

Escuela Politécnica Superior

20
21

Trabajo fin de grado

Sistemas de Reconocimiento Facial basados en Redes Neuronales Convolucionales usando Información de Género y Raza



Julia Simón Chico

Escuela Politécnica Superior
Universidad Autónoma de Madrid
C/ Francisco Tomás y Valiente nº 11

**UNIVERSIDAD AUTÓNOMA DE MADRID
ESCUELA POLITÉCNICA SUPERIOR**



Grado en Ingeniería Informática

TRABAJO FIN DE GRADO

**Sistemas de Reconocimiento Facial basados en
Redes Neuronales Convolucionales usando
Información de Género y Raza**

**Autor: Julia Simón Chico
Tutor: Rubén Vera Rodríguez**

junio 2021

Todos los derechos reservados.

Queda prohibida, salvo excepción prevista en la Ley, cualquier forma de reproducción, distribución comunicación pública y transformación de esta obra sin contar con la autorización de los titulares de la propiedad intelectual.

La infracción de los derechos mencionados puede ser constitutiva de delito contra la propiedad intelectual (*arts. 270 y sgts. del Código Penal*).

DERECHOS RESERVADOS

© 9 de junio de 2021 por UNIVERSIDAD AUTÓNOMA DE MADRID

Francisco Tomás y Valiente, n° 1

Madrid, 28049

Spain

Julia Simón Chico

Sistemas de Reconocimiento Facial basados en Redes Neuronales Convolucionales usando Información de Género y Raza

Julia Simón Chico

AGRADECIMIENTOS

En primer lugar quería agradecerse a Rubén Vera, mi tutor. Gracias a él he podido aprender mucho con este trabajo de fin de grado. Además, me ha estado ayudando todo este tiempo a pesar de las dificultades por el poco tiempo que tenía yo con el máster.

Quería agradecerse a toda mi familia que me ha estado apoyando a lo largo de estos años en todas las decisiones que he tomado. Pero sobre todo quería agradecerse en especial a mi madre que me ha estado ayudando desde la distancia para que yo pudiera acabar este trabajo. Sin ella no podría haberlo terminado y por eso le estaré siempre agradecida.

Gracias a Álvaro que hizo que mi extraño paso por informática fuera mucho más llevadero y que me enseñó algunas cosas básicas de informáticos.

También estoy muy agradecida a mis amigos de toda la vida que han estado ahí conmigo todos estos años a pesar de que en algunas épocas estuviera un poco desaparecida.

Por último, quería darle las gracias a todas esas personas que han estado ahí durante estos dos últimos años en los que además del máster estaba haciendo este trabajo de fin de grado. Esas personas que me han ayudado cuando lo he necesitado y sobre todo me han dado apoyo para poder hacerlo todo. Gracias a esas que estuvieron ahí en verano apoyándome y a las que han estado en esta última parte escuchándome cuando no tenía tiempo. En especial muchas gracias a Marta por ser mi gran apoyo este último año que a veces ha sido un poco duro.

RESUMEN

Una serie de estudios han demostrado que hay sesgos en los sistemas actuales de reconocimiento facial. La motivación principal de este trabajo es terminar con los sesgos de género y raza para conseguir tener una buena precisión en todos los grupos demográficos. Para lograrlo, el objetivo de este trabajo es añadir información de género y etnia en estos sistemas para mejorar el rendimiento.

Para añadir información de género y etnia se plantea que sobre las redes neuronales convolucionales ya existentes, Resnet-50 y VGG16, previamente entrenadas en reconocimiento facial, se añada una capa de triplet loss. Esta capa se entrena únicamente con personas de un cierto género, una etnia en concreto o una combinación de etnia y género concretos. Al entrenarla de esta forma, ese modelo se va a especializar en esa clase concreta, y por ello ese será el modelo que se utilizará para las personas que pertenezcan a ese grupo demográfico.

Para evaluar los resultados se utilizan diferentes bases de datos existentes que disponen de información de etnia o género. En concreto se utilizan VGGFace2, RFW y BUPT-Balancedface. Se ha observado que para las bases de datos de VGGFace2 y RFW los resultados eran satisfactorios, mejorando los resultados en casi todas las clases para ambos modelos base, con una disminución de aproximadamente un 2 % del EER, equal error rate, en algunas clases de Resnet50 y una reducción del EER mayor en VGG16, teniendo una disminución de un 4 % en el mejor de los casos. Para BUPT-Balancedface, se han visto mejoras en algunos grupos pero no en todos. Además, después de realizar diferentes experimentos se ha concluido que el número de triplets que se generan para entrenar la red influye en que haya unos mejores resultados.

Sin embargo, a pesar de mejorar en ciertos grupos demográficos el resultado obtenido por los sistemas ya existentes, se ha visto que sigue habiendo diferencias entre los diferentes grupos, obteniendo siempre un mejor resultado para hombres caucásicos.

PALABRAS CLAVE

Reconocimiento facial, triplet loss, redes neuronales convolucionales, bases de datos, sesgos, soft biometrics, género, etnia

ABSTRACT

A number of studies have shown that there are biases in current face recognition systems. The main motivation of this work is to end gender and race biases in order to achieve good accuracy across all demographic groups. To achieve this, the main goal of this work is to add gender and ethnicity information to those systems to improve their performance.

To add gender and ethnicity information, we propose to add a triplet loss layer to the existing convolutional neural networks, Resnet-50 and VGG16, previously trained in face recognition. This layer is trained only with people of a certain gender, a specific ethnicity or a combination of specific ethnicity and gender. By training it in this way, that model specializes in that particular class, and therefore that is the model that is used for people belonging to that demographic group.

To evaluate the results, we use different existing databases that have information on ethnicity or gender. Specifically, VGGFace2, RFW and BUPT-Balancedface are used. It has been observed that for the VGGFace2 and RFW databases the results were satisfactory, improving results in almost all classes for both base models, with a decrease of about 2% of the EER, equal error rate, in some Resnet50 classes and a larger EER reduction in VGG16, having a decrease of 4% in the best case. For BUPT-Balancedface, improvements were seen in some but not all groups. In addition, after performing different experiments, it has been concluded that the number of triplets generated to train the network influences better results.

However, despite improving the results obtained by the existing systems in certain demographic groups, it has been found that there are still differences between the different groups, always obtaining a better result for Caucasian men.

KEYWORDS

Facial recognition, triplet loss, convolutional neural networks, databases, biases, soft biometrics, soft biometrics, gender, etnia

ÍNDICE

1	Introducción	1
1.1	Motivación	1
1.2	Objetivos	2
1.3	Estructura de la memoria	2
2	Estado del arte	3
2.1	Sistemas de reconocimiento facial basados en redes neuronales	3
2.2	Sesgos en los sistemas de reconocimiento facial	4
2.3	Soft biometrics	5
2.4	Conclusiones	6
3	Diseño	7
3.1	Bases de datos	7
3.1.1	VGGFace2	7
3.1.2	DiveFace	8
3.1.3	RFW	8
3.1.4	BUPT-Balancedface	9
3.2	Modelo propuesto	9
3.2.1	Modelos preentrenados: Resnet-50 y VGG-16	10
3.2.2	Triplet loss	11
4	Desarrollo	15
5	Resultados	19
5.1	Resultados VGGFace2	19
5.2	Resultados RFW	21
5.3	Resultados BUPT-Balancedface	24
5.4	Resultados RFW con BUPT-Balancedface	28
6	Conclusiones y trabajo futuro	31
6.1	Conclusiones	31
6.2	Trabajo futuro	31
	Bibliografía	36
	Apéndices	37
A	Resultados RFW	39

LISTAS

Lista de figuras

3.1	Función Softmax	10
3.2	Arquitectura VGG-16	10
3.3	Bloque residual	11
3.4	Arquitectura Resnet-50	11
3.5	Ejemplo triplet	12
3.6	Estructura triplet	12
3.7	Método propuesto	13
3.8	Estructura modelo CNN.	14
4.1	Imagen antes y después de recortar el área de la cara.	16
4.2	Entrenamiento y validación del triplet loss.	17
5.1	Gráficas por género VGGFace2.....	20
5.2	Gráficas por género RFW.	21
5.3	Gráficas por etnia RFW.	22
5.4	Gráficas por género BUPT-Balancedface.	25
5.5	Gráficas por etnia BUPT-Balancedface.	26
5.6	Gráficas por etnia BUPT-Balancedface + RFW.	28
A.1	Gráficas por etnia para las mujeres en RFW.	39
A.2	Gráficas por etnia para los hombres en RFW.	40
B.1	Gráficas por etnia en mujeres BUPT-Balancedface.	41
B.2	Gráficas por etnia en hombres BUPT-Balancedface.	42

Lista de tablas

3.1	Información de género y etnia para la base de datos de RFW.	9
3.2	Información de género y etnia para la base de datos de BUPT-Balancedface.	9
5.1	Resultados de EER y AUC para el género en la base de datos de VGGFace2.	20
5.2	Resultados de EER y AUC para el género en la base de datos de RFW.	22
5.3	Resultados de EER y AUC para la etnia en la base de datos de RFW.	23

5.4	Resultados de EER y AUC para el género combinado con la etnia en la base de datos de RFW.	23
5.5	Resultados de EER y AUC para el género en la base de datos de BUPT-Balancedface.	25
5.6	Resultados de EER y AUC para la etnia en la base de datos de BUPT-Balancedface. .	26
5.7	Resultados de EER y AUC para el género combinado con la etnia en la base de datos de BUPT-Balancedface.	27
5.8	Resultados de EER y AUC para la etnia en la base de datos de BUPT-Balancedface + RFW.	29

INTRODUCCIÓN

1.1. Motivación

El reconocimiento facial consiste en reconocer o verificar la identidad de una persona utilizando los rasgos faciales. En los últimos años, los métodos que utilizan aprendizaje profundo, deep learning, han hecho que aumente mucho el rendimiento y la precisión de los sistemas de reconocimiento facial.

Estos sistemas de reconocimiento facial están cada vez más presentes en nuestra vida cotidiana. Podemos encontrarlos en aeropuertos para pasar el control de pasaportes, en sistemas de videovigilancia, para acceder a ciertas instalaciones o mucho más cerca, para desbloquear nuestro teléfono móvil.

Sin embargo, se ha probado que muchos de estos métodos obtiene peores resultados en mujeres que en hombres, y también en ciertas razas, teniendo unos mejores resultados en los de raza caucásica (origen europeo, América del norte o parte de América latina). Estos sesgos han hecho que se hayan quitado del mercado algunos de estos sistemas de reconocimiento facial ya que perjudicaban a ciertos grupos demográficos. En concreto, el pasado año, 2020, IBM retiró su sistema de reconocimiento facial para el uso de vigilancia masiva, y Amazon prohibió el uso policial de su software *Rekognition* por el sesgo racial que presentaba ¹.

Por este motivo, se quiere desarrollar un sistema de reconocimiento facial que no presente diferencias al reconocer personas de diferentes grupos demográficos. De esta forma conseguiríamos todas las ventajas que aporta el reconocimiento facial sin presentar sesgos.

Este trabajo se centrará en mejorar los sistemas ya existentes aportando información de etnia y género para así conseguir reducir los sesgos existentes en los sistemas actuales.

¹ Amazon e IBM reniegan de la tecnología de reconocimiento facial por su sesgo racista: https://cincodias.elpais.com/cincodias/2020/06/11/companias/1591881710_616230.html

1.2. Objetivos

El objetivo principal de este trabajo de fin de grado es mejorar el reconocimiento facial añadiendo información de género y raza para reducir el sesgo que existe entre hombres y mujeres y entre las diferentes razas.

Con esta finalidad se van a cumplir los siguientes objetivos:

- Estudio del estado del arte en redes neuronales aplicadas al reconocimiento facial.
- Desarrollar modelos específicos para cada una de las clases demográficas, de etnia y género, partiendo de los modelos base, Resnet-50 y VGG-16.
- Probar los modelos en las diferentes bases de datos.
- Hacer una comparativa entre los modelos base y los modelos específicos.

1.3. Estructura de la memoria

La memoria consta de los siguientes capítulos:

- Capítulo 1: Motivación, objetivos y organización de la memoria.
- Capítulo 2: Estado del arte en los sistemas de reconocimiento facial en redes neuronales.
- Capítulo 3: Bases de datos utilizadas y diseño del modelo.
- Capítulo 4: Desarrollo del modelo
- Capítulo 5: Resultados obtenidos en las diferentes pruebas
- Capítulo 6: Conclusiones y trabajo futuro

ESTADO DEL ARTE

En este capítulo primero se explican los sistemas de reconocimiento facial basados en redes neuronales, en la sección 2.1. Después se presenta el problema del sesgo en estos sistemas, en la sección 2.2. Por último, en la sección 2.3, se explican las características blandas, soft biometrics, y su efecto en los sistemas de reconocimiento facial.

2.1. Sistemas de reconocimiento facial basados en redes neuronales

En el reconocimiento biométrico de personas se puede distinguir entre rasgos fisiológicos como la huella dactilar, el reconocimiento facial, el iris, o la mano entre otros ([1] [2] [3] [4]), y los rasgos conductuales como la voz, la dinámica de tecleo, la forma de caminar, o la firma y escritura entre otros ([5] [6] [7] [8] [9] [10] [11]).

En este trabajo nos centramos en el reconocimiento facial, el cual es el proceso de identificar o verificar la identidad de una persona utilizando sus rasgos faciales. Para conseguirlo se realizan tres pasos: primero se detecta la cara en un vídeo o imagen; después se transforma en un vector de características la información facial; por último, se hace una comparación entre dos vectores de características para verificar que pertenecen a la misma persona. Este proceso ha ido ganando importancia y cada vez se utiliza en más sistemas de seguridad debido a su fácil implementación y a que no requiere interacción física del usuario.

Existen diversos métodos de reconocimiento facial basados en detectar puntos de interés en la cara como el presentado en [12] o en regiones [13], pero en los últimos años se han centrado en el uso de redes neuronales convolucionales, CNNs. Uno de los primeros métodos de reconocimiento facial con redes neuronales que se presentó fue DeepFace [14], que utiliza una CNN profunda entrenada para clasificar caras con una base de datos de 4 millones de imágenes y alrededor de 4000 usuarios. Cuando fue presentado era el que mejor rendimiento tenía en LFW, Labeled Faces in the Wild, [15], la base de datos utilizada para verificar el rendimiento de los sistemas de reconocimiento faciales. Otros

autores extendieron DeepFace creando DeepId que tenía mejores resultados [16] [17] [18].

En [19] se presentó un sistema nuevo llamado FaceNet en el que se utilizaron 200 millones de imágenes en el entrenamiento. Además, en el entrenamiento se añadió el uso de triplet loss. Esta CNN obtenía un nuevo récord de precisión respecto a las redes existentes.

El trabajo presentado en [20] muestran otra CNN que obtenía una precisión comparable con el estado del arte anterior y que había sido entrenada con una nueva base de datos que presentaban en ese mismo estudio que constaba con 2,6 millones de imágenes. Demostraban así que con el entrenamiento adecuado se podían obtener buenos resultados con menos imágenes.

Por otro lado, dos de las arquitecturas más importantes de redes neuronales profundas, VGG [21] y Resnet [22], han sido entrenadas con VGGFace2 [23] para su utilización en sistemas de reconocimiento facial.

2.2. Sesgos en los sistemas de reconocimiento facial

Hay diferentes artículos que han probado que determinados grupos demográficos son más difíciles de identificar o verificar por los sistemas de reconocimiento facial existentes [24] [25] [26].

La mayoría de sistemas de reconocimiento facial han sido entrenados con bases de datos desbalanceadas en cuanto a género y etnia. Por ejemplo, la base de datos VGGFace2 [23] que ha sido utilizada para entrenar muchas de las redes neuronales disponibles consta con un 59.7 % de hombres y 40.3 % de mujeres, siendo de raza caucásica un 74.2 % de los usuarios. Se cree que este es uno de los motivos que hace que ciertos grupos demográficos tengan peores resultados. Se va a cumplir el efecto raza cruzada, explicado en [27], en los sistemas de aprendizaje profundo. Este efecto consiste en que los humanos diferenciamos mejor personas de nuestra raza que de otra. Al entrenar las redes con más personas de una cierta raza, para esas redes es más fácil distinguir personas de esa raza. Sin embargo, el problema de los sesgos no se resuelve únicamente con bases de datos balanceadas. En [28] y [26] se demuestra que una base de datos balanceada en género, no implica que haya una precisión igual en ambos géneros, y que entrenar la red para que no haya diferencia entre la precisión en hombres y mujeres no implica mejores resultados finales en ambos géneros. Además, en [26] se hace un análisis de la desigualdad de género en los sistemas de reconocimiento facial.

Algunos sistemas de reconocimiento facial tenían este problema y eran discriminatorios, por ello algunos estudios se han centrado en crear sistemas de reconocimiento facial que no incluyan ciertos rasgos, los conocidos como soft biometrics, que se explican con detalle en la siguiente sección 2.3. En [29] se demostró que con un algoritmo se podían eliminar los sesgos. Más tarde, en [30] los autores presentaron un sistema neuronal que era capaz de eliminar ciertos rasgos como la etnia o el género.

Por otro lado, existen estudios que se han centrado en incorporar información de raza o género

para mejorar el rendimiento de los sistemas de reconocimiento, especialmente en las clases demográficas que tenían peor rendimiento. En [31] se presenta FaceGenderID que es un sistema basado en redes neuronales convolucionales profundas, DCNNs, que hace uso de la información de género para entrenar una última capa de la red basada en triplet loss. Ese sistema obtenía mejores resultados que las arquitecturas ya existentes, y que sistemas entrenados con bases de datos balanceadas en género.

Además, en los últimos años se han desarrollado bases de datos balanceadas [30] [32]. La presentada en [30] está balanceada en género y en raza mientras que la de [32] está únicamente balanceada en raza. Estas bases de datos serán utilizadas para entrenar los sistemas de aprendizaje profundo intentando evitar el sesgo producido al entrenar las redes con bases de datos desbalanceadas.

2.3. Soft biometrics

Las características blandas, soft biometrics en inglés, son rasgos de las personas que los humanos utilizamos para distinguir a unas de otras de manera sencilla. Estos rasgos pueden ser físicos, de comportamiento u otros elementos que nos permitan identificar a la persona. Algunos ejemplos de soft biometrics son los siguientes:

- Físico: color de piel, de los ojos o del pelo, presencia de barba o bigote, altura, peso, género o etnia.
- Comportamiento: forma de caminar, la pulsación de las teclas (fuerza) o la firma.
- Otros: tatuajes, ropa o accesorios.

En algunos estudios se han centrado en la identificación de algunas de estas características. En [33] se identificaba el género de las personas basándose en características físicas de cuerpo y cara. Por otro lado, en [34] se extraen diferentes características de imágenes simples, como la altura, la longitud de los brazos o el color de pelo.

Otra aplicación de los soft biometrics es la presentada en [35] donde se estudiaba la posibilidad de añadir estas características para el reconocimiento de personas.

En el reconocimiento facial estas características están presentes también, principalmente las relacionadas con el físico. Por ello, en diversos estudios los soft biometrics han sido utilizados para el reconocimiento facial de diferentes formas [36].

Primero, algunos sistemas proponían el uso de varios rasgos, únicamente soft biometrics, para reconocer a las personas [37] [38] [39] [40] [41]. Después, en otros estudios se propuso utilizarlos para reducir el campo de búsqueda [42] [43]. Por último, se han propuesto métodos que fusionan los soft biometrics, obtenidos de manera manual o automática, con un sistema de reconocimiento facial para

mejorar el rendimiento de ese sistema. En [44] se demostró que utilizar marcas faciales como pecas o cicatrices mejora el rendimiento en sistemas de reconocimiento facial. En [45] se utilizaron además de marcas faciales, el género y la etnia como rasgos para mejorar el rendimiento. Estos dos estudios no utilizaban redes neuronales como sistema base. El sistema propuesto en [31] va a añadir información de género para mejorar el rendimiento en una DCNN.

2.4. Conclusiones

Después de haber realizado un estudio del estado del arte se decidió crear y entrenar un modelo de reconocimiento facial que incorpore soft biometrics para mejorar el rendimiento. Los rasgos biométricos que se van a incorporar son información sobre el género y la raza. Para ello se va a añadir una nueva capa de triplet loss a modelos de aprendizaje profundo de manera similar a la utilizada en [31] y se va a experimentar sobre nuevas bases de datos.

DISEÑO

En este capítulo se explican las bases de datos empleadas y el diseño de los modelos utilizados. En primer lugar, en la sección 3.1 se presentan las diferentes bases de datos. En segundo lugar, en la sección 3.2 se expone el modelo propuesto, que consiste en añadir una capa de triplet loss sobre dos modelos baseline, Resnet-50 y VGG16.

3.1. Bases de datos

Para hacer diferentes pruebas con los modelos base y el modelo propuesto, se han seleccionado una serie de bases de datos con imágenes faciales que se explican en esta sección. La base de datos VGGFace2, sección 3.1.1, se utilizó al principio pero sólo disponía de información de género. Después para información de etnia y género se hizo uso de las bases de datos de DiveFace, sección 3.1.2. La base de datos de RFW, sección 3.1.3, nos proporciona información de raza también. Por último, se buscó otra base de datos que tuviera información balanceada respecto a las razas y se utilizó BUPT-Balancedface, sección 3.1.4. En estas dos últimas bases de datos manualmente se separaron los diferentes usuarios en hombres y mujeres para poder hacer un estudio conjunto de etnia y género.

3.1.1. VGGFace2

La base de datos VGGFace2 fue presentada en [23] en el año 2018. Contiene 3.31 millones de imágenes de 9131 usuarios distintos, con una media de 362.6 imágenes por usuario. Todas estas imágenes fueron extraídas de Google y representan una gran variedad de poses, edades e iluminaciones. La base de datos está dividida en train y test, teniendo 500 usuarios de test y el resto de train.

Además, se dispone información del género de los usuarios, teniendo un 59.7% de hombres y un 40.3% de mujeres. Al realizar el trabajo de fin de grado no se disponía de la información de etnia por lo que sólo se utiliza para la parte de género.

3.1.2. DiveFace

DiveFace esta formada por más de 150K, con 24K identidades con una media de 5.5 imágenes por cada usuario, teniendo cada uno un mínimo de 3 imágenes. Los usuarios están agrupados según su género, hombre o mujer, y según la etnia. Para la etnia se han agrupado en tres categorías principales, sabiendo que agrupar a todas las personas en 3 grupos étnicos tiene sus limitaciones.

- Grupo 1, este asiático: engloba las personas con origen en Japón, China, Corea y otros países de esa región.
- Grupo 2, África subsahariana y sur de India: personas con origen en África, India, Bangladesh o Bután entre otros.
- Grupo 3, caucásico: personas con origen europeo, américa del norte y américa latina (con origen europeo).

El objetivo de esta base de datos era crear una base de datos que estuviera balanceada tanto en género como en etnia. Por tanto, se tienen 6 grupos diferentes de las distintas combinaciones de etnia y género, donde cada uno de estos grupos dispone de 4000 usuarios.

Esta base de datos fue creada seleccionando imágenes de la base de datos de MegaFace2, sección 3.1.2, por el grupo BiDALab [46] que fue el que proporcionó las imágenes de esta base de datos.

MegaFace

La base de datos de MegaFace, cuya página web es [47], fue presentada para dos *challenges* distintos [48] [49], el primero en 2016 y el segundo en 2017. El conjunto de entrenamiento de esta base de datos cuenta con 4.7 millones de fotos de 672057 identidades diferentes, con una media de 7 fotos por persona, un mínimo de 3 y un máximo de 2469.

Esta base de datos proporcionaba información tanto de raza, englobando la población 3 grupos, como de género, estando dividida en 6 grupos diferentes formados por las combinaciones de raza y género.

3.1.3. RFW

La base de datos RFW, Racial Faces in-the-Wild, fue presentada por primera vez en [50], y se encuentra disponible bajo solicitud como se explica en la página de los creadores [51]. Fue creada como base de test para estudiar el sesgo, bias, racial que existe en los sistemas de reconocimiento facial.

Está dividida en 4 etnias, Caucasian (origen europeo, américa del norte y parte de américa latina), Asian (este asiático, China, Japón, Corea...), Indian (origen en India y países de alrededor) y African

(origen en África subsahariana) y cada una de estas contiene alrededor de 3000 individuos. Además, la base de datos tiene 6000 pares de imágenes para el reconocimiento facial.

Para poder realizar un estudio del género en conjunto con la etnia, manualmente se han separado los usuarios en hombres y mujeres. En la tabla 3.1 se muestra el porcentaje de usuarios que hay de cada etnia y género. Se puede observar que aunque en etnia no hay una gran diferencia entre las clases, en género está bastante desbalanceada.

	African	Asian	Caucasian	Indian	TOTAL
Mujeres	1.23 %	8.02 %	8.02 %	7.79 %	25.06 %
Hombres	25.02 %	13.8 %	17.9 %	18.23 %	74.94 %
TOTAL	26.24 %	21.82 %	25.92 %	26.02 %	100 %

Tabla 3.1: Información de género y etnia para la base de datos de RFW.

3.1.4. BUPT-Balancedface

Esta base de datos fue presentada en el año 2019 en [32] con el objetivo de eliminar el bias que tienen algunas de las bases de datos ya existentes previamente. Se encuentra disponible bajo solicitud como se explica en la página web de los creadores [52]. Contiene 1,3 millones de imágenes de 28K celeridades de 4 razas diferentes y está aproximadamente balanceada, con 7K imágenes por raza. Las imágenes fueron seleccionadas de la base de datos MS-Celeb-1M o descargadas directamente de páginas web siguiendo el atributo de nacionalidad.

Como en el caso de la base de datos de RFW, se han separado los usuarios también en hombres y mujeres, y la información de cada uno de estos usuarios se presenta en la tabla 3.2. Se puede observar que aunque la base de datos está balanceada en etnia, no está balanceada en género.

	African	Asian	Caucasian	Indian	TOTAL
Mujeres	6.48 %	6.22 %	7.59 %	6.06 %	26.35 %
Hombres	18.52 %	18.78 %	17.41 %	18.94 %	73.65 %
TOTAL	25 %	25 %	25 %	25 %	100 %

Tabla 3.2: Información de género y etnia para la base de datos de BUPT-Balancedface.

3.2. Modelo propuesto

Existen una serie de modelos preentrenados para el reconocimiento facial disponibles. Para este trabajo de fin de grado se propone usar dos de estos modelos preentrenados, sección 3.2.1, como modelos base, y a estos añadirles después una capa de triplet loss, explicada en la sección 3.2.2.

3.2.1. Modelos preentrenados: Resnet-50 y VGG-16

Existen numerosas arquitecturas preentrenadas para el reconocimiento facial. Para este trabajo se han utilizado dos muy conocidas, **Resnet-50** y **VGG-16**, entrenadas con la base de datos de VGGFace2 [23]. Además, para poder entrenarlas en el reconocimiento facial se les añadió al final de la red una capa de activación softmax para la decisión. Esta función se corresponde con la ecuación 3.1, y su representación es la que encontramos en la Figura 3.1.

$$f(z)_i = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}} \quad (3.1)$$

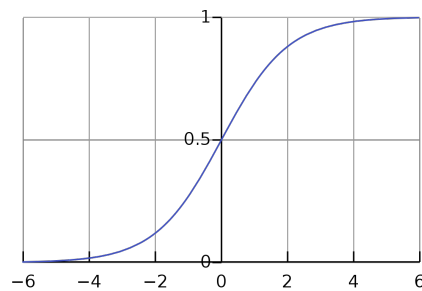


Figura 3.1: Función Softmax.

La arquitectura VGGNet, apareció con el objetivo de reducir el número de parámetros en las capas convolucionales de las redes para reducir el tiempo de entrenamiento. Hay varios tipos de redes VGG que varían en el número de capas de la red. Para este trabajo se ha utilizado VGG-16, Figura 3.2.

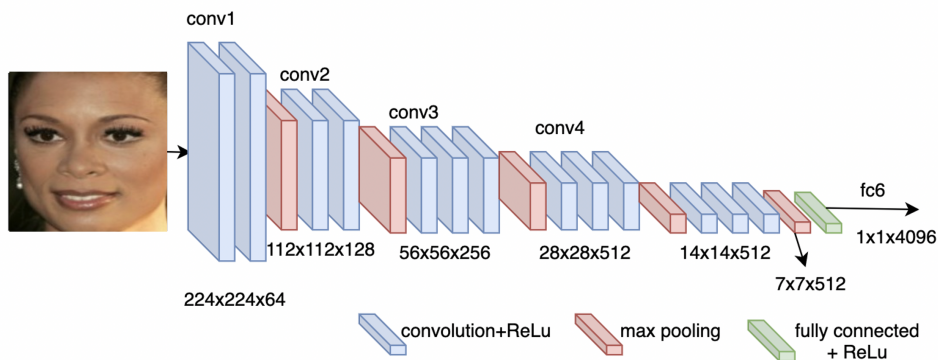


Figura 3.2: Arquitectura VGG-16.

La característica principal de las redes Resnet es que permiten encontrar un mapa más sencillo cuando se puede, haciendo saltos entre capas. Estos saltos se conocen como conexiones residuales [22]. En la Figura 3.3, se muestra como es un bloque residual, donde se ve que hay una conexión que permite saltar dos capas. Existen diferentes modelos de Resnet que tienen en común que constan con

diferentes etapas, y la diferencia es el número de capas convolucionales en cada una de estas etapas, haciendo más o menos profunda la red neuronal. La arquitectura de Resnet-50 está construida con bloques residuales que se repiten en la red como se muestra en la Figura 3.4.

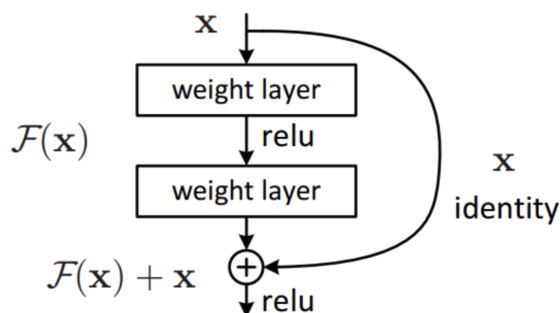


Figura 3.3: Bloque residual.

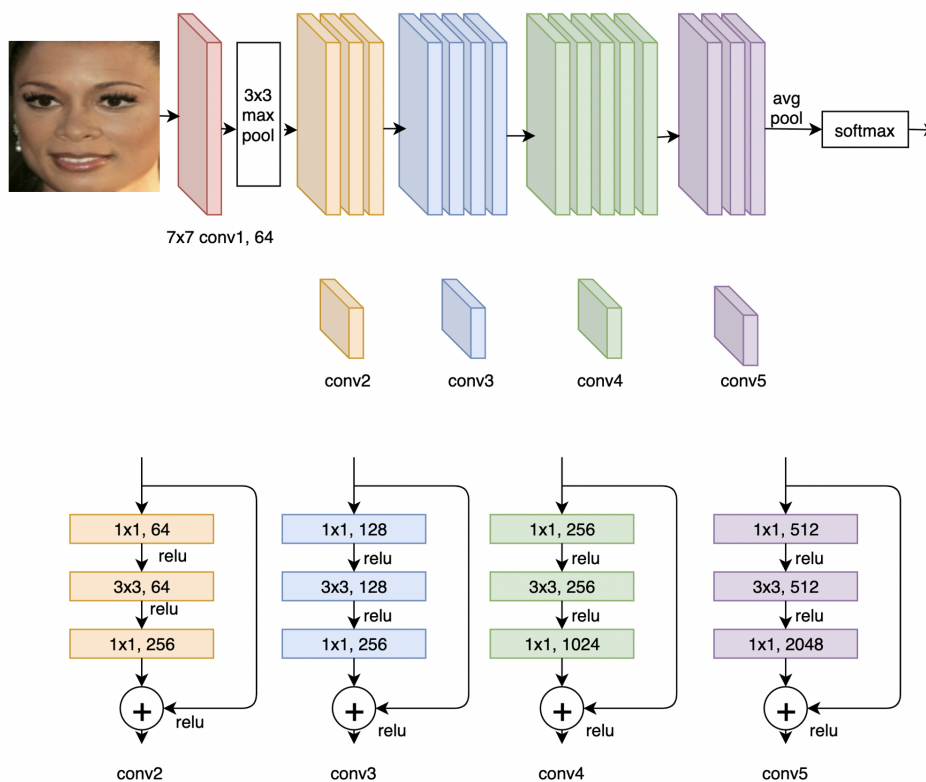


Figura 3.4: Arquitectura Resnet-50.

3.2.2. Triplet loss

Los modelos existentes de Resnet-50 y VGG-16 presentan bias en el reconocimiento facial, obteniendo mejores resultados, mayor precisión, en hombres caucásicos que en el resto de etnias y

géneros. Para solventar estas diferencias se ha propuesto añadir una capa de triplet loss al final de los modelos existentes y entrenar esta última capa con información de raza y género. Esto significa entrenarla para cada grupo, según raza y género, con imágenes que pertenezcan únicamente a personas de ese grupo. Por ello, se va a entrenar un modelo para cada una de las combinaciones. Una vez se entrenen estos nuevos modelos, para cada uno de los grupos demográficos se usa el modelo entrenado correspondiente a ese grupo.

El funcionamiento del triplet loss se basa en generar tríos de imágenes y compararlas. De estas imágenes, una es la imagen base o ancla, anchor, que se va a comparar con una imagen positiva y otra negativa. La imagen anchor y la positiva corresponden al mismo usuario mientras que la negativa corresponde con un usuario diferente. En la Figura 3.5 se muestra un ejemplo de un trío de imágenes donde la imagen anchor y positiva presentan diferencias de pose, peinado y maquillaje. El objetivo del triplet loss es minimizar la distancia entre la imagen anchor y la imagen positiva y maximizar la distancia entre el anchor y la imagen negativa. En la Figura 3.6 se muestra la estructura del triplet, donde X_A , X_P y X_N son los vectores de características de las imágenes anchor, positiva y negativa, y donde d_{PA} y d_{NA} son las distancias entre anchor e imagen positiva y anchor e imagen negativa.



Figura 3.5: Ejemplo de un conjunto de tres imágenes que constituye un triplet

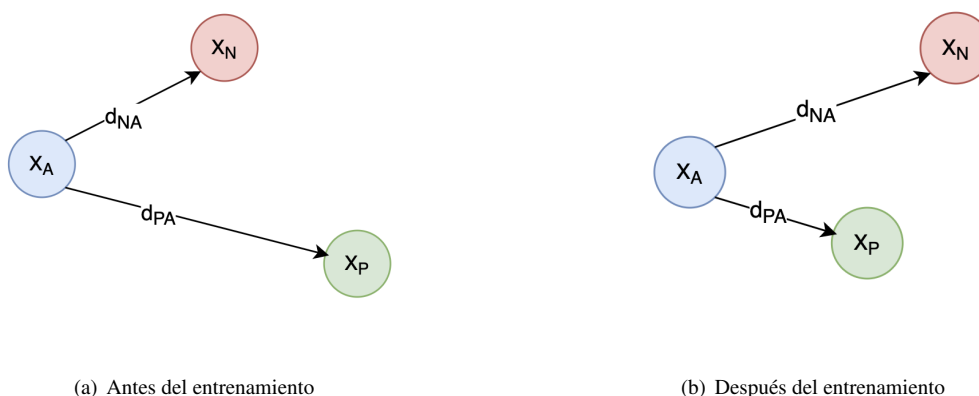


Figura 3.6: Estructura de un triplet.

Los triplets que se usarán tienen que cumplir que su diferencia esté por encima de un umbral. Para

ello se calcula la puntuación de similitud entre las imágenes, teniendo la distancia entre el positivo y el anchor (d_{PA}), y la distancia entre el negativo y el anchor (d_{NA}). Una vez obtenidas estas distancias se acepta el triplet si se cumple la ecuación 3.2 donde α es un umbral fijado en un valor de -0.2. Además, se busca que los triplets sean difíciles, esto significa que la distancia entre la imagen positiva y la anchor sea grande y entre la negativa y la anchor pequeña para poder entrenar la red y que aprenda de tal forma que esto cambie, pasando a ser menor la distancia con la imagen positiva como se muestra en la Figura 3.6.

$$\|d_{PA}^2 - d_{NA}^2\| > \alpha \quad (3.2)$$

Para calcular estas distancias se utiliza la distancia euclídea, ecuación 3.3, donde los features de las imágenes son X_A , X_P y X_N , para el anchor, positivo y negativo respectivamente.

$$\begin{aligned} d_{PA} &= (X_A - X_P)^2 \\ d_{NA} &= (X_A - X_N)^2 \end{aligned} \quad (3.3)$$

Para poder añadir nuevas capas que entrenar a los modelos existentes de Resnet-50 y VGG-16, se ha quitado la última capa de estas que se encargaba de la decisión para tener acceso al vector de características, features. En el caso del modelo Resnet-50 el vector de características que se genera es de 2048 mientras que en VGG-16 es de 4096.

En la Figura 3.7 se presenta el modelo propuesto. En la imagen se representan 3 CNN diferentes para simplificarlo. Sin embargo, el CNN de género tendrá dos modelos diferentes entrenados, uno para hombres y otro para mujeres. El CNN de raza hay cuatro diferentes, uno para asiáticos, africanos, indios y caucásicos. Por último, el CNN de raza y género representa las 8 combinaciones entre las diferentes razas y géneros.

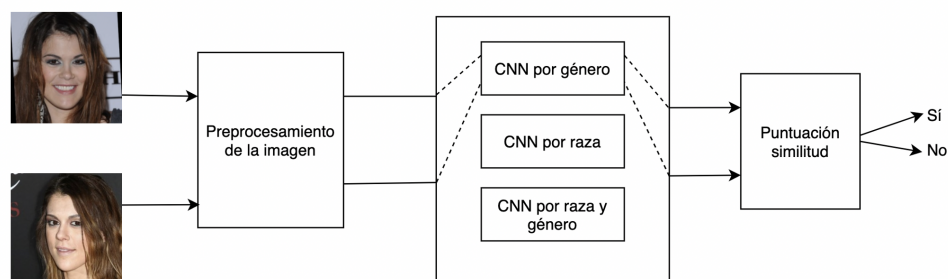


Figura 3.7: Método propuesto.

En la Figura 3.8 se representa la estructura simplificada de una CNN específica.

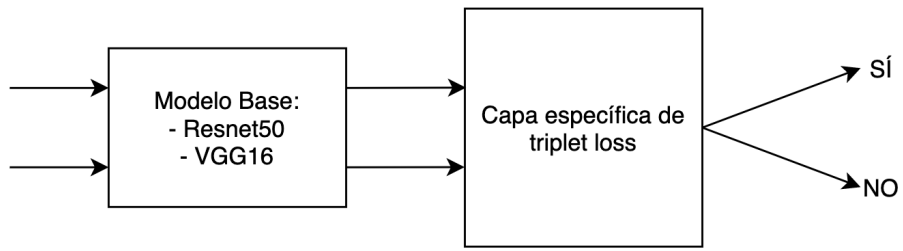


Figura 3.8: Estructura modelo CNN.

DESARROLLO

Para implementar el modelo se hace uso de *Python* con *TensorFlow* y *Keras*. *TensorFlow* es una plataforma de código abierta que se utiliza para machine learning y cuenta con un gran número de herramientas y librerías que se pueden utilizar. Por otra parte, *Keras* es una extensión de *TensorFlow* utilizada para deep learning que facilita el desarrollo de nuevas aplicaciones.

Los modelos baseline que se utilizan, Resnet-50 y VGG-16, se van a descargar de los modelos preentrenados en *Keras*, en la extensión de `vgg_face` [53]. De estos modelos se va a quitar la última capa de decisión para poder obtener los vectores de características que después se utilizarán para hacer la comparación entre las diferentes imágenes.

En el caso del modelo propuesto, se añade la capa de triplet loss tras la que obtiene los features. Esta capa es lineal y reduce el tamaño del vector de características a 1024.

Para generar los resultados presentados en el capítulo 5 se van a realizar los siguientes pasos:

- 1.– Separar los usuarios de las bases de datos en conjunto de entrenamiento, train, y de test en las bases de datos en las que sea necesario.
- 2.– Extraer los features del conjunto de entrenamiento.
- 3.– Generar los triplets.
- 4.– Entrenar la capa de triplet loss.
- 5.– Escoger los pesos de triplet loss para los resultados finales.
- 6.– Extraer los features del conjunto de test.
- 7.– Generar matches, parejas de imágenes, para comprobar el rendimiento.
- 8.– Probar los diferentes modelos en el conjunto de test.

El primer paso es **separar los usuarios en conjuntos de train y test**. En la base de datos de VGGFace2, los usuarios se encontraban divididos por lo que en esa este paso no es necesario, sin embargo en las otras bases de datos utilizadas no estaba hecha esta separación. Por ello, antes de hacer las pruebas, se han dividido las bases de datos. Para realizar esta división se ha hecho que haya el mismo porcentaje de test de cada grupo demográfico.

Una vez hecha la división, lo primero es **extraer el vector de características** de las imágenes de entrenamiento. Las imágenes primero se recortan, manteniendo únicamente la parte de la cara y después se redimensionan para que tengan el tamaño de entrada que tienen los dos modelos usados, Resnet-50 y VGG-16, de 224x224. Una vez hecho esto, las imágenes se pasan por los modelos que devuelven un vector de características.

Para recortar la zona de la cara, en las bases de datos de VGGFace y RFW se dispone de información de donde se encuentran las caras en la imagen. Por otro lado, en DiveFace las imágenes ya se encuentran recortadas por lo que no será necesario este paso. Por último, en la base de datos de BUPT-Balancedface no se disponía de información de la posición de las caras y para obtener la posición se ha usado el detector facial MTCNN [54]. Este detector está basado en redes neuronales y recibe ese nombre ya que utiliza Multi-task Cascaded Convolutional Networks. Se eligió este detector facial ya que tenía un accuracy del 95%. En la Figura 4.1 se muestra un ejemplo de una imagen antes y después de haberla recortado con este detector. Por otro lado, las imágenes en las que MTCNN no detectaba ninguna cara fueron descartadas y no utilizadas ni para el entrenamiento ni para el test.

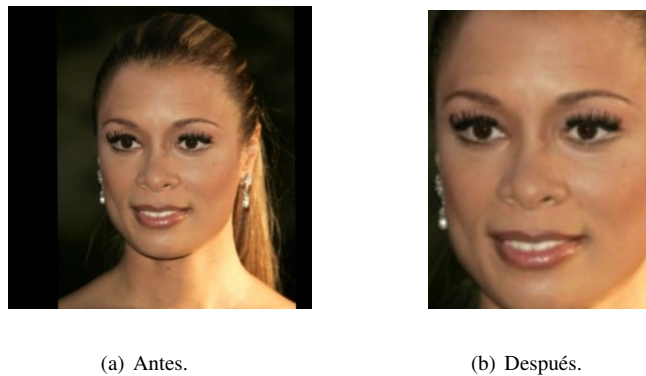


Figura 4.1: Imagen antes y después de recortar la imagen en el área de la cara.

El siguiente paso es **generar los triplets** para cada grupo demográfico. Primero se generan todos los posibles triplets, con las diferentes combinaciones entre las imágenes. Como se explicó en la sección 3.2.2, se generan triplets en los cuales la diferencia está por encima de un umbral. Una vez se tienen todos los posibles triplets, si hay más de 75K, se cogen los más difíciles hasta llegar a 75K y el resto se descartan. En los casos en los que haya menos nos vamos a quedar con todos los posibles.

Para **entrenar el triplet loss**, se crea un modelo lineal sencillo que tenga tres entradas, una para cada uno de los vectores de características del anchor, positiva y negativa. Se le añade una capa con la función de pérdidas, loss. Se utiliza *maxplus* que consiste en seleccionar el máximo entre 0 y el valor de pérdidas basado en distancia euclídea calculado como se muestra en la ecuación 4.1. Siendo d_{PA} la distancia entre los features del anchor y de la positiva y d_{NA} la distancia entre anchor y la negativa, y α el umbral.

$$\mathcal{L} = d_{PA}^2 - d_{NA}^2 + \alpha \quad (4.1)$$

Cuando tenemos el modelo creado vamos a entrenarlo con cada uno de los triplets que se habían generado anteriormente durante 100 épocas. Para el entrenamiento se utiliza un batch size de 128, esto es en cuanto dividimos los triplets de cada época.

Una vez entrenada, **se seleccionan los pesos de la mejor época**. Para ello se hace de dos formas diferentes según el tamaño de la base de datos. Para las bases de datos pequeñas, se han escogido los pesos de la época con un loss menor, mientras que para las que tenemos más imágenes, se ha hecho una comparativa del accuracy en el conjunto de validación en las diferentes épocas y se escoge el que devuelve mejores resultados. El conjunto de validación ha sido generado separando una parte del conjunto de train, un 15 %. En la Figura 4.2 se presenta la salida al pasar el conjunto de validación por la red en las diferentes épocas, marcando la época que mejores resultados aporta.

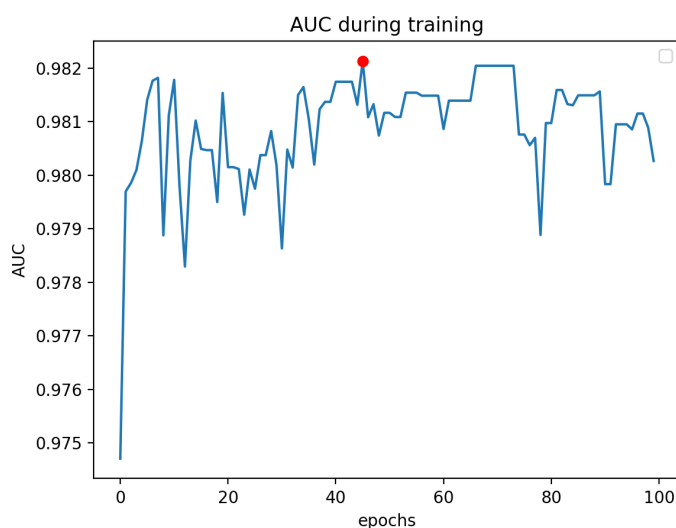


Figura 4.2: Entrenamiento y validación del triplet loss.

Se puede observar que la diferencia entre las diferentes épocas no es muy significativa por lo que se podría disminuir el número de épocas que se utilizan.

Los **features del conjunto de test** se han extraído de manera análoga a la forma de extraerlos en el conjunto de entrenamiento. Para **comprobar el rendimiento**, se han generado parejas de imágenes para comparar. Para hacer esto se han creado todas las posibles combinaciones de parejas, y después se han seleccionado un cierto número de parejas genuinas, pertenecen a la misma persona, y de impostores, son de diferentes personas. El número de matches de impostores se ha fijado en el doble del de genuinos.

Por último, una vez generadas las parejas para comprobar **se prueban los diferentes modelos** en el conjunto de test para ver si mejoran o no los resultados al añadir el triplet loss a los modelos baseline. En el capítulo 5 se muestran los resultados obtenidos para las diferentes bases de datos.

RESULTADOS

En este capítulo se muestran los resultados obtenidos con los diferentes modelos y las diferentes bases de datos. Los resultados se presentan en función de las diferentes bases de datos que han sido utilizadas. Primero en la sección 5.1 se muestran los resultados obtenidos para la base de datos de VGGFace2. En la sección 5.2 los que se obtienen con RFW. Por último, la sección 5.3 los obtenidos con la base de datos de BUPT-Balancedface. Para la base de datos de DiveFace los resultados de rendimiento y el AUC, Area Under the Curve, eran muy altos, con un AUC por encima del 99% en resnet en todas las clases demográficas. Al tener un buen rendimiento no había mucho margen de mejora y por ello se descartó el uso de esta base de datos.

En cada una de estas secciones se van a mostrar los resultados obtenidos con los sistemas base-line, Resnet-50 y VGG-16, y los obtenidos con los modelos creados añadiendo una capa final de triplet loss a estos modelos baseline.

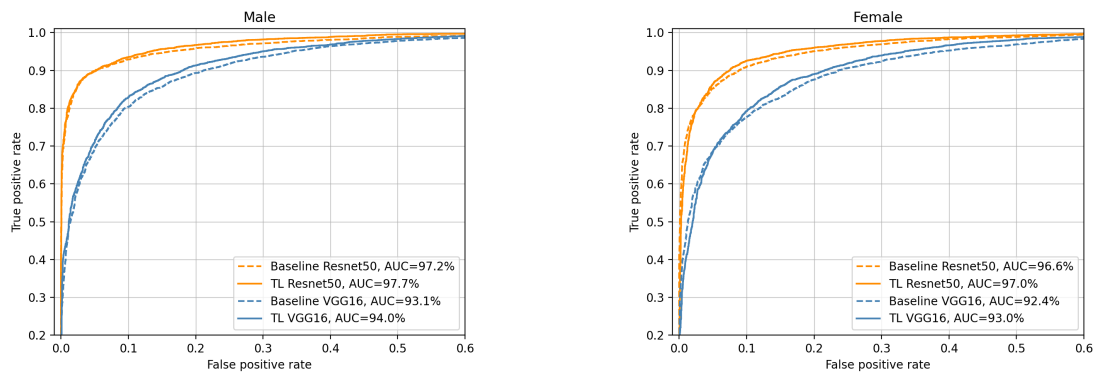
Para medir el rendimiento de los diferentes modelos se van a utilizar el valor de AUC, Area Under the Curve, y ERR, Equal Error Rate. AUC es un tipo de curva ROC, que representa la tasa de verdaderos positivos, TAR (True Acceptance Rate) frente a la tasa de falsos positivos, FAR (False Acceptance Rate). El EER se corresponde con el punto en el cual la tasa de falsa aceptación, FAR, es igual a la tasa de falso rechazo, FRR (False Rejection Rate).

5.1. Resultados VGGFace2

Para la base de datos de VGGFace2 únicamente se disponía de información del género, por ello sólo se va a probar la mejora en esa característica al añadir el triplet loss en los modelos baseline.

Para el conjunto de train se hizo una selección de 2500 mujeres y 2500 hombres en los cuales el número de imágenes por usuario era de 300 o más. De estos usuarios, al extraer los features se seleccionan las 240 primeras imágenes para train y las otras 60 para el conjunto de validación. Este conjunto se va a utilizar para probar la red reentrenada y escoger la que mejores resultados aportaba para utilizarla en los usuarios de test. Para el test se seleccionaron 500 usuarios de los cuales 202 eran mujeres y 298 hombres.

En la Figura 5.1 muestran los resultados de cada uno de los géneros con los modelos base y los entrenados con triplet loss.



(a) Hombres. Antes y después de entrenar con triplet loss.

(b) Mujeres. Antes y después de entrenar con triplet loss

Figura 5.1: Gráficas por género VGGFace2. Antes y después de entrenar el triplet loss.

Por otro lado en la tabla 5.1 se recogen los resultados de AUC y EER para cada género con los modelos baseline y los modelos entrenados en cada género.

EER	Resnet50	Resnet50 + TL	VGG16	VGG16 + TL
Mujeres	9.48 %	8.69 %	16.05 %	14.74 %
Hombres	8.14 %	7.9 %	14.16 %	13.48 %

(a) EER.

AUC	Resnet50	Resnet50 + TL	VGG16	VGG16 + TL
Mujeres	96.0 %	97.0 %	92.4 %	93 %
Hombres	97.2 %	97.7 %	93.1 %	94.0 %

(b) AUC.

Tabla 5.1: Resultados de EER y AUC para el género en la base de datos de VGGFace2.

Con los resultados obtenidos podemos ver que para ambos géneros y ambas bases de datos, el rendimiento de la red es ligeramente mejor al añadir la capa de triplet loss. El EER disminuye en todos los casos, teniendo una mayor disminución en VGG16, mientras que el AUC aumenta entre un 0.5 % y un 1 %.

Podemos observar que aunque la precisión es mejor para las mujeres que en los sistemas base, aún así se siguen obteniendo mejores resultados en los hombres.

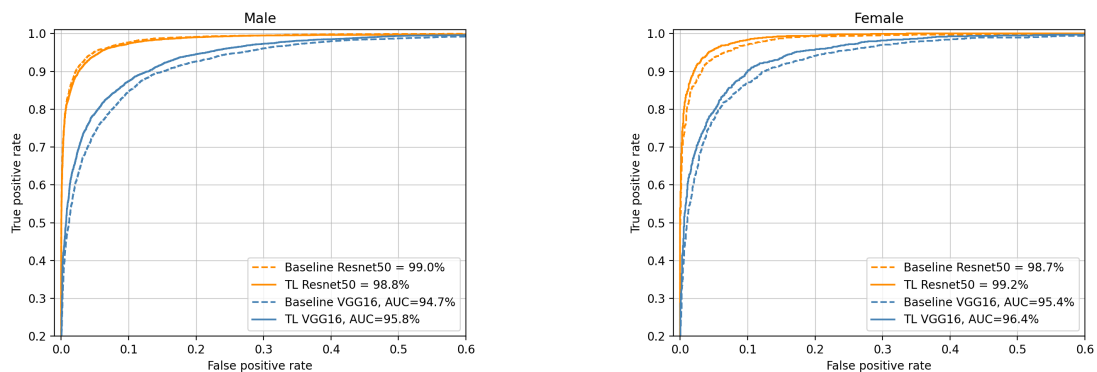
5.2. Resultados RFW

La base de datos de RFW estaba únicamente dividida en raza. Para probar los modelos también en género y en las combinaciones de género y raza, las imágenes de cada etnia fueron separadas manualmente entre hombres y mujeres.

Para generar los conjuntos de train y test, se han dividido en un 85 % train y 15 % test para cada combinación de género y raza. Como esta base de datos no está balanceada habrá más usuarios de ciertos grupos demográficos en el train y en el test. Para las combinaciones por raza o género, lo que se ha hecho es unir los conjuntos de train y test correspondientes.

Como se explicó en la sección 3.1.3, aunque la base de datos está balanceada en raza, no lo está en género. Por ello, el número de triplets que se generan para entrenar cada uno de los modelos va a ser diferente. En el caso de mujeres de etnia africana, el grupo de menor número de usuarios, se han generado un total de 3.6K triplets, sin embargo en todos los grupos de raza o género, se han generado 75K triplets que es el número máximo fijado.

En la Figura 5.2 muestran los resultados de cada uno de los géneros con los modelos base y los entrenados con triplet loss. En la Figura 5.3 se observan los resultados de cada una de las razas con los modelos base y los modelos reentrenados con triplet loss. Para esta base de datos, las etnias estaban separadas en cuatro clases, african, asian, caucasian e indian. Las gráficas correspondientes a cada una de las combinaciones de género y raza se encuentran en el apéndice A.



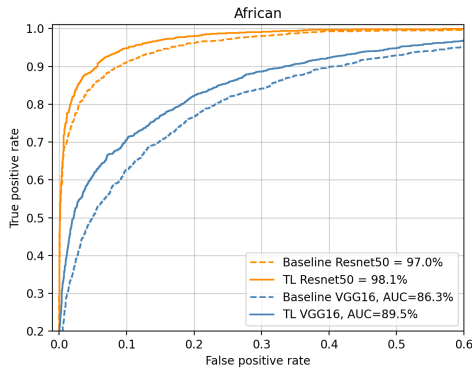
(a) Hombres. Antes y después de entrenar con triplet loss.

(b) Mujeres. Antes y después de entrenar con triplet loss

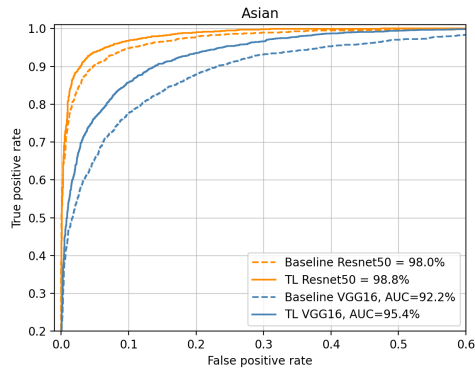
Figura 5.2: Gráficas por género RFW. Antes y después de entrenar el triplet loss.

En la tabla 5.2 se recogen los resultados de AUC y EER para cada género con los modelos baseline y los modelos entrenados en cada género. En la tabla 5.3 los correspondientes resultados a cada grupo étnico. Por último, en la tabla 5.4 los resultados de cada una de las combinaciones de etnia y género.

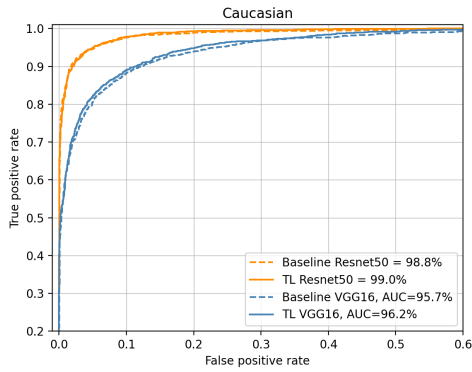
Se puede ver que para ambos géneros el resultado mejora, aumenta el AUC y disminuye el EER,



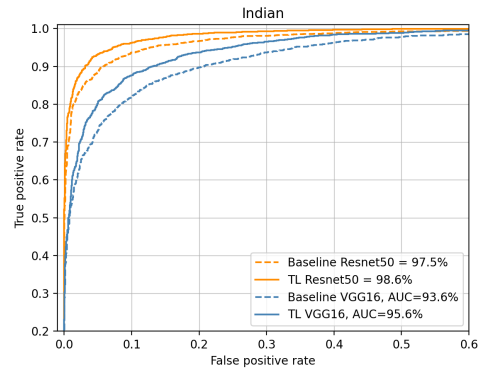
(a) African. Antes y después de entrenar con triplet loss.



(b) Asian. Antes y después de entrenar con triplet loss



(c) Caucasian. Antes y después de entrenar con triplet loss.



(d) Indian. Antes y después de entrenar con triplet loss

Figura 5.3: Gráficas por etnia RFW. Antes y después de entrenar el triplet loss.

EER	Resnet50	Resnet50 + TL	VGG16	VGG16 + TL
Mujeres	5.74 %	4.59 %	11.44 %	9.99 %
Hombres	4.88 %	5.4 %	12.18 %	11.2 %

(a) EER.

AUC	Resnet50	Resnet50 + TL	VGG16	VGG16 + TL
Mujeres	99 %	98.8 %	94.7 %	95.8 %
Hombres	98.7 %	99.2 %	95.4 %	96.4 %

(b) AUC.

Tabla 5.2: Resultados de EER y AUC para el género en la base de datos de RFW.

EER	Resnet50	Resnet50 + TL	VGG16	VGG16 + TL
African	9.33 %	7.3 %	21.41 %	19.0 %
Asian	7.2 %	5.74 %	15.7 %	12.06 %
Caucasian	5.2 %	5.0 %	10.86 %	10.54 %
Indian	8.02 %	6.03 %	13.85 %	11.33 %

(a) EER.

AUC	Resnet50	Resnet50 + TL	VGG16	VGG16 + TL
African	97 %	98.1 %	86.3 %	89.5 %
Asian	98 %	98.8 %	92.2 %	95.4 %
Caucasian	98.8 %	99 %	95.7 %	96.2 %
Indian	97.5 %	98.6 %	93.6 %	95.6 %

(b) AUC.

Tabla 5.3: Resultados de EER y AUC para la etnia en la base de datos de RFW.

EER	Resnet50	Resnet50 + TL	VGG16	VGG16 + TL
African female	12.99 %	15.58 %	16.88 %	19.48 %
African male	9.95 %	7.95 %	22.81 %	20.52 %
Asian female	9.13 %	8.68 %	18.94 %	15.33 %
Asian male	8.76 %	6.95 %	17.83 %	13.44 %
Caucasian female	5.79 %	6.27 %	10.77 %	10.93 %
Caucasian male	6.55 %	7.17 %	13.71 %	13.48 %
Indian female	10.96 %	8.75 %	17.44 %	16.78 %
Indian male	9.23 %	7.89 %	16.0 %	14.06 %

(a) EER.

AUC	Resnet50	Resnet50 + TL	VGG16	VGG16 + TL
African female	93.7 %	93.3 %	88.9 %	89.2 %
African male	96.8 %	97.7 %	85.1 %	88.1 %
Asian female	97.3 %	97.3 %	89.8 %	89.2 %
Asian male	97.2 %	98.3 %	90.6 %	94.4 %
Caucasian female	98.5 %	98.5 %	96.4 %	95.4 %
Caucasian male	98.3 %	98.2 %	94 %	94 %
Indian female	95.4 %	97.1 %	90.4 %	91.8 %
Indian male	97 %	97.8 %	91.7 %	93.5 %

(b) AUC.

Tabla 5.4: Resultados de EER y AUC para el género combinado con la etnia en la base de datos de RFW.

excepto en el caso de los hombres en Resnet50, que a pesar de que el AUC aumenta el EER también sube. Esto puede afectar según el objetivo del sistema de reconocimiento facial. Si se quiere que no haya muchos falsos positivos, será mejor un sistema que tenga un EER menor.

En las diferentes etnias, sin separarlas por género, los resultados son satisfactorios para las cuatro etnias, obteniendo las mayores mejoras en los de etnia africana, donde el AUC aumenta un 1.1 % en Resnet50 y un 3.2 % en VGG16.

Observando los resultados de las diferentes combinaciones de género y raza, se puede observar que la mejora es mucho mayor en los hombres que en las mujeres. Esto puede ser porque el número de imágenes utilizadas para entrenar los modelos de las mujeres eran mucho menores. Por ello, el número de triplets era menor y la red se va a entrenar menos. Cuánto mayor número de triplets tengamos, más triplets difíciles (en los cuales en la distancia del anchor con el negativo y el positivo no se aprecia diferencia) tendremos y mejor se podrá entrenar la red.

Además, también se observa que para los grupos demográficos en los que ya se tenían buenos resultados, una buena precisión, y un AUC alto, las mejoras son menores que en otros grupos demográficos. En los hombres caucásicos no mejora, e incluso empeora un poco en el caso de Resnet50 donde el EER aumenta en un 0.6 % mientras que en los hombres asiáticos la mejoría es bastante alta, disminuyendo el EER en Resnet50 en casi un 2 % y en VGG16 en un 4 %.

Por otro lado, se puede ver que VGG16 mejora más que Resnet50 al añadir el triplet loss, pero aún así el modelo sigue obteniendo peores resultados que el modelo baseline de Resnet50.

5.3. Resultados BUPT-Balancedface

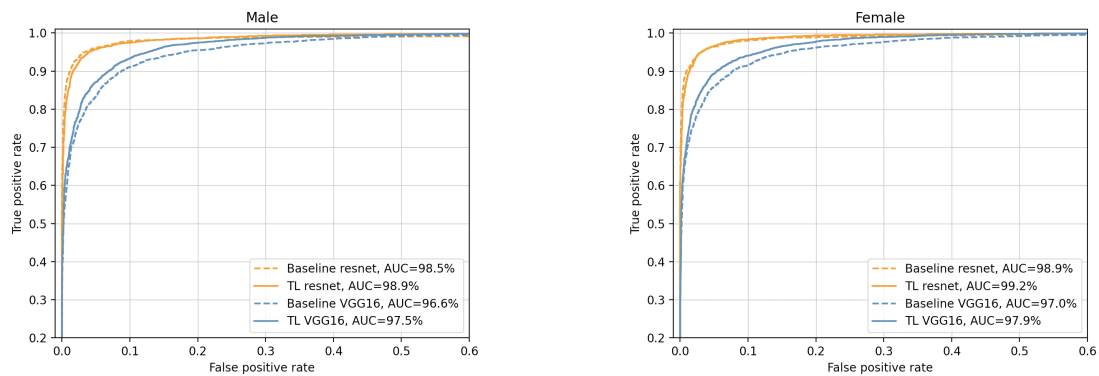
La base de datos de BUPT-Balancedface estaba únicamente dividida en raza. Como en el caso de RFW, para poder probar los modelos también en género y en las combinaciones de género y raza, las imágenes se separaron manualmente entre hombres y mujeres.

En esta base de datos no se tenía la información de la localización de la cara y al tener que usar MTCNN para detectarla el tiempo de ejecución para extraer los features aumentaba. Por este motivo, se decidió reducir el número de imágenes y usuarios utilizados para obtener una primera aproximación de los resultados en esta base de datos. Para las pruebas se utilizan 1000 usuarios de cada etnia y género y 10 imágenes de cada usuario. De esta forma, las pruebas fueron realizadas balanceadas en género y raza.

Una vez hecha la selección de usuarios se van a dividir en conjunto de train y test. Se van a dividir un 85 % train y un 15 % test. Por tanto, se tendrán 850 usuarios para el train y 150 para test. Al tener pocas imágenes en el conjunto de train, se disponen de 8.5K imágenes para cada combinación de etnia y género, el número de triplet que se generan es bajo. El número de triplets será de aproximadamente

34K para las pruebas de género, de 15K para las pruebas por etnia y solamente de 7.5K para las de combinación de etnia con género.

En la Figura 5.4 muestran los resultados de cada uno de los géneros con los modelos base y los entrenados con triplet loss. Esta base de datos, como la de RFW, se encontraba dividida en cuatro etnias, african, asian, caucasian e indian, y en la Figura 5.5 se presentan las gráficas con los resultados de cada una de estas etnias. Por último, las gráficas de cada una de las combinaciones de género y etnia se encuentran en el apéndice B.



(a) Hombres. Antes y después de entrenar con triplet loss.

(b) Mujeres. Antes y después de entrenar con triplet loss.

Figura 5.4: Gráficas por género BUPT-Balancedface. Antes y después de entrenar el triplet loss.

En la tabla 5.5 se recogen los resultados de AUC y EER para cada género con los modelos baseline y los modelos entrenados en cada género. En la tabla 5.6 los correspondientes resultados a cada grupo étnico. Por último, la tabla con los resultados de EER y AUC de cada una de las combinaciones de etnia y género se encuentran en la tabla 5.7.

EER	Resnet50	Resnet50 + TL	VGG16	VGG16 + TL
Mujeres	4.04 %	4.02 %	8.90 %	7.61 %
Hombres	4.33 %	4.56 %	9.38 %	8.18 %

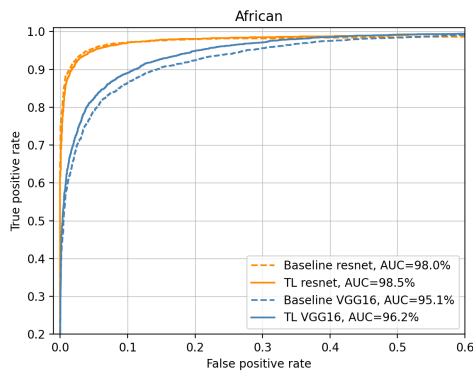
(a) EER.

AUC	Resnet50	Resnet50 + TL	VGG16	VGG16 + TL
Mujeres	98.9 %	99.2 %	97.0 %	97.9 %
Hombres	98.5 %	98.9 %	96.6 %	97.5 %

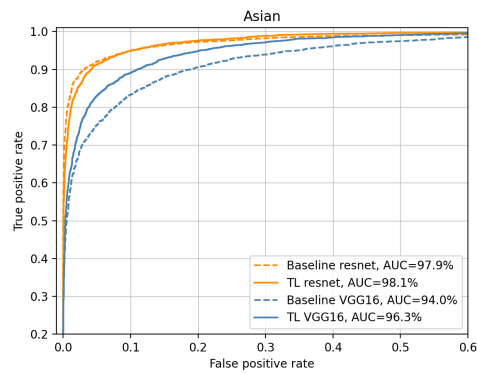
(b) AUC.

Tabla 5.5: Resultados de EER y AUC para el género en la base de datos de BUPT-Balancedface.

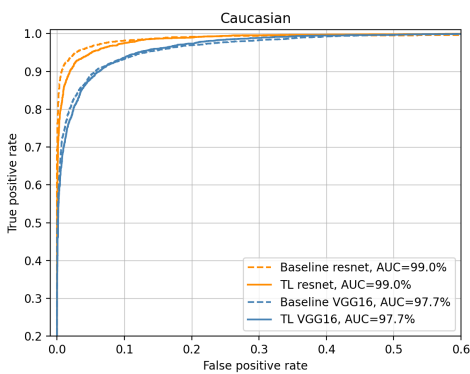
Se puede observar que para las parejas de test, los matches, generadas en esta base de datos, los modelos baseline tienen mejores resultados en mujeres que en hombres. Para ambos casos los resultados mejoran al añadir el triplet loss, bajando en todos los casos el EER y aumentando la AUC.



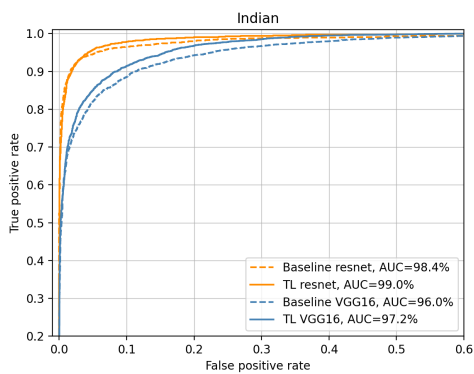
(a) African. Antes y después de entrenar con triplet loss.



(b) Asian. Antes y después de entrenar con triplet loss



(c) Caucasian. Antes y después de entrenar con triplet loss.



(d) Indian. Antes y después de entrenar con triplet loss

Figura 5.5: Gráficas por etnia BUPT-Balancedface. Antes y después de entrenar el triplet loss.

EER	Resnet50	Resnet50 + TL	VGG16	VGG16 + TL
African	4.64 %	5.04 %	11.83 %	10.48 %
Asian	6.79 %	6.96 %	13.43 %	10.50 %
Caucasian	4.08 %	4.96 %	7.9 %	7.93 %
Indian	5.25 %	4.66 %	10.6 %	9.16 %

(a) EER.

AUC	Resnet50	Resnet50 + TL	VGG16	VGG16 + TL
African	98.0 %	98.5 %	95.1 %	96.2 %
Asian	97.9 %	98.1 %	94.0 %	96.3 %
Caucasian	99.0 %	99.0 %	97.7 %	97.7 %
Indian	98.4 %	99.0 %	96.0 %	97.2 %

(b) AUC.

Tabla 5.6: Resultados de EER y AUC para la etnia en la base de datos de BUPT-Balancedface.

En la etnia todos mejoran el baseline de VGG16 menos en el caso de los caucásicos. Sin embargo, en Resnet50, únicamente los indios obtienen mejores resultados. Esto es debido a que al tener menos margen de mejora, al no tener muchos triplets para entrenar la red, con Resnet50 no se consiguen mejoras.

EER	Resnet50	Resnet50 + TL	VGG16	VGG16 + TL
African female	6.30 %	6.65 %	12.82 %	11.27 %
African male	5.46 %	6.88 %	14.81 %	13.30 %
Asian female	8.20 %	9.51 %	16.82 %	14.40 %
Asian male	7.95 %	8.44 %	15.25 %	13.15 %
Caucasian female	4.86 %	7.16 %	8.92 %	9.49 %
Caucasian male	4.52 %	7.72 %	9.19 %	10.92 %
Indian female	7.77 %	7.13 %	13.5 %	11.39 %
Indian male	6.73 %	8.27 %	12.8 %	13.35 %

(a) EER.

AUC	Resnet50	Resnet50 + TL	VGG16	VGG16 + TL
African female	97.6 %	98 %	94.2 %	95.4 %
African male	97.9 %	97.5 %	93 %	94.4 %
Asian female	97.2 %	96.9 %	91.2 %	93.3 %
Asian male	96.9 %	96.9 %	92.7 %	94.2 %
Caucasian female	98.8 %	98.0 %	97.0 %	96.7 %
Caucasian male	98.9 %	97.9 %	96.7 %	95.9 %
Indian female	98.4 %	98.1 %	94.0 %	92.6 %
Indian male	97.4 %	97.3 %	94.3 %	94.4 %

(b) AUC.

Tabla 5.7: Resultados de EER y AUC para el género combinado con la etnia en la base de datos de BUPT-Balancedface.

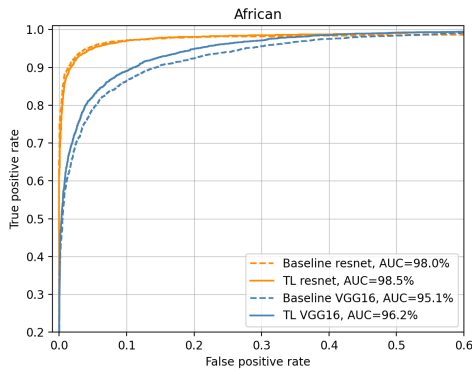
Por último, en el caso de la combinación de etnia y género, al tener menos triplets para entrenar la red sólo se consigue mejora en los modelos de VGG16 de los hombres y mujeres africanos, hombres y mujeres asiáticos y las mujeres indias. Para el modelo de Resnet50 con la nueva capa de triplet loss no se obtienen mejores resultados en ninguno de los casos, la precisión en estos casos disminuye.

El motivo por el que no mejoran estos sistemas puede ser debido a que tenemos pocas combinaciones de triplets por lo que se puede producir un overfitting en estos y no mejorar el sistema general. El overfitting consiste en que el sistema trabaja muy bien con esas imágenes pero no generaliza para las demás.

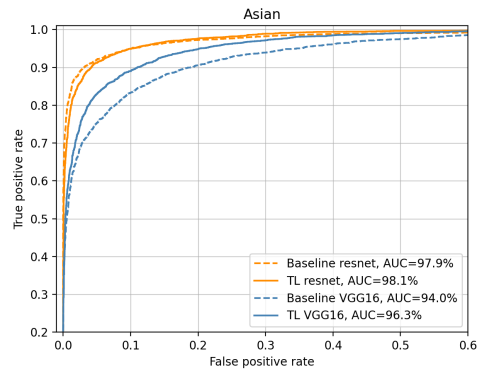
5.4. Resultados RFW con BUPT-Balancedface

Los creadores de ambas bases de datos proponían combinarlas, utilizando BUPT-Balancedface como base de datos de entrenamiento y RFW como base de datos de test [32] [52]. Por ello, se probó a hacerlo de esta forma. Los usuarios que en las pruebas de BUPT-Balancedface se utilizaban como test, en este caso se utilizaban como conjunto de validación para poder escoger la mejor época. Por otro lado, en RFW se van a generar nuevos matches, parejas de test, en todo el conjunto de cada uno de los grupos demográficos.

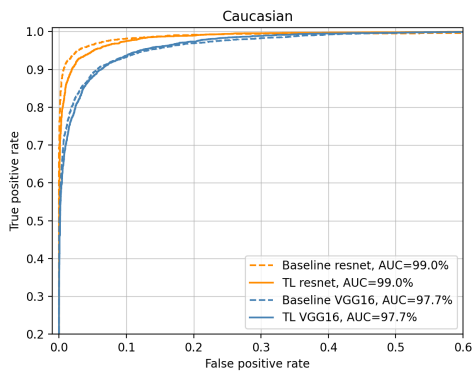
Las primeras pruebas se hicieron sobre los resultados de raza. En la Figura 5.6, podemos observar las gráficas de FPR, false positive rate, vs TPR, true positive rate, que representan la AUC, obtenidos para cada una de las razas, y en la tabla 5.8 se recogen los resultados de EER y AUC para cada una de las clases.



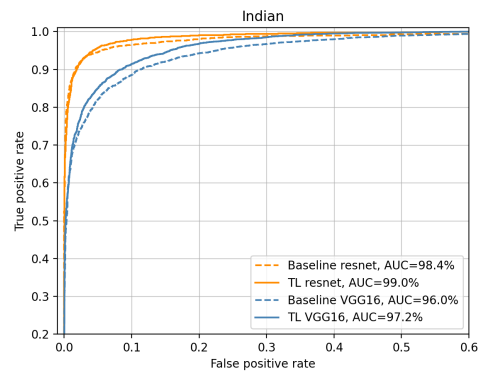
(a) African. Antes y después de entrenar con triplet loss.



(b) Asian. Antes y después de entrenar con triplet loss



(c) Caucasian. Antes y después de entrenar con triplet loss.



(d) Indian. Antes y después de entrenar con triplet loss

Figura 5.6: Gráficas por etnia BUPT-Balancedface + RFW. Antes y después de entrenar el triplet loss.

Tras realizar estas pruebas y ver que los resultados eran muy negativos, se descartó este experimento y no se hicieron las pruebas en género y en género combinado con etnia. El rendimiento bajaba

EER	Resnet50	Resnet50 + TL	VGG16	VGG16 + TL
African	9.33 %	10.12 %	21.41 %	28.73 %
Asian	7.20 %	6.87 %	15.70 %	24.89 %
Caucasian	5.2 %	6.47 %	10.86 %	19.56 %
Indian	8.02 %	7.72 %	13.85 %	23.22 %

(a) EER.

AUC	Resnet50	Resnet50 + TL	VGG16	VGG16 + TL
African	97 %	96.4 %	86.3 %	78.2 %
Asian	98.0 %	98.2 %	92.2 %	83.9 %
Caucasian	98.8 %	98.5 %	95.7 %	88.2 %
Indian	97.5 %	97.9 %	93.6 %	85.3 %

(b) AUC.

Tabla 5.8: Resultados de EER y AUC para la etnia en la base de datos de BUPT-Balancedface + RFW.

respecto a los baseline, especialmente en VGG16, y únicamente tuvo una ligera mejora en Resnet50 para los asiáticos y los indios.

Observando esta gran diferencia entre como bajaba el rendimiento en uno u otro modelo se puede concluir que Resnet50 es más adaptable a ciertos cambios en las bases de datos, por tanto va a ser mejor para realizar transferencias de entrenamiento.

El motivo de que baje tanto el rendimiento puede ser debido a que la forma de extraer las caras para obtener los features en cada una de las bases de datos es diferente. En RFW estaban marcadas mientras que en BUPT-Balancedface lo hemos hecho con el detector facial MTCNN. Además, las bases de datos fueron creadas de manera diferente por lo que la naturaleza de las imágenes puede ser distinta, lo que podría afectar al transferir el entrenamiento de una a la otra.

CONCLUSIONES Y TRABAJO FUTURO

6.1. Conclusiones

El objetivo de este trabajo de fin de grado era crear nuevos modelos de reconocimiento facial que permitieran obtener mejores resultados que el estado del arte en las diferentes etnias y géneros. Eliminando de esta forma los sesgos existentes en los modelos actuales relacionados estos rasgos.

Se ha diseñado un sistema que, añadiendo una nueva capa de triplet loss a los modelos de reconocimiento facial ya existentes, permite obtener mejores resultados que los modelos existentes en las diferentes etnias y géneros añadiendo información de estos rasgos. Se obtienen mayores mejoras en VGG16 ya que al tener una precisión menor tenía mayor margen de mejora. Lo mismo ocurre en las etnias y géneros en los que la precisión era menor. La mayor mejora en Resnet50 se obtuvo en la base de datos de RFW para mujeres indias, con un 1.7% y en VGG16 de un 3.8% para hombres asiáticos. Además, también se vieron grandes mejoras en las etnias, sin mirar género, con menor precisión de esa base de datos. Se obtuvieron mejoras del 3.2% en asiáticos y africanos, y del 2% en indios para VGG16 y de aproximadamente un 1% en los tres modelos correspondientes de Resnet50. Sin embargo, a pesar de las mejoras obtenidas, sigue habiendo sesgos entre las etnias y géneros, teniendo la mayor precisión en los hombres caucásicos.

Por otro lado, se ha visto que el conjunto de entrenamiento utilizado influye directamente en los resultados obtenidos. En las clases en las que se tenía menor número de imágenes para el entrenamiento y por tanto se tenía un menor número de triplets, las mejoras eran menores, incluso empeorando en algunos casos como ocurría con la partición de la base de datos BUPT-Balancedface que se utilizó. Por este motivo, podemos concluir que para obtener una mejora se tiene que tener un cierto número de imágenes de entrenamiento para poder crear suficientes triplets que entrenen la red.

6.2. Trabajo futuro

Como trabajo futuro se propone entrenar toda la base de datos de BUPT-Balancedface, para poder generar un mayor número de triplets por cada clase que nos permitan entrenar mejor esta última capa.

Además, se propone calcular de nuevo los vectores de características, features, de la base de datos de RFW pero, en lugar de utilizar las posiciones de la cara que aportan los creadores, utilizar el detector facial de MTCNN.

Tras realizar estos dos puntos se propone probar de nuevo a entrenar la capa de triplet loss con la base de datos completa de BUPT-Balancedface y como test utilizar RFW con los nuevos features calculados.

Por otro lado, se propone probar con otros soft biometrics en los que también existen sesgos como la edad para verificar si en otros tipos también se pueden obtener mejorías.

Por último, se propone realizar un estudio para ver si alguno de los modelos entrenados obtiene mejores resultados que los otros en todas las clases, ambos géneros, las diferentes etnias y las diferentes combinaciones de género y etnia.

BIBLIOGRAFÍA

- [1] P. Tome, J. Fierrez, R. Vera-Rodriguez, and J. Ortega-Garcia, "Combination of face regions in forensic scenarios," *Journal of Forensic Sciences*, vol. 60, pp. 1046–1051, May 2015.
- [2] T. Ali, P. Tome, J. Fierrez, R. Vera-Rodriguez, L. Spreeuwers, and R. Veldhuis, "A study of identification performance of facial regions from cctv images," in *5th International Workshop on Computational Forensics, IWCF 2012*, LCNS 8915, Springer, November 2015.
- [3] E. Gonzalez-Sosa, R. Vera-Rodriguez, J. Fierrez, and J. Ortega-Garcia, "Exploring facial regions in unconstrained scenarios: Experience on icb-rw," *IEEE Intelligent Systems*, 2018.
- [4] P. Tome, R. Vera-Rodriguez, J. Fierrez, and J. Ortega-Garcia, "Variability compensation using nap for unconstrained face recognition," in *Proc. 10th International Conference on Practical Applications of Agents and Multi-Agent Systems (PAAMS'12)*, vol. 151, pp. 129–139, Springer, March 2012.
- [5] S. Gabriel-Sanz, R. Vera-Rodriguez, P. Tome, and J. Fierrez, "Assessment of gait recognition based on the lower part of the human body," in *Proc. International Workshop on Biometrics and Forensics (IWBF'13)*, pp. 1–4, 2013.
- [6] R. Vera-Rodriguez, J. S. D. Mason, and N. W. D. Evans, "Automatic cross-biometric footprint database labelling using speaker recognition," in *International Conference on Biometrics (ICB)*, pp. 503–512, Springer LNCS-5558, June 2009.
- [7] A. Acien, J. V. Monaco, A. Morales, R. Vera-Rodriguez, and J. Fierrez, "Typenet: Scaling up keystroke biometrics," in *IEEE/IAPR Intl. Joint Conf. on Biometrics (IJCB)*, September 2020.
- [8] M. Santopietro, R. Vera-Rodriguez, R. Guest, A. Morales, and A. Acien, "Assessing the quality of swipe interactions for mobile biometric systems," in *IEEE/IAPR Intl. Joint Conf. on Biometrics (IJCB)*, 2020.
- [9] R. Vera-Rodriguez, R. Tolosana, M. Caruana, G. Manzano, C. Gonzalez-Garcia, J. Fierrez, and J. Ortega-Garcia, "Deepsigncx: Signature complexity detection using recurrent neural networks," in *Proc. 15th International Conference on Document Analysis and Recognition, ICDAR*, September 2019.
- [10] R. Tolosana, J. Gismero-Trujillo, R. Vera-Rodriguez, J. Fierrez, and J. Ortega-Garcia, "Mobi-letouchdb: Mobile touch character database in the wild and biometric benchmark," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, CVPRw*, June 2019.
- [11] R. Tolosana, R. Vera-Rodriguez, C. Gonzalez-Garcia, J. Fierrez, S. Rengifo, A. Morales, J. Ortega-Garcia, J. C. Ruiz-Garcia, S. Romero-Tapiador, J. Jiang, S. Lai, L. Jin, Y. Zhu, J. Galbally, M. Diaz, M. A. Ferrer, M. Gomez-Barrero, I. Hodashinsky, K. Sarin, A. Slezkin, M. Bardamova, M. Svetlakov, M. Saleem, C. L. Szücs, B. Kovari, F. Pulsmeier, M. Wehbi, D. Zanca, S. Ahmad, S. Mish-

- ra, and S. Jabin, "Icdar 2021 competition on on-line signature verification," *arXiv preprint arXiv:2106.00739*, 2021.
- [12] P. Tome, J. Fierrez, R. Vera-Rodriguez, and D. Ramos, "Identification using face regions: Application and assessment in forensic scenarios," *Forensic Science International*, no. 233, pp. 75–83, 2013.
- [13] H. Proença, M. Nixon, M. Nappi, E. Ghaleb, G. Özbulak, H. Gao, H. K. Ekenel, K. Grm, V. Struc, H. Shi, X. Zhu, S. Liao, Z. Lei, S. Z. Li, W. Gutfeter, A. Pacut, J. Brogan, W. J. Scheirer, E. Gonzalez-Sosa, R. Vera-Rodriguez, J. Fierrez, J. Ortega-Garcia, D. Riccio, and L. D. Maio, "Trends and controversies," *IEEE Intelligent Systems*, vol. 33, pp. 41–67, May 2018.
- [14] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1701–1708, 2014.
- [15] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, "Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments," in *Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition*, (Marseille, France), Erik Learned-Miller and Andras Ferencz and Frédéric Jurie, Oct. 2008.
- [16] Y. Sun, Y. Chen, X. Wang, and X. Tang, "Deep learning face representation by joint identification-verification," in *Advances in Neural Information Processing Systems* (Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Q. Weinberger, eds.), vol. 27, Curran Associates, Inc., 2014.
- [17] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting 10,000 classes," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014.
- [18] Y. Sun, D. Liang, X. Wang, and X. Tang, "Deepid3: Face recognition with very deep neural networks," 2015.
- [19] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 2015.
- [20] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *Proceedings of the British Machine Vision Conference (BMVC)* (M. W. J. Xianghua Xie and G. K. L. Tam, eds.), pp. 41.1–41.12, BMVA Press, September 2015.
- [21] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015.
- [22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015.
- [23] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "Vggface2: A dataset for recognising faces across pose and age," 2018.
- [24] A. Dantcheva, P. Elia, and A. Ross, "What else does your biometric data reveal? a survey on soft biometrics," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 3, pp. 441–467, 2016.

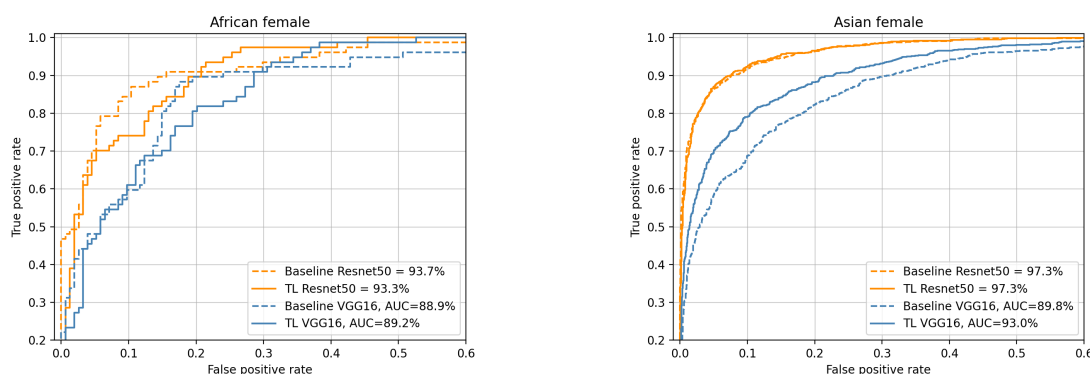
- [25] H. Khiyari and H. Wechsler, "Face verification subject to varying (age, ethnicity, and gender) demographics using deep learning," *Journal of Biometrics Biostatistics*, vol. 07, 01 2016.
- [26] V. Albiero, K. K. S., K. Vangara, K. Zhang, M. C. King, and K. W. Bowyer, "Analysis of gender inequality in face recognition accuracy," 2020.
- [27] R. Bothwell, J. Brigham, and R. Malpass, "Cross-racial identification," *Personality and Social Psychology Bulletin*, vol. 15, pp. 19–25, 03 1989.
- [28] V. Albiero, K. Zhang, and K. W. Bowyer, "How does gender balance in training data affect face recognition accuracy?," 2020.
- [29] M. Alvi, A. Zisserman, and C. Nellaker, "Turning a blind eye: Explicit removal of biases and variation from deep neural network embeddings," 2018.
- [30] A. Morales, J. Fierrez, R. Vera-Rodriguez, and R. Tolosana, "Sensitivenets: Learning agnostic representations with application to face images," 2020.
- [31] R. Vera-Rodriguez, M. Blazquez, A. Morales, E. Gonzalez-Sosa, J. C. Neves, and H. Proença, "Facegenderid: Exploiting gender information in dcnn face recognition systems," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 2254–2260, 2019.
- [32] M. Wang and W. Deng, "Mitigate bias in face recognition using skewness-aware reinforcement learning," *arXiv preprint arXiv:1911.10692*, 2019.
- [33] E. Gonzalez-Sosa, A. Dantcheva, R. Vera-Rodriguez, J. Dugelay, F. Bremond, and J. Fierrez, "Image-based gender estimation from body and face across distances," in *Proc. International Conference on Pattern Recognition, ICPR*, December 2016.
- [34] R. Vera-Rodriguez, P. Marin-Belinchon, E. Gonzalez-Sosa, P. Tome, and J. Ortega-Garcia, "Exploring automatic extraction of body-based soft biometrics," in *Proc. International Carnahan Conference on Security Technology (ICCST)*, October 2017.
- [35] P. Tome, J. Fierrez, R. Vera-Rodriguez, and M. Nixon, "Soft biometrics and their application in person recognition at a distance," *IEEE Transactions on Information Forensics and Security*, vol. 9, pp. 464–475, March 2014.
- [36] E. Gonzalez-Sosa, J. Fierrez, R. Vera-Rodriguez, and F. Alonso-Fernandez, "Facial soft biometrics for recognition in the wild: Recent works, annotation and cots evaluation," *IEEE Trans. on Information Forensics and Security*, vol. 13, no. 7, 2018.
- [37] P. Tome, R. Vera-Rodriguez, J. Fierrez, and J. Ortega-Garcia, "Facial soft biometric features for forensic face recognition," *Forensic Science International*, vol. 257, pp. 171–284, December 2015.
- [38] B. F. Klare, S. Klum, J. C. Klontz, E. Taborsky, T. Akgul, and A. K. Jain, "Suspect identification based on descriptive facial attributes," in *IEEE International Joint Conference on Biometrics*, pp. 1–8, 2014.
- [39] N. Kumar, A. Berg, P. N. Belhumeur, and S. Nayar, "Describable visual attributes for face verification and image search," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 10, pp. 1962–1977, 2011.

- [40] E. M. Rudd, M. Günther, and T. E. Boulton, "Moon: A mixed objective optimization network for the recognition of facial attributes," *Lecture Notes in Computer Science*, p. 19–35, 2016.
- [41] P. Samangouei and R. Chellappa, "Convolutional neural networks for attribute-based active authentication on mobile devices," in *2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pp. 1–8, 2016.
- [42] A. Dantcheva, C. Velardo, A. D'Angelo, and J.-L. Dugelay, "Bag of soft biometrics for person identification: New trends and challenges," *Multimedia Tools and Applications*, vol. 51, 01 2011.
- [43] H. Zhang, J. R. Beveridge, B. A. Draper, and P. J. Phillips, "On the effectiveness of soft biometrics for increasing face verification rates," *Computer Vision and Image Understanding*, vol. 137, pp. 50–62, 2015.
- [44] A. K. Jain and U. Park, "Facial marks: Soft biometric for face recognition," in *2009 16th IEEE International Conference on Image Processing (ICIP)*, pp. 37–40, 2009.
- [45] U. Park and A. K. Jain, "Face matching and retrieval using soft biometrics," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 3, pp. 406–415, 2010.
- [46] "Bida lab." <http://atvs.ii.uam.es/atvs/>.
- [47] "Megaface and mf2: Million-scale face recognition." <http://megaface.cs.washington.edu/index.html>.
- [48] I. Kemelmacher-Shlizerman, S. M. Seitz, D. Miller, and E. Brossard, "The megaface benchmark: 1 million faces for recognition at scale," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4873–4882, 2016.
- [49] A. Nech and I. Kemelmacher-Shlizerman, "Level playing field for million scale face recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017.
- [50] M. Wang, W. Deng, J. Hu, X. Tao, and Y. Huang, "Racial faces in the wild: Reducing racial bias by information maximization adaptation network," in *The IEEE International Conference on Computer Vision (ICCV)*, October 2019.
- [51] "Racial faces in-the-wild (rfw)." <http://www.whdeng.cn/RFW/testing.html>.
- [52] "Ethnicity aware training datasets." <http://www.whdeng.cn/RFW/Trainingdataste.html>.
- [53] "Oxford vggface implementation using keras functional framework v2+." <https://github.com/rcmalli/keras-vggface>.
- [54] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Processing Letters*, vol. 23, p. 1499–1503, Oct 2016.

APÉNDICES

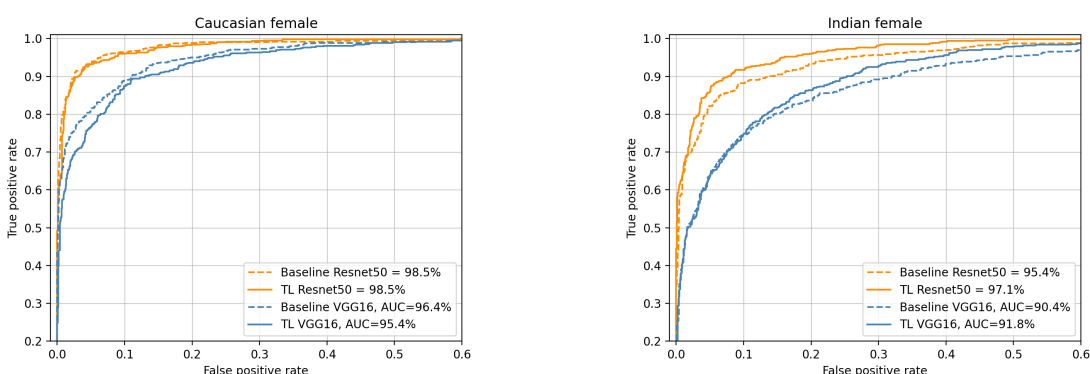
RESULTADOS RFW

En este apéndice se recogen las gráficas obtenidas para la base de datos de RFW para las clases específicas de las diferentes combinaciones de etnia y género. En la Figura ?? se encuentran los resultados para las diferentes razas, en mujeres. Por otro lado, en la Figura A.2 se muestran los resultados para las diferentes razas en hombres.



(a) African female. Antes y después de entrenar con triplet loss.

(b) Asian female. Antes y después de entrenar con triplet loss



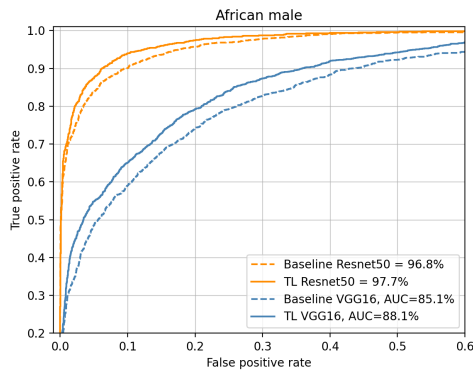
(c) Caucasian female. Antes y después de entrenar con triplet loss.

(d) Indian female. Antes y después de entrenar con triplet loss

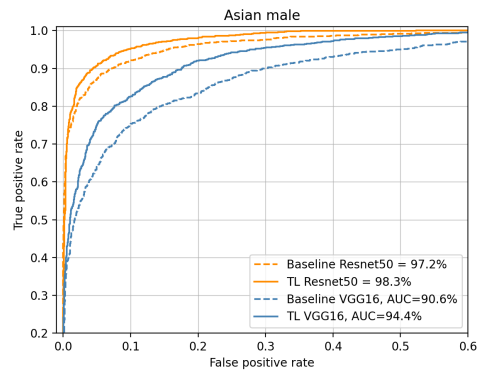
Figura A.1: Gráficas por etnia en mujeres en RFW. Antes y después de entrenar el triplet loss.

En el caso de los hombres se puede ver una mejora en todos menos para los hombres caucásicos en Resnet50. Esto es debido a que ese sistema ya aportaba un buen valor de precisión.

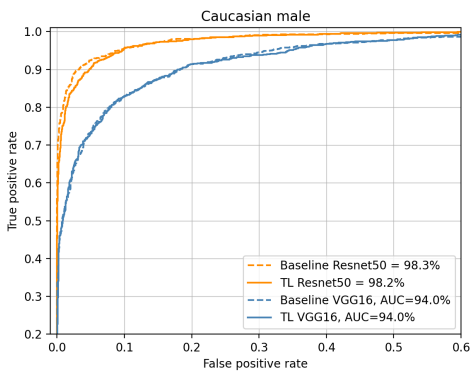
Por otro lado, en las mujeres se pueden observar saltos más bruscos en las gráficas, esto es debido al bajo número de matches, parejas de test, que tenemos en estos grupos demográficos. En concreto este efecto se puede observar en las mujeres africanas. Para las mujeres indias y asiáticas se ve una mejora tanto en VGG16 como en Resnet50. Sin embargo, para las mujeres caucásicas, al tener pocos datos de entrenamiento y ya disponer de un buen resultado, no se producen mejoras, de hecho en el caso de VGG16 el AUC empeora.



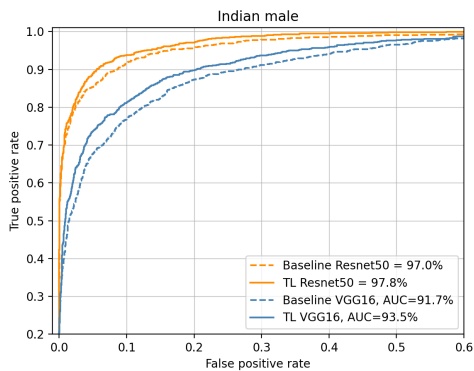
(a) African. Antes y después de entrenar con triplet loss.



(b) Asian. Antes y después de entrenar con triplet loss



(c) Caucasian. Antes y después de entrenar con triplet loss.

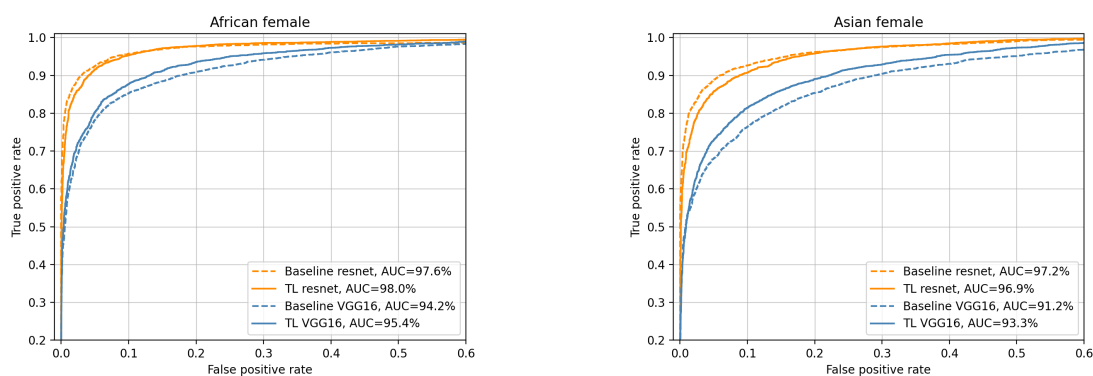


(d) Indian. Antes y después de entrenar con triplet loss

Figura A.2: Gráficas por etnia para los hombres en RFW. Antes y después de entrenar el triplet loss.

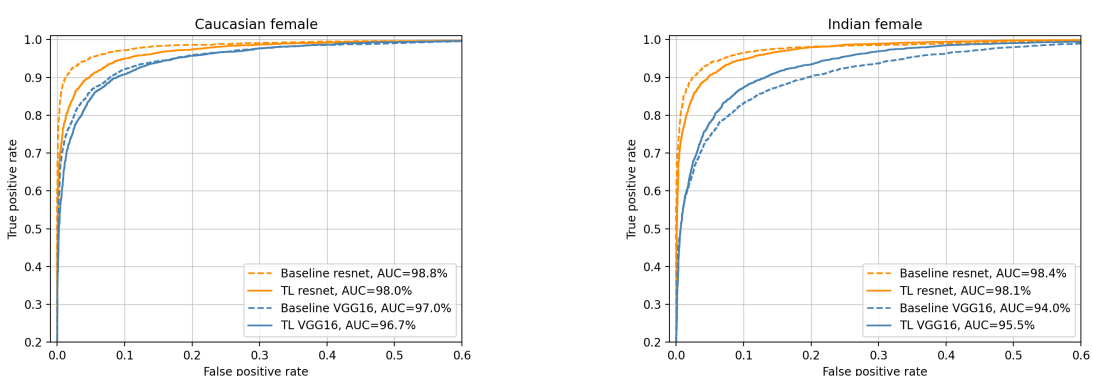
RESULTADOS BUPT-BALANCEDFACE

En este apéndice se recogen las gráficas obtenidas para la base de datos de BUPT-Balancedface para las clases específicas de las diferentes combinaciones de etnia y género. En la Figura B.1 se encuentran los resultados para las diferentes etnias para mujeres. Por otro lado, en la Figura B.2 se muestran los resultados para los hombres en las diferentes etnias.



(a) African female. Antes y después de entrenar con triplet loss.

(b) Asian female. Antes y después de entrenar con triplet loss



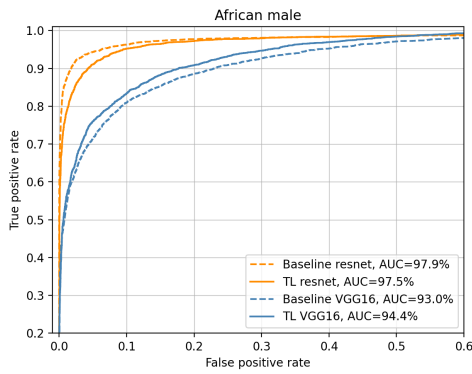
(c) Caucasian female. Antes y después de entrenar con triplet loss.

(d) Indian female. Antes y después de entrenar con triplet loss

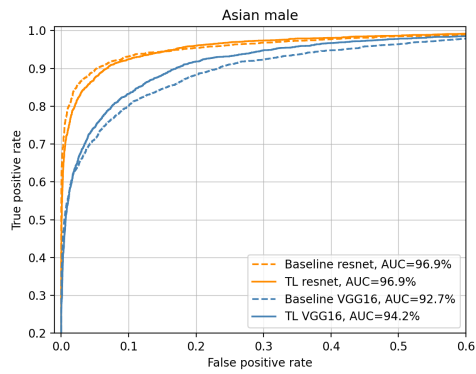
Figura B.1: Gráficas por etnia en mujeres BUPT-Balancedface. Antes y después de entrenar el triplet loss.

Se puede ver que para el modelo de Resnet50 no se producen mejoras en ninguna de las clases.

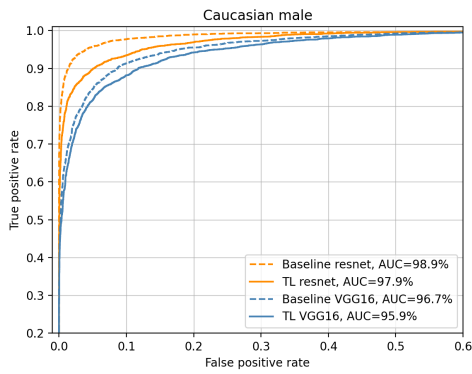
Para VGG16, sólo se producen pequeñas mejoras en hombres y mujeres asiáticos y africanos y en las mujeres indias.



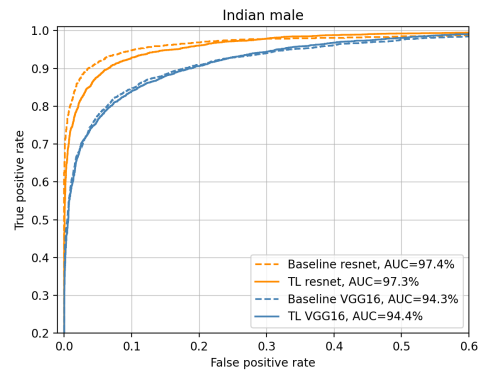
(a) African male. Antes y después de entrenar con triplet loss.



(b) Asian male. Antes y después de entrenar con triplet loss



(c) Caucasian male. Antes y después de entrenar con triplet loss.



(d) Indian male. Antes y después de entrenar con triplet loss

Figura B.2: Gráficas por etnia en hombres BUPT-Balancedface. Antes y después de entrenar el triplet loss.

UAM

UNIVERSIDAD AUTONOMA

DE MADRID