



Universitat
Oberta
de Catalunya

BIOMETRIA FACIAL i XARXES NEURONALS. *Aplicació notarial*

Norbert Garsot Borràs

Màster Universitari en Seguretat de les Tecnologies de la Informació i de les Comunicacions (MISTIC)

Anàlisi de dades

Enric Hernández Jiménez

Cristina Pérez Solà

29 de Desembre de 2020

Aquesta obra està subjecta a una llicència de

[Reconeixement-NoComercial-SenseObraDerivada](#)

[3.0 Espanya de Creative Commons](#)

FITXA DEL TREBALL FINAL

Títol del treball:	<i>Biometria Facial i Xarxes Neuronals. Aplicació notarial</i>
Nom de l'autor:	<i>Norbert Garsot Borràs</i>
Nom del consultor/a:	<i>Enric Hernández Jiménez</i>
Nom del PRA:	<i>Cristina Pérez Solà</i>
Data de lliurament (mm/aaaa):	<i>12/2020</i>
Titulació o programa:	<i>Màster Universitari en Seguretat de les Tecnologies de la Informació i de les Comunicacions (MISTIC)</i>
Àrea del Treball Final:	<i>Anàlisi de Dades</i>
Idioma del treball:	<i>Català</i>
Paraules clau	<i>Aplicació de reconeixement facial / Xarxes neuronals / Notariat</i>

Resum del Treball:

Aquest any 2020 queda i quedarà marcat, durant l'història i a nivell mundial, per la pandèmia produïda pel virus del SARS-CoV-2, més conegut com a Coronavirus o Covid-19. Aquesta pandèmia ha sacsejat la normalitat que coneixíem i hem hagut d'adaptar el nostre dia a dia a certes mesures que ens han anat limitant, per dir-ho d'alguna manera, la llibertat que abans coneixíem. Algunes d'aquestes mesures adoptades ha estat la limitació pel que fa a desplaçaments i reunions. Degut a la mesura de limitar les reunions, les feines relacionades amb l'àmbit notarial han perdut la reunió presencial (notari-interessat/s) i s'han hagut d'adaptar amb les tecnologies digitals actuals. Un dels requisits principals abans de desenvolupar la feina notarial és verificar que la persona interessada és qui diu ser. Aquesta verificació la realitza el notari fent ús del document d'identitat (e.g., DNI espanyol). El simple fet que s'hagi de realitzar la verificació durant una videoconferència i no presencialment, minva el grau d'incert i, per tant, augmenta la

probabilitat de frau. És per aquest motiu que aquest treball té com a finalitat estudiar les tècniques de reconeixement facial fent ús de l'emergent tecnologia coneguda com a xarxes neuronals, per tal de dissenyar i implementar una aplicació, amb estructura client-servidor, que tingui com a finalitat la identificació facial. Per tant, aquesta aplicació està pensada i té com a finalitat principal, donar suport al notari durant el procés de verificació de la persona en una videoconferència.

Per a la realització del treball, es farà un estudi de l'estat de l'art tant en l'àmbit de la biometria facial com de les xarxes neuronals, i la integració d'ambdós mons. Finalment, es dissenyarà i s'implementarà una aplicació de reconeixement facial durant una videoconferència i s'extraurà les conclusions que respondran a la pregunta de si ha estat possible solucionar el problema.

Abstract:

This year 2020 is and will be marked, throughout history and worldwide, by the pandemic caused by the SARS-CoV-2 virus, better known as Coronavirus or Covid-19. This pandemic has shaken the normalcy we knew, and we shall adapt our daily lives to certain measures that have been limiting the freedom we were used to. One of the main taken measures has been to limit the movements and the meetings. Due to the measure of limiting the meetings, the works related to the notarial field have lost the face-to-face meeting (notary-interested) and have had to adapt with the current digital technologies. One of the main requirements before developing the notarial work is to verify that the person concerned is who claims to be. This verification is carried out by the notary using an identity document (e.g., Spanish DNI). The fact that verification must be performed during a video conference and not face-to-face decreases the degree of accuracy and, therefore, increases the likelihood of fraud. For this reason, this work aims to study facial recognition techniques using the emerging technology known as neural networks, in order to design and implement an application, with client-server structure, which aims to facial identification. Therefore, this application is intended, and has as its main purpose, to support the notary during the process of verifying the person when using a video conference.

To carry out the work, a study of the state of the art in the field of facial biometrics and neural networks will be made, and the integration of both worlds. Finally, a facial

recognition application will be designed and implemented during a video conference and conclusions will be drawn that will answer the question of whether it has been possible to solve the problem.

Índex

1.	<i>Introducció</i>	10
1.1.	Context i justificació del Treball.....	10
1.2.	Objectius del Treball	11
1.3.	Enfocament, mètode seguit i tasques a realitzar.....	11
1.4.	Breu revisió de l'estat de l'art	12
1.5.	Breu descripció dels altres capítols de la memòria	12
2.	<i>Revisió i estudi de l'estat de l'art</i>	14
2.1.	Sistema de reconeixement facial	14
2.1.1	Etapas d'un sistema de reconeixement facial	16
2.1.2	Detecció facial i aïllament	16
2.1.3	Normalització i alineament.....	19
2.1.4	Extracció de característiques	20
2.1.5	Reconeixement (Identificació i verificació facial)	21
2.2.	Xarxes neuronals	22
2.2.1	Introducció a les xarxes neuronals.....	22
2.2.2	Aplicacions	22
2.2.3	Estructura, funcionament i topologies de les ANN	23
3.	<i>Reconeixement facial mitjançant xarxes neuronals</i>	29
3.1.	Implementacions més vigents (SoA).....	29
4.	<i>Procés de creació de l'aplicació</i>	32
4.1.	Requeriments funcionals i no funcionals de l'aplicació de reconeixement facial	32
4.2.	Tecnologies usades	35
4.3.	Disseny, arquitectura del software i lògica	36
4.4.	Implementació de l'aplicació	39
4.5.	Notes addicionals	41
5.	<i>Procediment experimental</i>	43
5.1.	Verificació del sistema de reconeixement (Identificació i verificació de cares)	43

5.2.	Verificació de l'aplicació notarial.....	49
6.	<i>Conclusions</i>	55
7.	<i>Glossari</i>	58
8.	<i>Bibliografia</i>	59
9.	<i>Annexos</i>	61
9.1.	Codi Font.....	61

Llista de Figures

Figura 1. Etapes comunes pel procés de reconeixement facial.....	16
Figura 2. Filtres Haar [3]	17
Figura 3. Arquitectura de la P-Net, R-Net i O-Net [4].....	18
Figura 4. Funcionament de la xarxa proposada en [4].....	18
Figura 5. Plantilla HOG d'una cara estàndard i HOG generat a partir d'una fotografia de la cara de l'actor Will Ferrell [6].....	19
Figura 6. Resultat d'una cara alineada [6].....	20
Figura 7. Estructura d'un exemple d'ANN on es mostra les neurones que formen la capa d'entrada, les ocultes i la de sortida, i la interconnexió existent entre elles [9].....	23
Figura 8. Exemple d'estructura d'una ANN [10]	24
Figura 9. Corbes d'entrenament i verificació obtingudes a partir de l'error i el nombre d'èpoques [10]	26
Figura 10. Exemple d'un esquema d'una CNN destinada al processament d'imatges [1]	27
Figura 11. Face landmarks - punts característics d'una cara [6]	30
Figura 12. Sistema de reconeixement facial segons protocol d'entrenament o tasca d'avaluació [1].....	35
Figura 13. Simple diagrama de l'arquitectura de l'aplicació	37
Figura 14. Estructura de carpetes del projecte. Captura estreta de la interfície de l'IDE PyCharm.....	37
Figura 15. Verificació entre cares de la mateixa persona (figura masculina).....	44
Figura 16. Verificació entre cares de la mateixa persona (figura femenina).....	45
Figura 17. Verificació entre cares de diferents persones (figures masculines i femenines).....	45
Figura 18. Verificació entre cares de diferents persones amb relació familiar de germanes (figures femenines)	46

Figura 19. Verificació entre cares de diferents persones amb relació familiar de germanes bessones (figures femenines)	46
Figura 20. Verificació entre el DNI i el propietari del mateix, així com també amb diferents persones.....	47
Figura 21. Verificació entre DNI i cares de diferents persones amb relació familiar de germanes (figures femenines)	48
Figura 22. Vista de l'inici de l'aplicació notarial.....	49
Figura 23. Interfície de l'aplicació on es pot observar l'apartat de la videoconferència i el menú de les 3 funcionalitats diferents.....	49
Figura 24. Interfície de FR - Persona propietària del DNI	50
Figura 25. Interfície de FR - Persona no propietària del DNI.....	51
Figura 26. Interfície de CL - Exemple que conté dos test que han estat passats correctament i dos tests que no han estat passats correctament	52
Figura 27. Interfície de FR_DB - Exemple que conté el resultat d'una identificació correcte.....	53

1. Introducció

1.1. *Context i justificació del Treball*

L'any 2020 ha estat una any sacsejat, a nivell mundial, per una forta pandèmia produïda pel virus del SARS-CoV-2 (Covid-19), conegut arreu pel nom de Coronavirus. Aquesta pandèmia ha portat amb ella un seguit de dificultats evidents i, per consegüent, un conjunt de mesures adoptades per la majoria dels governs, com pot ser la limitació de desplaçament i la capacitat de lliure reunió. És per aquest motiu que totes les persones i moltes feines s'han hagut d'adaptar a aquestes mesures. Una de les feines, entre moltes altres, que s'ha hagut d'adaptar són les relacionades amb l'àmbit notarial. Una part important en les feines notariales es realitzen amb reunions entre el notari i el/s interessat/s. Així doncs, degut a les mesures adoptades, aquestes reunions s'han hagut de realitzar a través d'una videoconferència. Tanmateix, sorgeix un problema evident, i és que, abans de realitzar qualsevol operació, el notari ha de verificar que l'interessat sigui qui diu ser. Per aquest motiu, el primer pas que realitza el notari és la correcta verificació de l'individu. Tot i que és un procés trivial quan es realitza presencialment, la verificació a través de videoconferència queda dificultada i pot arribar a carregar problemes de frau, si aquesta no es realitza correctament. És per aquest motiu que s'han de preparar eines amb l'objectiu d'assistir el procés de verificació al notari i de mitigar qualsevol risc.

Durant els darrers anys, els sistemes de detecció i d'identificació facial han estat una realitat en constant evolució. Aquests sistemes han anat evolucionant fent ús de tècniques cada cop més innovadores per tal de millorar-ne la fiabilitat i, sobretot, l'encert. Els sistemes de detecció i d'identificació facial que aporten millors resultats són els que han estat implementats mitjançant tecnologia de *Machine learning*, més concretament, les que han estat implementades amb la branca del *Machine learning* anomenada: *Deep learning* (xarxes neuronals). A [1] es presenta la visió general de la identificació facial implementada amb el *Deep learning* i la seva corresponent evolució.

Aquest treball es focalitzarà en la realització de l'estudi del SoA (State-of-the-Art) de les xarxes neuronals i la identificació facial per tal de ser capaços de presentar una solució al problema presentat. Com a conseqüència, es pretén dissenyar i implementar una aplicació que realitzi la identificació facial entregant-nos el nivell de semblança, entre la fotografia del DNI (document nacional d'identitat) i una persona, durant una videotrucada a temps real.

Per a extreure'n un veredictes ajustat, es realitzaran un seguit de proves que ens ajudaran a verificar la fiabilitat i qualitat del sistema (pros i contres).

1.2. Objectius del Treball

Els objectius principals es presenten en forma de llista ordenada i de forma seqüencial, pel que fa al seu desenvolupament, i són els següents:

- 1- Revisió del SoA en reconeixement facial.
- 2- Revisió del SoA en xarxes neuronals.
- 3- Revisió del SoA en sistemes de reconeixement facial fent ús de xarxes neuronals.
- 4- Creació d'una aplicació destinada a la identificació facial entre una persona i la fotografia present al DNI (*objectiu principal*):
 - a. Recerca dels llenguatges de programació i tecnologies més adients per a la implementació.
 - b. Disseny de l'arquitectura del software.
 - c. Implementació de l'aplicació:
 - i. Reconeixement facial (verificació).
 - ii. Prova de vida (*check life*).
 - iii. Identificació facial mitjançant una DB (*Database*).
- 5- Verificació i validació del reconeixement facial i funcionament de l'aplicació.

1.3. Enfocament, mètode seguit i tasques a realitzar

Per tal d'implementar una aplicació capaç de realitzar el reconeixement facial entre la fotografia del DNI i la cara d'una persona, s'ha de realitzar un estudi sobre la tecnologia vigent. Primer de tot, es realitzarà un estudi sobre xarxes neuronals i els diferents sistemes d'identificació facial ja existents. També es buscarà informació sobre la integració d'ambdós mons i quines tecnologies/llenguatges han estat els més usats.

Un cop realitzada aquesta primera part, es procedirà a fer la part pràctica, on es desenvoluparà el sistema, en forma d'aplicació, que integri la detecció i identificació facial mitjançant xarxes neuronals. Es realitzarà, com a pas previ a la creació de l'aplicació, un seguit

de proves per verificar la fiabilitat de la tecnologia usada. També s'anirà testejant l'aplicació per veure com es comporta en un entorn real, per tal d'anar extraient les conclusions.

Simultàniament, es redactarà la memòria i les conclusions extretes durant la recerca, el desenvolupament i les proves de verificació i validació de l'aplicació.

1.4. Breu revisió de l'estat de l'art

A mode introductori, ja que la major part de la recerca es realitzarà durant les següents i pròximes etapes del projecte, l'estat de l'art pel que fa a biometria facial i xarxes neuronals és el següent:

- Pel que fa a detector de cares, els sistemes que han presentat millors resultats han estat implementats amb xarxes neuronals. Més concretament, amb CNN (*Convolutional Neural Network*).
- Pel que fa al procés de reconeixement, aquest mateix tipus de topologia de xarxes neuronals són els que han presentat millors resultats.
- Un dels llenguatges més usats és *Python*. Aquest llenguatge presenta una varietat de llibreries com per exemple; *TensorFlow*, *Keras*, *SKlearn* o *dlib* pel que fa a la implementació de les xarxes neuronals, i *OpenCV*, *Scikit-image* o *Numpy* pel que fa al processat i tractament de dades.

Durant els últims anys, aquests dos camps; el de la detecció i el del reconeixement d'objectes (com pot ser el cas de les cares), han estat força al dia pel que fa a estudi, recerca i evolució. Així doncs, la literatura és força extensa i treballada, i per tant, trobem un gran ventall d'opcions pel que fa a llibreries ja implementades.

1.5. Breu descripció dels altres capítols de la memòria

Durant el transcurs del segon i tercer capítol s'introduiran els conceptes teòrics dels sistemes de reconeixement facial, les xarxes neuronals i la seva evolució. També s'introduiran les implementacions més vigents de la integració d'ambdós mons. En el quart capítol s'exposaran els requeriments funcionals de l'aplicació i l'estudi de la tecnologia a usar per acabar dissenyant i implementant l'aplicació de reconeixement facial. En el cinquè capítol

s'exposaran la bateria de proves que s'ha dut a terme per realitzar la verificació i validació del sistema. Finalment, es presentaran els resultats i s'exposaran les conclusions extretes durant el transcurs del projecte.

2. Revisió i estudi de l'estat de l'art

Al llarg d'aquest capítol s'introduiran els conceptes de sistemes de reconeixement facial i les xarxes neuronals, exposant la seva evolució. Finalment, es presentarà la motivació en l'àmbit tecnològic sobre la unió de les dues matèries. A més, s'exposaran, a mode de repàs, les implementacions de reconeixement facial més vigents.

2.1. *Sistema de reconeixement facial*

La tasca de reconèixer cares és aparentment ben senzilla si qui la realitza és una persona. Els humans som capaços d'identificar les persones, sempre i quan siguin conegudes o se'n tingui alguna referència. Aquest procés el realitzem quasi a l'instant. De fet, els humans som capaços d'identificar persones sense necessitat de veure'ls-hi la cara. En general, els humans som molt bons identificant objectes. Si és cert que podem fallar identificant alguna cara, però en gran mesura (excloent-hi persones que pateixen alguna dolència) això ve condicionat per circumstàncies adverses com podrien ser; la distància entre l'observador i la cara a identificar, l'exposició a la llum, clima, obstacles com ulleres de sol, barrets, mocadors, etc. Tanmateix, i en termes generals, podem afirmar que els humans disposem d'una gran fiabilitat i precisió per identificar cares a temps real.

Per contra partida, per a una màquina el procés d'identificar una cara no és tasca senzilla i sobretot realitzar-ho a temps real. Una màquina destina molts recursos, suposant un gran esforç i cost computacional, el simple fet de realitzar un reconeixement facial. Tanmateix, la recerca de mètodes, tecnologies i algorismes per realitzar el reconeixement està en el punt de mira del camp tecnològic, degut a la quantitat d'aplicacions que aquests tipus de sistemes poden tenir a l'abast. Per exemple, trobem que la majoria de telèfons mòbils d'última generació incorporen el reconeixement facial per a desbloquejar la pantalla, o per identificar persones en fotografies, o per accedir a alguna aplicació o fins i tot per realitzar pagaments, entre d'altres. Així doncs, si el reconeixement facial dut a terme per una màquina serveix per aspectes tant crítics com realitzar un pagament (temes relacionats amb la seguretat), aquests han de comptar amb un nivell de fiabilitat i precisió molt elevat perquè no suposi un risc per l'usuari. Per tant, els sistemes de reconeixement facial, des de la seva aparició, han anat patint una evolució vertiginosa ja que han despertat un gran interès social.

A mode de controvèrsia, tot i que no serà un tòpic fermament a tractar durant el transcurs d'aquest treball, és cert que aquests sistemes porten intrínsecament un gran problema, i és que poden representar una vulneració en la privacitat de les persones. Aquest reconeixement el realitza una màquina, per tant, això significa que aquesta màquina està registrant la cara de la persona per a realitzar una posterior comparació sobre una DB, i aquestes alhora contindran un gran conjunt de dades que correspondran a fotografies i, per tant, a identitats. A més, aquests sistemes units amb sistemes de busca, com podria ser un sistema GPS, permetrien identificar on es troba una persona en concret i realitzar un seguiment a temps real, fet que suposaria, si no hi ha un consentiment explícit, una clara vulneració de la privacitat.

Deixant de banda els problemes amb la privacitat, podem remarcar que la quantitat de possibilitats pel que fa a aplicacions i serveis que ens ofereixen i ens oferiran els sistemes de reconeixement facial és tant gran que val molt la pena seguir destinant recursos humans treballant en aquest camp per tal de millorar-lo i fer-lo seguir creixent.

Ràpidament ens adonem dels principals avantatges que els sistemes de reconeixement facial (identificació i/o verificació) ens aportaran, tal i com indica [2]. Aquests, entre d'altres, són:

- És el mètode més natural pels humans i amb el que estem acostumats.
- Sistema d'ús molt senzill i poc intrusiu ja que l'usuari només ha de presentar la cara.
- Gran quantitat d'aplicacions de les quals es pot destacar:
 - o Sistemes de control d'accés (autenticació/verificació).
 - o Identificació de persones en fotografies o vídeos (identificació/reconeixement d'identitat).

En el següent apartat s'exposaran les diferents etapes que presenta un sistema de reconeixement facial.

2.1.1 Etapes d'un sistema de reconeixement facial

En aquest apartat es descriuran les etapes que realitza tot sistema de reconeixement facial actual. Aquests sistemes estan compostats principalment per tres destacades etapes. Aquestes són:

- Detecció de cares i posterior aïllament.
- Normalització de la imatge i alineament.
- Extracció de característiques.

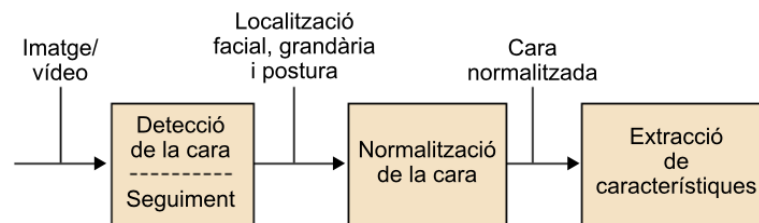


Figura 1. Etapes comunes del procés de reconeixement facial

Aquestes són totes les etapes que comparteixen les aplicacions que fan ús d'un sistema de reconeixement facial, ja sigui d'identificació o verificació. A partir d'aquí, depenent de l'aplicació, es realitzaria una acció o una altre. Per exemple, en un sistema d'identificació facial, es faria una comparativa de les característiques extretes al llarg d'una DB i en sortiria com a resultat, la identitat de la persona que hagués presentat una major semblança. Per altre banda, en un sistema de verificació, es faria una comparativa de les característiques extretes amb les que el dispositiu contingué. En cas de *match*, la verificació hauria estat exitosa i la identitat de l'individu validada. En les següents seccions s'exposaran les tres etapes en més detall i es farà una petita introducció a les etapes d'un sistema de reconeixement facial (identificació o verificació) introduint alguns dels mètodes més usats.

2.1.2 Detecció facial i aïllament

Correspon a la primera a etapa a realitzar. Els humans tenim la capacitat d'identificar tots tipus objecte a l'instant, però una màquina, per tal d'identificar la identitat d'una persona, primer n'ha de detectar la cara. Així doncs, aquesta fase consisteix en introduir una imatge com a input i obtenir una imatge amb la cara aïllada com a output. Aquest output serà l'input de la següent etapa de normalització.

Característiques Haar

Un dels sistemes més utilitzats des de la seva aparició i que va presentar excel·lents resultats va estar el proposat per Viola i Jones [3]. Aquest mètode consisteix en aplicar una successió de filtres Haar al llarg de tota la imatge per extreure'n les característiques Haar. Els resultats obtinguts durant l'aplicació dels filtres seran qui ens indicaran si en aquella posició de la imatge s'hi troba una cara o no. La Figura 2 mostra aquests filtres Haar. Notar que aquests poden variar les seves dimensions i formes per tal de tenir una variabilitat major durant el procés de detecció. Remarcar que aquest mètode ha quedat superat per altres tecnologies que s'exposaran a continuació.

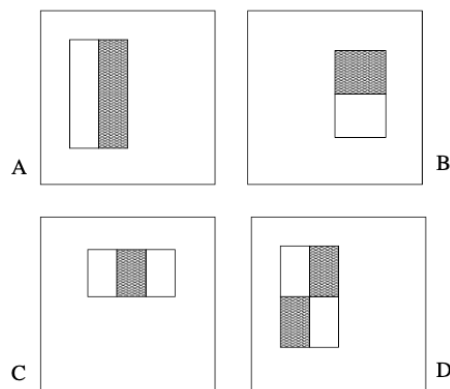


Figura 2. Filtres Haar [3]

Xarxes neuronals

Un enfocament més actual de la detecció facial és fent ús de xarxes neuronals tal i com indica [4] i [5]. Ambdós treballs obtenen resultats que els situa en el SoA. A mode de resum, els autors de [4] proposen una xarxa neuronal formada per 3 CNN en cascada. La primera xarxa, anomenada P-Net, s'encarrega de seleccionar possibles candidats. La segona xarxa, R-Net, s'encarrega d'eliminar molts falsos positius. La última xarxa, O-Net, és molt similar a R-Net, però en aquest cas ja és l'output de la xarxa i et retorna la cara detectada amb els punts de referència o *landmarks* (posicions dels ulls, nas i boca). Aquests punts de referència poden ser usats per a realitzar l'alineament de la imatge de la cara detectada. La Figura 3 mostra la topologia de cada una de les CNN connectades entre sí.

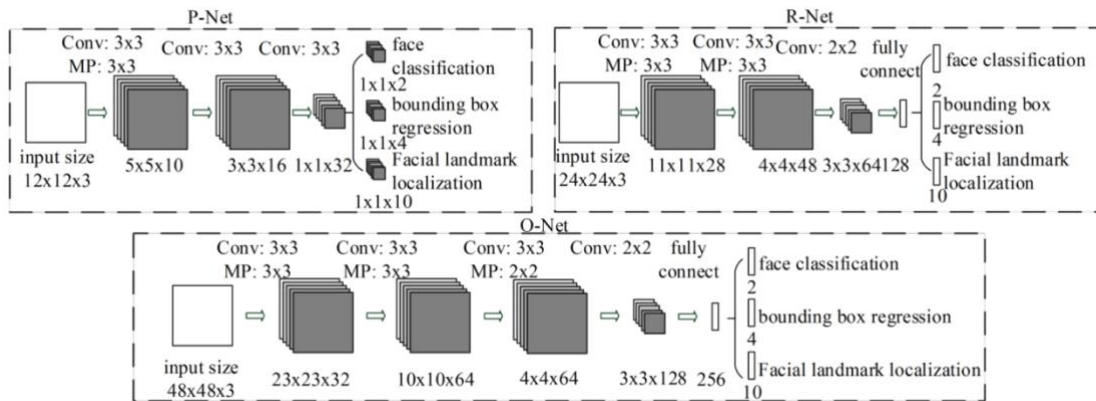


Figura 3. Arquitectura de la P-Net, R-Net i O-Net [4]

La Figura 4 mostra el funcionament de les diferents xarxes durant el procés de detecció facial.

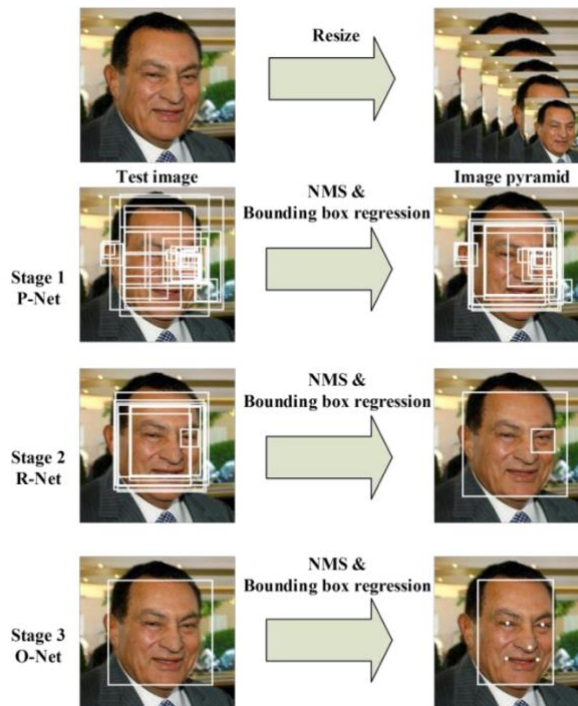


Figura 4. Funcionament de la xarxa proposada en [4]

Histogram of Oriented Gradients

Finalment, trobem un altre sistema també molt eficaç creat l'any 2005 anomenat HOG (*Histogram of Oriented Gradients*). Aquest sistema converteix la imatge a escala de grisos, i acte seguit mira cada píxel aïlladament i comprova quant fosc és aquest píxel en comparació als que té al voltant. Finalment, es crea una fletxa indicant cap a on es dirigeix la foscor de la

imatge. Així doncs, cada píxel es transforma en una fletxa. Aquestes s'anomenen *gradients*. Un cop es té creat el mapa amb totes les fletxes, per tal de no treballar amb tots els píxels, es creen finestres de subpíxels en les quals es calcula cap a on es dirigeixen la major part de les fletxes. Així doncs, la matriu amb les fletxes queda simplificada i haurem creat el HOG de la imatge. Finalment, el resultat es compara amb una plantilla que correspon a un HOG d'una cara estàndard. Si hi ha una zona en el HOG generat que obté una similitud molt alta amb la de la plantilla, es considerarà que la imatge analitzada conté una cara. La Figura 5 mostra la plantilla HOG i un exemple d'una imatge HOG generada a partir de cara de l'actor Will Ferrell.

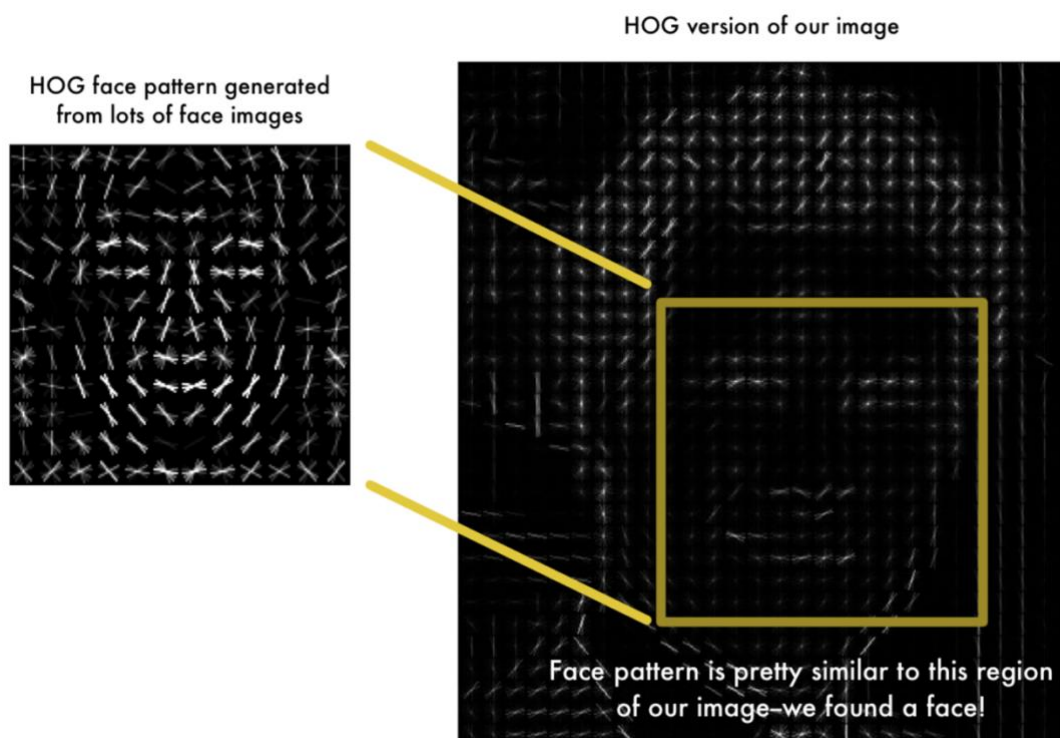


Figura 5. Plantilla HOG d'una cara estàndard i HOG generat a partir d'una fotografia de la cara de l'actor Will Ferrell [6]

2.1.3 Normalització i alineament

Aquesta etapa és usada per tal de normalitzar la imatge de la cara aïllada abans d'atacar a la següent etapa. Principalment, en aquesta etapa intermèdia, es reduirà la imatge a un nombre determinat de píxels, s'aplicaran tècniques per a normalitzar la intensitat (com podria ser el mètode d'equalització de l'histograma) i s'alinejarà la imatge de la cara. Aquest últim

procés s'aconsegueix rotant la imatge per tal que els ulls, el nas i la boca, quedin el més centrats possible. La Figura 6 mostra un exemple d'una cara alineada.

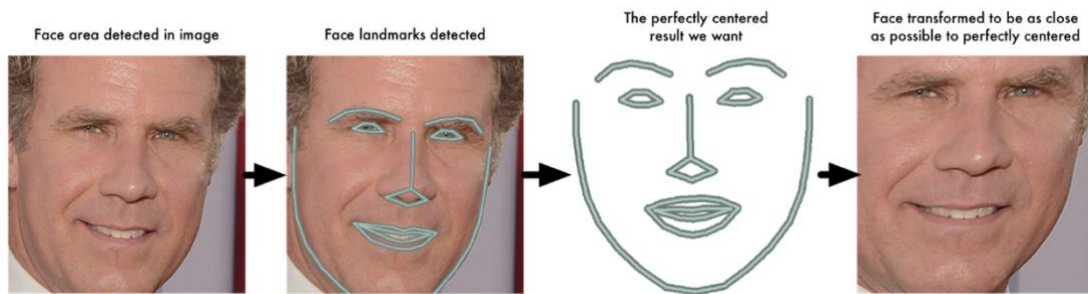


Figura 6. Resultat d'una cara alineada [6]

El procés d'alineament es pot aconseguir fàcilment després d'obtenir els punts de referència obtinguts durant el procés de detecció facial utilitzant la xarxa neuronal exposada a la secció 2.1.2 apartat Xarxes neuronals.

Remarcant que aquesta etapa pot ser opcional, però els resultats obtinguts en la etapa d'extracció de característiques es veurà minvat en quant a fiabilitat i precisió, en cas d'omissió de la mateixa.

2.1.4 Extracció de característiques

L'extracció de característiques és l'última etapa dins d'un sistema de reconeixement facial. Això no és del tot cert, ja que realment encara quedaria un últim pas i és el de fer el reconeixement facial en qüestió, ja sigui procés d'identificació o verificació de l'individu. Aquest últim procés, el de reconeixement facial, consistirà en fer una comparació de les dades extretes durant aquesta darrera etapa amb les dades guardades en el sistema o en una DB.

Entrant en matèria sobre l'etapa d'extracció de característiques trobem un gran ventall de mètodes existents que són usats per extreure les característiques facials més representatives o que seran usades com a representació de la cara analitzada [7]. Alguns d'aquests mètodes a destacar són: PCA (Anàlisi dels components principals) i l'ús de xarxes neuronals.

- **PCA:** El PCA és un mètode estadístic que serveix per simplificar la complexitat dels espais mostrals que presenten moltes dimensions obtenint així només les components principals de les dades multidimensionals. Tanmateix aquest mètode, tot i reduir la dimensió de les mostres, permet preservar la informació

en la seva totalitat o gran part d'ella. Per tant, el mètode PCA permet identificar i trobar patrons per tal de reduir la complexitat de les dades a analitzar i, aquests alhora, ens són útils per extreure les característiques facials amb una precisió elevada.

- **Xarxes neuronals:** Aquesta tècnica és actualment la més innovadora i la que està presentant millors resultats. Les xarxes neuronals, que solen ser CNN (xarxes neuronals convolucionals), són les encarregades d'extreure les característiques principals de la cara durant el processament de les dades mitjançant un conjunt de filtres que es troben al llarg de les diferents capes. Aquest mètode serà tractat més profundament a futures seccions.

2.1.5 Reconeixement (*Identificació i verificació facial*)

Un cop extretes les característiques facials principals, ja s'està preparat per a realitzar, en funció de l'ús o de l'aplicació, tant la identificació com la verificació facial. Aquest últim procés es realitza mitjançant tècniques de *matching*. Aquestes tècniques poden ser: **distància euclidiana**, **correlació (Coeficient de Pearson)** o **sistema de classificació** mitjançant xarxes neuronals, entre d'altres.

Independentment de la tècnica de *matching* usada, el sistema de reconeixement pot ser:

- **Identificació:** Es realitza una cerca al llarg de les dades guardades a la DB. Aquesta es realitza fent ús de qualsevol tècnica de *matching*. La identitat de la persona serà la que retorna el major resultat de similitud entre les característiques. Pot donar-se el cas que el sistema no identifiqui a ningú. En aquest cas, cap dels resultats ha superat un valor llindar de similitud del sistema.
- **Verificació:** Es realitza una comparació amb les característiques extretes amb les que es vol verificar. Si el resultat obtingut és inferior o superior al valor llindar (en funció del tipus de *matching*) definit en el sistema, es considerarà que d'identitat de la persona ha estat verificada. En qualsevol altre cas, no es considerarà la mateixa persona.

2.2. *Xarxes neuronals*

En aquesta secció s'introduirà i es presentarà, a mode de resum, el camp de les xarxes neuronals, així com els conceptes bàsics fent un repàs de les arquitectures més vigents.

2.2.1 *Introducció a les xarxes neuronals*

Les ANN (xarxes neuronals artificials) són una eina computacional que va ser creada per simular el comportament, l'aprenentatge i/o el processament del cervell humà. Per tant, una ANN forma part del món de les AI (intel·ligències artificials).

Una ANN està composta d'unes "neurons" interconnectades entre elles, essent aquestes les unitats mínimes de processament de la xarxa. Les dades flueixen al llarg de la xarxa a través d'aquestes "neurons" simulant els estímuls elèctrics que es produeixen en el cervell humà. Per tant, les ANN processen un input (un conjunt de dades) i retornen un output. El flux (processament) que seguiran les dades dins de l'ANN dependrà d'uns paràmetres, anomenats pesos, i funcions que definiran la xarxa. Aquestes funcions dependran de l'aplicació que se li vulgui donar i de la pròpia topologia de la xarxa.

A la següent secció es presentaran les aplicacions principals.

2.2.2 *Aplicacions*

Les ANN són usades principalment en el camp de la robòtica i de l'enginyeria. Per exemple, un dels camps on ha tingut més rellevància ha estat en els cotxes autònoms. Les ANN destinades al món automobilístic han estat entrenades per dirigir el vehicle en qüestió. Això involucra fer una verificació i seguiment exhaustiu de l'entorn, com identificar vianants, senyals, línies de carretera, vehicles, etc. També trobem altres aplicacions en les ANN com poden ser les de resoldre problemes de classificació o aproximació de funcions.

En qualsevol cas, i el que fa d'aquesta tecnologia quelcom molt interessant i innovador i del que es diferencia dels programes seqüencials (procediment ofimàtic seqüencial), és que les ANN pateixen un aprenentatge a partir d'exemples, i després són capaços de resoldre el problema sense haver hagut de definir unes regles per a resoldre'l. A grans trets, la pròpia xarxa és capaç de resoldre el problema per ella mateixa a partir d'un aprenentatge.

Notar doncs que quan una ANN ha estat entrenada per resoldre un problema, aquesta només serà capaç de resoldre aquest problema en concret. Per tant, per a cada aplicació caldrà una xarxa diferent. A la següent secció s'introduirà el funcionament de les ANN i conceptes com l'aprenentatge d'una xarxa.

2.2.3 Estructura, funcionament i topologies de les ANN

Totes les ANN processen un input i retornen un output. Tanmateix, aquest output no cobra cap sentit si aquesta no sap quin és el problema a resoldre. Abans però, s'introduiran els conceptes bàsics que formen l'estructura d'una ANN [8], [9] i [10].

Tota ANN està formada per un conjunt de neurones artificials formant capes. Aquestes es troben interconnectades entre elles. Trobem:

- **Capa d'entrada:** Són l'input de la xarxa.
- **Capa/es oculta/es:** Són les que processen i assimilen les dades.
- **Capa de sortida:** Són l'output de la xarxa.

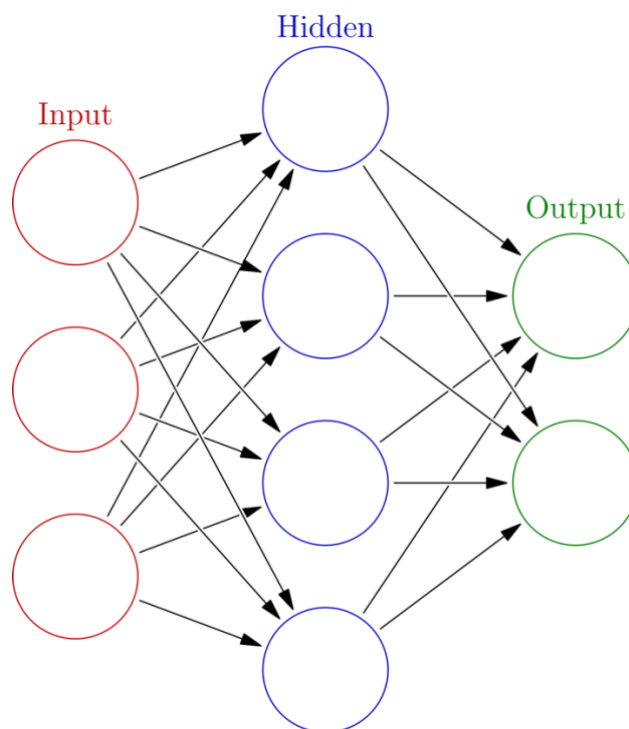


Figura 7. Estructura d'un exemple d'ANN on es mostra les neurones que formen la capa d'entrada, les ocultes i la de sortida, i la interconnexió existent entre elles [9]

El conjunt de neurones connectades entre elles formaran la ANN. La interconnexió de cada neurona vindrà definida per uns pesos (normalment expressat en la literatura com a ; w). Les neurones es poden agrupar en capes o nivells. A més, les neurones individualment també reben un input, el processen i retornen un output. La sortida de la neurona quedarà definida per:

- **Funció de propagació o excitació:** generalment consisteix en el sumatori de cada entrada multiplicada pel pes de la interconnexió entre la neurona anterior i l'actual.
- **Funció d'activació:** aquesta modifica l'anterior si és que existeix.
- **Funció de transferència:** és usada per limitar la sortida de la neurona. Algunes de les funcions més usades és la *sigmode* (valors compresos entre 0 i 1, ambdós inclosos) o la *tangent hiperbòlica* (valors compresos entre -1 i 1, ambdós inclosos), entre d'altres.

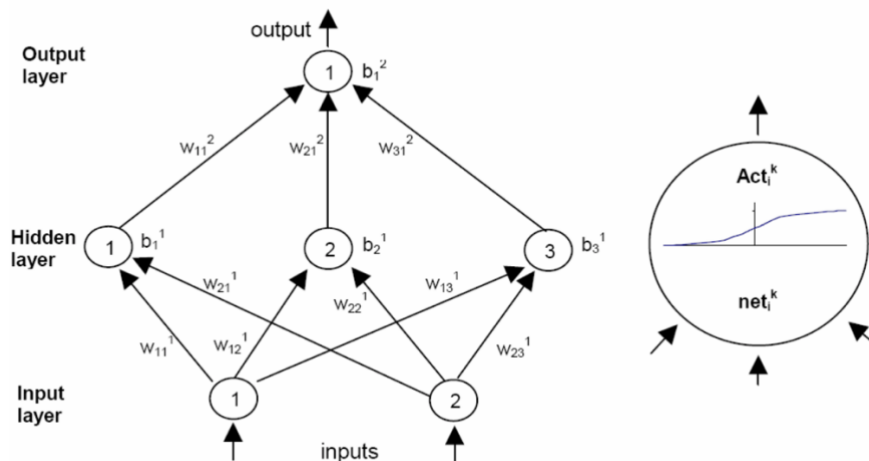


Figura 8. Exemple d'estructura d'una ANN [10]

L'estructura i topologia d'una ANN pot ser molt variada i dependrà de l'aplicació o el problema a resoldre. En qualsevol cas, no és tasca senzilla escollir la millor topologia per una aplicació en concret. Moltes vegades caldrà realitzar un procés de prova/error per tal d'escollir aquella topologia que ens sembli més adequada o que presenti millors resultats en la resolució del problema. El nombre de paràmetres d'entrada determinarà les dimensions espacials de la xarxa, i els paràmetres de sortida determinarà el nombre de solucions generades. A mode de resum, escollir una topologia o una estructura de xarxa, com pot ser les funcions d'activació, el nombre de capes ocultes, el tipus de xarxa, etc, quedarà determinat

per un procés de prova/error. Tanmateix, s'ha de tenir en compte que com més complexa sigui la xarxa (més gran significa més paràmetres a tractar) més cost computacional. A més, també implica una major dificultat en saber què està passant o com ha après a solucionar el problema.

Finalment, i per tal que la xarxa pugui resoldre un problema, aquesta ha de patir un aprenentatge. Aquest aprenentatge és el procés que fa variar els valors de les interconnexions, és a dir, dels pesos. Hi ha dos tipus d'aprenentatges:

- **Supervisat:** Aquest procés consisteix en entrenar la xarxa amb unes dades d'entrada i unes dades de sortida associades a les d'entrada. És a dir, es fa passar les dades per la xarxa i s'observa l'output. Acte seguit, se li indica quin era el resultat esperat, i la xarxa modifica els pesos a través de minimitzar una funció d'error. Es fa ús de tècniques de retropropagació per anar modificant tots els pesos de la xarxa. Aquest procés es va realitzant amb dades destinades a l'entrenament. Finalment, un cop s'ha assolit un valor d'error mínim, es considera que la xarxa ha estat entrenada. A partir d'aquest moment, la xarxa ja és funcional i els pesos queden fixats. Poden sorgir errors d'aprenentatge com l'*over-fitting*, en que la xarxa obté resultats molt bons amb les dades d'entrenament, però mals resultats en la fase de producció, on tracta amb dades mai vistes per la xarxa. En aquest cas, es considera que la xarxa ha après a resoldre el problema de memòria.
- **No supervisat:** En aquest tipus d'aprenentatges no es proporcionen exemples del comportament desitjat. És la pròpia xarxa qui va adaptant els pesos de les connexions.

Notar que, pel bon funcionament d'una ANN en una aplicació en concret, fa falta un bon aprenentatge, i aquest alhora, recau sobre la qualitat de les dades d'entrenament. Això significa que si no es disposa d'un bon data set de dades per a la fase d'entrenament, la xarxa final no oferirà grans resultats. A més, i com a punt a destacar, és important que les dades usades durant l'entrenament no siguin usades durant la fase de producció.

Durant el procés d'aprenentatge, es fa ús d'un altre grup de dades amb el propòsit de verificar l'aprenentatge de la xarxa. Així doncs, aquest nou set de dades ens permetran detectar l'aparició de l'*over-fitting* o el moment en el que la xarxa ha obtingut millors resultats.

La Figura 9 mostra la corba característica d'aprenentatge, per norma general, d'una ANN.

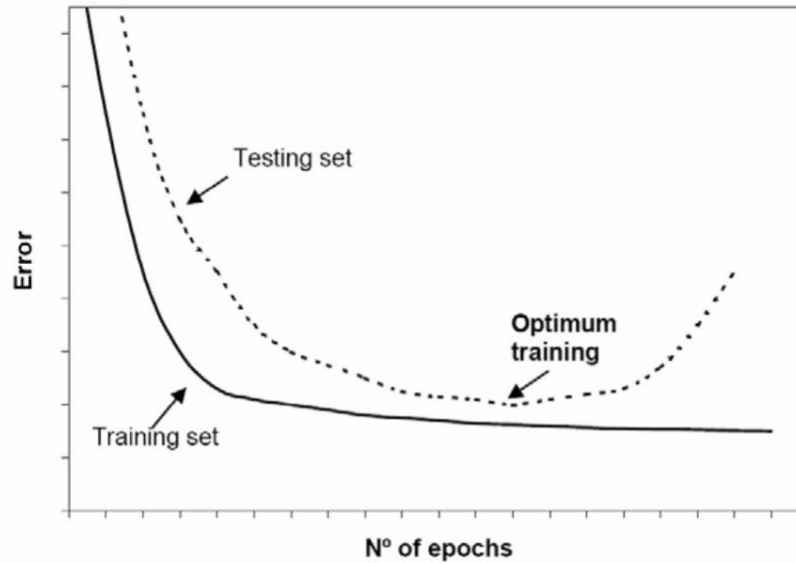


Figura 9. Corbes d'entrenament i verificació obtingudes a partir de l'error i el nombre d'èpoques [10]

La Figura 9 introdueix el concepte d'època. Es considera que ha passat una època en el moment en el qual el data set d'entrenament ha passat de manera completa per la xarxa. Aquest procés es va repetint per tal d'anar millorant el rendiment de la xarxa, i per tant, minimitzant l'error degut a la modificació del valor dels pesos. A la mateixa figura podem observar un parell de corbes. La corba contínua representa l'error obtingut en funció del nombre d'èpoques pel data set d'entrenament. En canvi, la corba discontinua representa el data set de verificació o *testing*. Podem observar com, i tenint en compte que la figura és un mer exemple, l'error pel data set d'entrenament sempre va disminuint. Tanmateix, això no passa en el cas del data set de verificació. En aquest cas, arriba un punt on l'error torna a créixer. Aquest punt és on comença el fenomen de l'*over-fitting*. Així doncs, el punt més baix o l'error mínim presentat per la corba de verificació serà el moment en el qual la xarxa presenta millors resultats.

Com s'ha mencionat anteriorment, hi ha diferents tipus de xarxes neuronals. Aquestes són, entre d'altres:

- **Perceptró multicapa (MLP):** Aquestes xarxes estan formades per una capa d'entrada, un nombre a determinar de capes ocultes i una capa de sortida. Cada capa conté un nombre a escollir de neurones. I totes les neurones estan connectades amb les veïnes de la capa anterior i la posterior. El mètode més comú d'aprenentatge és el de retropropagació, on es pretén minimitzar l'error

modificant els pesos de les connexions. Aquests tipus de xarxes són especialment útils per resoldre problemes de classificació. La Figura 7 mostra un exemple de MLP amb una sola capa oculta.

- **Xarxes neuronals convolucional (CNN):** Són un tipus de xarxes neuronals de complexitat diferent a les MLP. En aquest cas, aquest tipus de xarxa intenta distingir característiques o extreure patrons que presenten les dades a través d'aplicar un tipus de càlculs matemàtics anomenats operacions convolucional. Aquestes operacions fan la funció de filtre que es va propagant al llarg de les dades a processar. La successió d'aplicació de filtres al llarg de les diferents capes és qui permetrà anar extraient diferents patrons. Alhora, els paràmetres que conformen els filtres seran els que aniran canviant durant la fase d'aprenentatge. Per aquest motiu, aquests tipus de xarxes són especialment útils per sistemes de reconeixement d'imatges o processament de llenguatges, ja que s'aprofiten de la capacitat d'extracció de característiques. La Figura 10 mostra un exemple d'un esquema d'una CNN destinada al tractament d'imatges.

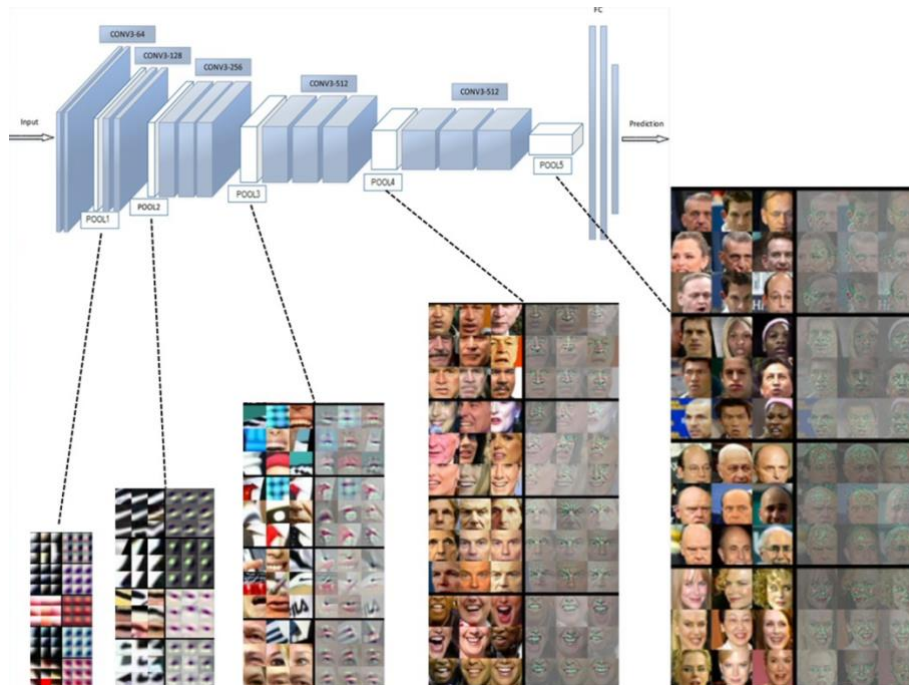


Figura 10. Exemple d'un esquema d'una CNN destinada al processament d'imatges [1]

Podem observar a l'estructura de capes presentada a la Figura 10, com aquestes cada cop es van fent més petites però gruixudes. Normalment, les CNN van extraient mapes de

característiques i aquests alhora són l'input de les següents capes. Així doncs, la quantitat de mapes extrets és major per cada capa successiva. Tanmateix, situem al capdamunt de cada nova capa un tipus de capa anomenada capes de *pooling*. Aquestes capes redueixen la dimensió de la matriu. Per aquests motius, observem com l'estructura característica d'una CNN és la de, per cada nova capa, créixer en volum però reduir en dimensió (de matriu).

Es pot notar ràpidament que, una clara diferència entre una MLP i un CNN, entre moltes altres, és que l'input i el processament de les dades a través d'una MLP sempre serà un vector pla, i en una CNN pot ser una matriu.

3. Reconeixement facial mitjançant xarxes neuronals

En el capítol 2 s'han introduït els conceptes generals dels sistemes de reconeixement i de les xarxes neuronals. A més, s'ha pogut notar com ambdós mons es poden nodrir, complementar i millorar. De fet, els sistemes de reconeixement facial que presenten millors resultats, fins a la data d'escriptura i revisió d'aquest treball, han estat implementats amb xarxes neuronals, més concretament amb ANN, que compten amb arquitectures de CNN. Al llarg d'aquest tercer capítol es citaran algunes de les ANN més actuals destinades al sistema de reconeixement facial, i es presentarà una de les implementacions de sistema de reconeixement facial mitjançant xarxes neuronals fent ús d'imatges en 2D. Aquesta implementació compta amb uns resultats obtinguts que es troben en el SoA d'aquest camp.

3.1. Implementacions més vigents (SoA)

Mei Wang *et al.* [1] fa un resum del SoA dels sistemes de reconeixement facial mitjançant xarxes neuronals des del seu inici. Algunes de les implementacions contemplades en el *paper* són; DeepID2 [11] o FaceNet (2015 *Google Inc.* Xarxa entrenada amb més de 500M de fotografies) [12], entre d'altres. Ambdós implementacions obtenen resultats sorprenents que superen el 99% d'*accuracy* i estan implementats amb arquitectura CNN. A més, usen un mètode d'aprenentatge mètric basat en distàncies euclidianes. Aquests tipus de mètodes representen les imatges en un espai euclidià on es comprimeix la intravariància i en canvi s'amplia la intervàriància (destacant el *clustering*). Tal i com cita [1], les funcions de pèrdua més usades d'aquest sistema d'aprenentatge són; *contrastive loss*, com és el cas de DeepID2, i *triplet loss*, cas de FaceNet.

En aquest apartat es presentarà un sistema de reconeixement facial que fa ús de la xarxa FaceNet per extreure les característiques facials principals, ja que una implementació d'aquesta serà la usada pel desenvolupament del nucli de l'aplicació a realitzar durant aquest treball. La implementació va a càrrec de l'autor Adam Geitgey amb la llibreria *face_recognition*¹ escrita en llenguatge *Python*. A [6] trobem un article escrit pel propi autor explicant la

¹ Repositori al portal *GitHub* a https://github.com/ageitgey/face_recognition i documentació a https://face-recognition.readthedocs.io/en/latest/face_recognition.html

implementació del sistema de reconeixement facial. A continuació s'exposaran els punts claus.

Un sistema de reconeixement facial, tal i com s'ha explicat en el **capítol 2** secció 2.1, consta de principalment 3 destacades etapes; detecció de cares, normalització i alineament i extracció de característiques. Finalment, quedaria una última etapa que consistiria en acabar realitzant el reconeixement facial, ja es tracti d'identificar o de verificar.

La implementació a tractar fa ús del mètode HOG per detectar les cares a les imatges. Sí és cert que l'extracció/detecció de les cares podria estar implementada amb xarxes neuronals, però el mètode HOG és força ràpid i precís. La Figura 5 mostra un exemple del procés de HOG. Un cop detectades les cares, s'han d'aïllar per poder realitzar l'alineament. Per realitzar l'alineament es fa ús d'una xarxa neuronal que s'encarrega d'extreure 68 punts que apareixen en qualsevol cara (ulls, boca, nas, etc). La Figura 11 mostra aquests punts característics.



Figura 11. Face landmarks - punts característics d'una cara [6]

Notar que aquests punts són retornats en forma de coordenades i es poden identificar en la imatge de la cara aïllada processada.

Aquests punts són usats per tal de centrar i orientar la cara d'una manera frontal. La Figura 6 mostra un exemple del procés d'alineament d'una imatge que hi apareix una cara que no està inicialment orientada frontalment. El fet d'alinejar la cara resulta en un nivell de precisió molt més elevat.

Un cop realitzat l'alineament, només queda realitzar l'extracció de les característiques que representin la cara a analitzar. En aquest cas, l'extracció és realitzada, com s'ha mencionat anteriorment, per una xarxa neuronal d'arquitectura CNN. Aquesta s'encarrega d'extreure 128 mesures de la cara. Així doncs, l'output de la CNN serà un vector de 128 dimensions/*features*. La xarxa neuronal és capaç d'extreure uns vectors amb distàncies euclidianes curtes per cares que són similars i distàncies llargues si no es tracta de cares similars.

Per tal d'extreure aquest vector, la xarxa ha estat entrenada fent ús del mètode anomenat *triplet loss*. Aquest mètode consisteix en passar-li dues cares que corresponen a la mateixa persona i una tercera imatge d'una cara que no correspon a l'individu de les altres dues fotografies. Acte seguit, la xarxa s'encarrega de realitzar les millores pertinents que obtinguin com a resultat que els vectors generats de les dues cares es trobin més pròxims i en canvi que el del tercer s'allunyi. Aquest procés és realitzat fent ús de milions de cares durant l'entrenament, fins que la xarxa és capaç de generar aquest vector de 128 mesures on es destaca que les distàncies més curtes es produeixen d'imatges de cares de la mateixa persona.

Aquesta tècnica d'extracció de característiques és de gran utilitat ja que d'una imatge en cru se n'extreu un vector de 128 característiques que alhora representen ser les característiques principals de la cara amb gran *accuracy* i fiabilitat. Així doncs, quan es tracta de realitzar una cerca en una DB, sempre serà més ràpid i de cost computacional molt menor, el simple fet de realitzar una simple comparació entre variables de poques dimensions (e.g. distància euclidiana entre dos vectors de 128 dimensions).

A partir d'aquí, segons quina sigui l'aplicació, és podrà realitzar una verificació o una identificació de la persona. En el cas d'identificador, es pot fer ús d'una altre xarxa neuronal classificadora que s'haurà entrenat amb un conjunt de personalitats a identificar.

4. Procés de creació de l'aplicació

En aquest quart capítol es definirà l'abast de l'aplicació, citant quins són els requeriments que es contemplen, i quines són les tecnologies usades pel disseny i la seva implementació.

4.1. Requeriments funcionals i no funcionals de l'aplicació de reconeixement facial

A mode introductori, l'any de realització d'aquest projecte (2020) ha estat marcat per la pandèmia produïda pel virus SARS-CoV-2. El virus ha sacsejat la societat i la normalitat a la que estàvem acostumats ja que, per tal de frenar la propagació massiva del virus, els governs estatals han hagut d'implementar un seguit de mesures dràstiques de les quals cal destacar, degut a l'impacte econòmic i social, el fet de restringir la lliure mobilitat. Així doncs, moltes de les feines s'han hagut d'adaptar, realitzar-se de manera telemàtica. Unes de les feines afectades han estat les relacionades amb l'àmbit notarial, on és imprescindible realitzar una reunió entre el notari i el/els interessat/s. El virus i les mesures preses han impossibilitat el fet de realitzar aquestes reunions de manera presencial. Així doncs, aquestes reunions s'han hagut de conduir a través d'una videoconferència. Aquest fet dificulta un dels principals requeriments dels notaris abans de procedir amb el tòpic de la reunió, i és el de verificar que la persona amb la qui va a tractar és qui diu ser. Aquesta verificació realitzada de manera presencial és relativament senzilla, però realitzar-la telemàticament presenta algunes dificultats. Així doncs, aquest projecte pretén crear una aplicació que doni assistència al notari durant tot el procés de verificació de la identitat de la persona. Al llarg d'aquesta secció s'exposaran els requeriments tan funcionals com no funcionals de l'aplicació a desenvolupar.

L'aplicació haurà d'establir una videoconferència via *Jitsi*, entre l'usuari (notari), que farà ús de l'aplicació, i el conferenciant (interessat), la persona que està sotmesa al reconeixement facial. El programa haurà de satisfer amb 3 funcionalitats principals:

- 1- **Verificació de la cara de persona amb la foto del DNI:** S'indicarà a l'aplicació quan capturar i processar un seguit de *frames* adquirits a temps real a través de la videoconferència. Acte seguit, el conferenciant haurà de presentar la cara i el DNI

davant de la càmera. El processament consistirà en realitzar les diferents etapes que componen un sistema de reconeixement facial. Aquestes seran les següents:

- a. Detecció i extracció de les cares (s'imposarà un màxim de dues cares a detectar per tal de processar el *frame*).
- b. Normalització i alineament de les cares.
- c. Extracció de característiques d'ambdues cares.
- d. Resultat de similitud entre les característiques avaluades.

Es mostraran per pantalla a l'usuari, tots els resultats generats; similitud instantània, similitud acumulada i millor resultat de similitud obtingut (resultats presentats en %).

Com a requeriment principal es demana que es pugui escollir un valor llindar que permeti discernir de si es tracta de la mateixa persona o no.

- 2- **Verificació de vida:** L'aplicació haurà de disposar d'un seguit de tests que permetin verificar si la persona que es troba davant de la pantalla es tracta d'una persona viva, i, per exemple, no pas d'un vídeo. Per a realitzar aquesta bateria de proves, l'usuari demanarà al conferenciant que realitzi un seguit de moviments amb el cap. El ventall de proves a conduir consistiran en moure el cap:

- a. Cap a l'esquerra,
- b. Cap a la dreta,
- c. Cap amunt i
- d. Cap avall

El test a realitzar ha de ser escollit per l'usuari, no pel conferenciant. Per tant, la bateria de tests no s'ha de realitzar en cap ordre determinat. Finalment, el resultat del test ha de ser consistent amb els moviments realitzats pel conferenciant.

- 3- **Identificació facial a través d'una DB:** L'aplicació haurà de capturar un *frame* de la videotrucada i processar-lo. El processament consistirà en realitzar les diferents etapes que componen un sistema de reconeixement facial. Aquestes seran les següents:

- a. Detecció i extracció de la cara (s'imposarà un màxim d'una cara a detectar per tal de processar el *frame*).
- b. Normalització, alineament de les cares.
- c. Extracció de característiques de la cara detectada.

d. Comparació amb tota la DB.

No es farà ús de sistema de classificació (e.g. mitjançant xarxes neuronals) ja que només es disposarà d'una fotografia facial per cada individu. S'haurà de ser capaç de definir un valor llindar que discerneixi entre un resultat favorable (es tracta de la mateixa persona) i un resultat desfavorable (no es tracta de la mateixa persona diferent) abans de retornar la identitat del millor resultat obtingut, al llarg de la cerca, juntament amb el nivell de garantia o de similitud en %.

Cal remarcar que definir un valor llindar de similitud és una tasca força complexa ja que es requereix d'una gran bateria de proves i verificacions fent ús d'un gran conjunt de mostres, per tal d'ajustar el valor escollit i que aquest alhora sigui el més adequat pel nostre sistema de reconeixement facial.

A més, i com a punt interessant a tractar, adonar-se que el sistema solament utilitza una cara i la fotografia del DNI o de la DB per a realitzar la verificació/identificació. És a dir, la majoria de sistemes de reconeixement facial disposen d'un conjunt d'imatges de la cara d'una mateixa persona. Aquestes imatges consten de diferents posicions, exposicions, etc. I, depenent del sistema, si la verificació ha quedat marcada com a correcte, la imatge de la cara capturada pot ser utilitzada com una mostra més o com a entrenament del sistema, per tal de fer-lo cada cop més robust, acurat i precís (aprenentatge autònom). En el cas de la nostre aplicació, ens trobem davant d'una verificació entre dues cares i se'ns demana extreure'n, com a resultat, la similitud entre ambdues mostres facials amb un nivell de garantia que ens permeti discernir entre *match* (es tracta de la mateixa persona) o *no-match* (no es tracta de la mateixa persona). Aquest fet aporta, a priori, grans dificultats que calen ser resoltes realitzant una verificació del sistema i determinar si aquest requeriment en concret, ha quedat satisfet. La Figura 12, extreta de [1], mostra els diferents sistemes de reconeixement facial segons la tasca d'avaluació.

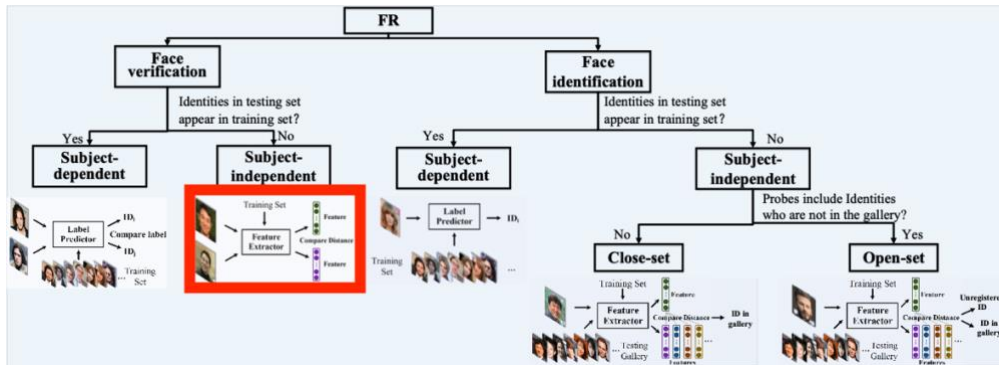


Figura 12. Sistema de reconeixement facial segons protocol d'entrenament o tasca d'avaluació [1]

A la Figura 12 queda remarcad en vermell el cas a tractar, tant pel que fa la verificació com la identificació sobre la DB.

Finalment, les funcionalitats hauran d'implementar una funció de *reset* que es podrà excitar a través de la interfície mitjançant un botó, o que s'activarà sol si algun punt del programa ha funcionat malament al llançar, per exemple, alguna excepció degut a un error de processament intern.

Per altre banda, es demana que l'aplicació estigui construïda mitjançant una estructura de client-servidor per tal de poder desplegar el servei de manera centralitzada. El codi ha d'estar escrit mitjançant classes per tal de tenir un codi modular i, per tant, de fàcil manteniment. A més, les classes, els mètodes i les funcions hauran d'estar comentades.

4.2. *Tecnologies usades*

Per tal de satisfer amb un dels requeriments establerts, l'aplicació estarà construïda fent ús d'una arquitectura de client-servidor. S'ha optat per fer ús d'aquesta estructura per tal de ser possible centralitzar el processat a un servidor especialitzat. Així doncs, com que es fa ús de xarxes neuronals, es pot optimitzar molt la velocitat de processament si es fa ús d'una GPU destinada a realitzar tots els càlculs. A més, gràcies a aquesta estructura, l'aplicació es pot posar en marxa a qualsevol navegador. Per tant, l'ús de l'aplicació serà independent al sistema operatiu del qual es disposi.

S'ha escollit el llenguatge de programació *Python* (v.3.8.3) per implementar el servidor ja que aquest ofereix llibreries com *Flask* i *Django* que permeten un fàcil desplegament i gestió d'un servidor. En aquest cas, es farà ús de *Flask* (v.1.1.2), ja que ens proporcionarà un servidor de test funcional fàcil de desplegar i suficient per l'aplicació que es vol implementar.

Pel que fa al tractament del reconeixement facial amb xarxes neuronals es farà ús de la llibreria *face_recognition* (v.1.3.0) exposada a la secció 3.1. Per tant, aquesta llibreria serà l'encarregada de detectar, alinear i extreure les característiques principals de les cares a analitzar. De fet, aquesta llibreria compta amb un petit conjunt de funcions que realitzen tot el procés en qüestió, facilitant-ne l'ús de la mateixa.

La DB estarà creada i gestionada per la llibreria *sqlite3* inclosa entre les *built-in* de *Python*.

El client constarà d'una “*single-page application*”. És a dir, totes tres funcionalitats estaran agrupades en un sol fitxer *HTML*. Es farà ús de; *Javascript* per controlar les peticions i la gestió de les respostes. Totes les peticions es faran a través de la tecnologia *AJAX* per fer-les lleugeres i transparents a l'usuari. A més, es farà ús del *framework Bootstrap* i *CSS* per dotar l'aplicació d'un cert estil visual.

Es farà ús de *Jitsi* per a realitzar la videoconferència, ja que aquest proporciona una API per interactuar i configurar els paràmetres de la videotrucada, entre d'altres aspectes, fent ús del llenguatge de programació *Javascript*. La documentació, per establir una videoconferència a través d'una aplicació web, pot ser consultada mitjançant el següent enllaç en l'apartat “*IFrame APP*”: <https://jitsi.github.io/handbook/docs/dev-guide/dev-guide-iframe>.

S'usarà l'IDE *PyCharm* pel desenvolupament de l'aplicació ja que aquesta eina ens facilita molt la tasca d'importar llibreries, de la depuració, d'escriure el codi i de mantenir un ordre durant el transcurs del projecte gràcies al control de versions que incorpora amb *Git*.

4.3. Disseny, arquitectura del software i lògica

Com s'ha comentat anteriorment, l'aplicació estarà estructurada seguint una arquitectura de client-servidor. L'estructura es mostra a la Figura 13.

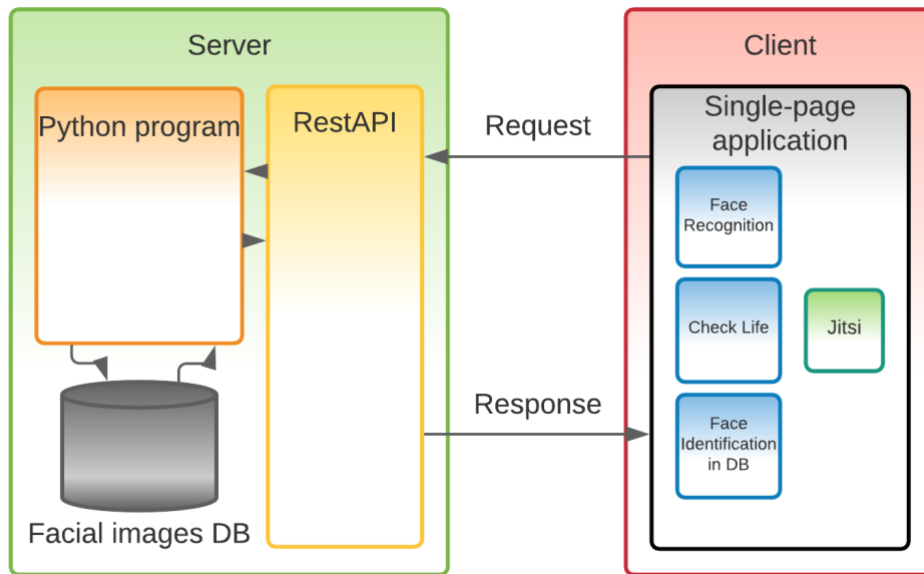


Figura 13. Simple diagrama de l'arquitectura de l'aplicació

La diferenciació client-servidor quedarà reflectida en l'estructura de carpetes que contindrà els diferents arxius que formaran el projecte. L'estructura de carpetes es pot observar a la Figura 14.

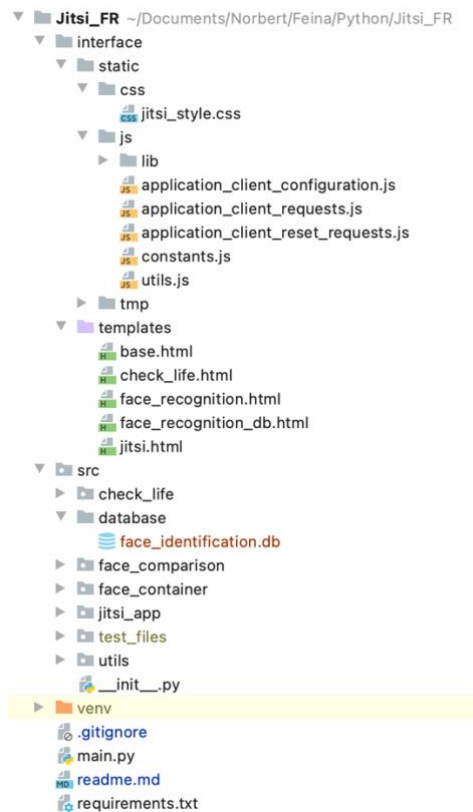


Figura 14. Estructura de carpetes del projecte. Captura extreta de la interfície de l'IDE PyCharm

Es pot observar en la Figura 14 com s'ha separat la part de la interfície gràfica (Client) amb la part del codi font (Servidor).

El fitxer `main.py` (*RestAPI*) és qui posa en marxa el servidor, accepta i respon les peticions realitzades a través del client. Així doncs, notar que el programa s'ha de posar en marxa en aquest punt.

S'ha fet un disseny i implementació per classes a fi d'aconseguir un codi el més modular possible i així poder facilitar-ne la depuració i el manteniment. A fi d'aconseguir la màxima portabilitat i facilitat a l'hora de mantenir o reutilitzar el codi, s'han comentat totes les classes i els seus mètodes.

Tal i com es pot observar a la Figura 14, s'ha creat un total de cinc classes; dues de les quals implementen les funcionalitats de cada una de les funcionalitat de l'aplicació, una que fa de "container" i és heretada per les altres dues classes anteriorment citades, una altre que simplement conté alguns mètodes útils per a les classes i finalment una altre conté les constants i lògica de l'aplicació.

En aquest cas, no és necessari mantenir una lògica per al correcte funcionament de l'aplicació, ja que tots els processos poden ser realitzats sense necessitat de seguir una màquina d'estats. Així doncs, la poca lògica del programa o les accions que es poden dur a terme són gestionades a través del propi client.

Principalment, l'aplicació funciona de la següent manera:

- 1- En cas de disposar del servidor (veure secció 4.5 per realitzar la instal·lació): S'inicia el programa executant el fitxer `main.py`. Aquest posa en marxa el servidor i obra un port on es serveix l'aplicació. Per defecte, si es fa ús del servidor que proporciona *Flask* i no es realitza cap tipus de modificació, el servei s'executa a <http://localhost:5000> o <http://127.0.0.1:5000/>.
- 2- Des del navegador s'obre l'aplicació del client. Acte seguit, és possible establir una videoconferència a través del servei de *Jitsi meet*, simplement fent clic a l'únic botó que proporciona la pròpia interfície de l'aplicació.
- 3- En aquest moment, l'usuari disposa de les 3 funcionalitats:
 - a. Verificació de la foto del DNI amb la cara de la persona.

- b. Verificació de vida.
 - c. Identificació facial a través d'una cerca a la DB.
- 4- La interacció que fa l'usuari amb l'aplicació és a través de botons, i aquests alhora realitzaran una *request* al servidor. Totes les *requests*, a excepció de les *requests* de *reset*, són enviades amb una captura o un conjunt de *frames* de la videotrucada. Aquestes són processades pel servidor. Un cop realitzats els càlculs i els processats pertinents, s'envia la resposta al client.
 - 5- El client verifica la resposta i plasma els resultats modificant els camps de la interfície.
 - 6- Es torna al punt 3.

Remarcar que és el servidor qui processa totes les dades. D'aquesta manera, si es centralitza el servidor en un ordinador potent, les peticions es podran gestionar més ràpidament. Notar que el processat és força costós degut al processat d'imatges i al desplegament de les xarxes neuronals.

4.4. Implementació de l'aplicació

La *RestAPI* que uneix el client i el servidor, gestionant-ne totes les peticions, ha estat implementada mitjançant el *framework Flask* de *Python*.

Per la banda del client, s'ha creat i establert la videoconferència fent ús de l'API pública del servei de *Jitsi*. Aquesta proporciona un seguit de guies i funcions que permeten establir i configurar la trucada. En aquest cas ens ha interessat solament crear i establir una sala per a gestionar la videoconferència i adquirir els *frames* quan pertoqui. Principalment, les funcions necessàries pel desenvolupament de l'aplicació han estat les següents:

- `api = new JitsiMeetExternalAPI(domain, options)` → Aquesta funció crea l'objecte que permetrà gestionar la videoconferència. El paràmetre *domain* permet definir la URL de la trucada, i el paràmetre *options* ens permetrà configurar-la, per exemple, afegint nom a la sala, definint el node del pare on estarà ubicat l'element *HTML iframe*, etc.

- `captureLargeVideoScreenshot().then(dataURL => {...})` → Aquesta funció ens permetrà capturar un *frame* de la videotrucada en format *dataURL*.

Pel que fa a la banda del servidor, com s'ha comentat en seccions anteriors, s'ha fet ús de la llibreria *face_recognition* (versió usada 1.3.0) de *Python* com a nucli del codi que implementa el sistema de reconeixement facial. Aquesta llibreria compta amb un nombre reduït de funcions a utilitzar. De fet, la llibreria es podria considerar, en bona part, com una interfície per *Python* de la llibreria *dlib* escrita en *C++* (llibreria que implementa la xarxa neuronal encarregada d'extreure les 128 característiques facials). Fent una ullada a la documentació, destaquem tres funcions degut a la utilitat per a dur a terme l'aplicació:

- `face_recognition.load_image_file(file, mode='RGB')` Aquesta funció és l'encarregada de carregar la imatge en una variable en format *numpy array*. De fet, carrega en una matriu, els píxels dels 3 canals de color, RGB.
- `face_recognition.face_landmarks(face_image, face_locations=None, model='large')` Aquesta funció s'encarrega d'extreure les localitzacions de les 68 característiques de qualsevol cara (posició del nas, boca o ulls, entre d'altres) punts/coordenades en la imatge.
- `face_recognition.face_encoding(face_image, known_face_locations=None, num_jitters=1, model='large')` Aquesta funció és l'encarregada de detectar, aïllar, alinear i extreure les 128 característiques representatives de la/es cara/es que apareixen a la imatge.

Aquestes funcions són les encarregades de processar els *frames* obtinguts a través de la videotrucada.

A demés, s'ha fet ús de la llibreria *numpy* de *Python* per calcular la **distància euclidiana** i el **coeficient de Pearson** com a mètriques de comparació per realitzar el *matching* entre els dos vectors que contenen les 128 característiques facials.

Finalment, s'ha creat una simple DB del tipus *SQL*. La DB consta de tres columnes: *id*, *name*, *encoding* (característiques extretes).

4.5. Notes addicionals

El codi de tot el projecte pot ser revisat i/o clonat a través del repositori *Git* penjat al portal *GitHub* amb enllaç https://github.com/ngarsot/Face_recongnition_Jitsi.git. En ell es pot trobar un `readme.md` i un `requirements.txt` amb les guies necessàries per instal·lar i posar en marxa l'aplicació.

En cas de revisió del codi font, notar que:

- La DB pot ser:
 - o un directori que conté un conjunt d'imatges de cares de persones en que el nom del fitxer és el propi nom de l'individu.
 - o una DB *SQL* mitjançant `sqlite3` de *Python*. En qualsevol cas, si es volgués usar un altre tipus de DB com podria ser una *NoSQL*, seria necessari fer els canvis pertinents per tal d'adaptar l'aplicació.
- Actualment, el servidor només accepta/serveix a un client a la vegada. En cas de voler acceptar més clients a la vegada, s'hauria d'adaptar el codi del servidor i fer-lo concurrent.

Notar que tot i que no seria una tasca complicada realitzar els canvis de les observacions anteriorment presentades, aquestes no han estat realitzades durant el transcurs del projecte.

Instal·lació del servidor i posada en marxa

No seria necessari realitzar la instal·lació de l'aplicació ja que el servei pot corre en un servidor i fer ús d'aquesta a través d'internet. Tanmateix, a continuació es presenten les guies bàsiques per instal·lar i fer corre el servidor en una màquina local:

- 1- Clonar el repositori en local mitjançant l'enllaç facilitat en aquesta mateixa secció 4.5.
- 2- Disposar de o instal·lar *Python* (la versió testejada ha estat *v.3.8*).
- 3- Opcional: Usar l'IDE *PyCharm*. Crear un projecte amb l'opció d'afegir un `venv` i afegir els fitxers del repositori a la carpeta `root` del projecte creat. Si s'ha realitzat aquest pas, passar al punt 5.
- 4- Opcional: Crear un `venv` amb *Python* en el repositori clonat.

- 5- Instal·lar les dependències:
 - a. `pip install -r requirements.txt`
 - b. `pip install cmake` (pot ocasionar algun error si no es disposa d'un compilador en C)
 - c. `pip install dlib`
 - d. `pip install face_recognition`
- 6- Crear una DB al directori anomenat `database/` mitjançant una de les dues vies següents:
 - a. La DB ha de ser del tipus SQL i ha de contenir una taula anomenada “`encoded_faces`”. A més, ha de disposar de tres columnes; `id`, `name`, `encoding`. El codi per a realitzar aquest pas es troba comentat al final del fitxer `utils/utils_db.py`.
 - b. Es pot afegir fotografies de les persones a identificar al directori. El fitxer ha d'estar nombrat amb el nom de l'individu. S'ha de realitzar alguna modificació al fitxer `main.py`. Simplement s'ha de modificar la funció `get_similitud_from_db()` (veure els comentaris que la pròpia funció proporciona).
- 7- Executar el fitxer `main.py` i *Flask* farà corre el servidor. Per defecte, el servei s'executarà a la ruta <http://127.0.0.1:5000/>.

5. Procediment experimental

En aquest darrer capítol s'exposarà la validació funcional del sistema de reconeixement facial usat i correcte funcionament de l'aplicació.

5.1. Verificació del sistema de reconeixement (identificació i verificació de cares)

En aquest apartat es presentarà l'anàlisi del ventall de proves realitzades per testejar les funcions usades de la llibreria *face_recognition*, i com aquest respon per a sistemes de reconeixement on queda involucrat per una banda la verificació i per l'altre la identificació quan només es disposa de dues mostres, és a dir, que la comparativa és d'una cara versus una altre (Figura 12). Per tant, en aquest sistema no es fa ús de sistemes de classificació on es disposa, per exemple, d'una xarxa neuronal entrenada amb diferents fotografies de l'individu.

La verificació del sistema de reconeixement s'ha dut a terme seguint el següent procediment:

- 1- S'han creat dues carpetes que contenen:
 - a. Un conjunt de fotografies de la mateixa persona. En aquest cas, les imatges corresponien a fotografies de la meua cara. El set d'imatges presentava les cares amb diferents condicions, és a dir, variabilitat en la il·luminació i exposició de llum, posició de la cara, expressions facials, ús d'ulleres graduades transparents, diferents edats o condicions físiques (pes i pèl facial, entre d'altres).
 - b. Un conjunt d'imatges de 10 persones diferents (membres de la família i amistats). El set d'imatges també presentava variabilitat en les condicions. A més, entre les mostres hi havia un parell de persones que eren bessones.

- 2- S'han extret les característiques facials de cada persona que apareixia en les fotografies i s'han enllaçat aquestes característiques amb la identitat. Per extreure les característiques s'ha fet ús de la llibreria *face_recognition*.

3- S'ha realitzat una mètrica comparativa entre cada una de les característiques extretes:

- a. De la mateixa persona:
 - i. Extreteres de la meva cara
 - ii. Extreteres d'altres persones
- b. De persones diferents

4- Anàlisi dels resultats

A continuació s'exposaran els resultats obtinguts i es presentaran algunes captures mostrant-ne alguns dels més representatius. Els resultats queden plasmats seguint la mateixa estructura; nom descriptiu i percentatge de comparació (en aquest cas fent ús de la mètrica → coeficient de *Pearson*).

La Figura 15 mostra els resultats extrets amb la mateixa persona sota diferents condicions i circumstàncies (llum, edat o posició del cap, entre d'altres).

```

=====Norbert_25anys_frontal_1.jpg=====
Norbert_13anys.png --> 91.61%
Norbert_16anys.png --> 93.6%
Norbert_18anys_batxillerat.jpg --> 94.22%
Norbert_18anys_batxillerat_1.png --> 94.46%
Norbert_18anys_batxillerat_2.png --> 95.0%
Norbert_20anys_sobrepes.png --> 93.67%
Norbert_25anys_BarbaDespentinada.png --> 96.24%
Norbert_25anys_capAbaix.jpg --> 95.21%
Norbert_25anys_capAbaix_noUlleres.jpg --> 95.43%
Norbert_25anys_capDreta_1.jpg --> 95.24%
Norbert_25anys_capDreta_2.jpg --> 97.03%
Norbert_25anys_capDreta_3.jpg --> 97.46%
Norbert_25anys_capDreta_4.jpg --> 96.94%
Norbert_25anys_capDreta_5.jpg --> 96.02%
Norbert_25anys_carota_focusLlumDesdeAbaix.jpg --> 95.66%
Norbert_25anys_focusLlumDesdeAdalt.jpg --> 96.63%
Norbert_25anys_frontal.jpg --> 96.53%
Norbert_25anys_frontal_24.png --> 95.98%
Norbert_25anys_noUlleres_frontal.jpg --> 96.93%
Norbert_25anys_pocalLlum_frontal.jpg --> 97.87%
Norbert_25anys_ullisMigTancats_frontal.jpg --> 99.3%
  
```

Figura 15. Verificació entre cares de la mateixa persona (figura masculina)

Es pot observar que els resultats són molt sorprenents per tractar-se de la comparativa entre dos cares, és a dir, dos vectors de característiques comparades amb el coeficient de *Pearson*. Les fotografies estan ordenades per edat i es pot observar com el percentatge més baix és obtingut quan es compara entre diferents edats, en aquest cas, entre 25 i 13 anys, amb un valor de 91,61%. No s'adjunten fotografies per temes de privacitat, però les

diferències facials són palpables: absència de barba, pes, ulleres, etc. Els valors més elevats són obtinguts quan es van comparar les cares corresponents a les mateixes edats. A més, el valor més elevat es troba als 99.3% i com a valor mitjà 96,56% en les fotografies d'edat 25 anys. S'ha pogut observar com la posició de la cara no afecta gaire. Tanmateix, arriba un punt de perfil que la llibreria no és capaç de detectar una cara. La exposició de la llum tampoc s'ha vist que afecti en gran mesura.

Els resultats obtinguts amb altres persones són molt similars als presentats anteriorment. La Figura 16 mostra un altre exemple dels resultats extrets amb la mateixa persona, en aquest cas, figura femenina, sota diferents condicions i circumstàncies (edat, maquillatge o posició del cap, entre d'altres).

```

=====Marta_26anys_frontal.jpg=====
Marta_23anys_frontal.png --> 93.95%
Marta_23anys_capInclinat.png --> 93.85%
Marta_24anys_frontal.png --> 96.22%
Marta_24anys_maquillada.png --> 95.46%
Marta_24anys_maquillada_1.png --> 97.11%
Marta_26anys_capGirat.png --> 95.96%
Marta_26anys_platja.png --> 96.52%
  
```

Figura 16. Verificació entre cares de la mateixa persona (figura femenina)

Per altre banda, la comparativa entre persones diferents ha presentat els següents resultats. La Figura 17 mostra la comparativa entre cares de diferents persones i condicions entre les imatges.

```

=====Norbert_25anys_frontal.jpg=====
Ada_17anys_frontal.jpg --> 87.54%
Angelina_83anys_frontal.png --> 86.05%
Angelina_83anys_frontal_1.png --> 82.45%
Arnau_8anys_capAlcat.png --> 87.4%
Arnau_8anys_frontal.png --> 86.9%
Cristian_23anys_frontal.jpg --> 85.35%
CristianoRonaldo_capInclinat.png --> 83.63%
Cristina_17anys_capDreta.png --> 83.07%
Cristina_18anys_baixaResolucio.png --> 84.18%
Cristina_18anys_maquillada.png --> 84.0%
Cristina_20anys_frontal.png --> 84.12%
Cristina_20anys_maquillada.png --> 84.2%
Ferran_16anys_baixaResolucio.png --> 89.83%
Ferran_18anys_capInclinat.png --> 90.23%
Ferran_18anys_frontal.png --> 88.93%
Ferran_19anys_capInclinat.png --> 88.5%
Ferran_20anys_frontal.png --> 89.11%
Germán_26anys_frontal.png --> 88.11%
Hector_21anys_frontal.png --> 85.77%
Hilari_80anys_capGirat.png --> 82.32%
Hilari_80anys_frontal.png --> 82.92%
Marta_23anys_capInclinat.png --> 86.59%
Marta_23anys_frontal.jpg --> 86.47%
Marta_24anys_frontal.png --> 85.96%
Marta_24anys_maquillada.png --> 84.5%
Marta_24anys_maquillada_1.png --> 86.16%
Marta_26anys_capGirat.png --> 86.46%
Marta_26anys_frontal.png --> 86.47%
Marta_26anys_platja.png --> 87.04%
Merce_51anys.png --> 83.97%
Merce_51anys_capInclinat.png --> 85.13%
Natalia_18anys.png --> 84.33%
Natalia_19anys_capInclinat.png --> 85.02%
Natalia_20anys_frontal.png --> 83.65%
Roger_20anys_baixaResolucio.png --> 87.99%
  
```

Figura 17. Verificació entre cares de diferents persones (figures masculines i femenines)

Es pot observar com cap valor ha presentat un valor superior al 90%. Tot i així, els valors són força elevats. Això es deu al fet que la CNN s'encarrega d'extreure les característiques principals d'una cara, i tota cara comparteix algunes característiques com són els ulls, la boca, el nas, la forma, etc. Tanmateix, si es comparen els resultats presentats a la Figura 15 i la Figura 16 amb les de la Figura 17, és pot observar com els percentatges són força diferents.

S'ha realitzat també una comparativa entre tres germanes, dues de les quals són bessones. Els resultats queden plasmats a la Figura 18 i Figura 19.

```

=====Marta_24anys_frontal.png=====
Marta_23anys_capInclinat.png --> 96.29%
Marta_23anys_frontal.jpg --> 96.22%
Marta_24anys_maquillada.png --> 95.53%
Marta_24anys_maquillada_1.png --> 96.22%
Marta_26anys_capGirat.png --> 97.06%
Marta_26anys_frontal.png --> 94.43%
Marta_26anys_platja.png --> 96.11%

Cristina_17anys_capDreta.png --> 92.0%
Cristina_18anys_baixaResolucio.png --> 92.08%
Cristina_18anys_maquillada.png --> 92.76%
Cristina_20anys_frontal.png --> 92.15%
Cristina_20anys_maquillada.png --> 93.26%
Natalia_18anys.png --> 92.48%
Natalia_19anys_capInclinat.png --> 93.51%
Natalia_20anys_frontal.png --> 92.94%
  
```

Figura 18. Verificació entre cares de diferents persones amb relació familiar de germanes (figures femenines)

```

=====Cristina_18anys_maquillada.png=====
Cristina_17anys_capDreta.png --> 94.78%
Cristina_18anys_baixaResolucio.png --> 97.95%
Cristina_20anys_frontal.png --> 97.09%
Cristina_20anys_maquillada.png --> 95.54%

Natalia_18anys.png --> 95.12%
Natalia_19anys_capInclinat.png --> 96.88%
Natalia_20anys_frontal.png --> 95.83%
  
```

Figura 19. Verificació entre cares de diferents persones amb relació familiar de germanes bessones (figures femenines)

Es pot observar a la Figura 18 que els valors obtinguts en el cas de les germanes és força elevat. Sí que és cert que quan es tracta de la mateixa persona, el valor és superior, però els percentatges de similitud són força superiors als mostrats a la Figura 17, donant com a resultat una diferenciació no tant clara en valor.

Com a cas extrem, els resultats obtinguts a la Figura 19 mostra com el sistema no és capaç, amb les mostres aportades, de diferenciar entre dues persones bessones. És a dir, no

seria possible definir un valor llindar en el sistema per diferenciar de si es tracta de la mateixa persona o no. *Remarcar de nou que la comparativa ha estat realitzada entre dues cares, i no s'ha provat en cap cas la comparació dins d'un sistema classificador amb xarxes neuronals.*

Amb els resultats i les proves realitzades s'ha escollit un valor de llindar de 93,5%. Aquest llindar presenta alguns falsos positius, com el cas de les germanes, i alguns falsos negatius, sobretot quan es comparen cares de la mateixa persona però corresponents a edats diferents. Tot i així, pot ser considerat com un bon valor llindar pel sistema de verificació per diferenciar si es tracta o no de la mateixa persona, ja que pretén minimitzar el nombre de falsos positius a costa d'algun fals negatiu. Pel sistema d'identificació, no serà necessari el valor llindar més que per decidir si la persona que presenta el valor més baix és o no la persona en qüestió.

Ja més focalitzat en l'aplicació notarial a implementar, s'ha realitzat un seguit de proves per mesurar quin seria el valor llindar quan es realitza la verificació entre una cara i la fotografia del DNI. A continuació es presenten alguns dels resultats obtinguts.

```

=====Norberty_dni.png=====
Angelina_capAbaix.png --> 85.68%
Angelina_capAmunt.png --> 79.86%
Angelina_capLateral.png --> 83.0%
Angelina_dni.png --> 86.45%
Angelina_dni_1.png --> 87.21%
Angelina_frontal.png --> 81.59%
Angelina_frontal_1.png --> 81.94%
Cristina_capAmunt.png --> 85.08%
Cristina_capLateral.png --> 84.61%
Cristina_capLateral_1.png --> 86.18%
Cristina_dni.png --> 84.31%
Cristina_dni_1.png --> 86.57%
Cristina_frontal.png --> 85.65%
Cristina_frontal_1.png --> 84.26%
Hilari_capAmunt.png --> 87.93%
Hilari_capLateral.png --> 84.05%
Hilari_dni.png --> 83.77%
Hilari_dni_1.png --> 84.18%
Hilari_frontal.png --> 83.66%
Hilari_frontal_1.png --> 85.11%
Hilari_frontal_2.png --> 86.09%
Marta_capAmunt.png --> 87.49%
Marta_capLateral.png --> 89.42%
Marta_dni.png --> 88.21%
Marta_dni_1.png --> 88.87%
Marta_frontal.png --> 88.19%
Marta_frontal_ulleres.png --> 88.29%
Merce_capAbaix.png --> 84.67%
Merce_capAmunt.png --> 87.61%
Merce_capLateral.png --> 87.06%
Merce_dni.png --> 88.25%
Merce_dni_1.png --> 88.41%
Merce_frontal.png --> 87.75%
Merce_frontal_1.png --> 87.42%
Norberty_capAbaix.png --> 93.24%
Norberty_capAmunt.png --> 91.71%
Norberty_capLateral.png --> 86.27%
Norberty_frontal.png --> 94.09%
Norberty_frontal_1.png --> 95.09%
Norberty_frontal_noUlleres.png --> 93.08%
    
```

Figura 20. Verificació entre el DNI i el propietari del mateix, així com també amb diferents persones

Es pot observar com els valors, quan la comparativa és amb diferents persones no propietàries del DNI, es troben per sota del llindar 90%. Tanmateix, depenent de la posició del cap (pròxim a l'estat del perfil o alçant el cap cap amunt) els valors obtinguts amb el propietari del DNI s'hi han trobat per sota. En qualsevol cas, els resultats obtinguts quan es trobava frontalment, han sigut similars als mostrats en la Figura 15 i la Figura 16.

Si realitzem la comparativa entre les germanes, els valors obtinguts no són tant clars.

```
=====Cristina_dni.png=====
Cristina_capAmunt.png --> 94.48%
Cristina_capLateral.png --> 94.74%
Cristina_capLateral_1.png --> 96.85%
Cristina_frontal.png --> 96.25%
Cristina_frontal_1.png --> 96.12%
Marta_capAmunt.png --> 92.09%
Marta_capLateral.png --> 89.78%
Marta_frontal.png --> 92.26%
Marta_frontal_ulleres.png --> 90.21%
```

Figura 21. Verificació entre DNI i cares de diferents persones amb relació familiar de germanes (figures femenines)

En aquest cas, es pot observar com els valors obtinguts amb la germana són força elevats, superant lleugerament el 92%. Tanmateix, els valors obtinguts per ella mateixa han resultat amb una mitjana de 95,69%

Així doncs, determino que un valor llindar per a l'aplicació per diferenciar si es tracta o no de la mateixa persona rondaria sobre els 92,5%. Aquest resultat ha aportat falsos positius i falsos negatius. Tanmateix, sembla ser que, en vista de les proves realitzades, és el llindar que minimitzava la major quantitat d'errors. **Tot i així, perquè els resultats cobrin pes, s'haurien de realitzar més d'una mesura per tal extreure el resultat mitjà, ja que, algunes posicions del cap o similitud entre cares de persones (cas de les germanes), podrien afectar al extreure conclusions amb una simple mesura.**

Remarcant i notant que per a realitzar aquestes proves i extreure uns resultats conclouents, seria necessari comptar amb un data set molt extens. Les proves realitzades han servit per extreure unes primeres conclusions.

5.2. Verificació de l'aplicació notarial

En aquest apartat es realitzarà una verificació funcional de l'aplicació web desenvolupada.

Per accedir a l'aplicació, primer de tot el servidor ha d'estar executant-se. Acte seguit, s'obre el navegador i s'accedeix a la URL on està corrent el servei. Un cop dins, l'aplicació es veu de la següent manera (mostrat a la Figura 22):

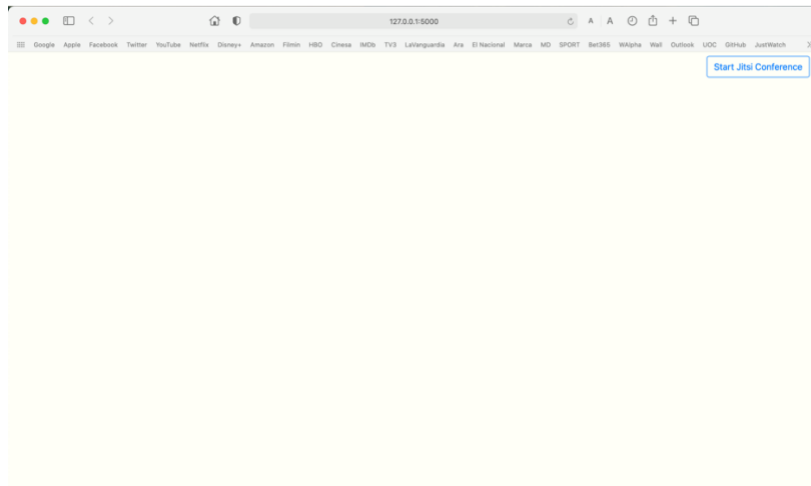


Figura 22. Vista de l'inici de l'aplicació notarial

Polsant el boto s'accedeix al menú inicial amb les 3 funcionalitats i la configuració inicial de la videotrucada fent ús de *Jitsi*. La Figura 23 mostra una captura de l'aplicació abans d'establir connexió de videotrucada.

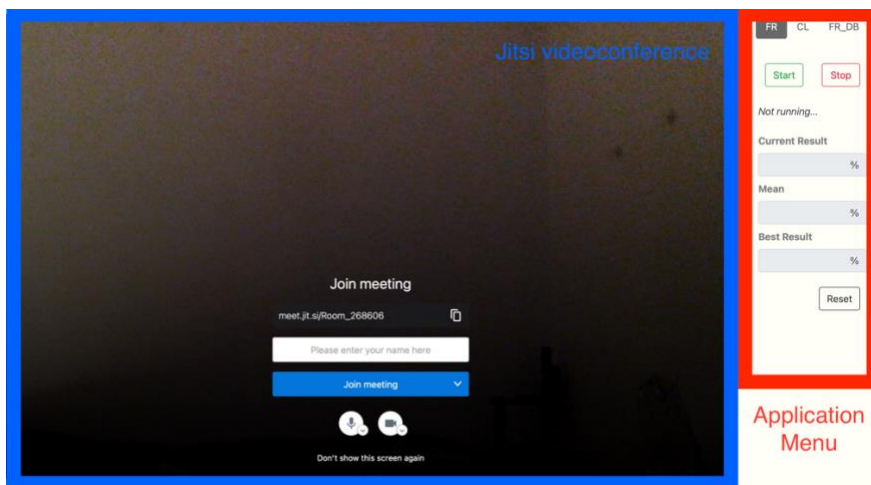


Figura 23. Interfície de l'aplicació on es pot observar l'apartat de la videoconferència i el menú de les 3 funcionalitats diferents

La videoconferència ha estat implementada de tal manera que simplement s'ha d'introduir el nom del conferenciant, compartir l'enllaç generat aleatòriament i polsar el boto "Join meeting" per establir-la. En qualsevol cas, és possible canviar i definir el nom de la sala que es desitgi.

Un cop establerta la videotrucada, ja es pot fer ús de les tres funcionalitats que l'aplicació ofereix: FR (*face recognition*), CL (*check life*) i FR_DB (*face recognition in DB*).

FR

Aquesta funció permet comparar la foto del DNI amb la cara de la persona que es troba en la videoconferència. El procés de comparació es pot realitzar polsant el boto *start* i es pot parar polsant el botó *stop*. Un cop comença la cerca, l'aplicació només detecta dues cares. En aquest cas, detecta qualsevol cara que estigui davant de la càmera, és més, la cara pot venir d'una imatge, vídeo, o una persona física. En qualsevol cas, un cop el procés de comparació està corrent, si es detecten dues cares (sigui quina sigui la font), el servidor retorna el valor de del resultat actual, la mitjana dels resultats de les cares analitzades i el millor resultat obtingut fins el moment o fins que es fa un *reset* de la funcionalitat. La Figura 24 i la Figura 25 mostren un parell d'exemples sense mostrar la imatge de la videoconferència.



Figura 24. Interfície de FR - Persona propietària del DNI

Els resultats plasmats a la Figura 24 han estat obtinguts comparant el DNI amb la cara de la persona propietària del document. Es pot observar que els resultats obtingut són superiors al valor llindar de 92,5%.



Figura 25. Interfície de FR - Persona no propietària del DNI

Els resultats plasmats a la Figura 25 han estat obtinguts comparant el DNI amb la cara d'una persona no propietària del document. Es pot observar que els resultats obtingut són inferiors al valor llindar de 92,5%.

Les proves anteriorment presentades han estat realitzades amb 10 persones diferents on s'ha fet ús del DNI de la pròpia persona i d'altres DNI. En cap cas, del petit ventall de proves realitzades, el resultat mitjà d'una persona que no ha utilitzat el seu DNI ha estat superior al llindar de 92,5%. Per tant, tot i que el set de proves sigui molt limitat, no s'han obtingut falsos positius (a diferència dels resultats obtinguts en la secció de verificació del sistema de reconeixement, on sí se'n havia obtingut alguns amb aquest llindar seleccionat). Tanmateix, en alguna ocasió, el "Current Result" obtingut d'una persona presentant el seu DNI ha estat inferior al valor llindar. Aquest fet es deu a problemes amb la imatge adquirida, exposició de la llum, reflexos de la fotografia del DNI, o altres. Aquest esdeveniment reforça el fet que el camp amb major importància és el que presenta el valor mitjà, seguit del millor resultat ("Best Result"). El valor mitjà d'una mateixa persona rondava els 94%, 2 punts percentuals més.

Com a limitació, notar que l'aplicació simplement mostra els resultats de qualsevol comparativa entre dues cares. Així doncs, queda com a requeriment indispensable la supervisió del notari. De fet, aquesta eina està pensada per a donar suport i afegir un major grau de confiança durant el procés de verificació de la persona.

A més, el set de proves realitzat és petit degut a la poca quantitat de gent amb la qual s'han pogut realitzar. Tanmateix, els resultats podrien ser fiables si es pogués realitzar amb una quantitat de persones fora del nostre abast.

CL

La funció de *check life* permet determinar si la persona que es troba en la videotrucada realitza un seguit de moviments comandats pel notari a través d'un seguit de tests. Per tant, serveix per determinar si realment hi ha una persona a davant i aquesta no es tracta, per exemple, d'un vídeo amb una persona que hi apareix. La Figura 26 mostra la interfície de la funcionalitat CL amb uns test en *pass* i altres com a *no-pass*.

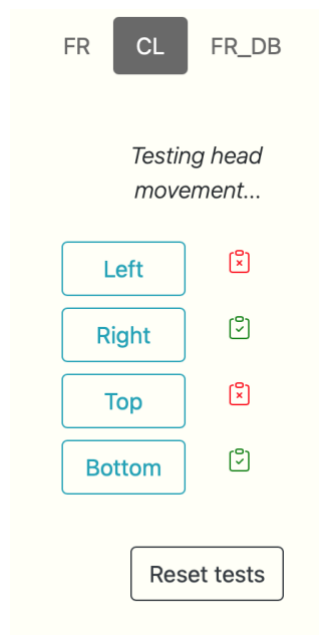


Figura 26. Interfície de CL - Exemple que conté dos test que han estat passats correctament i dos tests que no han estat passats correctament

La funció de CL serveix per verificar que la persona que està en la videotrucada reacciona correctament als moviments que li demana el notari per tal de verificar que no es tracta d'un vídeo o d'una persona amb absència de vida.

Notar que al tractar-se d'un sistema de reconeixement en 2D, la verificació del camp de profunditat és fa impossible. Per aquest motiu, és possible passar els tests fent ús d'una fotografia. De nou, és condició necessària la presència del notari o d'un supervisor.

Els tests poden presentar alguns falsos negatius. Aquests dependran de la velocitat de moviment del cap i la freqüència de mostreig (adquisició de *frames*), que dependrà fortament del processador de la màquina i la pròpia implementació. Aquest punt podria ser optimitzat. En qualsevol cas, no han aparegut falsos positius durant les proves realitzades.

FR_DB

La funcionalitat permet adquirir la cara d'una persona i identificar-la a través d'una DB. Un cop finalitzada la cerca, l'aplicació et mostra el nom de la persona identificada i el coeficient de similitud en %. La Figura 27 mostra un exemple de la interfície de la funcionalitat FR_DB.

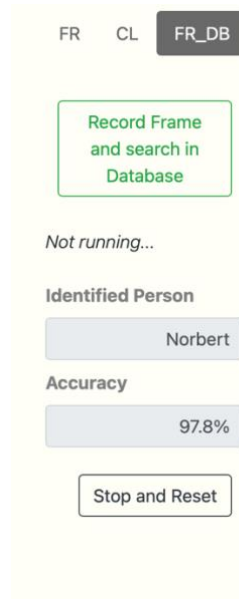


Figura 27. Interfície de FR_DB - Exemple que conté el resultat d'una identificació correcte

La Figura 27 mostra una identificació realitzada correctament.

El ventall de proves ha estat realitzat amb una DB que contenia 8 identitats diferents amb un vector de característiques per a cada una d'elles. En cap cas, sota diferents condicions d'exposició de llum, posicions de la cara o obstacles es va obtenir resultats incorrectes. En tots els casos provats i tests realitzats, la persona ha estat correctament identificada.

Finalment, s'han realitzat altres tipus de verificacions de caire funcional, com són el de fer ús dels botons de *reset* de cada una de les funcionalitats, o intentar trobar errors de funcionament que bloquegin l'aplicació. Fins ara, les proves realitzades determinen que l'aplicació sembla ser estable i està preparada per a ser usada.

Com a nota, l'aplicació web i el servidor han estat executats a la mateixa màquina, un MacBook Pro (Retina, 15-inch, Mid 2015) amb les següents característiques; Intel Core i7 de quatre nuclis a 2.2 GHz.

6. Conclusions

En aquest treball s'ha realitzat un estudi sobre l'estat de l'art dels sistemes de reconeixement facial i les xarxes neuronals i, com a conseqüència, els beneficis d'implementar aquests sistemes amb xarxes neuronals. Els sistemes de reconeixement facial, degut a la poca intrusió, facilitat i comoditat d'ús, cada cop prenen més força com a primera etapa per a verificar o identificar una persona. Així doncs, formaria part de la primera etapa per protegir un sistema o una aplicació. Aquests sistemes de reconeixement han de comptar amb un nivell de garantia elevat per tal d'evitar futurs riscos. L'estudi que hi ha darrera és força extens i grans empreses treballen per millorar constantment aquesta tecnologia. Actualment, l'estat de l'art dels sistemes de reconeixement facial estan implementats amb xarxes neuronals, i els resultats obtinguts han arribat a assolir nivells de precisió més elevats que els propis humans.

Aquest treball tenia com a objectiu principal dissenyar i implementar una aplicació de reconeixement facial que donés suport als notaris abans de procedir amb una reunió, fent ús de xarxes neuronals ja que, aquest any 2020 ha quedat marcat per la pandèmia produïda pel virus Covid-19 i les mesures preses pels governs per mitigar el contagi. Les mesures més destacades han estat la de limitar el lliure desplaçament i les reunions. Així doncs, el treball proposa una aplicació, com a alternativa, per a realitzar les reunions a través d'una videoconferència. L'aplicació pretén donar suport al notari durant tot el procés de verificació de la persona, és a dir, obtenir un major nivell de garantia per tal de formalitzar que la persona que està a l'altre banda de la videoconferència és qui diu ser. Procés que es duu a terme comparant la cara de la persona amb el seu DNI. A més, l'aplicació també permet verificar si la persona respon als estímuls del notari (proves de vida) i permet identificar a la persona a través d'una DB.

Com a conclusió principal, remarcar que s'han pogut assolir els objectius marcats, ja que gràcies a l'estudi de l'estat de l'art, s'ha pogut dissenyar i implementar una aplicació funcional que permeti donar suport al notari a l'hora de verificar i identificar una persona amb cert nivell de garantia, fent ús només de dues úniques mostres o cares.

S'ha realitzat una bateria de proves i el seu posterior anàlisi per tal de verificar les tecnologies usades per implementar el sistema de reconeixement. La verificació del sistema de reconeixement facial ha estat clau per definir la capacitat que té el nostre sistema per saber si la verificació o identificació de la persona és correcte o possible. Destacar que la definició d'un valor llindar en el nostre sistema ha vingut condicionat pel petit nombre de

dades/fotografies tractades. És complicat definir correctament un valor llindar que minimitzi la introducció d'errors al sistema, és a dir, que redueixi el nombre de falsos positius i falsos negatius. Notar que, per aquesta aplicació, és necessari minimitzar el nombre de falsos positius, per tal de no verificar algú que no és qui diu ser. En qualsevol cas, el valor llindar s'ha intentat escollir minimitzant l'error total tot i prioritant l'absència de falsos positius.

Tenint en compte que tant la verificació com la identificació es realitza mitjançant la comparació de només dues imatges de les cares, s'ha pogut observar que:

- Els valors de similitud superaven el 94% tot i les condicions d'exposició de llum o posicions de la cara (fins a un cert punt, e.g. de perfil), quan es comparaven dues imatges de la cara de la mateixa persona.
- Els valors de similitud baixaven lleugerament quan es tractava de la comparativa d'una cara i la fotografia del DNI (majoritàriament, degut a la qualitat de la fotografia i als reflexos de llum degut a les marques que hi ha al damunt de la fotografia). Tanmateix, el valor mitjà també rondava el 94%.
- Així doncs, ha estat possible escollir, amb les proves realitzades, un valor llindar pel sistema de reconeixement, que permet diferenciar si es tracta o no de la mateixa persona.
- El sistema de verificació pot introduir alguns falsos positius si les dues cares comparades són similars, com pot ser el cas analitzat de les germanes, tant bessones com no.
- En el sistema de verificació queda reforçat si es prenen diferents mesures i es va calculant la mitjana dels valors obtinguts.
- El sistema d'identificació ha retornat, amb les imatges de les que es disposava, la identitat correcta. Tanmateix, tal i com és el cas de les germanes bessones, el sistema podria haver introduït algun error perfectament, ja que dependrà de la imatge de la cara de la qual es disposi.
- Per tant, en el cas de les germanes bessones, es pot afirmar que el sistema (identificació i verificació) no ha estat capaç de diferenciar-les.
- Les proves de vida podrien ser passades com a vàlides si s'utilitza una fotografia, ja que el sistema de reconeixement facial usat crea una imatge 2D (no compte amb profunditat de camp). Tanmateix, sí podria identificar si es tracta o no d'un vídeo.
- En qualsevol cas, remarcar que el sistema requereix d'un responsable (com seria el cas del notari) que verifiqui els resultats que l'aplicació va proporcionant.

- Notar de la dificultat de definir un valor llindar ajustat i considerat com a plenament vàlid amb les poques mostres utilitzades per a dur a terme la verificació de l'aplicació i del sistema de reconeixement facial, en sí. Per poder obtenir uns resultats fortament conclusius, el ventall de mostres hauria de ser de dimensions només a les mans de grans empreses.

Finalment, destacar de nou que la verificació o la identificació fent ús de la comparació de només dues imatges de cares pot minvar l'èxit, la precisió o l'encert. Tanmateix, remarcar que aquest sistema de reconeixement facial obté uns resultats realment sorprenents tot i només realitzar la comparativa entre dues mostres.

7. Glossari

Aquest apartat llista, en forma de glossari, totes les paraules i acrònims usats durant el transcurs de la memòria. Paraules ordenades en ordre d'aparició.

- **DNI:** Document Nacional d'identitat
- **SoA:** *State-of-the-Art* → Estat de l'art
- **DB:** *Database* → Base de dades
- **CNN:** *Convolutional Neural Network* → Xarxes neuronals convolucions
- **HOG:** *Histogram of Oriented Gradient*
- **ANN:** *Artificial Neural Network* → Xarxes neuronals artificials
- **AI:** *Artificial Intelligence* → Intel·ligència artificial
- **MLP:** *Multilayer Perceptron* → Perceptró multicapa
- **CNN:** *Convolutional Neural Network* → Xarxa neuronal convolucional

8. Bibliografia

[1] Mei Wang, i Weihong Deng, “Deep Face Recognition: A Survey”, School of Information and Communication Engineering, Beijing University of Posts and Telecommunications.

[2] Francesc Serratos, “Reconeixement de les persones pels trets de la cara”, PID_00215058, UOC.

[3] Paul Viola, i Michael Jones, “Rapid Object Detection using a Boosted Cascade of Simple Features”, Cambridge.

[4] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, i Yu Qiao, “Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks”.

[5] Sachin Sudhakar, Mohammad Saberian, i Li-Jia Li, “Multi-view Face Detection Using Deep Convolutional Neural Networks”, Yahoo.

[6] Adam Geitgey, (Jul 2016), <https://medium.com/@ageitgey/machine-learning-is-fun-part-4-modern-face-recognition-with-deep-learning-c3cffc121d78>, “Machine Learning is Fun! Part 4: Modern Face Recognition with Deep Learning”.

[7] Gurpreet Kaur, Monica Goyal, i Navdeep Kanwal, (Març 2018), <https://www.ukessays.com/essays/computer-science/feature-extraction-recognition-review-4007.php>, “Feature Extraction in Face Recognition: A Review”.

[8] Idelfonso Jiménez Silva, (2018) “Reconocimiento facial basado en redes neuronales convolucionales”, Escuela técnica Superior de Ingeniería, Universidad de Sevilla.

[9] Wikipedia, (2020), https://en.wikipedia.org/wiki/Artificial_neural_network, “Artificial neural network” (Edició en Anglès i Català).

[10] UPCcommons. Portal del coneixement obert de la UPC, <https://upcommons.upc.edu/bitstream/handle/2099.1/6483/05.pdf?sequence=6&isAllowed=y>, “Introducció a les xarxes neuronals”, Universitat Politècnica de Catalunya (UPC).

[11] Yi Sun, Yuheng Chen, Xiaogang Wang, i Xiaoou Tang, (2014) “Deep Learning Face Representation by Joint Identification-Verification”, Department of Information Engineering, The Chinese University of Hong Kong.

[12] Florian Schroff, Dmitry Kalenichenko, i James Philbin, (2015) “FaceNet: A Unified Embedding for Face Recognition and Clustering”, Google Inc.

9. Annexos

9.1. *Codi Font*

El codi font del projecte pot ser revisat o clonat a través del repositori *Git* penjat al portal *Github* mitjançant l'enllaç → https://github.com/ngarsot/Face_recongntion_Jitsi.git.