmbit de la Postproducció audiovisual. Per tal de saber com funcionen concretament les tècniques de "Deepfake" i envelliment o rejoveniment de persones. I així entendre els processos que segueixen per tal d'aconseguir els resultats esperats.

A més, per aprofundir en l'estudi dels diferents mètodes, com a objectiu principal es realitzarà l'aplicació de tècniques d'envelliment i rejoveniment a diferents imatges mitjançant diverses eines.

També, per tal de aclarir i saber quina tècnica és millor, s'analitzarà i estudiaran el resultats obtingut per tal de comparar-los amb altres mètodes.

## 3 Metodologia

La metodologia que es farà servir per dur a terme aquest projecte combina elements de Scrum i Kanban. En particular, es basa en l'ús de sprints de Scrum, tot i que s'adapta la metodologia per acomodar el fet que aquest és un projecte individual, la qual cosa simplifica la jerarquia.

A més, per la part de Kanban, es gestiona i es representa l'evolució de les tasques de manera visual. Per a això, es fa servir l'eina Jira, que permet un control visual de les tasques i el seu progrés.

## 4 Planificació

En aquesta secció es mostren les diferents tasques a realitzar per tal de complir els objectius en els diversos terminis de lliurament durant la realització del projecte. Les tasques s'han realitzat en sprints tenint com a termini màxim les entregues parcials.

- Estudiar i comprendre les xarxes neuronals GAN.
- Comprendre l'ús de les GAN en els models de "Deepfake".
- Comprendre els models d'envelliment o rejoveniment.
- Analitzar diversos mètodes d'envelliment o rejoveniment.
- Fer un seguit de proves amb els mètodes escollits.
- Realitzar una base de dades amb persones de diferents edats per controlar el funcionament dels diferents mètodes.
- Comparació dels resultats entre els mètodes.
- Utilitzar detectors d'edat i reconeixement facial per comprovar la robustesa dels mètodes.
- Analitzar els resultats, estudiar millores possibles.
- Elaboració de dossier
- Entrega de l'informe final
- Preparar la presentació.

## 5 Estat de l'art

Des dels darrers anys, l'aplicació de tècniques d'IA en la postproducció ha concentrat una atenció considerable. Pel que fa als diferents mètodes per tal d'aplicar transformacions facials primer de tot és necessari un mètode per detectar la cara on es volen realitzar els canvis.

Un exemple molt conegut és l'algorisme Viola Jones[21], aquest mètode analitza una sèrie de característiques anomenades característiques de Haar, que són patrons rectangulars que representen canvis abruptes en la intensitat dels píxels a una imatge. L'algorisme utilitza el mètode "Adaboost" per seleccionar el conjunt de característiques més discriminatives per a detectar rostres. Per millorar l'eficiència, el mètode utilitza una estructura en cascada que aplica al classificador en varies etapes. Aquesta tècnica de detecció de rostre és molt eficient fer a la detecció en temps real.

Un altre model és el "Selective Refinement Network" [14], es tracta d'un classificador capaç de millorar l'eficiència en quan a nombre de falsos positius i la precisió a l'hora de localitzar les cares. Tal i com es pot veure a la Figura 1, aquest model consisteix dos mòduls. El primer, de classificació, encarregat de filtrar la majoria de mostres negatives mentre que el segon de regressió ajuda en la precisió per al que fa a la detecció del rostre.

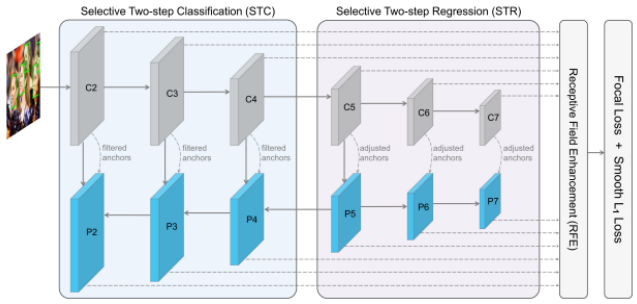



*Figura 1: Estructura de la Xarxa SRN.*

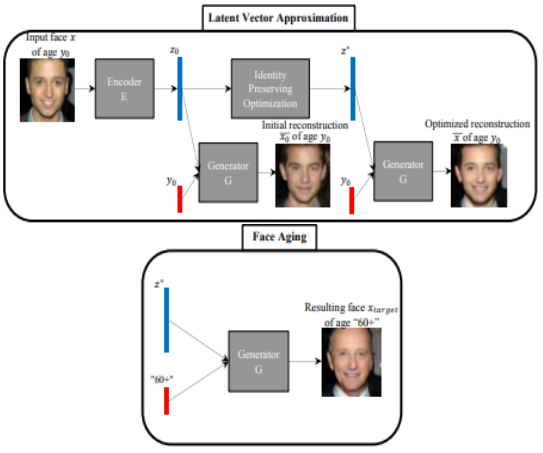



*Figura 2: Mètode d'envelliment amb cGAN. Primer es fa una aproximació del vector latent per reconstruir la imatge d'entrada i després es modifica l'edat a l'entrada del generador per realitzar l'envelliment.*

Per tal d'aplicar envelliment o rejoveniment, en destaquem l'ús de mètodes on s'apliquen "conditional Generative Adversarial Network" (cGAN), arquitectura a la Figura 2, que permeten modificar l'edat en imatges generades [5, 6, 7, 8].

Un dels inconvenient que tenen aquests models és que acostumen a produir imatges borroses o en les quals no es pot identificar a la persona. Per tal de solucionar-ho s'han implementat diferents mecanismes com un classificador d'identitat preentrenat [7, 9]. Una altra forma per tal de millorar els resultats obtinguts, ha sigut mitjançant "S²GAN" [10] que aplicaria transformacions específiques de l'edat a la base de l'envelliment codificat per tal d'obtenir i descodificar la representació de l'edat de la imatge d'entrada. Tot i així, aquest mètode té menys capacitat per representar el transcurs de l'edat.

Una altra proposta que busca millorar el resultat produït per les GAN és l'StyleGAN i la manipulació de l'edat basada en l'estil (SAM) [19]. Com es veu a la Figura 3, utilitzen una xarxa per predir l'edat i aquesta, guia al codificador encarregat de fer les transformacions pertinents, generant una ruta que representa la progressió de l'edat. A més, aquest mètode preserva la identitat facial.

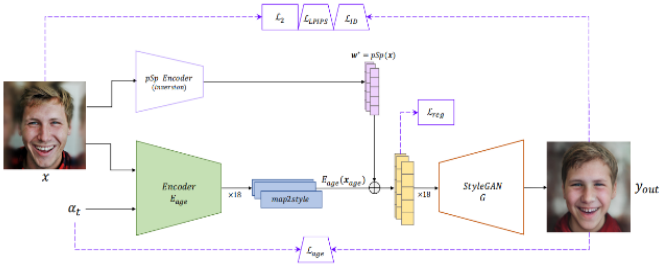



*Figura 3: Arquitectura del mètode SAM. La xarxa rep una imatge d'entrada i l'edat objectiu. El codificador d'envelliment té la tasca d'extreure les diferents característiques. Finalment, s'utilitza un StyleGAN preen-trenat per generar la imatge transformada.*

Per tal de generar imatges amb gran resolució es proposa un editor de l'edat facial capaç d'operar amb grans resolucions (HRFAE)[11]. Aquest mètode prepondera les característiques codificades mitjançant la sortida d'una única capa connectada. Un inconvenient és que limita el procés a l'envelliment i no al rejoveniment.

Un altre model, presenta un mètode per sintetitzar les transformacions de la cara en passar els anys (LATS)[12]. El descodificador d'aquest mètode realitza convolucions modulades en les característiques facials mentre implementa el vector d'edat objectiu après de la xarxa. Tot i això, LATS se centra en la cara obviant el fons de la imatge, fet que pot conduir a què apareguin objectes inesperats.

Per acabar, el "Re-Aging" GAN (REGAN)[13] permet conservar la identitat d'una persona canviant l'edat modificant les característiques facials entre diferents grups d'edat amb més eficàcia i, a diferència de LATS, no es centra únicament en la cara i permet mantenir el fons de la imatge, tal i com es veu a la Figura 4.

A més d'aquests mètodes, també podem aplicar la traducció d'imatge a imatge per a la transformació de l'edat facial, en especial la xarxa d'envelliment facial (FRAN)[3] capaç de proporcionar resultats d'envelliments estables a alta resolució en vídeos on es mostren cares amb diferents expressions i modificant la profunditat, el moviment i les condicions d'il·luminació. Un inconvenient és la incapacitat de modificar correctament les cares a edats inferiors als divuit anys.

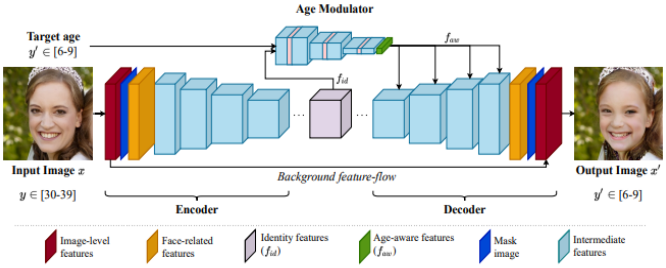



*Figura 4: Mètode REGAN. Donada una imatge d'entrada, el model la transforma a l'edat objectiu mitjançant diferents funcions que aporten informació sobre l'edat.*

Per acabar amb aquest llistat d'exemples, destacar l'aplicació mòbil FaceApp[], que té la capacitat de realitzar un munt de transformacions als rostres, des de corregir petites imperfeccions a afegir una barba o ulleres i fins i tot manipular l'edat del rostre. Aquesta aplicació utilitza un altre proposta de millora a les GAN, el CycleGAN està centrat en traducció d'imatge a imatge. En resum, aquesta arquitectura es basa en dos xarxes GAN connectades, com es pot veure a la Figura 5, les dades d'entrada de la segona xarxa corresponen a les dades de sortida de la primera.

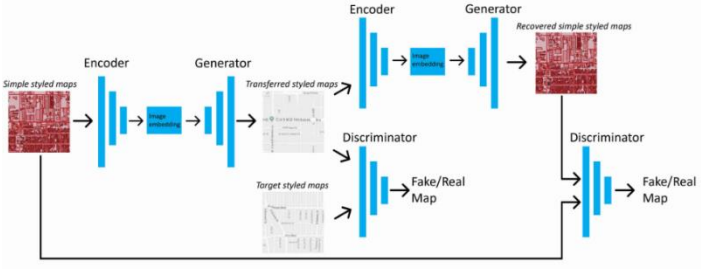



*Figura 5: Arquitectura CycleGAN*

Per tal de comprendre amb més exactitud els diferents funcionaments d'aquestes mètodes, és imprescindible comentar el funcionament de l'AutoEncoder i de les "Generative Adversarial Networks" (GAN) i com varia la definició d'espai latent per aquestes estructures.

Un AutoEncoder és una xarxa neuronal dissenyada per aprendre una representació compacta a partir d'unes dades d'entrada. Aquest està compost per dues parts principals: l'encoder i el decoder.

L'encoder és la part de la xarxa que pren les dades d'entrada i les transforma en una representació comprimida, això seria l'espai latent. En aquesta fase, les característiques més rellevants de les dades són extretes i comprimides en una representació més petita.

Per la seva banda, el decoder, pren aquests codi latent i intenta reconstruir les dades d'entrada originals a partir d'aquesta representació comprimida. Es podria dir que el decoder és l'encarregat de descomprimir el codi latent per recrear les dades originals.

En aquest cas, l'espai latent és el conjunt de totes les possibles representacions dels exemples d'entrada en el codi latent. Aquest espai és molt més compacte que l'espai

original de les dades, ja que conté només la informació essencial necessària per reconstruir els exemples d'entrada originals.

Pel que fa a les GAN, són un tipus de model generatiu que està format per dues xarxes competidores, el generador i el discriminador.

El generador pren un vector latent (una petita representació de les dades d'entrada) i intenta generar dades que es semblin a les reals. El discriminador, en canvi, s'entrena per distingir entre dades reals i dades generades. El seu objectiu és classificar les dades com a reals o falses.

Pel que fa a l'espai latent en una GAN, és l'espai dels vector d'entrada al generador. Aquest espai pot ser de dimensions més elevades i acostuma a tenir una estructura més complexa respecte al generat per l'AutoEncoder.

El que es fa amb aquest espai, és introduir un vector que servirà com a identificador de les dades que es vulguin generar, dins d'aquest espai latent. Per exemple, si a una xarxa se li dona el mateix vector dues vegades, generarà la mateixa sortida en els dos casos. Però si se li dona un de diferent, el resultat variarà. De la mateixa manera, si modifiquem una mica el vector d'entrada, el resultat serà una petita modificació.

## 6 DESENVOLUPAMENT

En aquesta secció es presenten les diferents tecnologies emprades en els experiments realitzats centrant-se en la detecció i posterior modificació de rostres en imatges.

### 6.1 Face Detection

Per tal de poder aplicar diferents canvis a les cares dels actors, és essencial poder detectar les cares on poder fer les modificacions.

Per tal de comprendre el funcionament del model de detecció de cares, s'ha realitzat un de propi. A la Figura 6, es veu el resultat de la detecció de cares en temps real.

El model ha estat entrenat amb una base d'imatges pròpia. Aquesta base de dades està formada per 3780 imatges per a l'entrenament, 840 per al testatge i 780 per a la validació.

Pel que fa a l'eficiència del model, aquest és capaç de detectar cares en diferents postures i paràmetres d'il·luminació. A més, té la capacitat de trobar cares en temps real.


*Figura 6: Detecció de cares en temps real.*

### 6.2 Face Recognition

Un altre mètode realitzat ha estat mitjançant la llibreria "face_recognition"[15]. Aquesta llibreria permet manipular i reconèixer cares i utilitza models del conjunt d'eines "dlib"[16]. El model, anomenat "dlib face recognition resnet model", està entrenat mitjançant la base de dades de "Labeled Faces in the Wild Home"[17] i té una precisió del 99,38%.

Mitjançant aquesta funcionalitat s'han dut a terme diferents proves per tal de familiaritzar-se amb l'ús de la llibreria. El primer mètode que s'ha testejat ha estat el de modificar i transformar aspectes de la cara mitjançant els "landmarks", com poden ser els llavis, les celles o els ulls.

*Figura 7: Comparativa d'imatges on s'han modificat aspectes de la cara.*

Com es pot apreciar a la Figura 7, aquesta funcionalitat té la capacitat d'obtenir els punts de referència facial i a partir d'aquest aplicar diferents màscares per tal de modificar la imatge inicial. Un aspecte important a tenir en compte per a aplicar les modificacions a la cara és procurar mantenir la identitat de la persona i les expressions d'aquesta.

Un altre mètode testejat és el reconeixement facial mitjançant el mètode de classificació "K-Nearest Neighbors" (k-NN)[18]. Aquest algoritme és un dels més simples pel que fa a la classificació i permet diferenciar i reconèixer vàries cares en una imatge i classificar-les pel nom de la persona.


*Figura 8: Reconeixement de persones en imatges.*

Com es pot observar a la Figura 8, la primera només reconeix una cara i a més la pot classificar, en la segona imatge, pot diferenciar dues cares i classificar-les correctament.

### 6.3 Envelliment i rejoveniment

Per realitzar les diferents transformacions a les cares dels actors, s'han utilitzat tres mètodes diferents. Dos d'ells basats en GAN, el mètode LATS i el mètode SAM. L'altre utilitza una estructura d'AutoEncoder, formada per un codificador, un modulador de funcions per a la selecció de l'edat i un descodificador, el mètode HRFAE

Per tal d'entendre el funcionament dels mètodes emprats per a l'envelliment o rejoveniment, és necessari comprendre el funcionament de les GAN. Tal i com s'ha

mencionat anteriorment, es tracta d'una arquitectura d'aprenentatge profund en la que dues xarxes neuronals competeixen entre elles. L'objectiu de la xarxa generadora és ser capaç de captar la distribució de les dades i a partir d'una sèrie de valors aleatoris generar una imatge que sigui similar en forma i contingut a les imatges del conjunt de dades que s'està utilitzant. Per altra banda, la tasca de la xarxa discriminadora serà jutjar i diferenciar entre si una imatge donada correspongui a una real o pel contrari generada mitjançant la xarxa antagònica.

Si observem aquest tipus de xarxa des del punt de vista matemàtic, es pot considerar com un joc de suma zero, en el qual el generador, tractarà de minimitzar la mateixa funció de pèrdua que el discriminador intentarà maximitzar.

En el cas específic per als mètodes que utilitzarem, la primera xarxa generarà una imatge facial amb l'edat modificada mentre que la segona comprovarà que la imatge resultant es pot determinar com a una imatge facial més envellida o rejovenida que l'original.

Els mètodes emprats, tant LATS com SAM, busquen una millora en l'ús de les GAN per a executar aquesta tasca. És per això que es basen en una arquitectura GAN, però prenen enfocaments diferents per tal d'obtenir resultats més bons.

En el cas de SAM, per modelar el procés d'envelliment, s'introdueix una arquitectura completa de traducció d'imatge a imatge combinant una xarxa de codificadors i un generador d'imatges pre-entrenat. El codificador, codifica directament la imatge i l'edat objectiu a un conjunt de vectors d'estil que capturen la transformació que es desitja. A partir d'aquest vectors d'estil, el generador s'encarrega de generar la imatge de sortida desitjada. En el model de SAM, a més, s'aplica durant l'entrenament del model una pèrdua de consistència del cicle (cycle consistency loss), que demostra ser més eficaç en tasques de traducció d'imatge a imatge [21]. Per guiar el codificador per generar els vectors d'estil adequats, es fa servir una xarxa pre-entrenada de regressió de l'edat que actua com una restricció addicional de pèrdua durant el procés de formació.

En el cas de LATS, es fa servir una arquitectura amb un codificador d'identitat, on a partir de la imatge d'entrada, s'extreuen les característiques d'identitat. Aquestes contenen informació de l'estructura de la imatge i la forma del rostre, elements importants en la generació posterior per tal de mantenir la identitat. També es fa servir una xarxa de mapeig que aprèn un espai latent d'edat òptim que permet una transició suau entre els diferents clústers d'edat, necessària per a realitzar transformacions d'edat continues. Per acabar, el descodificador pren un codi latent d'edat juntament amb les característiques d'identitat i produeix una imatge de sortida aplicant diferents processos per tal de reduir artefactes a les imatges.

En el cas d'StyleGAN (SAM), l'espai latent que genera és més ordenat i millor estructurat. Els vectors identificadors per a cada punt de l'espai constaran de 512 coordenades, és a dir, un espai de 512 dimensions. A partir del vector identificador d'una imatge, modificant certs aspectes del vector, podrem aconseguir diferents versions de la imatge. Permetent d'aquesta manera i fent servir l'espai latent, poder obtenir imatges on el que es modifica és l'edat aparent del rostre de la imatge original.

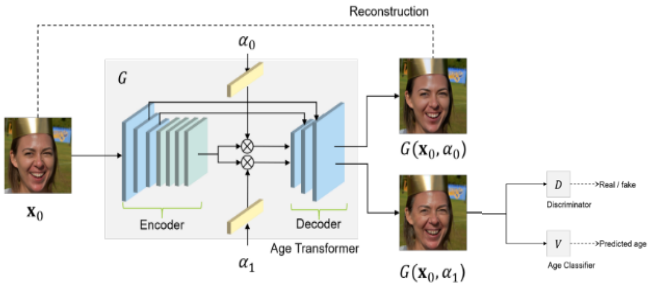



*Figura 9: Arquitectura del HRFAE.*

Pel que fa al model HRFAE, aquest utilitza un mòdul basat en un AutoEncoder, s'observa a la Figura 9. Aquest consta d'un codificador, que rep la imatge d'entrada i codifica les característiques en 128 canals. Seguidament es modulen les característiques del rostre per a la selecció de l'edat. Per fer-ho, l'edat objectiu es codifica com un vector que s'utilitzarà per preponderar les característiques de la imatge que surten del codificador. Per últim, el descodificador pren les característiques modulades com a valors d'entrada, més dues connexions que no han estat modificades per tal de preservar els detalls del rostre i mantenir la identitat, obtenint així la imatge de sortida que tindrà el rostre de la imatge inicial amb l'edat modificada.

### 6.4 Detector de l'edat

Per tal de corroborar el bon funcionament dels mètodes, s'han avaluat els resultats utilitzant "deepFace" [20]. Es tracta d'un entorn lleuger de reconeixement facial i anàlisi d'atributs (edat, gènere, emocions i raça). Consisteix en un marc híbrid entre diversos mètodes de reconeixement facial.

## 7 Experiments

En aquesta secció s'expliquen i s'interpreten els diferents resultats obtinguts en aplicar les diferents tecnologies. En concret, les encarregades d'envellir i rejovenir imatges facials, juntament amb mètodes de detecció de rostres i predictors d'edat, mencionats amb anterioritat a l'apartat de desenvolupament.

### 7.1 Tècniques d'envelliment i rejoveniment d'imatges facials amb diferents mètodes.

S'ha realitzat una base de dades amb imatges de diferents actors amb edats compreses entre els vint i els vuitanta anys. D'aquesta manera es podria comprovar l'eficàcia dels mètodes.

Tal com s'ha mencionat en l'apartat de desenvolupament, els mètodes utilitzats utilitzen diferents aproximacions per aconseguir modificar les imatges i així envellir o rejovenir la imatge inicial.

Per poder entendre les diferències entre els mètodes, cal destacar els punts forts o inconvenients de les tecnologies:

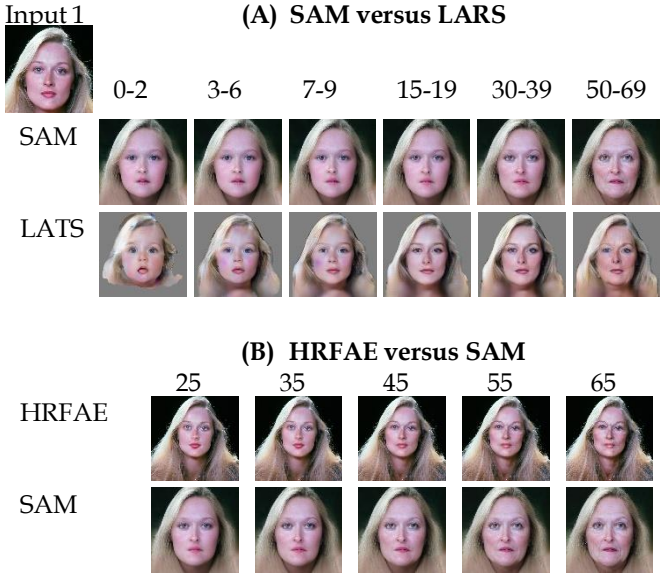
- "High Resolution Face Age Editing" (HRFAE): A diferència d'altres mètodes, permet aplicar transformacions a imatges de gran resolució (1024 x 1024). Un

gran inconvenient és el rang d'edats en el que opera, ja que només realitza modificacions entre els 20 i els 65 anys.

- "Lifespan Age Transformation Synthesis" (LATS): Es pot destacar la xarxa de mapatge, capaç d'aprendre l'espai latent d'edat òptim i d'aquesta manera generar una transició suau entre deferents grups d'edat, un punt necessari per a les transformacions d'edat contínues. Un pas del preprocessament de la imatge, és l'aplicació d'una màscara on el fons i la roba de la persona que apareix a la fotografia són eliminats, quedant només la cara. També, genera imatges modificades dintre d'unes franges d'edat establertes.
- "Style-based Age Manipulation" (SAM): Aquest mètode aprèn a codificar directament imatges facials en l'espai latent d'una GAN prèviament entrenada i subjectada a un canvi de l'envelliment determinat. Com s'ha mencionat abans, la xarxa de regressió s'utilitza per guiar al codificador en la generació dels codis latents corresponents a l'edat desitjada. A diferència d'altres mètodes que operen a l'espai latent usant un camí previ que controla l'edat. Aquest mètode aprèn un camí no lineal, permetent així una major edició de les imatges generades.

L'experiment ha consistit en modificar l'edat d'imatges de diferents actors mitjançant aquests tres mètodes i per tal de veure la seva eficàcia, es comprova l'edat de les imatges facials resultants amb un detector d'edat.

A causa de la limitació que té LATS (generar una imatge dintre de sis franges d'edat) i HRFAE (generar imatges entre els vint i els seixanta-cinc anys), es mostraran els resultats seguint aquestes restriccions per tal de poder-ho comparar.



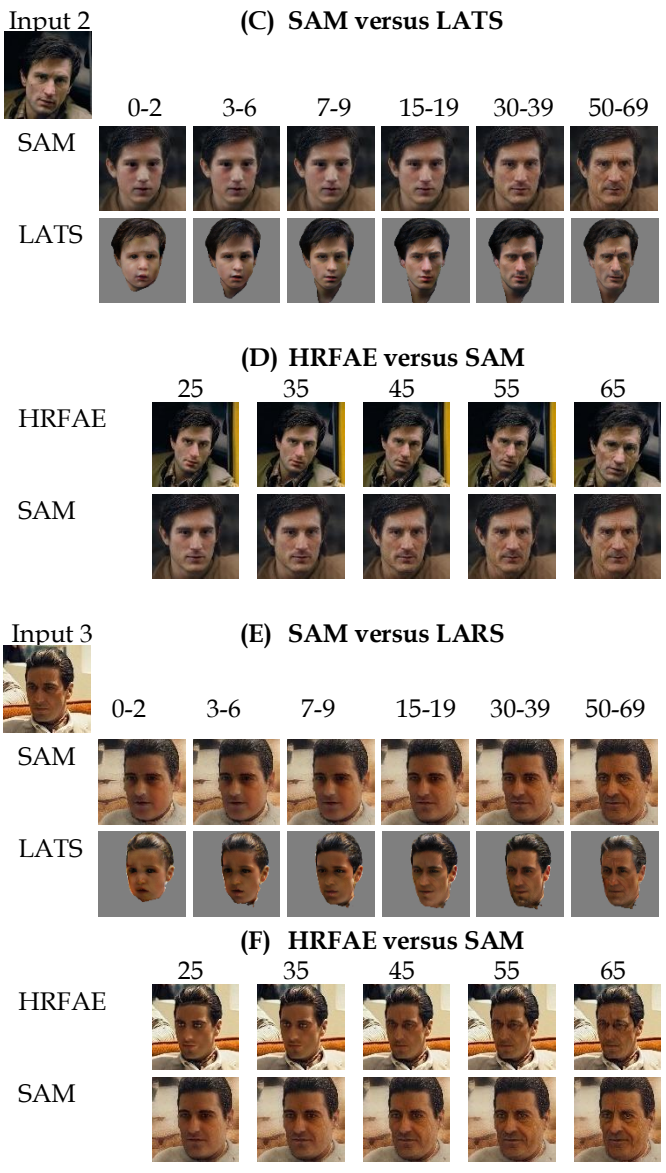




*Figura 3: Comparativa dels diferents mètodes. Les imatges generades per SAM (A, C, E) tenen com a edat objectiu la mitjana de la franja en la que actua LATS.*

Tal com es pot veure a la Figura 9 entre mètodes, LATS i SAM permeten generar i transformar amb una edat objectiu de zero anys. SAM, també permet realitzar modificacions des d'edats objectiu superiors als seixanta-cinc anys fins als cent. També es pot observar un inconvenient amb HRFAE on apareixen taques o objectes estranys a les imatges transformades.

### 7.2 Comprovació del resultats.

Per comprovar el bon funcionament dels diferents models, s'han realitzat una sèrie de proves per tal d'avaluar la seva eficàcia en quant a l'envelliment d'un rostre.

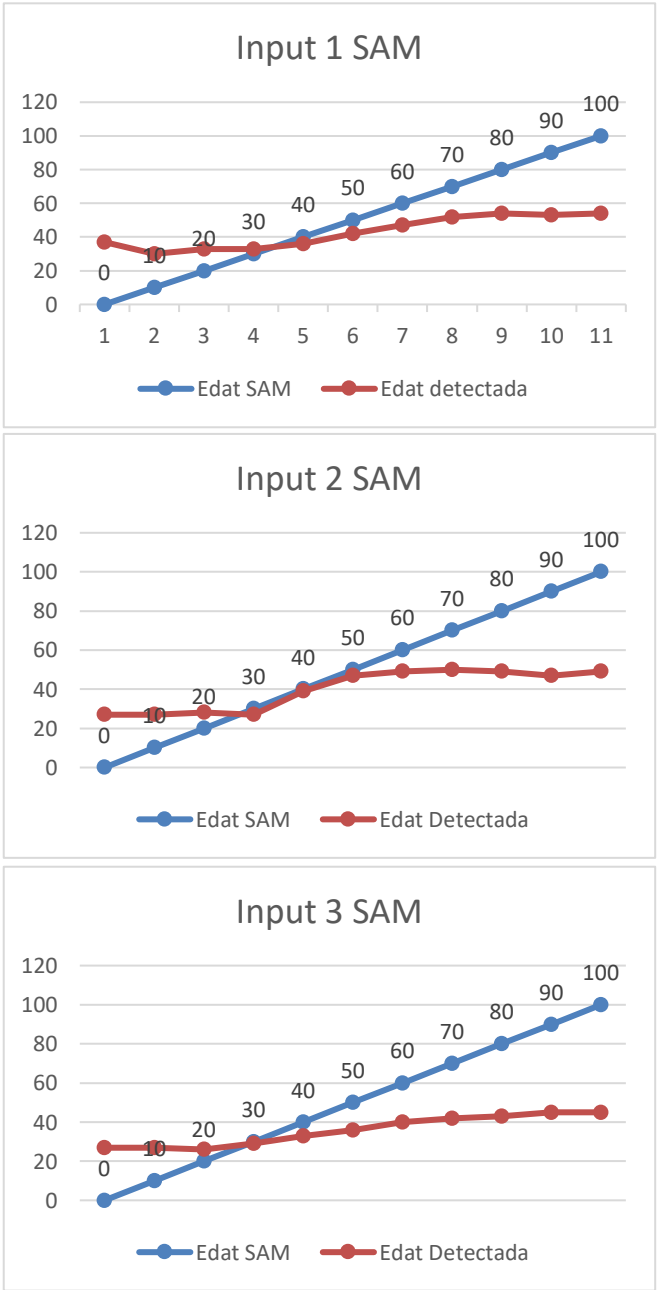
En el primer experiment, s'ha observat l'edat aparent de les imatges modificades per cada model i s'ha analitzat si aquesta concorda amb la que es capaç de detectar l'eina deepFace.

En el segon experiment, s'ha volgut comprovar si la identitat dels rostres es manté després de realitzar les

respectives modificacions per cada model. S'ha fet servir la llibreria de face_recognition i l'entorn de deepFace.

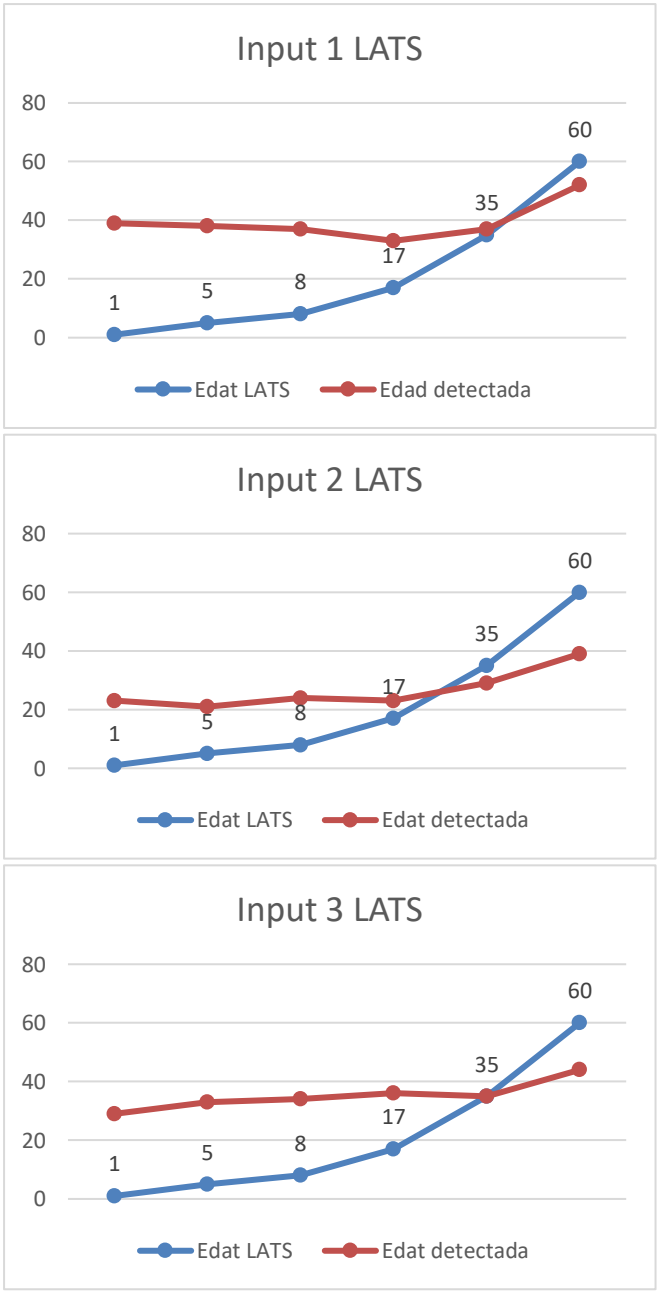
### *7.2.1 Detecció de l'edat*

Per corroborar que l'edat aparent de les imatges modificades pels models és correcta o semblant a l'edat objectiu i per tal d'avaluar-ho de la manera més objectiva possible, s'ha fet servir l'entorn deepFace, en concret la funcionalitat analyze que permet saber l'edat del rostre que detecta.
El model per l'edat d'aquest entorn té un ±4,65 MAE. Això vol dir que l'edat real es troba dins del segment del segment del MAE.
Un exemple. L'edat detectada per deepFace és de 40 anys, l'edat real del rostre es trobarà dins de 40 ± 4,65. Sent aquest rang entre 35,45 i 44,65 anys.




*Conjunt de gràfics 1: La Línia blava és l'edat objectiu de srotida del model SAM. La línia vermella és l'edat detectada per deepFace.*

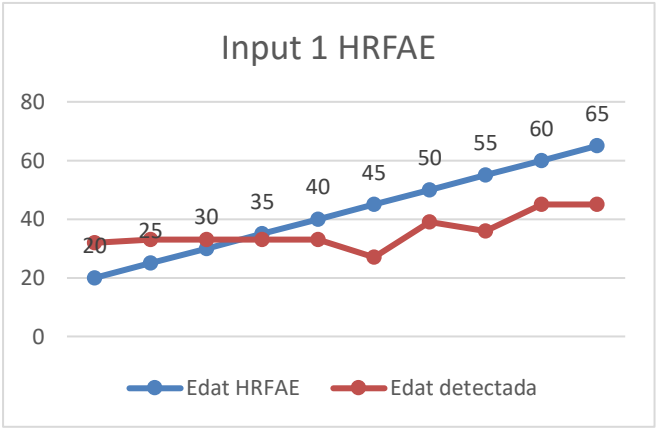
Es pot observar, al Conjunt de gràfics 1, com l'edat detectada pel deepFace a les imatges resultants del model SAM dels tres inputs diferents fluctua entre els 25 i els 45 anys. Les modificacions realitzades a les no son suficients per a que el detector d'edat tingui valors semblants a l'edat suposada que haurien de tenir aquestes. També es pot observar com entre l'edat objectiu dels trenta als cinquanta anys, el detector te resultats esperables dins de l'error absolut mitja.

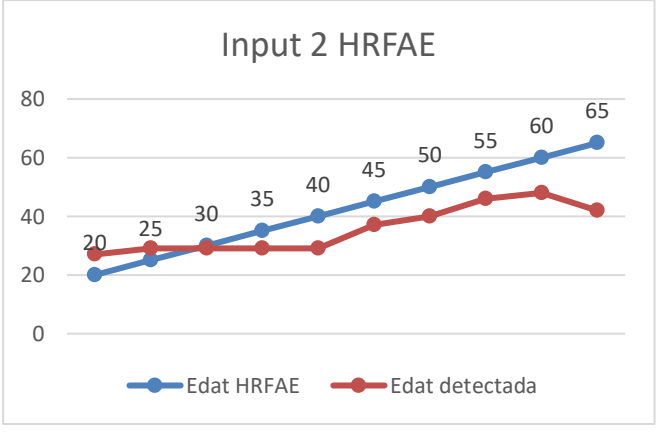



*Conjunt de gràfics 2: La Línia blava és l'edat objectiu de srotida del model LATS. La línia vermella és l'edat detectada per deepFace.*

Pel que fa als resultats obtinguts per LATS, al Conjunt de gràfics 2, en comparació amb l'edat detectada amb deepFace. Es pot destacar com aquest és capaç de detectar canvis en l'edat a partir que l'edat de sortida del model és

superior a l'edat de la fotografia original. És per això, que es pot afirmar, que les modificacions realitzades per rejovenir no apliquen un canvi substancial per tal que el detector sigui capaç de captar les diferents transformacions a les imatges. Per altra banda, pel que fa al envelliment, el detector d'edat és capaç de percebre aquestes modificacions però no s'apropa a l'edat objectiu que té el model.




*Conjunt de gràfics 3: La Línia blava és l'edat objectiu de srotida del model HRFAE. La línia vermella és l'edat detectada per deepFace.*

Per finalitzar, en el cas del model HRFAE, destacar que el detector d'edat deepFace no ha estat capaç de detectar el rostre per a les imatges resultants de l'input 3.

En els altres casos, es pot veure com el detector no capta canvis significatius en quant a l'edat en els cinc primers punts i és a partir del sisè en el que detecta modificacions considerables en l'edat amb un tendència positiva sobretot a l'input 2.

### *7.2.2 Percepció de la identitat a les imatges modificades.*

Per tal d'estimar si la identitat del rostre es pot reconèixer, s'han realitzat dues proves a les imatges resultants dels tres models. La primera s'ha fet mitjançant la funcionalitat "find" de l'entorn deepFace i la segona, utilitzant la llibreria face_recognition, que com s'ha explicat amb anterioritat, utilitza el model de dlib per reconeixement de rostres i un classificador K-NN.

D'aquesta manera es vol assegurar l'objectivitat dels models i es poden comparar les dues eines.

***Predicció d'identitat amb deepFace***

| | Input 1 SAM | SI | NO | Input 2 SAM | SI | NO | Input 3 SAM | SI | NO |
|---|---|---|---|---|---|---|---|---|---|
| 0 | | x | | | | x | | | x |
| 10 | | x | | | x | | | | x |
| 20 | | | x | | x | | | | x |
| 30 | | | x | | x | | | | x |
| 40 | | | x | | x | | | x | |
| 50 | | | x | | x | | | x | |
| 60 | | | x | | x | | | | x |
| 70 | | | x | | x | | | | x |
| 80 | | | x | | x | | | | x |
| 90 | | | x | | | x | | | x |
| 100 | | | x | | | x | | | x |

***Predicció d'identitat amb la llibreria face_recognition***

| | Input 1 SAM | SI | NO | Input 2 SAM | SI | NO | Input 3 SAM | SI | NO |
|---|---|---|---|---|---|---|---|---|---|
| 0 | | x | | | x | | | x | |
| 10 | | x | | | x | | | x | |
| 20 | | x | | | x | | | x | |
| 30 | | x | | | x | | | x | |
| 40 | | x | | | x | | | x | |
| 50 | | x | | | x | | | x | |
| 60 | | x | | | x | | | x | |
| 70 | | x | | | x | | | x | |
| 80 | | x | | | x | | | x | |
| 90 | | x | | | x | | | | x |
| 100 | | x | | | | x | | | x |

*Taula 1: Comparació de resultats per la prova d'identitat. Model SAM.*

***Predicció d'identitat amb deepFace***

| | Input 1 LATS | SI | NO | Input 2 LATS | SI | NO | Input 3 LATS | SI | NO |
|---|---|---|---|---|---|---|---|---|---|
| 0-2 | | x | | | | x | | | x |
| 3-6 | | x | | | | x | | | x |
| 7-9 | | x | | | | x | | | x |
| 15-19 | | | x | | x | | | | x |
| 30-39 | | | x | | x | | | x | |
| 50-59 | | | x | | x | | | | x |

***Predicció d'identitat amb la llibreria face_ recognition***

| | Input 1 LATS | SI | NO | Input 2 LATS | SI | NO | Input 3 LATS | SI | NO |
|---|---|---|---|---|---|---|---|---|---|
| 0-2 | | | x | | | x | | | x |
| 3-6 | | | x | | x | | | | x |
| 7-9 | | | x | | x | | | x | |
| 15-19 | | x | | | x | | | x | |
| 30-39 | | x | | | x | | | x | |
| 50-59 | | x | | | x | | | x | |

*Taula 2: Comparació de resultats per la prova d'identitat. Model LATS.*

***Predicció d'identitat amb deepFace***

| | Input 1 HRFAE | SI | NO | Input 2 HRFAE | SI | NO | Input 3 HRFAE | SI | NO |
|---|---|---|---|---|---|---|---|---|---|
| 20 | | | x | | x | | | | x |
| 25 | | | x | | x | | | | x |
| 30 | | | x | | x | | | | x |
| 35 | | | x | | x | | | | x |
| 40 | | | x | | x | | | | x |
| 45 | | x | | | x | | | x | |
| 50 | | | x | | x | | | | x |
| 55 | | x | | | x | | | x | |
| 60 | | | x | | x | | | x | |
| 65 | | | x | | x | | | | x |

***Predicció d'identitat amb la llibreria face_ recognition***

| | Input 1 HRFAE | SI | NO | Input 2 HRFAE | SI | NO | Input 3 HRFAE | SI | NO |
|---|---|---|---|---|---|---|---|---|---|
| 20 | | x | | | x | | | x | |
| 25 | | x | | | x | | | x | |
| 30 | | x | | | x | | | x | |
| 35 | | x | | | x | | | x | |
| 40 | | x | | | x | | | x | |
| 45 | | x | | | x | | | x | |
| 50 | | x | | | x | | | x | |
| 55 | | x | | | x | | | x | |
| 60 | | x | | | x | | | x | |
| 65 | | x | | | x | | | x | |

*Taula 3: Comparació de resultats per la prova d'identitat. Model HRFAE.*

Tal com es pot observar a la Taula 1, pel que fa al model SAM, en el cas de l'utilitat "find" de l'entorn deepFace, els resultat han sigut pobres excepte per l'input 2 on es pot observar que és capaç de reconèixer i predir la identitat de la persona. Pel que fa a la llibreria face_recognition, els

resultats són molt més positius, sent capaç de predir on les modificacions de rostre son molt potents ja que es vol rejovenir a una edat molt primerenca.

Els resultats pel que fa al model LATS, a la Taula 2, es pot observar com amb deepFace s'obtenen resultats dispars depenent de l'input. Si es comparen les prediccions per al primer rostre en contraposició amb el segon, es veu que prediu per segments d'edats contraris, pel primer detecta la identitat per fotos rejovenides mentre que pel segon ho fa per les envellides. En el cas en on es busca la identitat amb face_recognition, té més dificultat amb les imatges del primer i segon segment d'edat però millora els resultat respecte deepFace en les imatges de la resta de sectors.

Per acabar, al model HRFAE, a la Taula 3. En el cas en que es fa servir deepFace els resultats són similars a la resta de models pels inputs un i tres, pel segon conjunt d'imatges ha obtingut molt bons resultats. Els resultats obtinguts amb face_recognition han estat els més positius ja que per a totes les edats ha detectat la identitat dels rostres, sen en aquest model, a part del primer input pel model SAM, l'únic on ho ha aconseguit.

## 8 Conclusions

En aquest treball s'ha fet una comparativa i anàlisi de tres mètodes d'envelliment i rejoveniment d'imatges facials com a tècniques a aplicar en el camp de la Postproducció audiovisual.

Pel que fa a l'actuació dels tres models, destacar el rendiment objectiu del model SAM en quant al rang d'edat on actua. Sent l'únic capaç de retornar imatges facials modificades des del zero fins als cent anys. En segon lloc tindríem el model LATS, aquest actua en grups d'edat, en general entre els zero i els seixanta-nou anys. En últim lloc tindríem l'HRFAE que actua en una franja de quaranta-cinc anys (dels vint als seixanta-cinc).

En quant a la comparativa realitzada mitjançant l'entorn deepFace en la seva funcionalitat "analyze" capaç de detectar l'edat dels rostres de les imatges. Els resultats no han estat del tot adequats. Pel que fa al rejoveniment, el detector no és capaç de captar cap modificació que canviï l'edat aparent del rostre modificat i per tant detecta l'edat de la imatge original.

En quant a les imatges envellides, el detector si que capta certes modificacions i per tant l'edat aparent d'aquests rostres modificats varia en referencia a l'edat de l'original.

En resum per aquesta prova, la comprovació dels resultats en quant a la detecció de l'edat, no destaca cap dels tres models analitzats.

En últim punt, la comparativa dels models en quant a la percepció de la identitat, pel que fa a el anàlisis mitjançant l'entorn deepFace i la seva eina "find", els tres models obtenen resultats similars exceptuant el cas del segon conjunt d'imatges pel model HRFAE on ha detectat el rostre en totes les imatges.

En el cas de l'ús de la llibreria face_recognition, pel que fa a si es manté la identitat en les imatges rejovenides o envellides, l'actuació ha estat molt positiva. Els resultats obtinguts pels models SAM i especialment HRFAE demostren que la identitat es manté en les imatges modificades. En el cas de LATS els resultats són més mediocres però força millors que amb l'ús de deepFace.

Els avenços que es realitzen en aquest àmbit es produeixen cada cop amb major celeritat. Com a treball futur, es podria analitzar i estudiar la capacitat d'implementar un model d'envelliment i rejoveniment utilitzant la potencia de les IA generatives com Stable Difusion, ja que permeten crear i modificar imatges a partir d'un text d'entrada. L'habilitat de poder especificar quines modificacions realitzar al rostre i que aquestes siguin més editables permetria crear imatges més fidels, més realistes i amb resultats més exitosos.

## Agraïments

Vull agrair al meu tutor, Coen Antens, per deixar-me total llibertat alhora de realitzar el treball i guiar-me en les diferents etapes del projecte i poder assolir els resultats esperats.

També vull agrair a la meva família i parella per animar-me i donar-me suport durant la realització del projecte, motivant-me i enfortint-me en els moments més complicats.

## Bibliografia

[1] Bregler, Christoph & Covell, Michele & Slaney, Malcolm. (1997). Video Rewrite: Driving Visual Speech with Audio. Computer Graphics, SIGGRAPH 97 Annual Conference Series. 31. 353-360. 10.1145/258734.258880.

[2] Kim, Hyeongwoo & Elgharib, Mohamed & Zollhöfer, Michael & Seidel, Hans-Peter & Beeler, Thabo & Richardt, Christian & Theobalt, Christian. (2019). Neural style-preserving visual dubbing. ACM Transactions on Graphics. 38. 1-13. 10.1145/3355089.3356500.

[3] Zoss, Gaspard & Chandran, Prashanth & Sifakis, Eftychios & Gross, Markus & Gotardo, Paulo & Bradley, Derek. (2022). Production-Ready Face Re-Aging for Visual Effects. ACM Transactions on Graphics. 41. 1-12. 10.1145/3550454.3555520.

[4] Mehdi Mirza and Simon Osindero.(2014) Conditional generative adversarial nets. arXiv:1411.1784.

[5] Grigory Antipov, Moez Baccouche, and Jean-Luc Dugelay. (2017). Face aging with conditional generative adversarial networks. In 2017 IEEE International Conference on Image Processing, pages 2089–2093.

[6] Zhang, Zhifei & Song, Yang & Qi, Hairong. (2017). Age Progression/Regression by Conditional Adversarial Autoencoder.

[7] Tang, Xu & Wang, Zongwei & Luo, Weixin & Gao, Shenghua. (2018). Face Aging with Identity-Preserved Conditional Generative Adversarial Networks. 7939-7947. 10.1109/CVPR.2018.00828.

[8] Song, Jingkuan & Zhang, Jingqiu & Gao, Lianli & Liu, Xianglong & Shen, Heng. (2018). Dual Conditional GANs for Face Aging and Rejuvenation. 899-905. 10.24963/ijcai.2018/125.

[9] Li, Peipei & Hu, Yibo & Li, Qi & He, Ran & Sun, Zhenan. (2018). Global and Local Consistent Age Generative Adversarial Networks.

[10] He, Zhenliang & Kan, Meina & Shan, Shiguang & Chen, Xilin. (2019). S2GAN: Share Aging Factors Across Ages and Share Aging Trends Among Individuals. 9439-9448. 10.1109/ICCV.2019.00953.

[11] Yao, Xu & Puy, Gilles & Newson, Alasdair & Gousseau, Yann & Hellier, Pierre. (2020). High Resolution Face Age Editing.

[12] Or-El, Roy & Sengupta, Soumyadip & Fried, Ohad & Shechtman,

Eli & Kemelmacher, Ira. (2020). Lifespan Age Transformation Synthesis. 10.1007/978-3-030-58539-6_44.

[13] Makhmudkhujaev, Farkhod & Hong, Sungeun & Park, In. (2021). Re-Aging GAN: Toward Personalized Face Age Transformation. 3888-3897. 10.1109/ICCV48922.2021.00388.

[14] Chi, Cheng & Zhang, Shifeng & Xing, Junliang & Lei, Zhen & Li, Stan & Zou, Xudong. (2019). Selective Refinement Network for High Performance Face Detection. Proceedings of the AAAI Conference on Artificial Intelligence. 33. 8231-8238. 10.1609/aaai.v33i01.33018231.

[15] Geitgey, Adam. (2020). Face-Recognition. PyPI. https://pypi.org/project/face-recognition/

[16] Dlib C++ Library. http://dlib.net/

[17] LFW Face Database : Main. https://vis-www.cs.umass.edu/lfw/

[18] Wirdiani, Ayu & Hridayami, Praba & Widiari, Ayu & Rismawan, Diva & Candradinata, Putu & Jayantha, I. (2019). Face Identification Based on K-Nearest Neighbor. Scientific Journal of Informatics. 6. 150-159. 10.15294/sji.v6i2.19503.

[19] Alaluf, Yuval & Patashnik, Or & Cohen-Or, Daniel. (2021). Only a matter of style: Age transformation using a style-based regression model. ACM Transactions on Graphics. 40. 10.1145/3450626.3459805.

[20] Serengil, Sefik & Ozpinar, Alper. (2021). HyperExtended LightFace: A Facial Attribute Analysis Framework. 1-4. 10.1109/ICEET53442.2021.9659697.

[21] Viola, Paul & Jones, Michael. (2001). Rapid Object Detection using a Boosted Cascade of Simple Features. IEEE Conf Comput Vis Pattern Recognit. 1. I-511. 10.1109/CVPR.2001.990517.