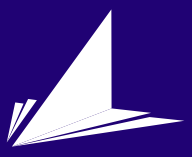


Escuela
Politécnica
Superior

Estimación de la calidad de imágenes de caras para aplicaciones de autentificación biométrica usando Deep Learning

Grado en Ingeniería en Sonido e Imagen en Telecomunicación

Trabajo Fin de Grado

Autor:
Luis López Payá
Tutor/es:
José García Rodríguez
Pedro Córdoba y Angela Sánchez

Julio 2022

Universitat d'Alacant
Universidad de Alicante

Estimación de la calidad de imágenes de caras para aplicaciones de autentificación biométrica usando Deep Learning

Autor
Luis López Payá

Tutor/es
José García Rodríguez
Departamento de Tecnología Informática y Computación
Pedro Córdoba y Angela Sánchez
Facephi Research

Grado en Ingeniería en Sonido e Imagen en Telecomunicación

Escuela
Politécnica
Superior

Universitat d'Alacant
Universidad de Alicante

ALICANTE, Julio 2022

Preámbulo

"El desarrollo de este trabajo está motivado por el interés en profundizar sobre el conocimiento y el estado del arte en la rama de imagen de la titulación. Se ha elegido este proyecto con el fin de conocer, experimentar y aprender sobre nuevas técnicas de estimación de calidad de imágenes para su uso en aplicaciones de reconocimiento biométrico mediante Deep Learning siendo una rama nueva de aprendizaje. "

"Tras mucho trabajo, reuniones y situaciones de verdadera incertidumbre sobre el desenlace de este trabajo, aquí está."

Agradecimientos

Este trabajo no habría sido posible sin el apoyo y la confianza de mi tutor, José García, con el que escogí este tema sin conocer de nada. Con su supervisión e ímpetu por que realizara este trabajo debido a su confianza depositada en mí, ha logrado que este trabajo se realice cuando yo no lo daba por hecho. Darle las gracias por hacerme ver que todo es posible, por apoyarme, por animarme y por ser la persona que es.

También me gustaría agradecer a la empresa Facephi [1]la oportunidad dada de realizar mi Trabajo Final de Grado (TFG) de manera conjunta con ellos, recibiendo ayuda y consejos de parte de todo el equipo de FacePhi. Me gustaría destacar el trabajo de Ángela Sánchez y Pedro Córdoba, trabajadores de Facephi. Gracias a ellos he podido sacar adelante este trabajo con sus charlas sobre este tema desconocido para mí en el área de la Inteligencia Artificial. Jamás olvidaré los cuatro meses de reuniones semanales en las que me habéis ayudado, apoyado y aportado información, además de haber tenido que aguantarme cuando mis expectativas eran tan negativas y cada vez lo veía más difícil, de verdad, muchísimas gracias.

No puedo terminar sin agradecer a mi familia, el apoyo y el esfuerzo realizado durante toda la vida para que yo realizase un estudio universitario. Fueron ellos los que cuando no tenía nada claro en mi vida me empujaron a dar el salto a la Universidad y los que día tras día han confiado en mí. Han sido años difíciles, pero ellos han hecho que sean mucho más fáciles. Agradecer a mis padres el trabajo realizado por el que hoy soy quien soy y por empujarme a la Universidad cuando yo no lo tenía claro, de verdad, siempre os estaré agradecido.

Es a ellos a quien dedico este trabajo.

[1]FacePhi es una empresa experta en verificación de identidad digital de usuarios, especializada en onboarding digital y soluciones biométricas de autenticación. Nació con el objetivo de crear procesos digitales más seguros, accesibles y libres de fraude. Para conseguirlo apuesta por la innovación con inteligencia artificial y machine learning, aplicando tecnología blockchain e introduciendo la identidad digital descentralizada.

A mi padre, mi madre y a mi hermano,
sin los cuales hubiese sido imposible realizar este trabajo

Cuanto mayor es la dificultad,
mayor es la gloria

Marco Tulio Cicerón.

Índice general

Índice de figuras

Índice de tablas

1 Introducción

En los últimos años la autenticación biométrica ha experimentado un gran crecimiento en aplicaciones de dominio público con el objetivo de verificar la identidad de personas de una manera rápida y segura debido al incremento de fraude de documentos, robos de identidades o amenazas internacionales. Hasta hace poco era una técnica utilizada únicamente en espacios de alta seguridad, pero debido al incremento de delitos con relación a la falsificación de la identidad se ha incrementado su uso en espacios más comunes como en ordenadores, smartphones, etc.

La biometría, del griego "bios" (vida) y "metron" (medida), es la identificación automática de los individuos en función de sus características biológicas y de su comportamiento. Se basa en las medidas biológicas o características físicas que se pueden utilizar para identificar personas. Este tipo de verificación personal es el proceso por el cual se comparan los datos biométricos de una persona con las características de esa persona previamente obtenidas. De modo que, al realizar la obtención de características biométricas y comparar mediante un modelo con las almacenadas se determina que es la misma persona, el reconocimiento de identidad puede verificar la identidad de la persona.

Para que se lleve a cabo un reconocimiento biométrico, la muestra capturada para realizar el reconocimiento debe ser de calidad. Una muestra de calidad en la autentificación biométrica significa que esa muestra posee una gran usabilidad para llevar a cabo un reconocimiento biométrico. En este caso, el reconocimiento biométrico que se lleva a cabo en el estudio es el reconocimiento facial. Para poder verificar la identidad de una persona se necesita que la muestra capturada sea de alta calidad para que el sistema de reconocimiento facial sea capaz de llevar a cabo una serie de comprobaciones hasta determinar si se trata de la misma persona o no.
Durante este trabajo se llevarán a cabo una serie de procesos para la estimación de la calidad de una muestra, para un posterior reconocimiento facial y se estudiará cómo esa calidad afecta a la precisión en el reconocimiento.

1.1 Métodos de autenticación biométrica

Es conocido el interés que los sistemas automáticos de identificación biométrica han suscitado en el ámbito empresarial y cotidiano, interés que va incrementándose día a día por el desarrollo de aplicaciones en el ámbito de la seguridad. La biometría se refiere a la aplicación automatizada de técnicas biométricas a la certificación, autenticación e identificación de personas en sistemas de seguridad. Las técnicas biométricas se utilizan para medir características corporales o de comportamiento de las personas con el objeto de establecer una identidad.

Se trata de un reconocimiento único y personal que aporta una mayor seguridad a la hora de proteger información personal como el teléfono móvil, control de presencia, control de acceso, etc.
Para llevar a cabo la autenticación biométrica, primero se debe registrar a los individuos que van a hacer uso del sistema. Para el registro se utiliza un dispositivo biométrico para examinar el atributo físico o de comportamiento elegido. La autenticación posterior se realiza cuando el individuo presenta su rasgo corporal o muestra su comportamiento ante un dispositivo biométrico. En el caso de la identificación, la persona no informa al sistema biométrico sobre cuál es su identidad. El sistema tan solo captura el rasgo característico de la persona y lo procesa para crear el modelo en vivo. Luego el sistema procede a comparar el modelo en vivo con un conjunto de modelos de referencia para determinar la identidad de la persona. El rendimiento de una medida biométrica se define generalmente en términos de tasa de falso positivo (False Acceptance Rate o FAR), la tasa de falso negativo (False Non Match Rate o FNMR, también False Rejection Rate o FRR), y la tasa de fallo de alistamiento (Failure-to-enroll Rate, FTE o FER). Una de las medidas más comunes de los sistemas biométricos reales es la tasa en la que el ajuste con el cual acepta y rechaza los errores es igual: la tasa de error igual (Equal Error Rate o EER), también conocida como la tasa de error de cruce (Cross-over Error Rate o CER). Cuanto más bajo es el EER o el CER, se considera que el sistema es más preciso.

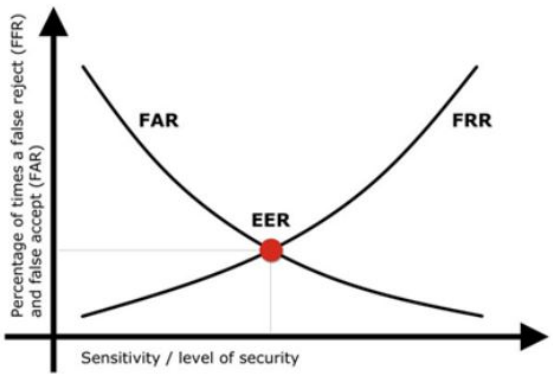


Figura 1.1: Ejemplo de gráfica de la obtención de seguridad de un sistema de identificación

No obstante, existe una gran preocupación por el robo de identidades, ya que, los rasgos biométricos son rasgos biológicos que no se pueden cambiar. De modo que, si la identidad de una persona fuese robada no se podrían cambiar sus rasgos biológicos (como se suele hacer cuando te roban una contraseña) y el daño que podría producir este robo teniendo acceso a todos los datos de la persona sería irreversibles.
Existen diferentes tipos de autentificación biométrica que dependen de las características biológicas de cada individuo, pero los rasgos comunes a todas las técnicas son:

- **Universalidad:** Todo el mundo debe poseer esa característica.
- **Distintividad:** Dos personas deben ser suficientemente diferentes en términos de ese

rasgo.

- **Estabilidad:** El rasgo debe permanecer invariable en el tiempo durante el periodo de autentificación.
- **Evaluabilidad:** La característica debe poder ser medida cuantitativamente.
- **Rendimiento:** Debe ser razonable y no depender de características del entorno.
- **Aceptabilidad:** Los usuarios deben estar de acuerdo en emplear ese rasgo.
- **Fraude:** Los sistemas basados en ese rasgo deben ser lo suficientemente robustos para evitar ser engañados.

El tipo de identificación biométrica se puede clasificar según las características que se midan, Lu (2008). Pueden ser físicas como el reconocimiento de huellas dactilares, retina, iris, geometría de la mano, etc. O características de comportamiento como el reconocimiento de una firma, el reconocimiento de tecleo o el reconocimiento de pasos. Algunas técnicas biométricas como la voz comparten aspectos físicos y de comportamiento.
En este trabajo se van a tratar diferentes métodos de biometría utilizados para el reconocimiento, pero se centrará la investigación en la biometría facial y cómo poder llegar a obtener imágenes de alta calidad para la autenticación facial. Se estudiarán los parámetros de calidad de las imágenes y se comentarán tanto las técnicas clásicas para obtener la calidad de imágenes válidas como las técnicas actuales con Deep Learning.

Se describen a continuación algunas de las técnicas biométricas más relevantes:

ADN: Es el método más común en aplicaciones forenses para reconocimiento. Además, la información que se puede extraer a partir del ADN de una persona puede revelar discapacidades u otras características que el usuario no desee hacer públicas. El ADN es único para cada individuo, excepto para el caso de gemelos monocigóticos.
El análisis de ADN se realiza mediante marcadores genéticos conocidos como regiones polimórficas. Estas regiones se caracterizan por la variación del número de veces que se repite en tándem una secuencia determinada. Estas repeticiones permiten realizar una autentificación de ADN.

Dinámica de tecleo: Es la información de tiempo detallado que describe exactamente cuando cada tecla es presionada y soltada por una persona cuando escribe en un teclado de computadora. Este rasgo biométrico es de tipo conductual y por lo tanto muy variable en el tiempo. Para su captura basta con emplear secuencias del tecleo del usuario, por lo que no es intrusivo. Es poco distintivo, pero puede ser utilizado para identificación en casos sencillos.

Escáner de retina: El escaneo de retina se realiza dirigiendo un rayo imperceptible de luz infrarroja de baja energía hacia el ojo de la persona cuando esta mira a través de la pieza ocular del escáner, como quien mira por un microscopio. Ese rayo de luz traza una ruta estandarizada sobre la retina. Como los vasos sanguíneos de la retina son más

Figura 1.2: Ejemplo de obtención de la dinámica de tecleo

absorbentes de esa luz que el resto del ojo, la cantidad de luz reflejada varía durante el escaneo. El patrón resultante de las variaciones es convertido a código informático y se guarda en una base de datos, Deng, Guo, Ververas, y cols. (2020).

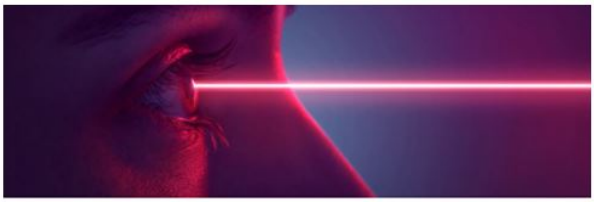


Figura 1.3: Ejemplo de escáner de retina en autenticación biométrica

Firma: Este tipo de biometría se encarga de reconocer la identidad de la persona mediante la firma. Este sistema permite firmar cualquier tipo de documento electrónico identificando tu identidad como firmante a través de la captura de tus datos biométricos. El tipo más específico de firma biométrica permite firmar documentos a través de la firma manuscrita en un dispositivo móvil como una Tablet o un Smartphone utilizando un bolígrafo digital, identificando al firmante a través de la captura de una serie de parámetros específicos de la persona que firma.
En el proceso de firma se capturan una serie de datos biométricos asociados al firmante de forma única, como son la velocidad de escritura, el número o duración de los trazos, la presión ejercida en la escritura, cambios de dirección o la aceleración. Este tipo de datos son los que, agrupados, permiten identificar de forma inequívoca a la persona a través de su firma manuscrita electrónica realizada en el dispositivo móvil.

Forma de caminar: Este tipo de biometría se basa en reconocer la forma de andar humana. Se trata de una técnica que se encarga de identificar a una persona por su forma de caminar, siendo así una técnica no invasiva.
Para llevar a cabo esta técnica se analizan las imágenes de un vídeo para obtener datos que reflejen la identidad de una persona.

Geometría de la mano: Implementado en sistemas de biometría. Se tratan de los sistemas más rápidos dentro de la biometría. Los sistemas que utilizan la geometría de la mano se basan en la extracción de un conjunto de características geométricas de la mano entre las que se pueden mencionar: el ancho de los dedos y localización, ancho de la palma, longitud de los dedos, etc.

Reconocimiento de iris: El reconocimiento del iris es la técnica de identificación más fiable

ya que el iris de las personas no cambia con el tiempo. Dado que el iris es diferente entre el ojo izquierdo y el derecho, el reconocimiento se puede realizar por separado en cada ojo.
Para este tipo de reconocimiento se utiliza una cámara de infrarrojos, que hace una fotografía al ojo, y a través de ella se obtienen los detalles del iris. Debido al uso de una cámara infrarroja, el reconocimiento está disponible incluso de noche o en la oscuridad, Orozco-Rosas y cols. (2012).
Como inconvenientes de este sistema cabe destacar que se trata de una tecnología muy

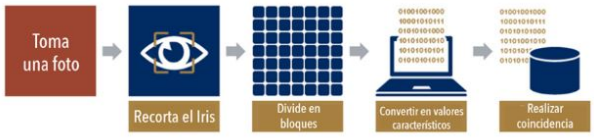


Figura 1.4: Diagrama de identificación biométrica mediante iris

costosa y es necesaria una cooperación por parte de la persona. El uso de gafas puede dificultar este tipo de reconocimientos.

Oreja: El reconocimiento biométrico mediante las orejas ha cobrado importancia en los últimos años aunque es un rasgo difícil de analizar en una fotografía debido a las condiciones variantes que existen al tomar una fotografía. Se pueden dar difrentes condiciones de iluminación, un enfoque de la cámara erróneo, etc. Para detectar y analizar las orejas en este tipo de autenticación se utilizan redes neuronales convolucionales.

Rostro: El reconocimiento de rostro es el proceso mediante el cual un sujeto puede ser reconocido mediante una foto o un vídeo, Kasar y cols. (2016). Este tipo de reconocimiento se utiliza para acceder a aplicaciones, salas de alta seguridad o a un sistema. Con los avances actuales este tipo de reconocimientos también se incluyen en smartphones. Es un tipo de identificación biométrica que se sirve de medidas corporales, en este caso la cara y cabeza, para verificar la identidad de una persona. El proceso de reconocimiento recoge un conjunto de datos biométricos únicos de cada persona asociados a su rostro y expresión facial para identificar, verificar y autenticar a una persona.
Los sistemas de reconocimiento facial capturan una imagen de entrada desde un dispositivo con cámara de forma bidimensional o tridimensional en función de las características del dispositivo.

Una vez capturada la imagen, se compara con una base de datos, previamente definida con imágenes del sujeto para que se pueda identificar. Este procedimiento necesita de una conexión a internet, dado que la base de datos no se puede encontrar en el dispositivo capturador debido a la gran cantidad de información que se requiere, si no que se aloja en servidores.

En esta comparación de rostros, se analiza matemáticamente y sin margen de error la imagen entrante y se verifica que los datos biométricos se corresponden con la persona que debe hacer uso del servicio o está solicitando un acceso.

Gracias al uso de las tecnologías de Inteligencia Artificial (IA) y DL, los sistemas de reconocimiento facial pueden funcionar con los más altos estándares de seguridad y fiabilidad. Con la integración de estos algoritmos y técnicas informáticas, el proceso puede llevarse a cabo en tiempo real.
En los siguientes capítulo el trabajo se centrará en las técnicas de estimación de calidad de imágenes para el reconocimiento facial.

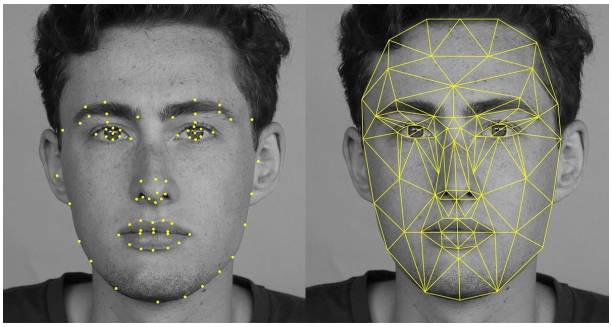


Figura 1.5: Obtención de puntos claves para el reconocimiento facial

Termogramas: Este tipo de reconocimiento se basa en una cámara termográfica la cual mide la temperatura corporal del sujeto sin tener ningún contacto físico. Estas cámaras detectan la energía infrarroja emitida, transmitida o reflejada por un cuerpo y convierten el dato en temperatura o un termograma que es un imagen térmica visible en la pantalla indicando la radiación calorífica emitida por un objeto, animal o persona. Una de las ventajas que aporta este reconocimiento es la efectividad que tiene alrededor de 99.7% y la comparación de rostros con mascarilla de alta precisión en menos de un segundo, Rani y cols. (2022).

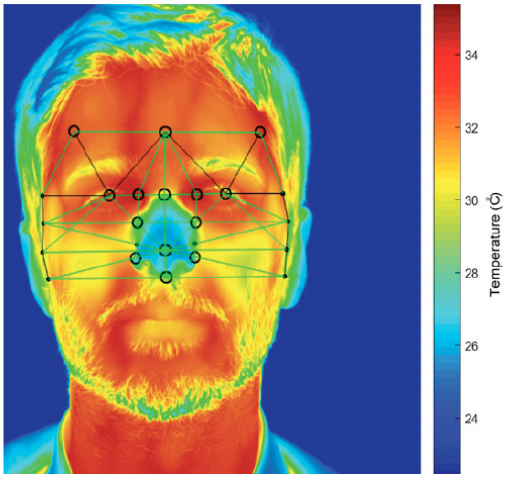


Figura 1.6: Obtención de la temperatura corporal para el reconocimiento del sujeto

Voz: El reconocimiento de voz es la capacidad de una máquina o programa para identificar palabras y frases en lenguaje hablado y convertirlas a un formato legible por máquina.

Reconocimiento de la huella dactilar: El reconocimiento de huellas dactilares es una técnica probada para verificar la identidad de los individuos y por lo tanto es una de las tecnologías biométricas más utilizadas. Una huella digital se compone de crestas y valles que están en la superficie del dedo. Los factores que dificultan este tipo de reconocimiento son varios tipos de ruidos como arrugas, manchas y agujeros que pueden dañar una imagen de huella digital. La calidad de la imagen de la huella digital no se puede mejorar, se requiere usar un algoritmo de mejora que puede mejorar la claridad de crestas y estructuras de valle de imágenes de huellas dactilares en las regiones recuperables y enmascarar las regiones irrecuperables.
Este reconocimiento permite almacenar en la base de datos más de un dedo por persona y compara la información alineando las características de los dedos del sujeto. Destacar que, la huella digital biométrica tiene una tasa de rechazo falso y una tasa de aceptación falsa relativamente baja, Bruno y cols. (2017).

Este TFG se centrará en el reconocimiento biométrico del rostro del que se estudiarán las técnicas existentes para la estimación de calidad de las imágenes para este tipo de biometría. En concreto se estudiarán las técnicas tradicionales y las técnicas que usan el aprendizaje profundo con Deep Learning.

2 Marco Teórico

En este capítulo se van a repasar conceptos teóricos en los que se basa el estudio realizado, revisando desde las bases de la estimación de calidad de imágenes, pasando por el concepto de Deep Learning o Aprendizaje profundo y su arquitecturas más populares, además de revisar los modelos de estimación da calidad más exitosos.

2.1 Técnicas de calidad para imágenes biométricas para el reconocimiento facial

Con el aumento de los sistemas biométricos basados en las imágenes faciales, ha surgido la necesidad de definir un estándar, que establezca los requisitos indispensables para que las imágenes que se utilizan tengan valor identificativo y que permita la interoperabilidad entre diferentes sistemas.

La norma internacional ISO/IEC 19794-5: 2005 y la ISO/IEC 19794-5: 2011 (Hernandez-Ortega y cols. (2021)), creadas por el Comité Internacional para los Estándares de Información y Tecnología, y adoptada por el Organización de Aviación Civil Internacional (ICAO) Ferrara y cols. (2012), tiene como objetivo establecer los requisitos de las imágenes de rostros para aplicaciones de reconocimiento de personas y definir un formato para el almacenamiento e intercambio de las fotografías.

Tras el incremento del uso de la biometría en aplicaciones diarias se han tenido que definir una serie de premisas que definen la calidad de una imagen. Cuando se habla de calidad de una imagen se refiere la usabilidad de la misma a la hora de llevar a cabo el reconocimiento facial. Para estimar la calidad de una imagen existen diferentes vertientes dentro del procesado de la imagen.

Una forma de realizar esta estimación es mediante el procesamiento de la imagen, de modo que, se comprueben aspectos como la pose, el ruido, la iluminación, etc de forma que para cada imagen se haga un procesado. Este tipo de estimación se realiza en paralelo lo que dificulta la implementación de esta técnica en tiempo real. No obstante, aunque es una técnica usada, cada vez tiene menos presencia en el mercado.

Otra forma de estimar la calidad de una imagen es mediante el aprendizaje profundo. Se trata de crear un modelo de Deep Learning el cuál se pueda entrenar con conjuntos de datos etiquetados para que sea capaz de saber si se trata de una imagen de buena o mala calidad en función de su usabilidad. Estos tipos de modelos devuelven un "score" que es el grado de calidad que el modelo indica que tiene una imagen. Con ese "score" obtenido se puede testear un modelo de reconocimiento facial y observar si el modelo es capaz de reconocer a

una persona según el grado de calidad que dispone la imagen.

2.1.1 Deep Learning

El aprendizaje profundo o Deep Learning es un subcampo del aprendizaje automático o Machine Learning.

El Deep Learning se encarga de entrenar a una computadora para que realice tareas o funciones como las haría una persona. Esto es debido a que el Deep Learning se construye mediante redes neuronales que asemejan a las conexiones de las neuronas en el cerebro humano. El aprendizaje se lleva a cabo entre las capas de la red. A medida que se avanza por la red neuronal el sistema es capaz de asimilar la información.

En un sistema de Deep learning, el sistema aprende de los datos brutos y puede aumentar su precisión si se le proporciona más datos. Es decir, cuantos más estímulos reciba del objetivo, mejor será su predicción. Por tanto, la diferencia entre un sistema de machine learning y un sistema de Deep learning radica en el nivel de implicación humana. Para enseñar a un sistema de machine learning hay que marcar cuáles son las características principales para reconocer el objeto. Sin embargo, en el caso de Deep Learning, no hace falta describir al sistema las características del objeto, sino que se le alimenta de imágenes para que aprenda por sí solo. Las imágenes que se le proporcionan al sistema suelen ir acompañadas de etiquetas para que el aprendizaje sea mejor.

2.1.2 Redes Neuronales

El Deep learning se basa en redes neuronales que intentan imitar el cerebro humano al analizar continuamente datos con una estructura lógica dada.

Las redes neuronales funcionan con un elevado número de unidades interconectadas denominadas neuronas y organizadas en capas. Normalmente una red neuronal está compuesta por tres partes: la capa de entrada, con neuronas que representan los campos de entrada; una o más capas ocultas; y una capa de salida con una o más unidades que representan el resultado computado por la red, Mehlig (2021).

Estas redes neuronales identifican patrones y clasifican diferentes tipos de información. Las diferentes capas de las redes neuronales sirven como filtro, yendo desde los elementos más generales a los más sutiles, aumentando la probabilidad de detectar y generar un resultado correcto. Por tanto, cuando un sistema de deep learning tiene que reconocer un objeto, lo compara con aquellos que ya conoce.

El deep learning se basa en el uso de las redes neuronales artificiales. Dentro de las redes neuronales de tipo profundo hay 3 tipos que son los más usados: Redes neuronales convolucionales (CNN), redes neuronales recurrentes (RNN), redes generativas antagónicas (GAN),

Sharkawy (2020).

2.1.2.1 Redes neuronales convolucionales

Las redes neuronales convolucionales son redes neuronales artificiales que han sido diseñadas para procesar matrices estructuradas, como imágenes. Se encargan de clasificar imágenes basándose en los patrones y objetos que aparecen en ellas, Coşkun y cols. (2017).

Por ello, las CNN suelen usarse en computer vision (visión artificial), ya que pueden operar con imágenes y no necesitan hacer un procesamiento previo. Esta cualidad hace que sean muy útiles para aplicaciones visuales de clasificación de imágenes.

Las redes neuronales convolucionales usan varias capas convolucionales (sobre unas 20-30 capas de media). Las capas convolucionales son un tipo especial de capa que, al apilarse unas encima de otras, permite reconocer formas más sofisticadas. Por ejemplo, con unas 3-4 capas convolucionales se pueden reconocer dígitos escritos a mano y con 25 se reconocen caras humanas. Este tipo de capas imitan a la estructura del cortex visual humano, ya que una serie de capas procesan una imagen entrante e identifican características cada vez más complejas.

2.1.2.2 Redes neuronales recurrentes

Las redes neuronales recurrentes son redes neuronales que usan datos secuenciales o series temporales. Este tipo de redes solucionan problemas ordinales o temporales, como la traducción de idiomas, reconocimiento de voz , procesamiento de lenguaje natural y captura de imágenes.

Las redes neuronales recurrentes difieren de otras redes neuronales artificiales en que tienen memoria. Las RNN toman información de inputs anteriores para influenciar los inputs y outputs actuales.

2.1.2.3 Redes neuronales antagónicas

Las redes generativas antagónicas consisten en usar 2 redes neuronales artificiales y oponiendo la una a la otra para generar nuevo contenido o datos sintéticos que pueden hacerse pasar por reales.

Una de las redes genera y la otra funciona como discriminadora. La red discriminatoria (también conocida como red antagónica) ha sido entrenada para reconocer contenido real y hace de controlador para que la red que genera contenido haga contenido que parezca real.

2.1.3 Técnicas tradicionales

Las técnicas tradicionales de estimación de calidad en imágenes para la biometría facial se basan en el procesamiento de la imagen. El procesamiento se define de acuerdo a la norma ISO/IEC 19794-5 ya que, establece los requisitos de las imágenes de rostros para aplicaciones de reconocimiento de personas.

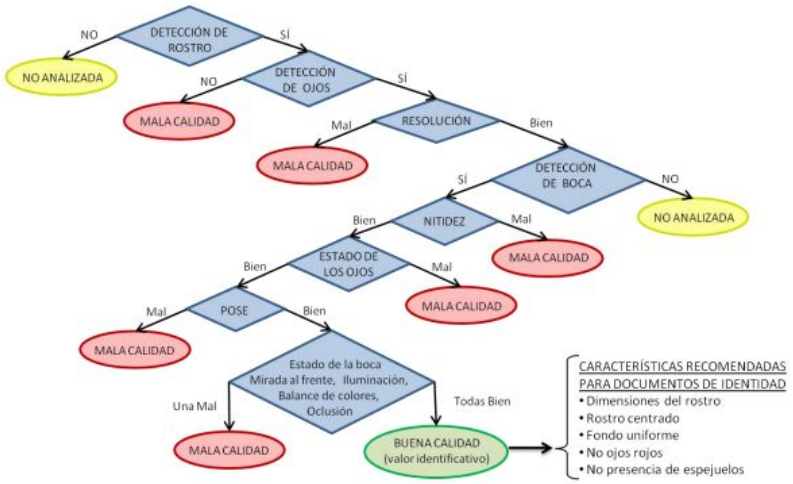


Figura 2.1: Ejemplo de arquitectura en árbol para técnicas tradicionales, Hernández-Durán y Plasencia-Calaña (2016)

Este tipo de técnicas tienen una estructura de árbol de modo que mediante una serie de estimadores de calidad se define la arquitectura, Hernández-Durán y Plasencia-Calaña (2016). En el caso de que la imagen pase por todos los identificadores y sea aceptada por todos ellos, la imagen tendrá una buena calidad para ser utilizada en el reconocimiento facial. La ventaja de este tipo de arquitecturas arbóreas es que en el caso de que un estimador detecte una baja calidad en el aspecto que está comprobando, la imagen es desestimada automáticamente sin tener que pasar por el resto de estimadores. Esto hace que la imagen sea válida o no, a diferencia de las técnicas recientes de DL que son capaces de estimar la calidad de una imagen con un grado de calidad, normalmente comprendido entre 0 y 1. Este hecho hace que también sea un inconveniente, ya que, estos tipos de estimadores no son capaces de dar un grado de calidad para una imagen.
Además, estas técnicas poseen el inconveniente de la percepción humana. Este tipo de estimadores se realizan mediante una serie de comprobaciones de ruido, brillo, contraste, etc que el ser humano define como correctas. Una correcta imagen para el ser humano no significa que la imagen sea correcta para un posterior reconocimiento facial. Este hecho hace que la figura del ser humano intervenga de manera muy directa sobre la estimación de la calidad. Con los modelos de DL que se muestran a continuación, se busca todo lo contrario. Con los modelos de DL se busca que la percepción del ser humano sea nula y los modelos sean capaces de estimar una calidad de imagen en función de su usabilidad en un posterior reconocimiento facial.

	Especificaciones
Resolución	La distancia entre los ojos debe ser mínimo de 60 píxeles
Nitidez	Las imágenes no deben estar borrosas ni desenfocadas
Estado de los ojos	El iris y la pupila deben estar claramente visibles
Pose	La imagen debe ser lo más frontal posible
Boca	Debe estar cerrada y con expresión neutral
Ojos	Deben estar mirando al frente
Iluminación	No deben aparecer sombras y la luz debe estar distribuida uniformemente
Color	Deben tener un contraste adecuado, sin estar saturados
Oclusiones	No pueden ocultar ninguna parte del rostro
Dimensión del rostro	El largo de la cabeza debe comprender entre el 70%-80% del alto de la imagen
Rostro	Centrado de manera horizontal. Los ojos entre el 50%-70% del alto de la imagen
Fondo	Debe ser uniforme con transición suave
Ojos rojos	No se admiten ojos rojos
Gafas	Deben de ser perfectamente visibles las pupilas sin reflejos

Tabla 2.1: Aspectos más importantes en el procesamiento de imágenes de técnicas tradicionales

Los estimadores que se utilizan en este tipo de técnicas, que son los definidos y adoptados por el ICAO, son los mostrados en la tabla 2.1.

Existen otros tipos de estructuras en los que tras aplicar un procesamiento a la imagen, se define un valor de calidad de la imagen independientemente de no cumplir con el ICAO en alguna de las premisas, pero no lo es usual.

El mayor inconveniente que poseen este tipo de técnicas tradicionales son el tiempo de procesado. La estimación de calidad debe ser un proceso rápido para poder verificar a una persona en tiempo real. El hecho de realizar comprobaciones sobre la imagen con procesamiento de imagen ralentiza el proceso y hace que la verificación sea más lenta, dificultando así la aplicación en tiempo real de este tipo de sistemas.

2.2 Técnicas mediante Deep Learning

Las técnicas mediante Deep Learning son las técnicas más novedosas y eficientes que existen actualmente. Se basan en el entrenamiento de un modelo, organizado en capas con varias neuronas por capa, con conjuntos de datos etiquetados. De esta forma, al estar etiquetados, el modelo es capaz de aprender mediante los pesos de las neuronas cuál es su función. En este caso, como es la calidad de la imagen, los modelos entrenan para obtener a la salida un valor de calidad apropiado para la imagen, Méndez-Vázquez y cols. (2012). Para el entrenamiento es importante que el modelo se entrene con todo tipo de imágenes relacionadas con lo que va a aprender. En este caso, es importante la variedad de imágenes en cuanto a pose, iluminación, ruido, exposición y todo tipo de factores que afecten a la imagen, Hernández y cols. (2016). Para analizar las técnicas existentes sobre este tipo de estimación de calidad, se realiza un estudio sobre el estado del arte para conocer los modelos con mejor funcionamiento en este

ámbito. Los modelos sobre las que se trabajará más detenidamente son los correspondientes a los trabajos como SER-FIQ Terhörst y cols. (2020), FaceQnet Hernandez-Ortega y cols. (2020), Magface Meng y cols. (2021) y SDD-FIQA Ou y cols. (2021).

2.2.1 Técnica de estimación de SER-FIQ

La técnica propuesta por SER-FIQ propone la obtención de calidad de una imagen a partir de un modelo de FR que debe estar entrenado con **Dropout**.
La técnica utilizada por SER-FIQ se basa en calcular las variaciones de los patrones estocásticos provenientes de subredes aleatorias (**Dropout**) en un modelo de reconocimiento facial. **Dropout** es un método que desactiva un número de neuronas de una red neuronal de forma aleatoria que se produce en la parte densa de la red no en la parte convolucional. En cada iteración de la red neuronal el dropout desactiva diferentes neuronas, las neuronas desactivadas no se toman en cuenta para el forward propagation ni para el backward propagation lo que obliga a las neuronas cercanas a no depender tanto de las neuronas desactivadas. Se define la magnitud de estas variaciones como una medida de robustez y, por lo tanto, calidad de imagen.
Es un algoritmo basado en la robustez de los patrones estocásticos para la calidad de imágenes debido a que la percepción humana puede no ser la correcta para la calidad de imágenes. Si una imagen produce pequeñas variaciones en los patrones significa que posee una mayor robustez lo que se traduce por una alta calidad de imagen. En cambio, si las variaciones de los patrones son altas, significa que la robustez de la imagen no es alta y se tratará de una imagen de baja calidad.

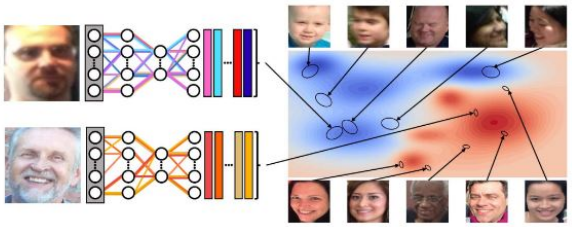


Figura 2.2: Ejemplo de estimación de calidad para el modelo de SER-FIQ, Terhörst y cols. (2020)

Una imagen como la de la figura 2.2 que produce pequeñas variaciones en los patrones (abajo a la izquierda), demuestra una alta robustez (áreas rojas a la derecha) y, por lo tanto, una alta calidad de imagen. Por el contrario, una imagen que produce altas variaciones en los patrones (arriba a la izquierda) provenientes de subredes aleatorias, indica una baja robustez (áreas azules a la derecha). Por tanto, se considera de baja calidad.

Para ver el funcionamiento del modelo se muestra la figura 2.3 que representa cómo se comportan las capas del modelo.

El algoritmo predice la calidad de una imagen de entrada "I" mediante un modelo de reconocimiento facial, Insightface en este caso. En la figura 2.3 se aprecia el uso del modelo de FR METRO que es el que utilizan en el artículo. La imagen se envía a diferentes subredes

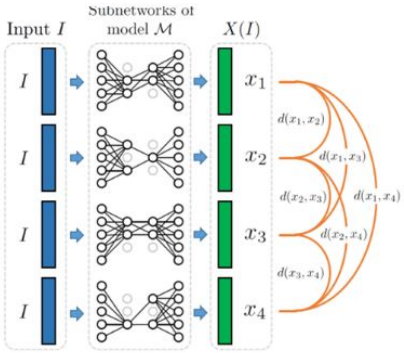


Figura 2.3: Ejemplo de estimación de arquitectura para el modelo de SER-FIQ, Terhörst y cols. (2020)

aleatorias del modelo de reconocimiento facial y cada red produce una inserción estocástica diferente, Xs. Por último, utilizando las distancias por pares se calculan las variaciones de las incrustaciones y se obtiene la calidad de la imagen I. En conclusión, mediante **Dropout**, se producen diferentes inserciones para una misma imagen, con estas inserciones SER-FIQ propone un método para calcular la robustez de la predicción para dicha imagen, es esta medida de robustez la que se toma como puntuación (score) de calidad. A partir de si las variaciones son altas o bajas, el modelo proporciona un "score" de calidad que define la calidad de la imagen tras pasar por el modelo.

2.2.2 Modelo de estimación de MagFace

La técnica de Magface propone una función de pérdida que permita entrenar modelos de FR para que ese modelo aporte un vector de características de la imagen, conocido como embeddings, y un valor de calidad para la imagen. Magface define el valor de calidad como la magnitud del embedding extraído. Se basa en la distancia a la cuál el modelo de FR sitúa la imagen en el hiperespacio. Cuanto más lejos se sitúe la imagen en el hiperespacio del origen, mayor calidad poseerá esa imagen.
El modelo de estimación de calidad de Magface introduce una técnica novedosa mediante la cual el modelo aprende sobre las distribuciones de las características de una imagen dada en el hiperespacio. Esta técnica se basa en la longitud que existe entre el origen y el centro de la clase de una imagen. Mediante el coseno del ángulo que forma el origen con las magnitudes calculadas, se calcula la longitud desde el origen hasta la muestra. Esta longitud aumenta debido al coseno del ángulo que se forma entre el origen y la muestra a analizar. Cuanto mayor sea la longitud que existe entre el origen y la magnitud de la imagen, más fácil será de reconocer esa imagen en un sistema de reconocimiento facial, es decir, mayor calidad poseerá esa imagen para poder ser usada correctamente.
Como se puede apreciar en la figura 2.4 al modelo se le pasan una serie de imágenes de diferentes calidades. El modelo de estimación de calidad es capaz de representar las magnitudes de cada imagen en el hiperespacio analizando la longitud desde el origen hasta la magnitud de la imagen analizada. Como se ve en la figura 2.4 una imagen que posee una mayor calidad,

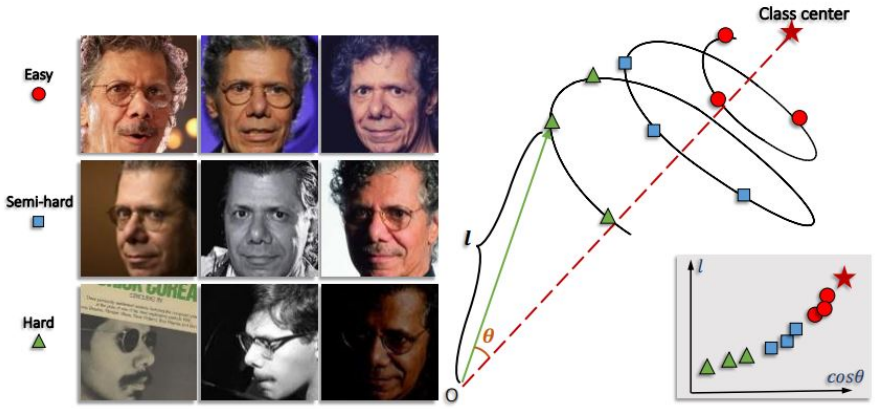


Figura 2.4: Ejemplo de estimación de la calidad para imágenes con el modelo de Magface, Meng y cols. (2021)

mayor usabilidad, posee una distancia respecto al origen superior que una imagen de baja calidad, baja usabilidad. El modelo agrupa las imágenes de mayor calidad en el hiperespacio cerca del centro de la clase mientras que las imágenes de baja calidad son agrupadas cerca del origen.
Una de las ventajas que posee este modelo de calidad de imágenes es que el cálculo de la calidad es totalmente independiente del ser humano. Muchos modelos anteriores han utilizado el etiquetado del ser humano para entrenar el modelo. Estas etiquetas hacían referencia a factores como la luminancia, la pose, los reflejos, etc. Estos factores muchas veces no eran sinónimos de buenos resultados ya que lo que para el ser humano puede ser una imagen no significa que para un modelo de reconocimiento facial lo sea. Al eliminar la presencia de etiquetas por parte del ser humano, surgió el modelo de Magface para poder obtener la calidad de una imagen solamente con las características adecuadas en referencia a un modelo de reconocimiento facial, sin la supervisión ni etiquetado del ser humano.

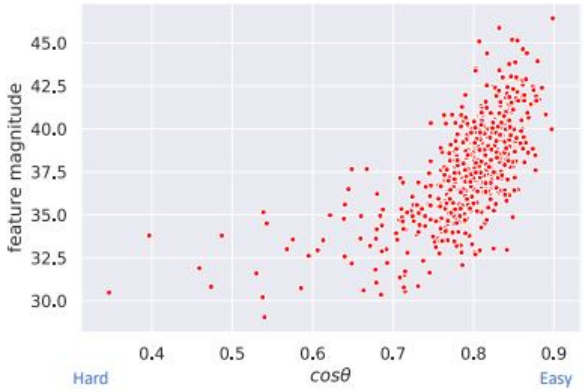


Figura 2.5: Ejemplo de distribución de la calidad para imágenes con el modelo de Magface, Meng y cols. (2021)

Como se puede apreciar en la figura 2.5, Magface optimiza la característica con margen

adaptativo y regularización en función de su magnitud. Bajo esta pérdida, es claro observar que existe una fuerte correlación entre las magnitudes de las características y sus similitudes de coseno con el centro de clase. Las imágenes situadas en la esquina superior derecha son las de mayor calidad. A medida que la magnitud se vuelve más pequeña, los ejemplos se desvían más del centro de la clase. Esta distribución respalda firmemente el hecho de que la magnitud de la característica aprendida por MagFace es una buena métrica para la calidad de la cara.

2.2.3 Modelo de estimación de FaceQnet

FaceQnet propone un modelo capaz de evaluar la calidad de una imagen en referencia a una imagen ICAO. El desarrollo del modelo se lleva acabo con la selección de una imagen válida según el ICAO y con el uso de diferentes modelos de FR. Con los diferentes modelos lo que pretenden es ver cómo una imagen de entrada se parece a la imagen ICAO que ellos seleccionan. Pretenden ver cómo de parecida es una imagen de entrada con una imagen ICAO que seleccionan como perfecta. A partir de esos resultados entrenan un modelo de DL capaz de obtener la calidad de una imagen. Este es el modelo que se va a utilizar para el análisis del trabajo.
FaceQnet es un sistema de evaluación de calidad de extremo a extremo sin referencia para el reconocimiento facial basado en Deep Learning. El sistema consta de una red neuronal convolucional para predecir la idoneidad de una imagen de entrada específica que vaya a ser utilizada en aplicaciones de reconocimiento facial.

Este modelo de estimación de calidad emplea Biolab-ICAO para etiquetar las imágenes que se proporcionan al modelo con información de calidad relacionada con su cumplimiento de la norma OACI. Las etiquetas de calidad de GroundTruth se obtienen utilizando FaceNet,Schroff y cols. (2015), para generar puntuaciones de comparación. Además, la red utilizada es una CNN basada en arquitectura ResNet-50. Este modelo emplea Biolab-ICAO para obtener automáticamente puntuaciones de cumplimiento de la OACI evitando así la introducción de sesgo por parte del ser humano.

En este modelo para la creación del GroundTruth se asume como perfecta una imagen ICAO perfectamente compatible. Al partir de esta suposición, se escoge una una imagen perfectamente compatible según el ICAO y se procede a realizar una serie de comparaciones. Se compara esta imagen asumida perfecta con otra imagen de entrada. Si la comparación entre ambas imágenes es alta, significa que la imagen de entrada es una imagen cercana a una imagen perfecta ICAO. En el caso contrario, si la comparación es baja, la calidad de la imagen es lejana a una imagen ICAO, lo que corresponde a una imagen de baja calidad. Este proceso se realiza junto a Biolab-ICAO que genera una puntuación entre 0 y 100 para cada imagen en relación con el cumplimiento de la OACI.

Como se puede ver en la figura 2.6 el modelo se entrena con un subconjunto de 300 sujetos de la base de datos VGGFace2. Se usan las puntuaciones de cumplimiento de la OACI para seleccionar una imagen de galería para cada usuario. Después de eso, se emplea FaceNet, modelo previamente entrenado para la extracción de características, y se obtienen todas las puntuaciones acopladas usando la distancia euclidiana entre los patrones de las imágenes de la galería compatible con la OACI y el resto de imágenes del mismo sujeto. Las puntuaciones de comparación se utilizan como medidas reales de calidad de las imágenes que no son de la

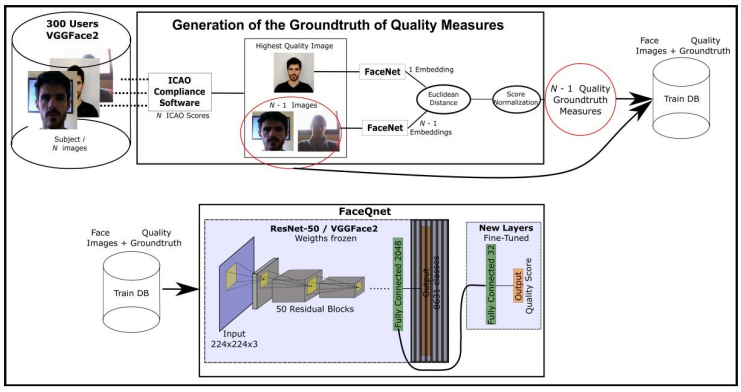


Figura 2.6: Ejemplo de entrenamiento del modelo FaceQnet, Hernandez-Ortega y cols. (2020)

OACI. FaceQnet se basa en la arquitectura ResNet-50, reemplazando la capa de clasificación con dos nuevas capas diseñadas para la regresión. Se realiza el entrenamiento solo de las nuevas capas manteniendo congelados los pesos del resto de capas ya que el modelo es una variación de FaceNet.

2.2.4 Modelo de estimación de SDD-FIQA

El modelo de estimación de calidad SDD-FIQA es un nuevo método FIQA no supervisado que incorpora la distancia de distribución de similitud para la evaluación de la calidad de la imagen facial. La propuesta de SDD-FIQA es el cálculo de los scores de calidad de una imagen para ser capaces de implementar un modelo que estime la misma calidad de imagen que calculan. Lo llevan a cabo con la información interclase. La mayoría de los trabajos previos ignoran la valiosa información de la interclase, que es para estimar la reconocibilidad de la imagen de la cara. Este modelo se basa en que una imagen facial de alta calidad debe ser similar a sus muestras intraclase y diferente a sus muestras interclase. SDD-FIQA genera pseudoetiquetas de calidad mediante el cálculo de la Distancia de Wasserstein entre las distribuciones de similitud intraclase y las distribuciones de similitud interclase. Con estas pseudoetiquetas de calidad, se entrena una red de regresión para la predicción de calidad.

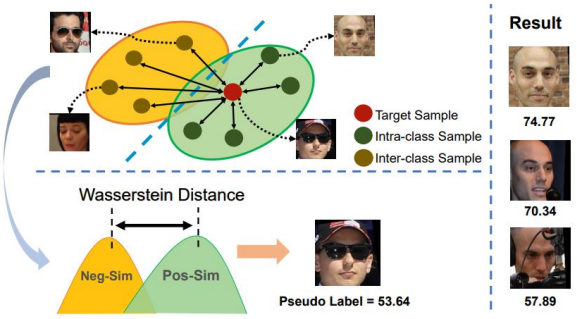


Figura 2.7: Cálculo del valor de calidad mediante la distancia de Wasserstein, Ou y cols. (2021)

Como se puede ver en la figura 2.7 el modelo considera simultáneamente las similitudes

de la imagen (punto rojo) con las muestras intraclase (señaladas con puntos verdes) y con muestras intraclases (puntos amarillos). La distancia de distribución entre Pos-Sim y Neg-Sim es calculada como la pseudo-etiqueta de calidad. A la derecha de la figura se muestran los resultados de calidad obtenidos.

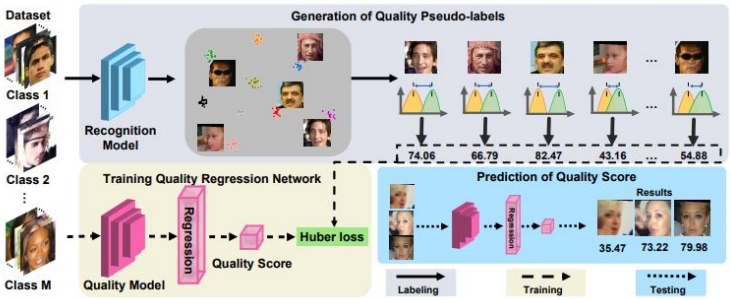


Figura 2.8: Estructura del modelo de SDD-FIQA, Ou y cols. (2021)

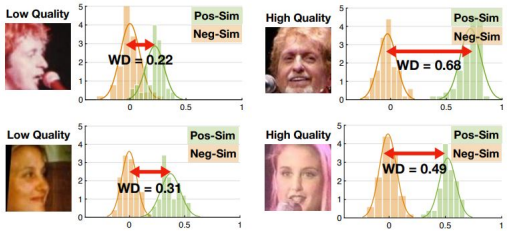


Figura 2.9: Obtención de calidad con SDD-FIQA, Ou y cols. (2021)

En la figura 2.8 se refleja la estructura del modelo de estimación de calidad SDD-FIQA. Esta estructura representa el paso de de la imagen a través del modelo obteniéndose las distancias Pos-sim y Neg-sim representadas sobre la figura 2.9. Mediante la distancia de Wasserstein se obtiene la etiqueta de calidad de la imagen y durante el entrenamiento, la red de regresión de calidad se entrena bajo la restricción de pérdida de huber para FIQA.

2.3 Modos de representar los resultados obtenidos

Un sistema de reconocimiento biométrico se basa en que, dada una muestra, el sistema biométrico es capaz de tomar la decisión correcta, tanto sea en verificación (si es o no es la persona) como en identificación (de qué persona se trata) de una persona.
En realidad, un sistema biométrico es un sistema de reconocimiento de patrones que inevitablemente toma decisiones incorrectas. Por esta razón, es importante saber por qué un sistema biométrico comete errores e interpretarlos para saber la magnitud de los errores.

Los métodos para representar los resultados que se van a explicar en el apartado 2.3.1 van a ser utilizados para visualizar los resultados tras el paso del conjunto de datos que se va a utilizar en el trabajo por el modelo de Face Recognition. A este modelo se le pasarán todas las

imágenes del conjunto y se irán filtrando por similitudes entre sujetos y valores de calidad que se estimen mediante el desarrollo del trabajo. Con los siguientes métodos de representación se podrá ver como afecta la calidad de la imagen que se obtiene de cada modelo y como de bueno será ese modelo junto al modelo de Face Recognition.

2.3.1 FNMR y FMR

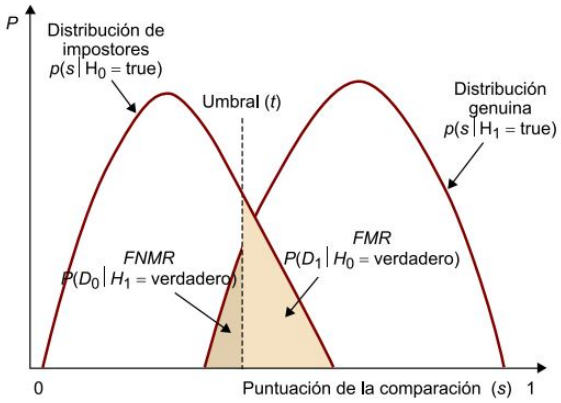


Figura 2.10: Representación de FNMR y FMR en función de un umbral t

Para poder evaluar la precisión de un sistema de autenticación biométrica es necesario obtener un número elevado de comparaciones entre muestras y plantillas de la misma persona (distribución genuina) y un número de comparaciones elevado entre muestras y plantillas de diferentes personas (distribución impostora). Esta recopilación de datos también es conocida como matriz de similitudes en las que se obtienen los pesos de todas las imágenes de un conjunto de datos enfrentadas contra todas las imágenes del conjunto. Esto hace que cuando se esté tratando la misma imagen, el valor de similitud sea alto mientras que cuando son de personas diferentes este valor sea más pequeño.

El FNMR es la tasa a la que las personas autorizadas no son reconocidas durante una comparación de características. Es decir, son la misma persona pero el sistema indica que la similitud es inferior al umbral fijado para el reconocimiento. La FMR es la tasa a la que las personas no autorizadas sean falsamente reconocidas durante una comparación de características. Es decir, no son la misma persona pero el sistema indica que la similitud supera el umbral fijado para el reconocimiento.

Como se puede apreciar en la figura 2.10, las distribuciones de FNMR y FMR varían según el umbral definido. Si se reduce el umbral 't' para volver al sistema más tolerante a las variaciones de entrada y al ruido, entonces la FMR aumenta. Esto quiere decir que el sistema es más vulnerable y menos seguro. Del contrario, si se aumenta el valor del umbral 't' para hacer el sistema más seguro, se aumentará la FNMR y disminuye la FMR. Esto significa un aumento de seguridad en el reconocimiento facial y una menor vulnerabilidad. La definición del umbral 't' se explica en el apartado 5.4

3 Objetivos

El objetivo principal de este TFG es analizar y estudiar las técnicas existentes para determinar la calidad de una imagen facial biométrica. Estas técnicas pueden ser mediante el procesamiento digital de la imagen o mediante Deep Learning. La técnica de Deep Learning utiliza aprendizaje profundo usando redes neuronales mediante las cuales se entrena a los modelos para que sean capaces de predecir si la calidad de la imagen es buena o es mala a partir de una serie de entrenamientos. La evaluación de la calidad del rostro tiene como objetivo estimar la idoneidad de una imagen facial para su reconocimiento. A lo largo del trabajo se explicará el uso del Deep Learning frente al procesamiento digital tradicional de las imágenes y cómo afecta esta técnica a la autenticación biométrica.

El segundo objetivo del TFG es determinar los métodos y métricas más adecuados para la estimación de la calidad de una imagen facial mediante la obtención de la calidad de las imágenes con modelos de Deep Learning. Estos modelos se utilizan ya entrenados para ser capaces de obtener el máximo rendimiento a esa puntuación de calidad para un posterior reconocimiento facial. Se realizan inferencias sobre tres modelos diferentes que usan Deep Learning y se obtienen los valores de calidad de las imágenes para cada modelo.

El tercer objetivo del TFG es la experimentación de los modelos y la proposición de las medidas más eficientes para cada modelo.

El cuarto objetivo es realizar un estudio sobre cómo los valores de calidad que se obtienen de cada imagen afectan a un modelo entrenado de Face Recognition. Para ello, se crea una aplicación para calcular las métricas de los modelos de estimación de calidad.

El quinto objetivo, y último, es validar la propuesta de trabajo sobre datasets públicos y el estudio de cada modelo por separado en función del sesgo que posee. El sesgo de un modelo de estimación de calidad puede afectar a la calidad de las imágenes, lo que afecta a un posterior reconocimiento facial.

Este trabajo pretende poner de manifiesto la importancia del factor calidad a la hora de utilizar una imagen para el reconocimiento facial. Es muy importante que esa calidad sea la correcta y se encuentre dentro de las normas en las que se basa el ICAO para poder aportar la seguridad que se pretende mediante estos sistemas de autenticación biométrica. El ICAO es un organismo cuya misión es “Servir” de foro mundial de Estados para la aviación civil internacional. El ICAO elabora políticas y normas, lleva a cabo auditorías de cumplimiento, realiza estudios y análisis, presta asistencia y crea capacidad en materia de aviación mediante otras muchas actividades y la cooperación de sus Estados miembros y partes interesadas. Es el organismo que elabora las normas de referencia a las que se adhieren muchos países para

asegurar una cierta calidad en el transporte aéreo. A su vez el ICAO basa sus directrices que recoge en la norma en las normas ISO. En este caso son la ISO/IEC 19794-5:2005 y la ISO/IEC 19794-5:2011.

4 Metodología

Para llevar a cabo el análisis completo del trabajo se utiliza el modelo de FR de Insightface disponible públicamente, Deng, Guo, Liu, y cols. (2020). El objetivo es ver como en base a un modelo de FR, se definen los scores de calidad en función de como de buena es esa calidad para las predicciones del modelo. Como se menciona en el apartado 2, las técnicas de Magface y SER-FIQ utilizan sus propios modelos de FR para llegar a obtener una puntuación de calidad de una imagen de entrada. De este modo se consigue un mejor análisis ya que se analizan los modelos propios con el modelo de Insightface. Para el caso de SER-FIQ, se obtienen los valores de calidad del conjunto de datos mediante el modelo de Insightface, ya que, este modelo es entrenado con **Dropout**. En este caso, al obtener los valores de calidad con el mismo modelo al que se compara posteriormente para el análisis de métricas, se estaría analizando la técnica de SER-FIQ en 'same model', en el mismo modelo.
Por otro lado, se utiliza la técnica de Magface. Magface define una técnica mediante la cual se define la calidad de una imagen acorde a la situación de la misma en el hiperespacio según su modelo de FR. El modelo de FR de Magface utilizado también se encuentra disponible públicamente. Como la técnica de Magface ya utiliza un modelo de FR para obtener la calidad de una imagen, se analiza el comportamiento del sistema cuando el modelo de reconocimiento facial es el mismo que el utilizado para la extracción de calidad, 'same model'. Además, para ver cómo afectan los valores de calidad de Magface sobre un modelo neutral, se analizan las métricas de los valores de calidad obtenidos con Magface sobre el modelo de FR de Insightface, 'on top model'. Magface es un modelo de FR en sí, pero se compara con el modelo de Insightface para ver cómo afectan los scores de calidad obtenidos según el modelo de FR usado.
También se hace uso del modelo de FaceQnet. Este estimador de calidad no hace uso de ningún modelo de FR para la estimación de calidad de una imagen. La evolución de las métricas de este modelo sobre el conjunto de datos utilizados también es 'on top model'.

En conclusión, para las siguientes secciones se hará uso de un modelo de reconocimiento facial, el de Insightface, y tres técnicas diferentes de estimación como son SER-FIQ, FaceQnet y Magface para ver como estas técnicas de estimación de calidad afectan a un sistema de reconocimiento facial.
Destacar que Magface y SER-FIQ son técnicas que estiman la calidad de la imagen con sus propios modelos de FR.

4.1 Descripción del trabajo

Para obtener un buen análisis sobre un modelo de estimación de calidad de autenticación biométrica, se necesita un modelo de Reconocimiento Facial (Insightface). Este modelo de FR

es el encargado de mostrar cómo de buena es la estimación de calidad de un modelo sobre una imagen mostrando la usabilidad que tiene esa imagen posteriormente para el reconocimiento facial. El objetivo principal de los modelos de reconocimiento facial es la extracción de características de de una imagen , a la que se denomina patrón o embedding. Cada patrón posee un valor característico para poder comparar una muestra facial tomada con cualquier imagen. El funcionamiento del modelo se basa en la extracción de características de cada imagen para la posterior creación de una matriz de similitudes. Una matriz de similitudes engloba todos los datos del conjunto de datos proporcionado al modelo de FR. Esta matriz engloba todas las comparaciones posibles de imágenes dentro del conjunto de datos con valores de -1 a 1. Las filas de esta matriz pertenecen a cada imagen del conjunto y las columnas son las encargadas de recopilar los valores de similitud para las imagenes de target de todo el conjunto siendo la primera columna la primera imagen para comparar, la segunda columna la segunda imagen para comparar y así sucesivamente. De este modo, si el conjunto posee 20 mil imágenes, la matriz de similitudes tendrá un tamaño de 20.000 x 20.000 obteniéndose así 400 millones de similitudes para el conjunto. Los valores de las similitudes están comprendidos entre -1 y 1 ya que las características extraídas por el modelo de FR están normalizadas. Es de prever que las similitudes que se encuentren cercanas a 1 o siendo 1 serán para un par de imágenes que según el modelo de FR sí se tratan de la misma persona. En cambio, si las similitudes son de valor bajo, el modelo estará interpretando que no se trata de la misma persona. Muchas veces se obtendrán valores de similitud altos cuando realmente no se trate de la misma persona y otras veces se obtendrán valores bajos cuando sí se trate de la misma persona. Es aquí donde entran los modelos de estimación de calidad. El objetivo de estos modelos es que cada par de imágenes comparadas por el modelo de FR contengan una similitud acorde con si son o no la misma persona. Para ello, se necesita que las estimaciones de calidad sean lo mejores posibles. Además, el modelo de FR debe ser adecuado para poder extraer las características de cada imagen de forma adecuada siendo bien entrenado previamente. Destacar que se hace uso de la matriz de similitudes para comprobar el funcionamiento del sistema en base al conjunto de datos utilizado. Si el sistema se basara en un sistema real la imagen de entrada se compararía con la base de datos de imágenes que el sistema tuviese almacenada.

Para llevar a cabo el trabajo se dispone de tres modelos de estimación de calidad mediante DL. El primer paso a realizar es el estudio de los modelos a utilizar y la realización de inferencias sobre cada modelo. Una inferencia a un modelo es el proceso mediante el cual un modelo de DL es capaz de extraer una serie de resultados a partir de una entrada definida al modelo. El fin de realizar esa inferencia es obtener un conjunto de datos, imágenes en este caso, procesado para poder proporcionar al modelo como entrada y obtener un valor de calidad de cada imagen según el modelo. El valor de calidad vendrá definido por la usabilidad que estime el modelo que posee esa imagen. Se realizará una inferencia sobre el modelo de estimación de calidad de FaceQnet, Magface y SER-FIQ. En el apartado 5 se detallan todas las inferencias.

Para analizar los modelos de estimación de calidad se utiliza un conjunto de datos (Apartado 4.2) sobre el cual ningún modelo se ha entrenado. Ni el modelo de reconocimiento facial ni los modelos de estimación de calidad. El uso de un conjunto de datos imparcial hace que el análisis sea más completo a la hora del análisis de la usabilidad de los modelos de estimación

y de reconocimiento facial, ya que, los modelos de Deep Learning siempre funcionan mejor con conjuntos de datos con los que han sido entrenados que con conjuntos nuevos. Se puede analizar cómo funciona cada modelo de los mencionados sin tener ellos ninguna referencia sobre el conjunto de datos. Esto permite ver el funcionamiento de cada modelo de estimación de calidad y su usabilidad para una posterior autenticación biométrica.

Con el fin de realizar un análisis más profundo de los modelos de calidad, se realiza el análisis sobre los tres modelos de calidad en base al modelo de FR y se realiza el análisis del modelo de Magface en base a su propio modelo. Esto es posible debido a que, en la extracción de la puntuación de calidad, el modelo de Magface también proporciona un vector de características de la imagen mediante el cual se puede crear una matriz de similitudes nueva para contrastar resultados. Al igual que un modelo de estimación de calidad funciona mejor si se ha entrenado con el mismo conjunto de datos que con el que es testeado, un modelo de reconocimiento facial también funciona mejor si el valor de calidad de cada imagen ha sido extraído con su valor de calidad. De esta forma, se analiza el funcionamiento de un sistema de reconocimiento facial en base a tres modelos externos él y se analiza el funcionamiento del modelo de reconocimiento facial de Magface en base a su propio modelo de estimación de calidad. Todos los modelos con un conjunto de datos desconocido respecto a sus entrenamientos.

Además del análisis del funcionamiento del modelo de FR para un valor de calidad determinado, se analiza como afecta a la seguridad de un sistema biométrico. La seguridad de un sistema biométrico viene definida por el comportamiento del modelo de FR mediante el análisis de las curvas de FNMR y FMR explicado en el apartado 2.3.1. Se analiza la seguridad y la vulnerabilidad que presenta cada modelo de estimación de calidad en función de cada modelo de reconocimiento facial empleado según las similitudes extraídas.

También se analiza la distribución de los 'scores' de calidad que cada modelo de estimación extrae para corroborar el funcionamiento de un posterior reconocimiento facial. Cada modelo presenta una distribución diferente de las puntuaciones de calidad del conjunto de datos en función de cómo predice el modelo de estimación la usabilidad de cada imagen.

Por último, se realiza un estudio sobre cada modelo de estimación de calidad en función de los 'scores' estimados para cada imagen según las características de la imagen. Este proceso es conocido como el análisis de sesgo de un modelo. Este análisis muestra la probabilidad de la estimación de los 'scores' en función del valor del 'score'. Este análisis permite analizar si un modelo funciona mejor para un tipo de imágenes que para otro. En este estudio, se realiza el estudio del sesgo según el color del pelo de la persona, las oclusiones faciales que pueda presentar o la edad y el género de la persona de la imagen. De esta forma se observa que modelo de estimación de calidad es menos discriminatorio y es capaz de estimar la calidad de forma correcta para todo tipo de imágenes.

4.2 Conjunto de datos

El conjunto de datos utilizado para el análisis es el conjunto de CelebFaces (CelebA) Z. Liu y cols. (2015), disponible públicamente para fines de investigación no comercial.

Se trata de un conjunto de datos de atributos faciales a gran escala con más de 200 mil imágenes de celebridades con 40 anotaciones de atributos por imagen.
Las imágenes en este conjunto de datos cubren grandes variaciones de pose y fondos heterogéneos. Posee imágenes de calidades muy diversas siempre entendiendo la calidad como la usabilidad de la imagen
El conjunto de datos contiene 10.177 identidades diferentes y 202.599 imágenes de rostros de estas identidades. Además, cada imagen contiene 5 anotaciones sobre puntos de referencia de la cara.
CelebA se puede emplear tanto como conjunto de entrenamiento como prueba para las siguientes tareas de visión por computadora: reconocimiento de atributos faciales, reconocimiento facial, detección de rostros, localización de puntos de referencia (o partes faciales) y edición y síntesis de rostros.

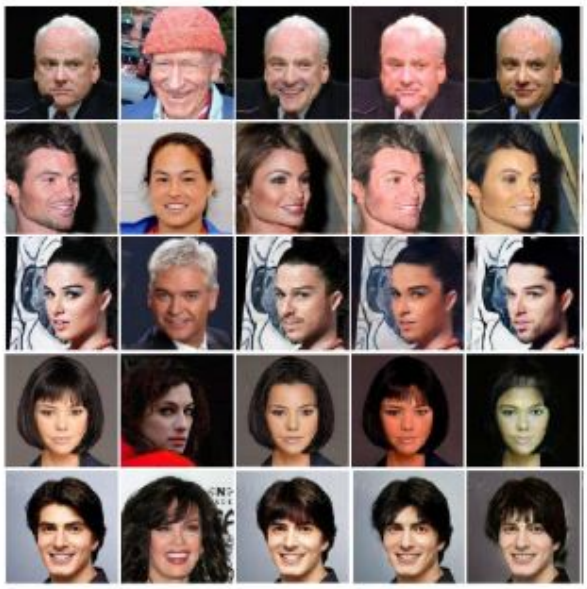


Figura 4.1: Muestra de la diversidad de imágenes y poses que se incluyen en el conjunto.

Como el objetivo del análisis es estudiar cómo afecta un determinado valor de calidad a un modelo de reconocimiento facial, se crea un conjunto de datos más pequeño debido al coste computacional, ya que, no se utiliza GPU durante el trabajo.
Se crea un conjunto que consta de 10 mil imágenes con un total de mil identidades. Se realiza así para poder contener al menos 10 imágenes de cada identidad. De este modo, los resultados serán más fiables a la hora de hacer el estudio del modelo de reconocimiento facial que si se creara un conjunto de datos de 10 mil imágenes con 5 mil identidades. Las identidades se seleccionan aleatoriamente haciendo uso de la librería numpy de Python para asegurar la di-

versidad de rasgos faciales en el conjunto. Una vez seleccionadas las identidades se seleccionan 10 imágenes por identidad. El tener un mayor número de imágenes por identidad permitirá obtener mejores resultados al realizar el reconocimiento facial debido a que los resultados obtenidos no se basarán en una única imagen, se tendrán 10 valores de calidad con los que se podrán analizar 10 similitudes de la misma persona dentro del conjunto de datos y ver como afecta ese valor de calidad en un posterior reconocimiento facial.
Debido a que la elección de identidades e imágenes del conjunto de CelebA es aleatorio para que el factor humano no influya en los resultados y el análisis sea los más real posible, existen identidades para las cuales el conjunto no posee 10 imágenes para cada identidad. Por esta razón el tamaño definitivo del conjunto de datos sobre el cual se va a realizar todo el análisis queda reducido a 9811 imágenes.

5 Desarrollo

En este capítulo central del trabajo se presentan las inferencias realizadas sobre los diferentes modelos del estado del arte y se presentan las arquitecturas para la obtención de métricas que permiten una estimación de calidad de las imágenes.

5.1 Preparación del conjunto de datos para las inferencias de los modelos

Para la realización de las inferencias es idóneo que las imágenes se encuentren recortadas y alineadas para un mejor análisis de la imagen que solo tenga en cuenta la región facial exclusivamente. Debido a que todos los modelos funcionan mejor con las imágenes recortadas y con un tamaño de 112x112 píxeles, se realiza esta inferencia para disponer del conjunto de datos listo para todas las inferencias posteriores. De no realizarse esta inferencia, en cada modelo se debería de extraer la región facial de cada imagen antes de pasar la imagen al modelo una por una lo que conllevaría un mayor tiempo para la estimación de los 'scores' de calidad y un mayor coste computacional.
Se procesa el conjunto de datos utilizado con un modelo de detección facial. A las imágenes utilizadas se les aplica un modelo de Deep Learning de reconocimiento de caras el cual se encarga de identificar cinco puntos clave de la cara y recortar la cara en un tamaño de 112 x 112 píxeles centrados en la cara. El modelo utilizado para este procesamiento es el modelo de detección de rostros de MTCNN de FaceNet para TensorFlow implementado con Keras, Schroff y cols. (2015). El detector devuelve una lista de objetos JSON en las que se incluyen el cuadro delimitador del rostro, la probabilidad de que el cuadro delimitador coincida con la cara y los puntos clave que detecta en el rostro. Estos puntos clave son el ojo izquierdo, el ojo derecho, la nariz y dos puntos de la boca (izquierda y derecha). El modelo utilizado se encuentra disponible públicamente.

```
input_imgs=r"D:\base de datos celeb a\\milidentidades"
dst_dir= r"D:\base de datos celeb a\\milidentidades_recortadas"
# alineación de la imagen
files = os.listdir(input_imgs)
for file in files:
    imagen=input_imgs+"/"+file
    imagen=cv2.imread(imagen)
    aligned_img = aplicar_mtcnn(imagen)
    if aligned_img is None:
        cv2.imwrite(os.path.join(dst_dir,file),imagen)
    else:
        cv2.imwrite(os.path.join(dst_dir,file),aligned_img)
```

Figura 5.1: Inferencia para el modelo de detección facial.

En la figura 5.1 se muestra la preparación de los datos para cada imagen a la hora de

aplicarle el detector facial. La función aplicar_mtcnn() contiene la extracción de puntos característicos de la cara y su posterior recorte. Finalmente las imágenes alineadas y recortadas se almacenan en otra carpeta para su posterior uso.
Este procesado hará que los modelos de Deep Learning que se van a utilizar funcionen mejor y obtengan un "score" de calidad más fiel a la realidad ya que las imágenes estarán centradas en la cara del sujeto. El objetivo de esta aplicación del detector facial es que la imagen de entrada al modelo sea totalmente de la región facial que necesita el modelo para extraer correctamente el valor de calidad. Para este trabajo como tanto las inferencias según las técnicas de SER-FIQ y Magface necesitan el mismo tamaño de imagen de entrada, realizando este procesado se consigue un ahorro computacional a la hora de estimar la calidad de las imágenes. Disminuyendo así el tiempo de estimación al no tener que recortar la imagen el modelo de predicción cada vez que se estima la calidad de una imagen.
Una vez recortadas las imágenes se almacenan para poder realizar las inferencias sobre los

Figura 5.2: Imágenes del conjunto de datos sin el detector facial del procesado

modelos.
Durante el estudio se van a realizar cuatro inferencias distintas sobre modelos de Deep

Figura 5.3: Imágenes del conjunto centradas en el rostro con un tamaño de 112 x 112

Learning. Se realizarán tres inferencias relacionadas con la estimación de calidad de imágenes en las que se usarán los modelos pre-entrenados de FaceQnet, MagFace y SER-FIQ, disponibles públicamente. La última inferencia que se realizará será sobre un modelo de Face Recognition, en este caso el modelo de InsightFace también disponible públicamente. Para obtener los diferentes resultados que se necesitan, cada modelo necesitará un procesado de imagen antes de evaluar la imagen con el modelo. Estos procesados ya son más simples e individuales para cada modelo. Se comentan en los apartados de las inferencias de cada modelo.

5.2 Realización de las inferencias sobre modelos de DL

5.2.1 Inferencia sobre el modelo de FaceQnet

Actualmente se encuentran disponibles públicamente dos versiones de FaceQnet, FaceQnet v0 (Hernandez-Ortega y cols. (2019)) y FaceQnet v1 (Hernandez-Ortega y cols. (2020)), siendo FaceQnet v1 la versión más reciente de FaceQnet. Para el estudio que se realiza se utiliza la versión 1 de FaceQnet mediante un modelo pre-entrenado que se encuentra disponible públicamente ('FaceQnet_v1.h5'). Este modelo es capaz de estimar la calidad de una imagen a través de las capas de neuronas que forman el modelo como un único valor entre 0 y 1. Para el uso del modelo pre-entrenado se necesita hacer uso de Keras y TensorFlow además de una serie de librerías de Python para el tratamiento de las imágenes y los modelos.

El procesado que se le aplica a cada imagen antes de ser utilizada por el modelo de FaceQnet se basa en la lectura de todo el conjunto de imágenes leyéndolas con la función imread() de OpenCV y el posterior cambio de tamaño de la imagen a 224 x 224 píxeles ya que este modelo ha sido entrenado y definido para este tamaño de imágenes. Mediante la función resize() de OpenCV se lleva a cabo el cambio de tamaño. Por último, todas las imágenes se almacenan en un array para su posterior uso.

Para obtener los resultados se utiliza el conjunto de datos explicado anteriormente el cual ha sido procesado para extraer la región del rostro. Tras cargar el conjunto de imágenes con el path correspondiente y aplicar el procesado anteriormente explicado, se carga el modelo mediante la función load_model() de Keras. Este modelo viene codificado en UTF-8, es necesario decodificar el modelo para hacer uso de él. Una vez se dispone de las imágenes almacenadas en un array y el modelo cargado, se procede a obtener la estimación de la calidad en cada imagen.

Mediante la función predict(), válida sólo para modelos implementados con Keras, se obtiene un valor de calidad para cada imagen del conjunto almacenando estos valores de calidad en un array y guardándolos mediante numpy.save() en archivos binarios en los que poder disponer de los valores de calidad posteriormente. Los valores de calidad deben de estar almacenados para obtener los resultados con el modelo de Face Recognition.

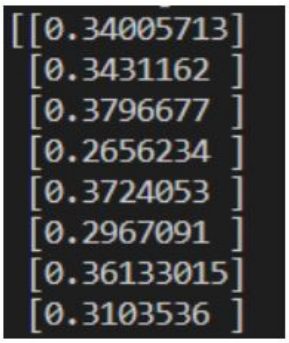


Figura 5.4: Demostración de la obtención de puntuación de calidad mediante FaceQnet. Cada valor pertenece a una imagen.

5.2.2 Inferencia sobre el modelo de MagFace

Para realizar la inferencia sobre el modelo de Magface es necesario tener preparado todos los datos que se van a necesitar. Primero se definen los argumentos a nivel global ya que son utilizados para toda la inferencia. Entre estos argumentos globales se incluye la arquitectura del modelo a utilizar (en este caso se trata de iresnet100), el archivo o lista en el cual se va a indicar que imágenes leer (9811 imágenes en la lista), el tamaño del array de embeddings que se va a calcular (se fija a 512 valores), el modo de operación lo definimos como el modo cpu ya que por limitaciones no se puede hacer uso de la gpu y por último, se define el modelo que se va a utilizar. En este caso, el modelo se carga desde un archivo .pth disponible públicamente en el repositorio de MagFace. El modelo utilizado es el "magface_iresnet100_quality.pth" entrenado para extraer la estimación de calidad de imágenes para aplicaciones biométricas. El modelo está implementado en Pytorch y se hace uso del mismo.

Una vez se dispone de las variables globales que definen la inferencia, se procesan las imágenes que se van a proporcionar al modelo de estimación de calidad de Magface. Puesto que el número de muestras que se le pasa al modelo es de 256 imágenes (*batch size*=256 debido a que es el valor típico para este tipo de inferencias), en el procesado de imagen se carga una lista con 256 imágenes las cuales deben de tener un tamaño de 112x112 píxeles (Ya poseen este tamaño del recorte facial). Se lee la lista con todas las imágenes y se leen mediante la función *imread()* de OpenCV. Además se le da la vuelta a la imagen que se lee para pasarle al modelo tanto la imagen original como la volteada para obtener mejores resultados.

Una vez definido el modelo que se utiliza, se carga en una variable mediante *"DataParallel()"* de Pytorch. Con los argumentos anteriormente explicados, se define el tamaño de los embeddings y se hace uso del modelo mediante la función *"eval()"* de Pythorch para modelos implementados con Pytorch. Con el conjunto de imágenes ya cargados según la lista definida en los argumentos globales, se evalúa el modelo. Al modelo se le pasan las imágenes de una en una las 9811 imágenes de 112 x 112 píxeles. Una vez termina el modelo con el análisis, se obtiene una matriz de 9811 filas que son las imágenes evaluadas, y 512 columnas, que son los embeddings que ha calculado el modelo para cada imagen. Hasta ahora no se dispone de la calidad de la imagen. Para calcular la calidad de la imagen, se recorren todas las posiciones de cada fila y se obtiene un único valor de calidad mediante la función *"linalg.norm()"* de la librería Numpy de Python. De esta forma, se obtiene un índice de calidad para cada imagen. MagFace obtiene la calidad de la imagen como el módulo de los vectores, de este modo, su valor de calidad no se encuentra definido en ningún rango. Esto es debido a que el modelo tiene una cierta libertad cuando entrena no limitando los valores ya que en este modelo lo que realmente importa es la calidad relativa de la imagen respecto al resto de imágenes.

Debido al análisis tanto del 'score' obtenido con Magface respecto a un modelo neutral (Insightface) como el análisis del mismo 'score' respecto al mismo modelo de Magface, es necesario la obtención de los vectores de características de cada imagen y su correcta normalización para su posterior uso en la matriz de similitudes. Una vez el modelo ha extraído los vectores de características de todas las imágenes se deben de normalizar todos los vectores para poder crear una matriz de similitudes comprendida entre -1 y 1. Como Magface utiliza la distancia del coseno para extraer las características, la matriz de similitudes del modelo

propio de Magface estará comprendida entre -1 y 1 debido a que se mide la distancia mediante el coseno. Los pares de imágenes que sean la misma persona obtendrán valores de similitud próximos a 1 mientras que los valores de similitud entre parejas de imágenes que no sean la misma persona contendrán similitudes cercanas a 0 o negativas. Destacar que el modelo puede obtener valores de similitud altos para parejas que no son la misma persona y valores de similitud bajos para parejas que sí son la misma persona. Para ellos se realiza el posterior análisis de explicado en el apartado 5.4.1 y mostrado en el apartado 6.
La normalización de los vectores de características extraídos por el modelo es diferente a los demás modelos debido al tipo de extracción que utiliza Magface. Se normaliza teniendo en cuenta las dos imágenes de la comparación, en lugar de tener en cuenta todas las características como el resto de modelos, y delimitando su similitud entre -1 y 1. La normalización se puede ver en la figura 5.5
Para mostrar el aspecto de la matriz de similitudes real(9811x9811) se visualiza una matriz

```
mag=np.load('magnitudesmagface.npy')
similarity=np.zeros((len(mag),len(mag)))
#mag=mag@mag.T
for i in range(len(mag)):
    for j in range(len(mag)):
        aux=np.sum(np.multiply(mag[i],mag[j]))
        normaliza=np.linalg.norm(mag[i])*np.linalg.norm(mag[j])
        similarity[i][j]=np.clip(aux/normaliza,-1.,1.)
```

Figura 5.5: Normalización de los vectores de características de Magface

```
[ 1.         -0.00822886  0.12487952
 -0.0430352 ]
[-0.00822886  0.99999988  0.06103645
  0.00556739]
[ 0.12487952  0.06103645  1.
```

Figura 5.6: Matriz de similitudes reducida 3x3 imágenes

de 3x3 en la que se define cómo es una matriz de similitudes en la figura 5.6. En las filas se encuentran las imágenes del conjunto de datos y en las columnas las mismas imágenes del conjunto de modo que para cada imagen (fila) se obtendrá un valor de similitud para todas las imágenes del conjunto (columnas). Como se puede observar en la figura 5.6, las comparaciones entre la misma imagen (misma persona) dan como resultado valores cercanos a 1 mientras que cuando no se trata de la misma persona, la similitud disminuye a valores cercanos a 0 o incluso negativos por la forma de estimar las similitudes con la longitud del coseno anteriormente explicada. Como en el conjunto de datos de 9811 imágenes para cada identidad hay 10 imágenes distintas con diferentes poses y diferentes calidades, en esas comparaciones la similitud se verá afectada bajando su valor de 1 según el modelo detecte si la imagen posee las mismas características que otra imagen de la misma persona. Es aquí donde entra la parte del análisis.

Tras la inferencia realizada al modelo, se obtiene un vector columna de 9811 posiciones en los que se encuentra el grado de calidad que el modelo estima para cada imagen en un rango abierto, no definido. Para el estudio de la distribución de los scores, se limitan de 0 a 1 para ver su distribución en comparación con el resto de estimadores. De esta forma ya se dispone de los datos necesarios para hacer uso del Modelo de Face Recognition y obtener las métricas del modelo.

5.2.3 Cálculo de la calidad de mediante la técnica de SER-FIQ

Se realizan una serie de inferencias al modelo de Insightface con la técnica de SER-FIQ, ya que está entrenado con Dropout. En el artículo de SER-FIQ se realiza una alineación de las imágenes y un recorte facial para que el modelo sólo tenga que estimar la calidad de la región facial. Esto conlleva a un mayor coste computacional a la hora de establecer la calidad para cada imagen. En este caso durante la inferencia no será necesario ya que para todo el estudio se hace uso del conjunto de datos alineado y recortado centrado en la cara de 112 x 112 píxeles que se obtiene de la inferencia al modelo de detección facial. En este caso debido a limitaciones computacionales, todos los cálculos relacionados con el modelo se realizan mediante cpu. Además, se especifica que la técnica de SER-FIQ proporciona valores de calidad de alta usabilidad cuando la calidad se extrae con una tasa de repetición del cálculo del 'score' de 100. Cada vez que se calcula una nueva calidad se van activando o desactivando una serie de neuronas aleatorias. Por coste computacional también se ha reducido ese valor a 10. La capa Dropout es una máscara que anula la contribución de algunas neuronas hacia la siguiente capa y deja sin modificar todas las demás. Es una técnica común de los modelos de DL pero que también se utiliza para la fase de test.

Para extraer los resultados primeramente se cargan todas las imágenes del conjunto sobre un array. Se leen todas las imágenes de la carpeta y se realiza una lista con el nombre de todas las imágenes a utilizar. Una a una se van leyendo con imread() de OpenCV y se le pasa al modelo la imagen transpuesta en espacio de color RGB mediante la función COLOR_BGR2RGB(), ya que, imread() por defecto lee las imágenes en espacio BGR y el modelo rebibe como entrada una imagen en RGB. Para la extracción de la estimación de la calidad para cada imagen se calcula la distancia euclidiana a partir de los resultados del modelo obtenido. La distancia euclidiana entre dos puntos en el espacio euclidiano es la longitud de un segmento de línea entre los dos puntos.

En este caso, el modelo viene definido en dos archivos, un archivo .params y otro archivo .json mediante los cuales se estimará la calidad de las imágenes. Los archivos de modelo utilizado para la inferencia son *' insightface-0000.params'* e *'insightface-symbol.json'* . Se carga y se almacena en una variable para poder hacer uso del mismo. Para obtener el 'score' de calidad se comprueba que la imagen contenga las tres capas RGB. Una vez preparada la imagen se le pasa al modelo y se extrae la calidad de la imagen a partir de la distancia euclidiana. Los valores de calidad que estima el modelo están comprendidos entre 0 y 1.

	Precisión
LFW	99.53%
CALFW	94.68%
CPLFW	89.75%
AGEDB_30	95.20%
CFP_FF	99.54%
CFP_FP	96.30%
VGG2_FP	94.84%

Tabla 5.1: Precisión del modelo utilizado con diferentes conjuntos de datos públicos

5.3 Realización de las inferencias sobre el modelo de FR

Tras las inferencias realizadas a los distintos modelos de DL para la estimación de calidad de imágenes para la autenticación biométrica y la obtención de sus respectivos 'scores', se procede a realizar la inferencia sobre el modelo de FR de Insightface, W. Liu y cols. (2017) como documentación para entender el modelo de FR. Esta inferencia se realiza con el fin de comprobar cómo de buenos son los estimadores de calidad utilizados (los tres métodos de DL) a la hora de calcular la calidad de una imagen para un posterior reconocimiento facial.
Para llevar a cabo una inferencia sobre un modelo de FR se necesita un conjunto de datos y un modelo de DL pre-entrenado para obtener el reconocimiento facial. En este caso, se utiliza el conjunto de datos explicado en el apartado 4.2 y un modelo pre-entrenado de Insightface basado en Arcface, Deng y cols. (2018) y Deng, Guo, Liu, y cols. (2020), disponible públicamente.
El modelo pre-entrenado del que se dispone se compone en dos archivos que produce TensorFlow al guardar un modelo entrenado. El modelo se almacena en un archivo '.meta' desde el cual se cargará el modelo con sus pesos. El modelo concreto utilizado para esta inferencia es el modelo *best-m-1006000.meta*. Este modelo ha sido entrenado mediante una serie de conjunto de datos públicos obteniendo los porcentajes de precisión que se muestran en la tabla 5.1.
Para la evaluación del conjunto de datos utilizado en el análisis, se prepara una inferencia sobre el modelo de FR. Esta inferencia se lleva a cabo mediante Python y TensorFlow.
Primero se definen las rutas de todos los archivos que se van a utilizar además de las imágenes sobre las cuales se va a realizar la inferencia. Se carga un archivo *'.yaml'* mediante *yaml.load()*, que no es más que un archivo en el que se definen variables para la inferencia. Destacar que en este archivo de definen variables globales como la arquitectura del modelo (*resnet_v2_m_50*) para este modelo, el tamaño de los embeddings (512 de forma general) y el tamaño de la imagen que se define a 112 que es el tamaño de las imágenes tras recortarlas centradas en la cara como se explica en el apartado 5.1. Seguidamente se define un tensor mediante *'tf.placeholder()'* en el cual se define la estructura del tensor para las imágenes del conjunto. Este tensor se definirá como un tipo flotante y de definirá el tamaño de la imagen en las 3 capas ya que se trabaja con imágenes en espacio RGB.
Una vez se tiene todo preparado para al inferencia se pasa a definir la arquitectura que va a poseer el modelo a utilizar y su posterior carga para hacer uso del mismo. Para este modelo la arquitectura definida es 'resnet_v2_m_50' y se configuran las capas de la arquitectura con su *'dropout'* y terminando la salida de la penúltima capa con *'fully_connected'*. Como

el modelo está almacenado en un archivo '.meta', se carga el modelo mediante TensorFlow con *'train_import_meta_graph()'* y se cargan los pesos del modelo entrenado mediante la función *'restore()'* pasándole como parámetro de entrada el archivo en el que se encuentran guardados los pesos.
Una vez se dispone del modelo, se debe realizar el preprocesado de las imágenes antes de predecir con el modelo. El procesado que se lleva a cabo es la lectura de todas las imágenes del conjunto desde una lista y la normalización de todas las imágenes del conjunto entre (-1,1). Al modelo también se le pasan las imágenes del conjunto recortadas en la región facial para que el resultado sea más óptimo.

Por último, se procede a la extracción de los embeddings de cada imagen según el modelo pre entrenado. Según el *'batch_size'* de imágenes que se defina (en este caso 100), el modelo calcula un vector de 1x512 embeddings por imagen mediante una *'Session()'* definida anteriormente para la extracción de embeddings a partir del modelo cuando se le pasa una imagen. Una vez se han extraído todos los embeddings de todas las imágenes, se normalizan todos los embeddings mediante la función *linalg.norm* de la librería de numpy.
Tras la extracción de todos los embeddings del conjunto de datos utilizado, ya se puede obtener la matriz de similitudes del conjunto de datos (explicada en más detenimiento en el apartado 5.4), enfrentando todas las imágenes del conjunto con todas las imágenes del conjunto. Para obtener la matriz de similitudes del a partir de los vectores de características, se multiplican todos los vectores por los propios vectores transpuestos obteniendo así, una matriz de 9811 x 9811 posiciones, 96.255.721 comparaciones.

5.4 Obtención de las métricas del modelo de FR para analizar los resultados

Con el fin de visualizar y obtener los resultados del estudio realizado, se realiza un programa mediante el cual se obtienen las métricas de los modelos de Deep Learning analizados y del algoritmo propio basado en el ICAO.

Ya que se necesita diversidad en cuanto a las imágenes de los sujetos, el conjunto de imágenes que se va a utilizar es de 10.000 imágenes y 1.000 identidades, lo que significa, que se tienen 10 imágenes para cada identidad. Esto hará que los resultados sean lo más reales posibles ya que existen diversas imágenes que corresponden a una identidad. Debido a que la obtención de las identidades es aleatoria, se crea el conjunto de imágenes eligiendo aleatoriamente 1.000 identidades y 10 imágenes de cada una también aleatorias. Debido a las limitaciones del conjunto de CelebA utilizado, existen identidades para las cuales no existen 10 imágenes para cada identidad. Por esta razón, el conjunto de datos utilizado para la obtención de las métricas es de 9811 imágenes. Como existen más de dos imágenes para cada identidad, esta reducción de tamaño no tendrá un impacto considerable en la obtención de las métricas.

Para la obtención de las métricas se van a seguir una serie de pasos. Una vez se tienen los embeddings del modelo de Face Recognition, se calcula la matriz de similitudes del conjunto de imágenes. Esta matriz no es más que el grado de similitud que tiene cada imagen contra

todas las imágenes del conjunto utilizado. Se obtiene mediante los embeddings calculados. Para hallar esta matriz, se multiplica el vector de los embeddings por el mismo vector transpuesto de todas las imágenes para obtener el grado de similitud de todas las imágenes frente a todo el conjunto. De este modo, resulta una matriz de 9811 x x 9811 valores de similitud. Cada fila corresponderá a una imagen del conjunto y las columnas a todas las imágenes del conjunto a las que se enfrenta para comparar y obtener el grado de similitud con cada una.

Una vez se obtiene la matriz de similitudes, se necesita obtener la matriz de etiquetas del conjunto. Esta matriz es una matriz creada con las mismas dimensiones toda llena de ceros y que para cada imagen que se analice, coloca el valor '1' en las posiciones de la matriz en las que esa imagen se enfrenta a las 10 imágenes de la misma identidad. Esta matriz de etiquetas sirve para poder analizar posteriormente si el modelo de Face Recognition ha sido capaz de identificar a la persona en las 10 imágenes que se le han proporcionado. Como resultado de esta matriz se obtendrá para cada imagen que se analice, 10 posiciones con valor '1' que corresponderá a las 10 imágenes de la misma identidad a las que se mide la similitud de la imagen analizada.

Usando la matriz de similitudes del conjunto de datos y la matriz de etiquetas del mismo, se procede a la evaluación de los modelos y la generación de resultados. Esta evaluación consiste en analizar cómo de eficientes son los modelos de calidad que se han utilizado a la hora de eliminar las imágenes de menos usabilidad (baja calidad) para el modelo de Face Recognition. Se evaluará visualmente como de buenos son estos modelos de estimación de calidad a la hora de que el modelo de Face recognition reconozca a una persona. Para la evaluación de los modelos se van a utilizar dos tipos de métricas, la FNMR y la FMR. Estas métricas se basan en las similitudes extraídas en el modelo de Face Recognition. A modo general, se podría definir la FMR como el número de parejas de imágenes que no son la misma persona pero el modelo de Face Recognition define una similitud igual o mayor al umbral que se define. En cambio, la FNMR define el número de parejas de imágenes que sí son la misma persona pero el modelo de Face Recognition define una similitud inferior al umbral definido indicando que no se trata de la misma persona.

Se ha hablado del término 'umbral' pero todavía no se ha definido. El umbral se escoge para que el modelo de Face Recognition posea un FNMR específico. Los valores típicos del FNMR suelen ser 0.5%, 1%, 2% y 5% ya que son los umbrales que utiliza el NIST para representar sus resultados. Para hallar los umbrales bastará con ordenar las similitudes que pertenecen a parejas de imágenes de la misma persona buscando el el porcentaje de imágenes que se deben rechazar para cada umbral de fnmr y fijando el umbral a dicho valor, esto asegura que el modelo de Face Recognition predice ese porcentaje de similitudes que pertenecen a parejas de imágenes de la misma persona pero son de personas diferentes.

Una vez definida la FNMR que se desea para la evaluación, se pasa a ver cómo afecta la calidad de la imagen en los resultados obtenidos. Este análisis se basa en la eliminación de imágenes del conjunto de datos según los 'scores' de calidad para analizar cómo mejora el modelo de Face Recognition a la hora de identificar a las personas. Para la eliminación de las imágenes se ordenarán los 'scores' de calidad de todo el conjunto de datos (9811 imágenes) y

se irán eliminando las imágenes con peor 'score' del conjunto. Se irán eliminando porcentajes de imágenes para ver como se refleja en los resultados la ausencia de éstas imágenes de baja calidad.

Para evaluar cual es el impacto de estas imágenes de baja calidad no es necesario volver a calcular los embeddings del conjunto de imágenes resultante a la eliminación de las imágenes de peor calidad. Como las imágenes se almacenan mediante etiquetas, al eliminar las imágenes de peor calidad, se recorre la matriz de similitudes eliminando las filas (que corresponden a las imágenes del conjunto) y eliminando la información de las filas que pertenezcan a las imágenes eliminadas debido a la baja puntuación de calidad. De esta forma se agiliza la comprobación de los resultados.

Por último, se realiza el cálculo de las métricas. Este es el último paso a realizar para extraer los resultados. Con las similitudes ya filtradas, teniendo en cuenta sólo aquellas imágenes que no se han eliminado, se obtendrá el valor de FNMR que posee el conjunto de imágenes con los umbrales que se han fijado anteriormente como se ha explicado en esta sección. Para poder mostrar las curvas de la FNMR se almacenará en un vector el porcentaje de imágenes rechazadas en cada iteración, y en otro vector, el valor de FNMR para cada uno de los porcentajes de imágenes rechazadas. El objetivo es representar las métricas de acuerdo al estándar que realiza el NIST.

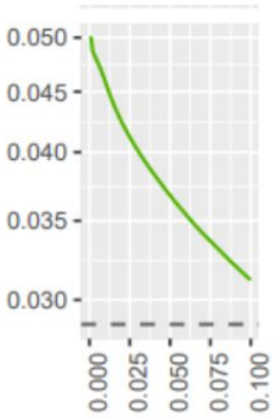


Figura 5.7: Método de graficación utilizado por el NIST

Como se observa en la figura 5.7, la representación en el eje 'y' muestra la métrica de FNMR, en este caso fijada inicialmente a 0.05 (5%). En el eje 'x' se muestra el porcentaje de imágenes rechazadas en un rango de 0% hasta un 100% de las imágenes. Como se ve en la figura a medida que el número de imágenes rechazadas es mayor, el FNMR disminuye lo que significa que la métrica es mejor. Una disminución de la FNMR indica que el número de parejas de imágenes que son de la misma persona pero el modelo de Face recognition detecta que no lo son disminuye, aumentando así la seguridad del sistema de autenticación.
A la vista de los resultados obtenidos para esta gráfica de la figura 5.7, se obtiene una métrica

fijada en un 5% de FNMR y la pendiente de la gráfica disminuye notoriamente lo más vertical posible a medida que se van rechazando imágenes con peor calidad. Este tipo de gráficas es la ideal para un sistema de reconocimiento biométrico ya que a medida que se rechazan las imágenes con peor calidad, la FNMR mejora rápidamente al disminuir su valor, aumentando así la seguridad del sistema. La FNMR representa el falso rechazo que posee un modelo de reconocimiento facial, es decir, el número de imágenes que sí son la misma persona pero el modelo define una similitud inferior a la similitud del conjunto para una FNMR fija. Por esta razón, a medida que la FNMR sea menor, mejor será el sistema de reconocimiento facial.
Para mostrar los resultados del estudio realizado se obtendrán las métricas de los tres modelos de DL. Se mostrarán tanto la evolución de las curvas FNMR como de las curvas FMR mostrando así la usabilidad del modelo junto a su seguridad y tolerancia.

5.4.1 Obtención de curvas FNMR y FMR

Como se ha comentado a lo largo del trabajo, la evaluación de los sistemas de reconocimiento facial se realiza mediante las métricas de FNMR y FMR. Estas métricas permiten analizar la evolución del modelo de FR a medida que se van rechazando imágenes del conjunto de datos. Se empiezan eliminando las imágenes de peor calidad según cada modelo de estimación. De modo que, el conjunto de datos original (9811 imágenes) se divide en pequeños conjuntos de imágenes con su puntuación de calidad ordenado de menor a mayor y se van eliminando esas imágenes de peor calidad de la matriz de similitudes. Cuantas más imágenes se rechazan, las imágenes que se obtienen para obtener la métrica son de mayor calidad. Así pues, para porcentajes de rechazo de imágenes del conjunto altos, las métricas han de mejorar.
Las métricas dependen de la estimación de calidad que cada modelo predice para una misma

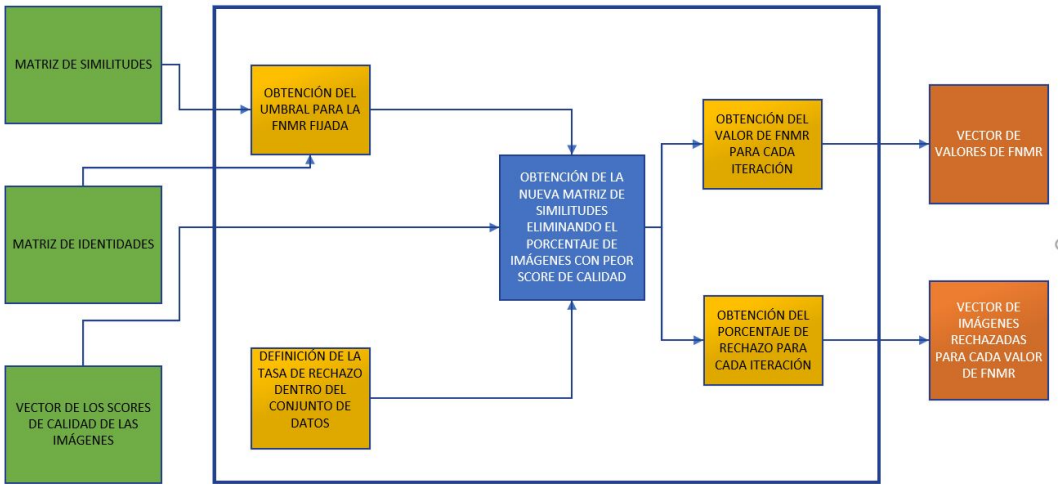


Figura 5.8: Diagrama de bloques simplificado para la obtención de las curvas de FNMR y FMR

imagen. No obstante, también es muy importante el modelo de reconocimiento facial que se emplea para extraer las métricas. En este trabajo se utiliza el modelo de FR de Insightface el cual extrae un vector de 512 características para cada imagen. Para realizar un análisis

mayor, también se emplea el modelo de FR de Magface, que extrae un vector de 512 características para cada imagen cuando el modelo estima la calidad de cada imagen. Se utilizan dos modelos diferentes para analizar cómo de importante es a su vez un modelo de reconocimiento facial. Para un mismo conjunto de datos, cada modelo de reconocimiento facial puede extraer una serie de características diferentes. Estas características que extrae para cada imagen afectan en la posterior creación de la matriz de similitudes. El rango de similitud puede verse afectado por cada modelo. De modo que, el valor de similitud de dos imágenes para los dos modelos puede ser diferente en rango. Este factor también influye en cómo de bueno es el modelo, las similitudes extraídas aportan información en cómo de bueno es ese modelo según el valor de similitud de las parejas de imágenes que no son la misma persona.

En las figuras 5.9 y 5.10 se muestra un pequeño fragmento de las similitudes extraídas de cada modelo ya que cada matriz posee un tamaño de 9811 x 9811. Como se puede apreciar en la figura 5.9, el modelo de Insightface extrae unas similitudes altas cuando la comparación de las imágenes no son de la misma persona. En cambio, en la figura 5.10, se puede apreciar como el modelo de Magface es más estricto a la hora de extraer las características de las imágenes y obtiene unas similitudes bajas, incluso negativas (utiliza la distancia del coseno para extraer las características, se encuentran entre -1 y 1), para los pares de imágenes que no son la misma persona. Se darán casos en los que se obtendrán valores bajos de similitud cuando se trate de la misma persona y viceversa pero eso forma parte del análisis.

```
[1.0000001  0.80189395 0.73908097
[0.80189395 1.0000001  0.7322967
[0.73908097 0.7322967  1.
```

Figura 5.9: Similitudes extraídas por el modelo de InsightFace

```
[ 1.         -0.00822886  0.12487952
 -0.0430352 ]
[-0.00822886  0.99999988  0.06103645
  0.00556739]
[ 0.12487952  0.06103645  1.
  0.02916997]
```

Figura 5.10: Similitudes extraídas por el modelo de Magface

Como las métricas de FNMR y FMR vienen relacionadas entre sí debido al umbral que se elige para el análisis y el umbral viene definido por la matriz de similitudes, se puede decir que las métricas vendrán definidas por la matriz de similitudes y a su vez por el modelo de reconocimiento facial. La FNMR se define como el falso rechazo, es decir, las imágenes que el sistema rechaza cuando las dos imágenes sí son de la misma persona. Esto ocurre cuando la similitud entre dos imágenes se encuentra por debajo del umbral de similitud fijado para la extracción de métricas. La FMR se define como la tasa de aceptación del sistema para imá-

genes que no son la misma persona pero el modelo estima una similitud superior al umbral definido para la extracción de métricas. Cuanto menor sea la FMR mayor seguridad obtendrá el sistema y menos vulnerabilidad presentará. De este modo, cuanto mejores sean las similitudes extraídas por los modelos de FR mejores serán las métricas obtenidas. Esto es debido a la definición del umbral de similitud fijado para el cálculo de las métricas. Si el modelo define umbrales altos para comparaciones de imágenes en las que no son la misma persona, la métrica se verá afectada al tener en cuenta imágenes que no se tienen que tener en cuenta idealmente. El objetivo de un modelo de FR es que esas comparaciones altas de imágenes de diferentes personas sean las mínimas posibles y que todas las imágenes que superen el umbral fijado de similitud para el análisis sean comparaciones de la misma persona. A medida que ese umbral de similitud para el cálculo de métricas se incremente, el sistema se volverá menos tolerante y se volverá más seguro, lo que influye en una disminución de la FMR. En cambio, si el umbral de similitud es pequeño, el sistema será más vulnerable y menos seguro, la FMR aumentará.
Para realizar el análisis de las métricas de los modelos de estimación de calidad de imágenes, se calculan las métricas para valores de FNMR fijados al 0.5%, 1%, 2%, 5%, 15% y 20% para ver la evolución de los sistemas de estimación a medida que el sistema se vuelve más seguro. Para extraer los umbrales que definen la FNMR a un valor determinando, se dispone de las matrices de similitudes extraídas de los dos modelos de FR, tanto Insightface como Magface. también de dispone de una matriz de 'etiquetas' que es la encargada de contener la información en la matriz de qué posiciones corresponden a comparaciones de la misma persona. Cuando se trate de comparaciones de la misma persona, el valor de la matriz de 'etiquetas' en esa posición de la matriz valdrá 1, del contrario, valdrá 0. También de tiene el vector de 'scores' de cada modelo a analizar y un vector de similitudes. En este vector de similitudes se incluyen únicamente las similitudes de la matriz que corresponden a comparaciones de la misma persona. Este vector de similitudes se ordena de menor a mayor y juntamente con el porcentaje al que se quiere fijar la FNMR, se obtiene el valor del umbral de similitud que se debe de fijar. Esto asegura que la FNMR esté fijada al porcentaje que se desea. Para cada porcentaje de FNMR que se desea, se selecciona el valor de similitud que corresponde a ese porcentaje dentro del vector de similitudes.
En la tabla 5.2 y en la tabla 5.3 se muestran los umbrales de similitud fijados para el cálculo de las métricas. Como se puede apreciar, el modelo de Insightface al obtener unas similitudes para parejas que no son la misma persona como las mostradas en la figura 5.9 y al definir los umbrales de las que sí son la misma persona, existen muchas similitudes de personas diferentes que superan ese umbral. Esto afectará al análisis de las métricas y permitirá analizar los dos modelos de FR que se utilizan en el trabajo.

Una vez se dispone de los umbrales de similitud fijados, se procede al cálculo de las métricas. Para el cálculo se utiliza una tasa de rechazo del conjunto de imágenes de 0.25%. Es decir, se calcula el valor de FNMR eliminando en cada iteración un 0.25% de las imágenes totales del conjunto utilizado. El rechazo de imágenes se realiza acorde al 'score' de calidad que define cada modelo de estimación de calidad. Con un rechazo de un 0.25%, en un rango de 0% a 100%, se obtendrán 400 valores de FNMR que permitirán analizar las métricas de cada modelo. Para calcular el valor de FMR se realizará el mismo procedimiento. Se almacenarán los valores de FNMR y FMR en vectores junto al porcentaje de imágenes rechazadas para su

	Umbrales
FNMR al 0.5%	0.633517
FNMR al 1%	0.650046
FNMR al 2%	0.667672
FNMR al 5%	0.693307
FNMR al 15%	0.731972
FNMR al 20%	0.744700

Tabla 5.2: Umbrales de similitud extraídos por el modelo de InsightFace

	Umbrales
FNMR al 0.5%	-0.049396
FNMR al 1%	-0.002956
FNMR al 2%	0.064628
FNMR al 5%	0.350117
FNMR al 15%	0.537298
FNMR al 20%	0.573020

Tabla 5.3: Umbrales de similitud extraídos por el modelo de Magface

posible representación.

Para el cálculo de las métricas se calcula el valor de FNMR y de FMR para cada subgrupo dentro del conjunto de datos a medida que se van rechazando un 0.25% de las imágenes en cada iteración. Para poder obtener las métricas de acuerdo a las ecuaciones 5.1 y 5.2 se definen los siguientes parámetros:

- **True Positive:** Mide el número de comparaciones en las que se trata de la misma persona y el modelo de Face Recognition indica que es la misma persona. Es decir, el valor de similitud entre las dos imágenes comparadas supera el umbral de FNMR al que se fijan las métricas.
- **False Positive:** Mide el número de comparaciones en las que no se trata de la misma persona pero el modelo de Face Recognition indica que es la misma persona.
- **True Negative:** Mide el número de comparaciones en las que no se trata de la misma persona y el modelo de Face Recognition indica que no es la misma persona.
- **False Negative:** Mide el número de comparaciones en las que sí se trata de la misma persona y el modelo de Face Recognition indica que si son la misma persona.

$$FNMR = \frac{FalseNegatives}{FalseNegatives + TruePositives} \tag{5.1}$$

$$FMR = \frac{FalsePositives}{FalsePositives + TrueNegatives} \tag{5.2}$$

5.5 Evaluación del sesgo para cada modelo de estimación de calidad de imágenes

Un sistema de Face Recognition necesita de una seguridad y un funcionamiento de uso acorde a la funcionalidad que se le aplica a este tipo de modelos. Debido a las aplicaciones en seguridad que poseen este tipo de modelos, debe comprobarse la funcionalidad de los mismos para todo tipo de situaciones en los que un rostro se puede ver afectado. Este factor viene relacionado con los modelos de estimación de calidad. Un modelo de FR funciona mejor a medida que la imagen posee una mayor calidad. Esto significa que cuanto mejor sea la estimación de calidad de la imagen, mejor debe funcionar el modelo de FR. Para analizar el correcto funcionamiento de los modelos de FR, se realiza un estudio sobre los modelos de estimación de calidad en cuanto a sesgos faciales u oclusiones.
Es importante ver cómo de bueno es cada modelo utilizado a la hora de estimar la calidad de una imagen según las características que posee la imagen. En base a los posibles sesgos de los modelos de estimación de calidad, se estudiará si los modelos distinguen entre diferentes color de pelo, si obtienen una mejor calidad para oclusiones sobre la región del rostro o si predicen una mejor calidad sobre personas de diferente sexo o edad.
Para ello, se realiza un estudio sobre el sesgo de los modelos de FaceQnet, SER-FIQ y Magface en base a las siguientes características:

- **Color de pelo:**
  - Pelo negro
  - Pelo rubio
  - Pelo canoso
  - Pelo castaño
- **Oclusiones:**
  - Usuarios con gafas
  - Usuarios con barba
  - Usuarios sin barba
  - Usuarios con bigote
- **Edad:**
  - Personas jóvenes
  - Personas adultas
- **Sexo de la persona:**
  - Masculino
  - Femenino

La evaluación del sesgo se lleva a cabo gracias a las anotaciones de atributos que proporciona la base de datos de CelebA que es la utilizada durante el trabajo. El sesgo se calcula seleccionando las imágenes del conjunto que se utilizan y extrayendo la información necesaria para

los atributos que se analizan. Una vez escogidos los atributos y los vectores que indican que atributo contiene cada imagen del conjunto utilizado, se selecciona el 'score' de calidad de esa imagen para poder trazar una curva de probabilidad con todos los 'scores' del conjunto que poseen esas características y así obtener cómo funciona cada modelo según las características de cada imagen. Se estudia si algún modelo de predicción está sesgado o no.

6 Resultados

En el apartado de resultados se realiza un análisis profundo sobre los diferentes modelos de estimación de calidad en referencia a cómo son capaces estos modelos de estimar la calidad de una imagen. Se lleva a cabo la extracción e impresión de métricas obtenidas gracias al índice de calidad obtenido para cada modelo en relación a un modelo de Face Recognition. Por último se evalúa cada modelo de estimación de calidad por separado, analizando posibles sesgos de cada modelo.

6.1 Distribución de los Scores de calidad definidos por cada modelo

Para ver el funcionamiento de cada modelo, se representa la distribución de los 'scores' de calidad de cada modelo. Este análisis permitirá ver cómo de bueno es el conjunto de datos que se le proporciona al modelo de estimación de calidad y cómo afectan estos índices de calidad a un posterior modelo de Face Recognition. El 'score' de calidad va ligado con el modelo de FR. Es importante realizar una predicción correcta de la calidad para una imagen pero también se debe tener un modelo de FR que funcione correctamente para poder sacar el máximo rendimiento a la calidad de cada imagen. En el apartado de los resultados de las métricas se verá como los modelos de FR también afectan al rendimiento del sistema. El caso ideal es una buena estimación de calidad para un modelo de FR que funcione correctamente a la hora de realizar un reconocimiento facial para imágenes que posean un mínimo de calidad.

6.1.1 Distribución de los scores según FaceQnet

Como se puede apreciar en la figura 6.1, el modelo de FaceQnet realiza la estimación de calidad para el conjunto utilizado simulando una distribución gaussiana. Este modelo estima que la calidad del conjunto no es alta ni baja, sino que la mayoría de las imágenes se encuentran con una calidad media con un valor de calidad entre 0.3 y 0.5. La estimación de calidad de FaceQnet abarca una puntuación entre 0 y 1 pero a la vista de los resultados obtenidos, se ve cómo este modelo no estima que las imágenes sean de muy mala calidad, rango de 0 a 0.2, ni de alta calidad, rango de 0,8 a 1 para este conjunto de datos utilizado. Esto es debido a que en índices cercanos a 0.7 dejan de haber imágenes con calidad superior según este modelo. Según los resultados obtenidos, como FaceQnet estima la calidad de una imagen basándose en las similitudes que tiene la misma con una imagen ICAO, se obtiene que la similitud más grande entre las mejores imágenes del conjunto de datos y la de una imagen ICAO es de 0.7 aproximadamente sobre 1. Esto refleja que las imágenes del conjunto de datos no son del todo adecuadas según el estándar ICAO ya que la mayoría de imágenes

Figura 6.1: Distribución del score de calidad para el modelo de FaceQnet

obtienen una calidad entre 0.3 y 0.5 sobre 1.

6.1.2 Distribución de los scores según SER-FIQ

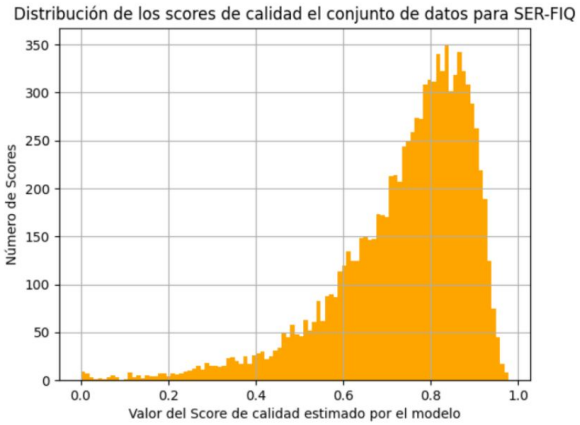


Figura 6.2: Distribución del score de calidad para el modelo de SER-FIQ

En el caso del modelo de SER-FIQ, ilustrado en la figura 6.2, distribución de los scores de calidad se muestra en valores cercanos a 1. Esta técnica posee un rango de calidad de 0 a 1 y tras la estimación de calidad, este modelo estima que el conjunto de datos proporcionado posee en gran mayoría imágenes de alta calidad. Tal y como se ve sobre la figura 6.2 la distribución de imágenes para 'scores' de baja calidad es muy baja, apareciendo un rango de imágenes considerable a partir de puntuaciones superiores a 0,6.
Las puntuaciones de calidad con esta técnica han sido extraídas para 10 estimaciones del modelo diferente. En la técnica de SER-FIQ se aconseja que sean 100 estimaciones las realizadas para poder obtener una buena calidad de imagen. Por límites computacionales no se ha podido llevar a cabo. Aún así la obtención de la calidad es válida para el estudio y mediante esta

técnica se observa que gran parte del conjunto de datos obtiene grandes valores de calidad. Esto se debe a las pequeñas variaciones entre patrones estocásticos de las imágenes que era el análisis que realizaba la técnica de SER-FIQ.

6.1.3 Distribución de los scores según Magface

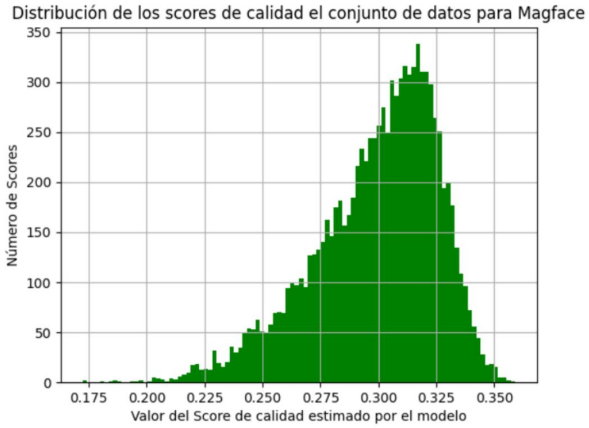


Figura 6.3: Distribución del score de calidad para el modelo de Magface

El modelo de Magface, figura 6.3, presenta una mayor restricción a la hora de estimar la calidad para este conjunto de datos. Este modelo, no abarca un rango definido ya que estima la calidad en función de la distancia a la que se encuentra la imagen en el hiperespacio. A la vista de la figura 6.3, se puede observar como las imágenes con mayor calidad alcanzan el valor de 0,35. Este modelo es más restrictivo que el resto de modelos utilizados representando una distribución de calidad de 0.175 a 0.35 aproximadamente. El pico de mayor calidad de encuentra en torno al 0.31 con un ancho dentro de la banda de 0.30 a 0.325 menor que los demás modelos. Como este modelo se evalúa respecto a su propio modelo de Face Recognition, se apreciará la importancia de que el modelo de estimación de calidad y de FR vayan de la mano. Cuanto mejor sea el modelo de FR y mejor sea el modelo de estimación, mejor será la métrica obtenida.

6.1.4 Distribución de los scores de todos los estimadores de calidad juntos

Por último, se muestra la distribución de todas las puntuaciones de calidad extraídos por cada modelo en una misma gráfica, figura 6.4. Sobre esta figura se puede observar como la se distribuye la calidad de imagen para cada modelo y cómo son las transiciones entre imágenes de peor calidad a imágenes de mejor calidad. Se puede apreciar cómo la transición entre FaceQnet y SER-FIQ es más suave mientras que la transición en el modelo de Magface es más abrupta.

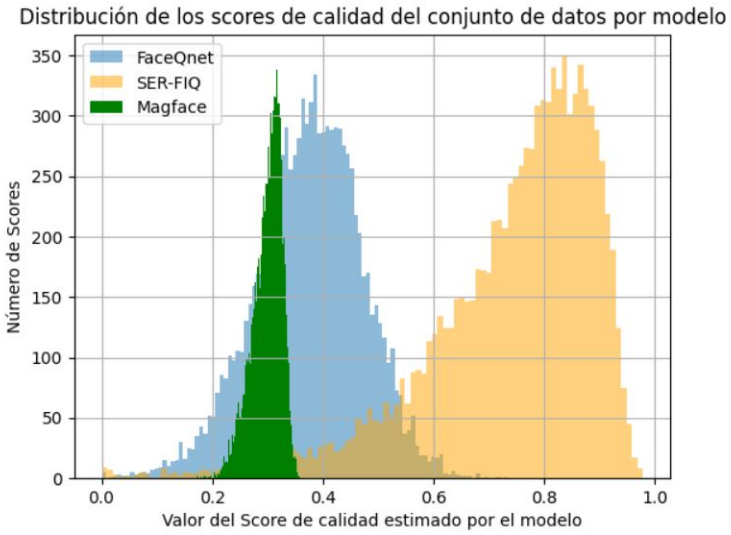


Figura 6.4: Distribución del score de calidad para todos los modelos

6.2 Evolución de FNMR según los modelos de Face Recognition

Para evaluar la calidad de un sistema de Face Recognition es necesario evaluar cómo se comporta el modelo según la FNMR. La evolución de la FNMR mostrará cómo de bueno es un modelo y cuál es el grado de seguridad que posee. La representación de las curvas de FNMR viene afectada por el modelo de FR y por el 'score' de calidad que cada modelo estima para cada imagen. En este apartado se evalúan dos modelos de FR, Insightface y Magface, y tres modelos de estimación de calidad FaceQnet, SER-FIQ y Magface. Se evalúan los modelos de calidad respecto al modelo de Insgightface para ver cómo funcionan los modelos sobre un modelo de FR imparcial y se evalúa el modelo de Magface respecto a las puntuaciones de calidad de Magface para ver cómo afecta que el modelo de FR utilice el modelo propio de estimación de calidad. Los sistemas biométricos funcionan mejor cuando son probados con conjuntos de datos con los que se han entrenado. Esto no sucede en aplicaciones reales en la vida cotidiana. Por esta razón se evalúa el funcionamiento de los modelos de FR y de estimación de calidad sobre un conjunto de datos nuevo para todos los modelos.
El objetivo es que todas las curvas presenten una disminución del valor de FNMR lo más pronto posible y lo más recto posible. Una curva de FNMR ideal se basa en una disminución de FNMR rápida en los primeros porcentajes de rechazo de imágenes para ser capaz de tener una FNMR de valores cercanos a 0 para porcentajes altos de rechazo de imágenes. Esto significaría que el modelo no posee falso rechazo debido a que la mayoría de pares de imágenes se encuentran sobre el umbral definido de similitud para una fijada. Cuando el falso rechazo es menor, significa que el sistema es más seguro y su funcionamiento es el correcto ya que es muy poco vulnerable.
Para evaluar el funcionamiento de los sistemas de reconocimiento facial, se calculan las curvas de FNMR fijadas a un valor de FNMR del 0.5%, 1%, 2%, 5%, 15% y 20%. Inicialmente sólo se iban a mostrar las métricas para valores de FNMR del 0.5%, 1%, 2%, 5% que son los que utiliza el NIST para sus métricas, pero ante la mejora del sistema a medida que aumenta la

FNMR, se decide representar también las métricas para FNMR 15% y 20%.

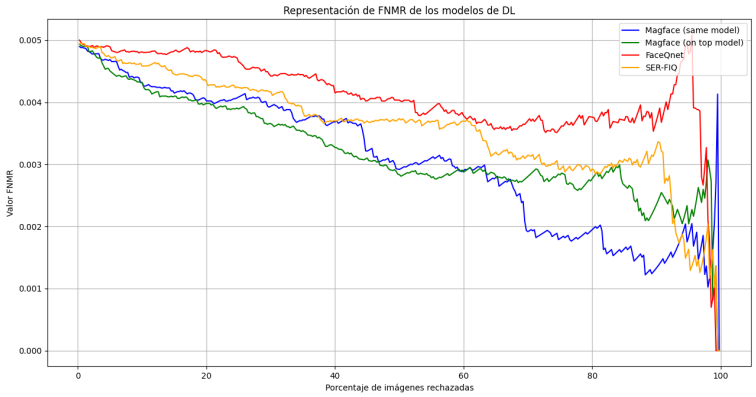


Figura 6.5: Evolución de la FNMR fijando el umbral al 0.5%

Para una FNMR fijada inicialmente al 0.5% tal y como se muestra en la figura 6.5, el sistema posee un umbral de similitudes muy bajo. Esto hace que el sistema se más vulnerable y menos seguro. Con un porcentaje tan pequeño (mostrado en la tabla 5.2 y 5.3), el sistema es permisivo en cuanto a ruido e imágenes de mala calidad. Este factor hace que los cálculos de FNMR empeoren ya que el sistema contiene una alta tasa de falso rechazo (personas que sí son las misma persona, pero el modelo determina que no lo son debido a la mala calidad de la muestra) y con ellos la usabilidad del sistema. Como se puede apreciar en la figura 6.5, la diferencia entre modelos de calidad es pequeña. Todos los modelos funcionan igual hasta el punto de rechazo del 50% de las imágenes. En este punto, ya las imágenes que formar la curva de FNMR son las que poseen una mejor calidad y afecta notoriamente sobre las métricas. En el caso de FaceQnet, la métrica mejora pero sigue siendo la peor curva respecto a los otros modelos. En el caso de SER-FIQ, con un rechazo de aproximadamente el 90% de las imágenes, presenta una mejora importante en su métrica. Cuanta mejor calidad poseen las imágenes, mejor funciona SER-FIQ en este caso. En cuanto a Magface, se obtiene la curva de los 'scores' extraídos respecto al modelo de Insightface (on top model) y la curva respecto al propio modelo de FR de Magface. Se puede apreciar como ambos modelos funcionan casi a la par hasta el punto en el que las imágenes aumentan de calidad según el modelo de Magface. Con el aumento de calidad, el modelo de FR propio de Magface funciona mejor obteniendo una métrica con hasta un 0.1% respecto a la curva de FNMR on top model.

Con el aumento del umbral definido para la FNMR fijada al 1%, se puede apreciar la evolución de las curvas de FNMR en la figura 6.6. Esta gráfica es muy similar a la gráfica 6.5, mostrando la tendencia anterior. FaceQnet debido a los 'scores' obtenidos continua mostrando una peor curva de FNMR mientras que el modelo de SER-FIQ continua mejorando la métrica para porcentajes de imágenes rechazadas menor al anterior con un umbral de FNMR fijado al 0.5%. Para esta nueva configuración del umbral, SER-FIQ mejora respecto a Magface ya que su curva de FNMR corta a la de Magface tanto para same model como para top

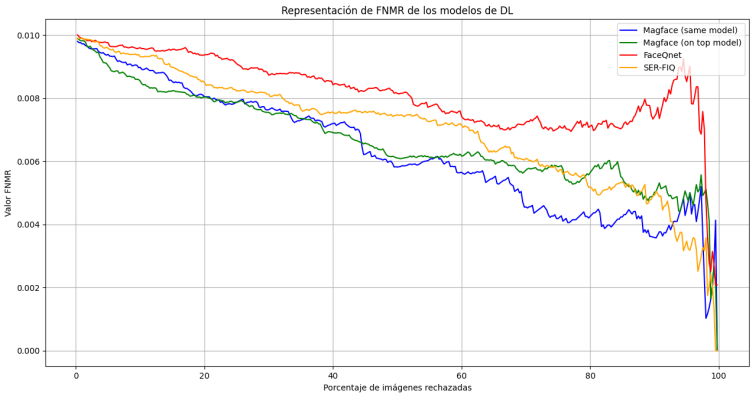


Figura 6.6: Evolución de la FNMR fijando el umbral al 1%

model, presentando así una mejor respuesta del sistema respecto a los demás modelos. Se puede apreciar como el modelo ya define un menor número de imágenes con falso rechazo. La calidad de las muestras están aumentando.

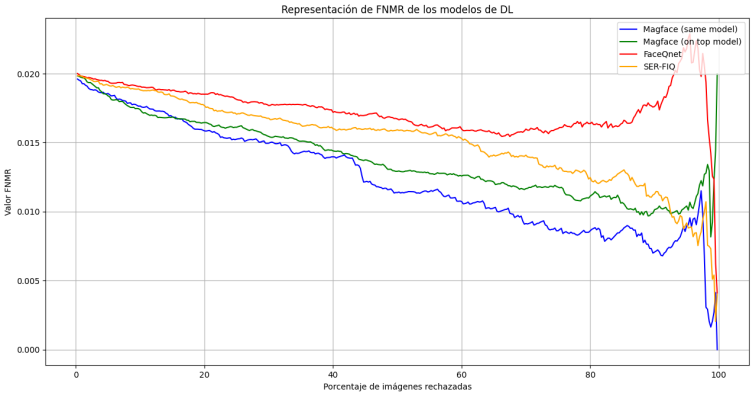


Figura 6.7: Evolución de la FNMR fijando el umbral al 2%

Para un umbral de FNMR fijado al 2% tal y como se muestra en la figura 6.7, la métrica de los modelos varía respecto a las anteriores. En este caso, el modelo de Magface, same model, presenta una métrica mejor al resto de modelos. Presenta una disminución de la FNMR mayor que ningún otro modelo. Para este caso, el modelo de SER-FIQ vuelve a mejorar la métrica de Magface respecto al modelo de Insightface pero únicamente cuando el porcentaje de rechazo de imágenes es muy alto, alrededor del 95%. Para este valor fijado de FNMR el modelo de Magface empieza a mostrar lo que sería un buen funcionamiento, disminuyendo el falso rechazo a medida que se van rechazando las primeras imágenes con peor calidad. Este funcionamiento se ve favorecido debido a las similitudes que extrae el modelo de Magface.

Cuanto mayor sea el umbral de similitud, menos falso rechazo debe poseer el modelo si la matriz de similitudes ha sido extraída con un buen modelo de FR.

A medida que aumenta el umbral fijado para la FNMR el sistema se vuelve más seguro y

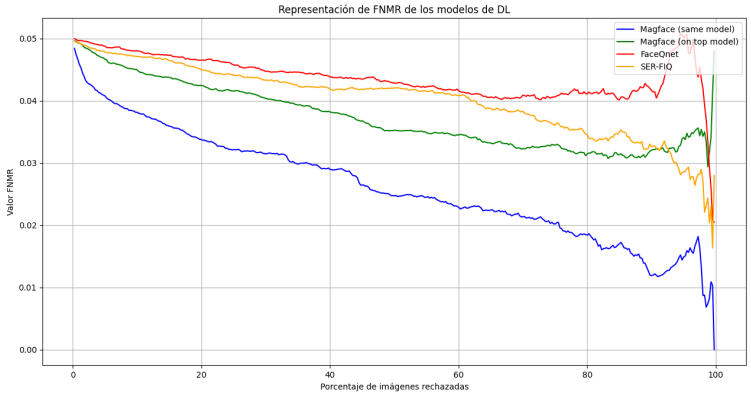


Figura 6.8: Evolución de la FNMR fijando el umbral al 5%

menos tolerante. A partir de estos umbrales se va a ver cómo funcionan realmente los modelos de estimación de calidad. Para un umbral de FNMR fijado al 5% como se muestra en la figura 6.8, se aprecia un cambio notorio en las métricas. Por una parte, FaceQnet continua representando una métrica demasiado alta para el sistema con poca disminución del valor de FNMR. Por otro lado, SER-FIQ también presenta una FNMR con poca pendiente excepto cuando el número de imágenes rechazadas es mayor. Por último, el modelo de MAgface presenta dos curvas diferentes con las mismas puntuaciones de calidad. Mientras que la métrica on top model sigue siendo mejor respecto a FaceQnet y a SER-FIQ (excepto para porcentajes de rechazo altos donde SER-FIQ funciona mejor), el modelo de Magface respecto a su propio modelo de FR presenta un mejor funcionamiento respecto al resto de modelos. Como se puede apreciar en la figura 6.8, el modelo de Magface (same model) presenta una disminución de la FNMR bastante notoria para los primeros porcentajes de imágenes rechazadas. Idealmente esa curva debería de decrecer de la misma forma para todos los porcentajes de imágenes rechazadas pero se ve como la curva no decrece tan rápido con los rechazos de las imágenes. Aún así, se representa la mejor curva de FNMR de todos los modelos. Esto se debe a que el modelo de FR se utiliza con los propios 'scores' de Magface. Aquí se ve la importancia de que el modelo de FR funcione adecuadamente de la mano con el modelo de estimación de calidad de la imagen.

Para analizar al sistema cuando la seguridad aumenta se decide aumentar el umbral de FNMR fijado para ver la evolución de los modelos al 15% tal y como se muestra en la figura 6.9. Esta gráfica muestra como evoluciona la FNMR para umbrales altos de similitud, sistema menos permisivo a las entradas. Como se puede apreciar, el modelo de FR de Insightface funciona mejor con los 'scores' extraídos por el modelo de Magface. Las curvas que se obtienen con este modelo no son las ideales debido a las similitudes que extrae este modelo respecto al conjunto de datos. Por otro lado, se obtiene la métrica del modelo de Magface (same model) que mejora con el aumento de seguridad del sistema. Se aprecia como la curva disminuye

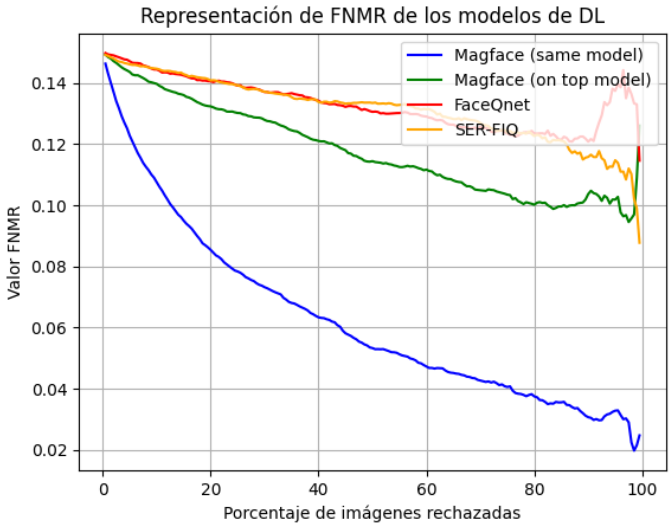


Figura 6.9: Evolución de la FNMR fijando el umbral al 15%

en valor de FNMR de forma considerable respecto al porcentaje de imágenes rechazadas del conjunto. El modelo de Magface obtiene una métrica mejor respecto al resto de modelos en un 6% de FNMR aproximadamente.
A la vista de los resultados obtenidos en las figuras 6.9 y 6.10 se ve como el modelo de Magface obtiene una métrica mejor que el modelo de FR de Insightface. Este modelo es capaz de disminuir la FNMR en los primeros rechazos de imágenes, disminuyendo la FNMR. Esta disminución de FNMR representa la caída del falso rechazo, es decir, existen menos imágenes que sí son la misma persona con similitudes inferiores al umbral definido. En definitiva, el modelo de Magface estima las similitudes entre pares de imágenes de forma más correcta que el modelo de Insightface.
Por último, se realiza el análisis para un valor de FNMR fijado al 20% como se muestra en la figura 6.10. Para este mayor umbral de similitud, el sistema se comporta como para el umbral de 15%. El modelo de FR de Insightface presenta unas curvas de FNMR mejorables. Esto se debe a que el modelo no es óptimo en la extracción de similitudes que se realiza para obtener la matriz de similitudes. Aún así, el modelo de estimación de calidad de Magface sigue siendo el que mejor funciona para este modelo de Insightface. Por otro lado, se obtiene la curva del modelo de Magface (same model) que sigue presentando la mejor métrica de todos los modelos debido la óptima extracción de similitudes por parte del modelo de FR y el uso de los propios puntajes de calidad que proporciona el modelo.

6.3 Evolución de FMR según los modelos de Face Recognition

La FMR va ligada a la FNMR. La FMR mide la tasa de aceptación que posee el sistema. De modo que, si el sistema posee una FNMR baja, el sistema posee una mayor tolerancia, una mayor tasa de aceptación, una mayor FMR. Por el contrario, cuando el sistema se vuelve más exigente (más seguro), la tasa de aceptación decrece ya que el umbral de similitud aumenta.

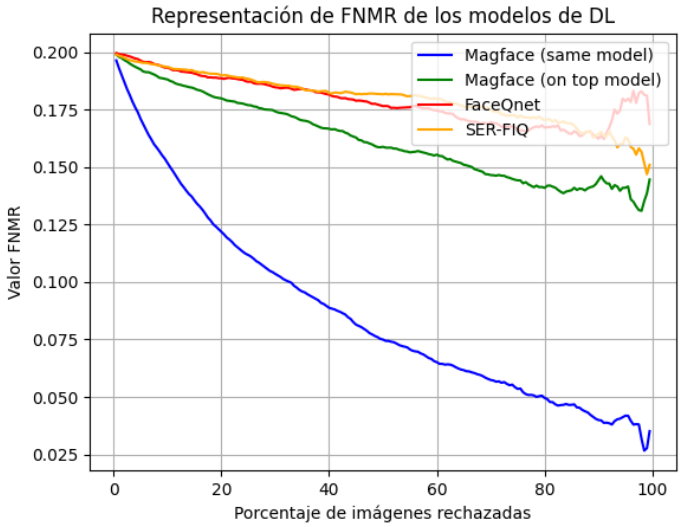


Figura 6.10: Evolución de la FNMR fijando el umbral al 20%

Al aumentar la FNMR disminuye la FMR.
Para ello, se muestra la evolución de la FMR para cada modelo de estimación de calidad respecto al valor de FNMR seleccionado para el análisis en el apartado 6.2. Tal y como se

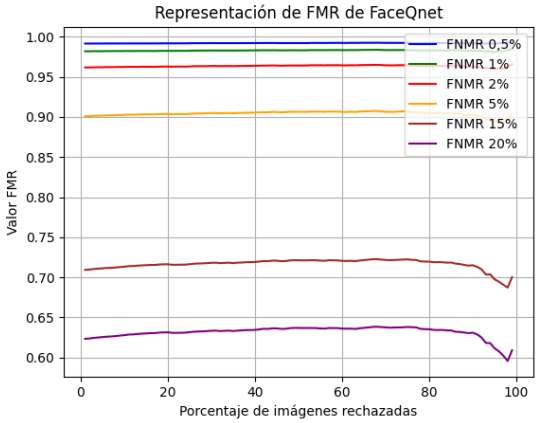


Figura 6.11: Evolución de la FMR para el modelo de FaceQnet

muestra en la figura 6.11 el modelo de FaceQnet reduce la tasa de aceptación a medida que la FMR aumenta. Este modelo consigue rebajar la FMR de casi un valor de 1 hasta un valor de 0.62 aproximadamente con una FNMR de un 20%. Como se aprecia en la figura, a medida que el sistema es menos tolerante, la tasa de aceptación decrece. Aunque el sistema decrece la FMR, no son valores óptimos para una autenticación facial.
En la figura 6.12 se puede observar como la FMR para el modelo de SER-FIQ sigue la misma tendencia que para el modelo de FaceQnet. Las gráficas de aceptación de los modelos son

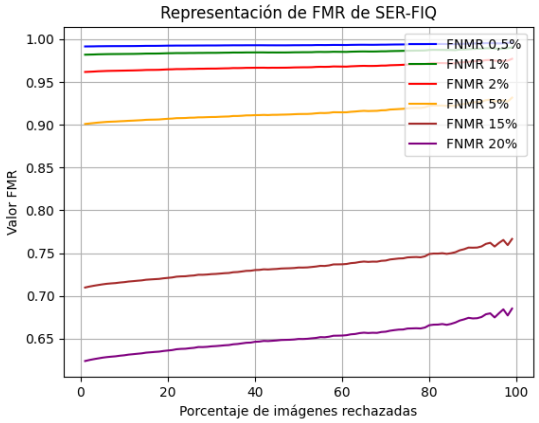


Figura 6.12: Evolución de la FMR para el modelo de SER-FIQ

muy parecidas debido a las similitudes que también poseen las curvas de FNMR de ambos modelos. En la figura 6.13 se puede apreciar el comportamiento de aceptación del modelo de

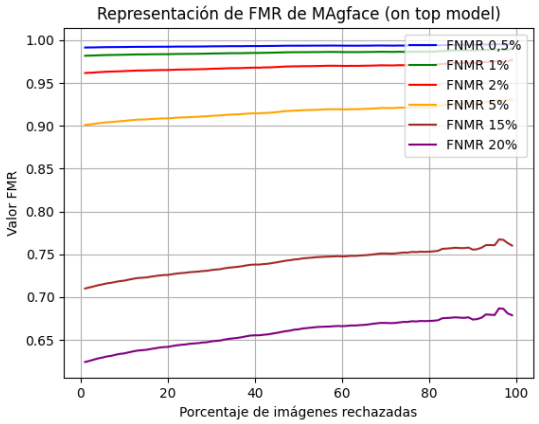


Figura 6.13: Evolución de la FMR para el modelo de Magface (on top model)

Insightface respecto a los 'scores' de Magface. Se puede apreciar como el sistema se comporta igual que los dos modelos anteriores ya que sus curvas de FNMR son muy similares. Se puede apreciar como con el aumento de la seguridad, el sistema decrece la tasa de aceptación. Para una FNMR del 20% la tasa de aceptación está entorno al 65%, lo que significa que para esa FNMR el modelo de Insightface rechaza un 35% de las imágenes debido a su calidad si el sistema estuviese funcionando en una aplicación real.

Por último, en la figura 6.14, se muestra el comportamiento del modelo de Magface cuando la FNMR aumenta. Este modelo reduce mucho la aceptación debido al correcto funcionamiento del sistema. Con valores de FNMR bajos ya el sistema tiene una tasa de aceptación cercana a 0.8, lejana del 0.99 del resto de modelos. En cuanto aumenta la FNMR el modelo reduce la tasa de aceptación considerablemente hasta valores aproximados a 0. En esta gráfica se

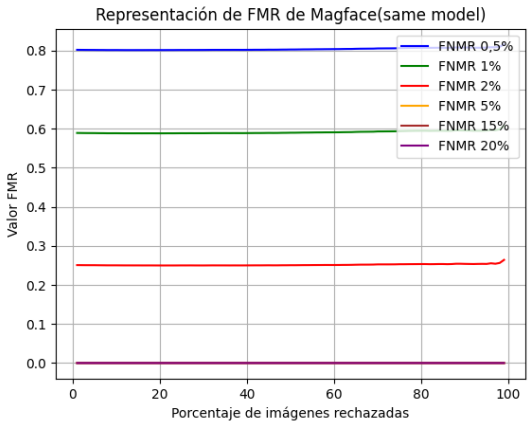


Figura 6.14: Evolución de la FMR para el modelo de Magface (same model)

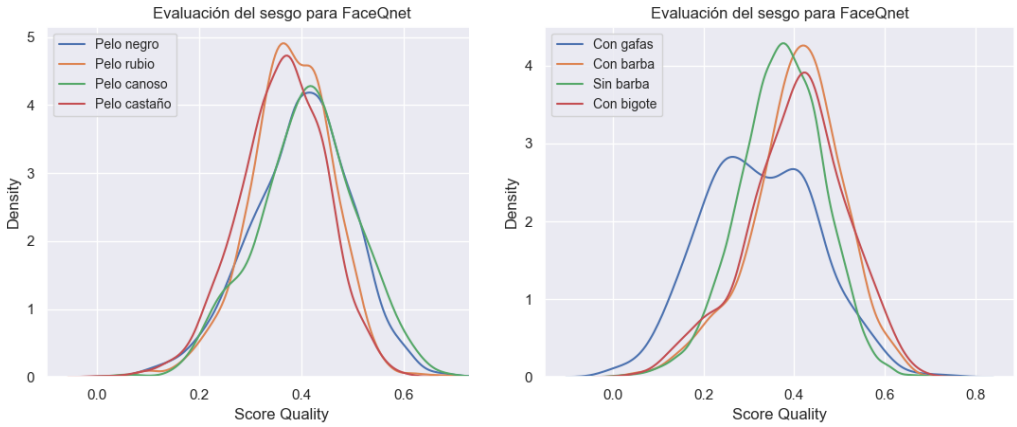
puede observar el correcto funcionamiento de un modelo de Face Recognition, bajando la tasa de aceptación a niveles de casi 0 para valores de FNMR superiores al 15%. Este modelo representa el mejor funcionamiento también en cuanto a curvas de FMR.

6.4 Evaluación del sesgo para cada modelo de estimación de calidad

Una vez analizada la importancia que tiene un un modelo de estimación de calidad sobre un modelo de Face Recognition, se procede a evaluar la posible existencia de sesgos dentro de cada modelo. Se define el sesgo de un modelo como el hecho de obtener un mejor o peor puntuación de calidad en función de las características que la propia imagen posee. El posible sesgo de un modelo también afecta a la calidad del modelo a la hora de utilizarse en sistemas de autenticación biométrica.
Para evaluar el sesgo de cada modelo, se estudia la distribución de 'scores' de calidad que presenta cada modelo con cada uno de los atributos a analizar.
El análisis se llevará a cabo fijándose en el color del pelo de la persona, las posibles oclusiones sobre la cara, el sexo y la edad de la persona.
El análisis se lleva a cabo gracias a las anotaciones de atributos que proporciona el conjunto de datos de CelebA.

6.4.1 Evaluación del sesgo sobre FaceQnet

En el modelo de FaceQnet se estudia el sesgo debido al color de pelo como se muestra en la figura 6.15a. Como se puede observar, el modelo de FaceQnet se encuentra sesgado. Cuando el modelo recibe una imagen de personas con el pelo de color rubio o castaño, estima una calidad de imagen inferior a personas con pelo negro y canoso. Las personas con pelo negro y canoso reciben una estimación de calidad superior a las personas con pelo castaño y rubio.

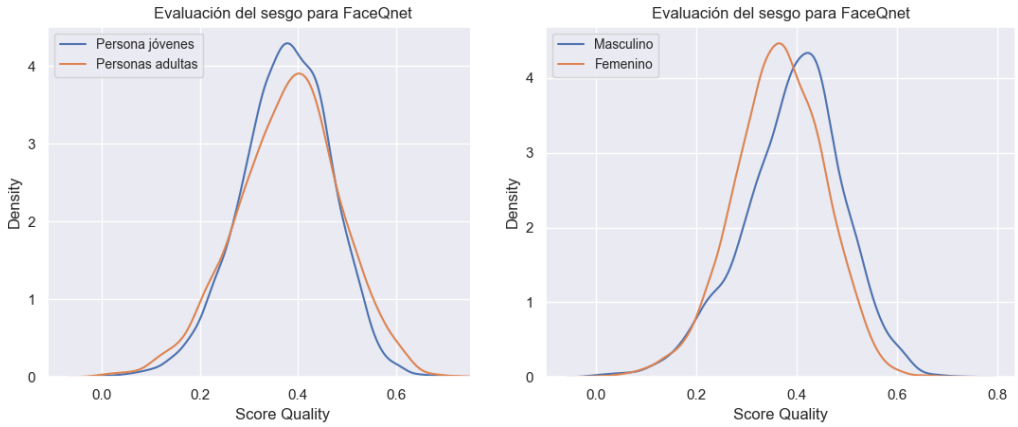


(a) Según el color de pelo de la persona **(b)** Según las oclusiones faciales

Figura 6.15: Evaluación del pelo y oclusiones para el modelo de FaceQnet

En cuanto a la estimación entre personas con pelo canoso y pelo negro, el modelo estima calidades similares pero funciona mejor para personas con el pelo canoso.
En la figura 6.15b se muestra las diferentes curvas de estimación de calidad según las oclusiones faciales que contiene la imagen. En este caso se ve claramente que el modelo de FaceQnet estima calidades bajas para personas con gafas. Esto puede ser un problema ya que al estimar una baja calidad para este tipo de oclusiones, cuando un sistema sea bastante seguro, correrá el riesgo de rechazar todas las imágenes de personas con gafas, algo que sería erróneo ya que por llevar gafas no se debe de estimar una peor calidad. Este tipo de problemas suelen darse debido a los reflejos que se dan en las gafas al realizar la fotografía. Por otro lado, el modelo funciona peor para personas sin barba que para personas con barba, la cual cosa hace indicar que el modelo ha sido más entrenado con imágenes de usuarios con barba que sin barba. Por último, el modelo funciona de forma similar respecto a usuarios con barba y con usuarios con bigote, pero estima una mejor calidad para personas con barba. El modelo contiene un sesgo importante con el cual las personas con gafas se verán muy a afectadas a la hora de que el sistema valide su imagen como buena en cuanto a calidad para el posterior reconocimiento facial.

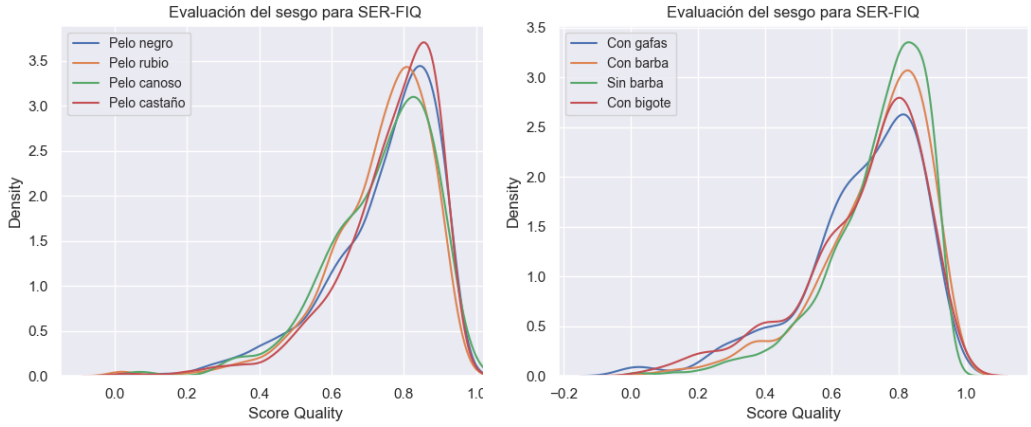
En la figura 6.16a también se puede apreciar otra muestra de sesgo en el modelo de FaceQnet. El modelo estima calidades de imagen inferiores para personas jóvenes que para personas adultas.
Por último, en la figura 6.16b se puede observar cómo afecta el sexo de la persona en la posterior estimación de calidad en el modelo de FaceQnet. Como se puede apreciar, existe un sesgo grande entre sexos. El modelo estima calidades de imagen superiores para el sexo masculino que para el sexo femenino.
A la vista de los resultados, este modelo de estimación de calidad presentará bajos 'scores' de calidad para personas que lleven gafas y sean del sexo femenino. Un sesgo importante para un posterior reconocimiento facial.



(a) Según la edad de la persona **(b)** Según el sexo de la persona

Figura 6.16: Evaluación de la edad y el sexo de la persona para el modelo de FaceQnet

6.4.2 Evaluación del sesgo sobre SER-FIQ

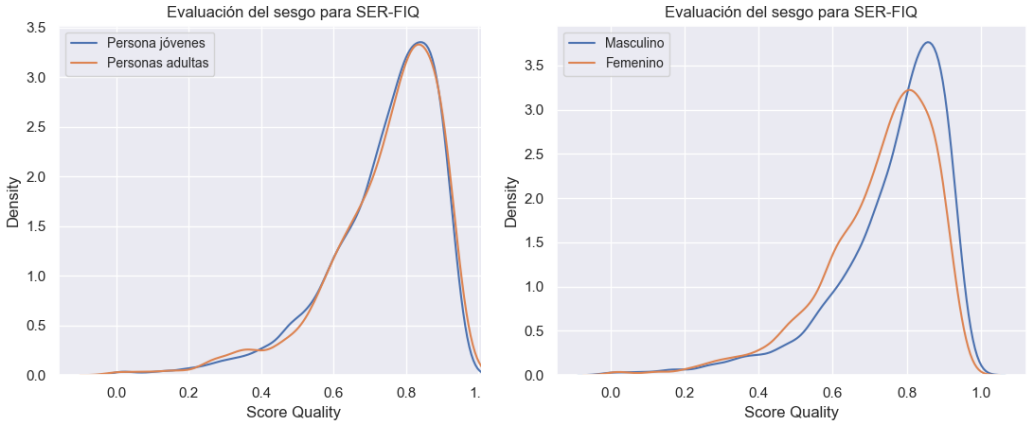


(a) Según el color de pelo de la persona **(b)** Según las oclusiones faciales

Figura 6.17: Evaluación del pelo y oclusiones para el modelo de SER-FIQ

En cuanto al modelo de SER-FIQ, en la figura 6.17a se observa como el modelo presenta un comportamiento similar a todos los tipos de cabellos de los usuarios pero hace una distinción muy clara con los usuarios de pelo castaño. Para este tipo de pelo, el modelo de SER-FIQ estima puntuaciones de calidad mayores que para el resto de colores de pelo. Por otro lado, están los usuarios de pelo canoso sobre el que los usuarios de pelo negro obtendrán una mayor calidad y lso usuarios de pelo rubio una peor calidad.
En la figura 6.17b se representa el comportamiento del modelo de SER-FIQ respecto a una serie de oclusiones. Para las oclusiones, el modelo de SER-FIQ presenta un sesgo importante entre las diferentes oclusiones. Para la oclusión que peor puntuación de calidad estima es

para usuarios con gafas. No es un sesgo tan importante como en FaceQnet pero sí afecta este factor de las gafas en cuanto a la calidad. El modelo de SER-FIQ estima una mayor calidad para usuarios sin barba, el hecho de que no exista oclusión en la región facial hace que este modelo funcione mucho mejor. Por último, el modelo funciona de forma similar ante usuarios con bigote y usuarios con barba, estimando una mayor calidad para usuarios con barba.



(a) Según la edad de la persona **(b)** Según el sexo de la persona

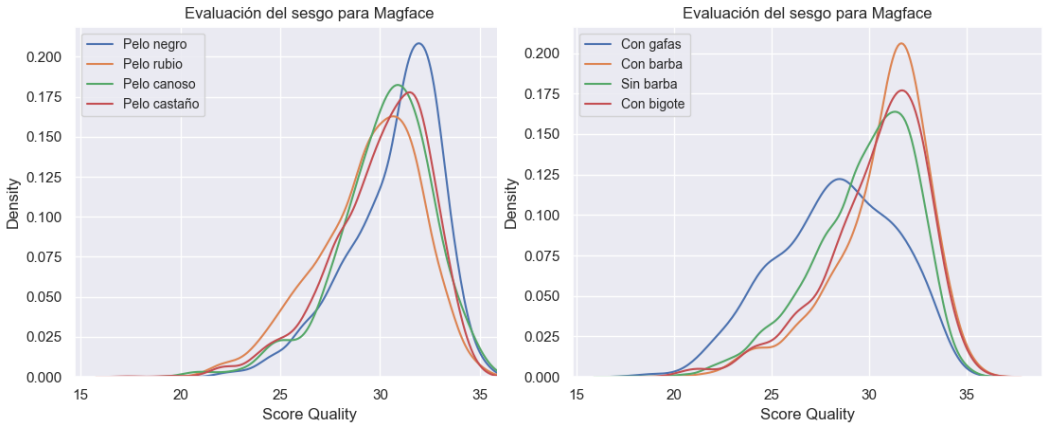
Figura 6.18: Evaluación de la edad y el sexo de la persona para el modelo de SER-FIQ

Sobre la figura 6.18a se puede apreciar como el modelo de SER-FIQ no se encuentra sesgado en cuanto a la edad de la persona que aparece en la foto. Tal y como se ve en la gráfica el modelo estima casi el mismo valor de calidad para personas jóvenes que para personas adultas. No se encuentra sesgado en este aspecto.
En cambio, la problemática del sesgo entre el sexo de la persona viene mostrada sobre la figura 6.18b. Al igual que el modelo de FaceQnet, SER-FIQ obtiene mejores puntuaciones de calidad para imágenes en las que aparecen hombres que en imágenes en las que aparecen mujeres. Se puede apreciar un sesgo importante en la curva de la evaluación del sesgo.

El modelo de SER-FIQ funciona correctamente sobre personas de cualquier edad. En cambio, obtiene un mayor sesgo entre personas de diferente sexo y acentúa una pérdida de calidad sobre personas que utilizan gafas.

6.4.3 Evaluación del sesgo sobre Magface

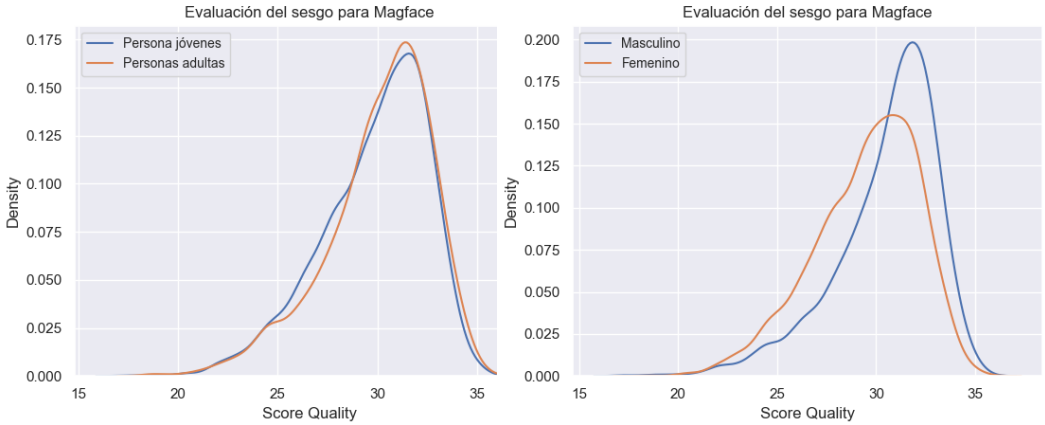
Para evaluar el sesgo del modelo de Magface se emplea el mismo criterio que en los modelos anteriores. En la figura 6.19a se muestra la distribución de los 'scores' de calidad del modelo Magface sobre el color del pelo de usuario. Este modelo obtiene una mayor puntuación de calidad sobre imágenes de personas con el pelo negro. En el caso del pelo canoso y castaño el modelo se comporta de forma similar, obteniendo peores resultados de calidad que para usuarios con pelo negro. En este caso, el modelo de Magface funciona peor con personas de



(a) Según el color de pelo de la persona **(b)** Según las oclusiones faciales

Figura 6.19: Evaluación del pelo y oclusiones para el modelo de Magface

pelo rubio. Para este tipo de personas el modelo estima puntuaciones de calidad inferiores al resto de colores de pelo. El modelo hace una clara distinción entre los diferentes colores de pelo, existe un sesgo importante en este aspecto.
Para la comprobación del modelo según las diferentes oclusiones, se muestra la gráfica 6.19b. Como en el resto de modelos, Magface realiza una distinción a la hora de la calidad sobre personas con gafas, otorgándoles una menor calidad a ese tipo de imágenes. En este caso, la diferencia es muy notoria al ver la figura 6.19b. En cuanto a personas con barba o sin ella y personas con bigote, el modelo también presenta estar sesgado aunque no en gran medida como sí lo está con las personas que utilizan gafas. Esta diferencia de calidad con respecto a las personas que usan gafas pueden causar una problemática importante a la hora de estimar la calidad para un posterior reconocimiento facial.



(a) Según la edad de la persona **(b)** Según el sexo de la persona

Figura 6.20: Evaluación de la edad y el sexo de la persona para el modelo de Magface

En cuanto a la edad de los usuarios, según la figura 6.20a, existe un pequeño sesgo en el modelo. Para personas jóvenes y adultas el modelo se comporta de manera similar. No obstante, para personas jóvenes el modelo estima una calidad mayor.
Por último, se evalúa cómo afecta el sexo de la persona en cuanto a la calidad de la imagen definida por el modelo. En la figura 6.20b se muestra el comportamiento de Magface. Este modelo presenta un sesgo significativo en cuanto al sexo de la persona. Se puede ver como el modelo estima valores de calidad altos cuando el sexo es masculino. En cambio, si el sexo es femenino, la puntuación de calidad del modelo es inferior.
Magface presenta sesgos importantes a la hora de estimar la calidad de una imagen. Uno de ellos es el hecho de poseer un sexo masculino o femenino. Si se posee un sexo femenino, la imagen tendrá más posibilidades de ser rechazada antes de realizar el reconocimiento facial. Por otra parte, el modelo hace una clara distinción entre personas con gafas y las que no. El valor de calidad de una imagen se ve afectado en gran medida por el hecho de llevar gafas. Este sesgo dificulta un buen reconocimiento facial.
Debido a que las puntuaciones de calidad de Magface son extraídas a partir de los vectores de características que el modelo extrae de cada imagen, viendo los sesgos presentes en el modelo de Magface en las figuras 6.19 y 6.20, se puede apreciar que el modelo de Face Recognition de Magface también se encuentra sesgado para los mismos aspectos. Esto ocurre porque el modelo de FR de Magface es el encargado de determinar la puntuación de calidad final de cada imagen.

7 Conclusiones

Una vez realizado el trabajo sobre los diferentes modelos de estimación de calidad con DL y sobre el modelo de FR cumpliendo los objetivos establecidos al comienzo del trabajo, se procede a las conclusiones.

Tras el análisis realizado sobre las diferentes técnicas de estimación de calidad para imágenes de caras, se puede concluir que los modelos de estimación que utilizan DL son mejores para implementarlos en aplicaciones en tiempo real. Esta mejora que presentan los modelos de DL respecto a los métodos tradicionales de procesamiento de imagen es sobre todo el tiempo de estimación que conlleva una imagen. La estimación de calidad de una imagen utilizando DL supone un menor coste computacional y un ahorro de tiempo para poder realizar un reconocimiento facial en tiempo real. Otra mejora que aportan los modelos de DL son las técnicas novedosas que presentan los modelos como FaceQnet, SER-FIQ, Magface, etc. Esta variación de técnicas hacen que los modelos sean más útiles respecto a las comprobaciones básicas que se realizan con técnicas tradicionales.

A la vista de las distribuciones de los valores de calidad que estiman los modelos de cada imagen, se puede ver como Magface presenta un modelo más restrictivo en cuanto a la puntuación de calidad de cada imagen para este conjunto ya que el modelo de Magface no tiene un rango definido de calidad. SER-FIQ, podría presentar una mejora en la estimación de calidad si se utilizase la tasa de dropout que se recomienda sobre su artículo. En este trabajo se ha realizado con una tasa de repetición de la extracción del valor de calidad mediante *dropout* de 10 mientras que en el trabajo se recomienda una tasa de repetición de 100. Esta disminución de la tasa de repetición se realiza debido a limitaciones computacionales. Con una tasa de 100 repeticiones por imagen el modelo tardada en predecir la calidad de todo el conjunto 363 horas. Una tasa mayor podría haber mejorado la calidad extraída para cada imagen.

Tras el análisis de las métricas obtenidas de todos los modelos de estimación mediante un modelo de FR se puede concluir que un sistema de estimación de calidad debe ser muy preciso cuando se enfrenta a un modelo de reconocimiento facial desconocido. Esto sucede cuando el modelo de FR no se ha entrenado con el modelo de estimación de calidad utilizado. Pero el estudio se realiza así para simular un funcionamiento de una aplicación real en la que un modelo de reconocimiento facial no tiene por qué estar entrenado con el modelo de estimación, aunque si lo estuviese se obtendrían mejores resultados. Fijándose en las métricas tanto de FNMR como en las de FMR se ve como el modelo de Magface resulta ser el mejor modelo de estimación de los tres utilizados para el modelo de FR.

El análisis del modelo de FR de Magface se realiza para mostrar la mejora que presenta un sistema de reconocimiento facial cuando el modelo utiliza un estimador en el que está basado el modelo. Como se ha visto en el apartado de resultados, cuando la FNMR aumenta (umbral de similitud más alto), el modelo de Magface de FR (análisis same model) funciona mejor disminuyendo la FNMR fijada al inicio. Esta disminución de FNMR representa la caída del

falso rechazo del sistema, lo que supone que el sistema es más seguro restrictivo. El sistema funciona mejor con menor rango de equivocación en el reconocimiento facial. El modelo de Magface presenta la mejor métrica de todos los modelos. Si se tuviese que implementar un modelo para un reconocimiento facial real, este sería el modelo elegido.

Los modelos de estimación de calidad utilizados durante el trabajo son modelos con un funcionamiento correcto que no han podido mostrar su funcionamiento en base al modelo de FR de Insightface. Este modelo de FR presenta un funcionamiento lejano al esperado, extrayendo valores de similitudes demasiado altos para todas las imágenes del conjunto. Sobre otro modelo de FR los modelos de estimación de calidad funcionarían mejor como se ha visto con el estimador de Magface y su propio modelo de FR.

Los modelos han sido probados sobre un conjunto de datos público sobre el que ningún modelo había sido entrenado. Este hecho dificulta la extracción de la calidad para cada imagen. Con conjuntos de datos con los cuales se hayan entrenado los modelos de estimación de calidad serán más precisos. Esto también repercute en una mejora de las métricas de los modelos.

Por último, el análisis de sesgos en los modelos ha sido importante en este trabajo. Gracias a este análisis se ha podido demostrar que los modelos de estimación presentan sesgos sobre las características que posee una imagen. La mayoría de las problemáticas en los modelos ha sido con personas que llevan gafas. Este sesgo es un factor importante ya que actualmente más del 50 % de la población mundial utiliza gafas. También se ha visto como según si una persona era hombre o mujer se estimaban valores de calidad más altos o más bajos. Magface representa la mejor métrica de todos los modelos cuando se ha utilizado con su propio modelo de reconocimiento facial. El hecho de que Magface extraiga la calidad a partir del vector de características que extrae el modelo, hace concluir que el modelo de FR de Magface también se encuentra sesgado en el sexo de la persona, el color del pelo de la persona o en si lleva o no gafas esa persona.

7.1 Trabajos futuros

En cuanto a trabajos futuros se dejan planteados una serie de trabajos que se podrían realizar para continuar el trabajo realizado.
Un trabajo futuro podría ser el análisis y estudio de la técnica de SER-FIQ sobre el modelo de Insightface (que es el que utiliza esta técnica) pero con diferentes números de predicciones por imagen, probando a obtener la calidad de una imagen aumentando el número de predicciones que extrae esta técnica ya que en este trabajo se ha visto reducida debido al alto coste computacional. El estudio de cómo afecta el número de predicciones que extrae la técnica en relación a la mejora de las métricas obtenidas. A la vista de los resultados obtenidos con Magface, las métricas mediante la técnica de SER-FIQ (ya que sería también un same model) mejorarían y optimizarían su uso en una aplicación real.
Otra ampliación del trabajo podría ser el análisis de otros modelos de estimación de calidad y contrastar los resultados con los obtenidos con estos tres modelos. Un modelo que se propone para un análisis futuro es SDD-FIQA (comentado en el marco teórico pero no analizado) o

el modelo de PCNet, Xie y cols. (2020), que es otro estimador de calidad de imágenes faciales.

A la vista de los sesgos existentes en los modelos de estimación de calidad, se podría realizar un estudio más a fondo sobre los sesgos de los modelos que imposibiliten el correcto funcionamiento de los modelos. Es decir, aquellos rasgos que impidan realizar un reconocimiento facial cuando la imagen realmente sí tiene la calidad necesaria. Se propone un análisis de sesgos más en función de los rasgos faciales, razas (asiáticos, africanos, caucásicos) y color de la piel.
Además, se podría realizar una ampliación basándose en la mecánica de este trabajo pero con diferentes datasets. Así se podría ver como se comportan los modelos de FR con diferentes datasets ante una mayor variación de imágenes en el test. Analizar el rango de calidad que se predicen a medida que los datasets proporcionan una mayor cantidad de imágenes diferentes en cuanto a iluminación, pose, etc.
Un trabajo de ampliación más técnico podría ser el reducir un modelo de los estudiados para usarlo en 'devices' con restricciones computacionales altas y probar su funcionalidad en el reconocimiento facial. Se propone implementar el modelo de SDD-FIQA ya que se centra en el score de calidad sin tener que utilizar modelos de FR para obtener la calidad de una imagen. Analizar su funcionalidad y el coste computacional.

Bibliografía

Bruno, M., Alarcón-Paredes, A., y Alonso Silverio, G. (2017, 09). Análisis de lectores biométricos de huella dactilar implementados en una raspberry pi..

Coşkun, M., Uçar, A., Yildirim, ., y Demir, Y. (2017). Face recognition based on convolutional neural network. En *2017 international conference on modern electrical and energy systems (mees)* (p. 376-379). doi: 10.1109/MEES.2017.8248937

Deng, J., Guo, J., Liu, T., Gong, M., y Zafeiriou, S. (2020). Sub-center arcface: Boosting face recognition by large-scale noisy web faces. En *Proceedings of the ieee conference on european conference on computer vision.*

Deng, J., Guo, J., Ververas, E., Kotsia, I., y Zafeiriou, S. (2020). Retinaface: Single-shot multi-level face localisation in the wild. En *Cvpr.*

Deng, J., Guo, J., Xue, N., y Zafeiriou, S. (2018). *Arcface: Additive angular margin loss for deep face recognition.* arXiv.

Ferrara, M., Franco, A., Maio, D., y Maltoni, D. (2012). Face image conformance to iso/icao standards in machine readable travel documents. *IEEE Transactions on Information Forensics and Security*, *7*(4), 1204-1213. doi: 10.1109/TIFS.2012.2198643

Guo, J., Deng, J., Lattas, A., y Zafeiriou, S. (2021). *Sample and computation redistribution for efficient face detection.* arXiv.

Hernandez-Ortega, J., Fierrez, J., Gomez, L. F., Morales, A., Gonzalez-de Suso, J. L., y Zamora-Martinez, F. (2021). *Faceqvec: Vector quality assessment for face biometrics based on iso compliance.* arXiv.

Hernandez-Ortega, J., Galbally, J., Fierrez, J., y Beslay, L. (2020). *Biometric quality: Review and application to face recognition with faceqnet.* arXiv.

Hernandez-Ortega, J., Galbally, J., Fierrez, J., Haraksim, R., y Beslay, L. (2019). *Faceqnet: Quality assessment for face recognition based on deep learning.* arXiv.

Hernández, M., Plasencia-Calaña, Y., y Vazquez, H. (2016, 01). Metric learning in the dissimilarity space to improve low-resolution face recognition.

Hernández-Durán, M., y Plasencia-Calaña, Y. (2016, 03). Aprendizaje de métrica para el reconocimiento de rostros a partir de imágenes de baja resolución. *Revista Cubana de Ciencias Informáticas*, *10*, 124 - 133.

Kasar, M., Bhattacharyya, D., y Kim, T.-H. (2016, 03). Face recognition using neural network: A review. *International Journal of Security and Its Applications*, *10*, 81-100.

Liu, W., Wen, Y., Yu, Z., Li, M., Raj, B., y Song, L. (2017). *Sphereface: Deep hypersphere embedding for face recognition.* arXiv.

Liu, Z., Luo, P., Wang, X., y Tang, X. (2015, December). Deep learning face attributes in the wild. En *Proceedings of international conference on computer vision (iccv).*

Lu, X. (2008, 07). Image analysis for face recognition.

Mehlig, B. (2021). *Machine learning with neural networks.* Cambridge University Press.

Meng, Q., Zhao, S., Huang, Z., y Zhou, F. (2021). Magface: A universal representation for face recognition and quality assessment.

Méndez-Vázquez, H., Chang, L., Rizo-Rodríguez, D., y Morales-González, A. (2012, 06). Evaluación de la calidad de las imágenes de rostros utilizadas para la identificación de las personas. *Computación y Sistemas*, *16*, 147 - 165.

Orozco-Rosas, U., García-Vázquez, M., y Ramirez, A. (2012, 06). Algoritmos de procesamiento del iris para un sistema de reconocimiento biométrico..

Ou, F.-Z., Chen, X., Zhang, R., Huang, Y., Li, S., Li, J., … Wang, Y.-G. (2021). Sdd-fiqa: Unsupervised face image quality assessment with similarity distribution distance.

Rani, K., Kalra, M., y Kumar, R. (2022, 01). Infrared thermography-based facial classification using machine learning. En (p. 275-284).

Schroff, F., Kalenichenko, D., y Philbin, J. (2015, jun). FaceNet: A unified embedding for face recognition and clustering. En *2015 IEEE conference on computer vision and pattern recognition (CVPR).* IEEE.

Sharkawy, A.-N. (2020, 08). Principle of neural network and its main types: Review. *Journal of Advances in Applied Computational Mathematics*, *7*, 8-19. doi: 10.15377/2409-5761.2020.07.2

Terhörst, P., Kolf, J. N., Damer, N., Kirchbuchner, F., y Kuijper, A. (2020). *Ser-fiq: Unsupervised estimation of face image quality based on stochastic embedding robustness.* arXiv.

Wasnik, P., Raja, K. B., Ramachandra, R., y Busch, C. (2017). Assessing face image quality for smartphone based face recognition system. En *2017 5th international workshop on biometrics and forensics (iwbf)* (p. 1-6). doi: 10.1109/IWBF.2017.7935089

Xie, W., Byrne, J., y Zisserman, A. (2020). *Inducing predictive uncertainty estimation for face recognition.* arXiv.

Lista de Acrónimos y Abreviaturas

DL	Deep Learning.
FMR	False Match Rate.
FNMR	False Non-Match Rate.
FR	Face Recognition.
IA	Inteligencia Artificial.
ICAO	Organización de Aviación Civil Internacional.
IEEE	Institute of Electrical and Electronics Engineers.
NIST	Instituto Nacional de Estándares y Tecnología.
TFG	Trabajo Final de Grado.