N° tesis: jcb

PROYECTO FIN DE CARRERA

Presentado a

LA UNIVERSIDAD DE LOS ANDES
FACULTAD DE INGENIERÍA
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA Y ELECTRÓNICA

Para obtener el título de

INGENIERO ELECTRÓNICO

por

José Daniel Pantoja Bernal

SISTEMA DE SEGURIDAD A PARTIR DE UNA BASE DE DATOS MEDIANTE MACHINE LEARNING, BASADO EN RECONOCIMIENTO DE ROSTROS

Sustentado el día 5 de diciembre de 2023 frente al jurado:

Composición del jurado

- *Asesor*: Fernando Enrique Lozano Martínez, Profesor Asociado, Universidad de Los Andes
- *Co-Asesor*: Juan José García Cárdenas, Profesor Instructor, Universidad de Los Andes
- *Jurado:* Fredy Enrique Segura Quijano, Profesor Asociado, Universidad de Los Andes

Contenido

1 INTRODUCCIÓN 3
2 MARCO TEÓRICO, CONCEPTUAL E HISTÓRICO 4
2.1 Marco Teórico 4
2.1.1 Deep Learning (Aprendizaje profundo) 4
2.1.2 Reconocimiento Facial 4
2.1.3 Transfer Learning (Transferencia de Aprendizaje) 5
2.2 Marco Conceptual 6
2.2.1 Detección de Rostros 6
2.2.2 Clasificación de Rostros 8
2.2.3 Detección de Rostros Artificiales (Deepfakes) 10
2.3 Marco Histórico 12
2.3.1 Antecedentes internos 12
2.3.2 Antecedentes externos 12
3 DEFINICION Y ESPECIFICACION DEL TRABAJO 13
3.1 Definición 13
3.2 Especificaciones 14
3.2.1 Detección de rostros 14
3.2.2 Clasificación de rostros 14
3.2.3 Detección de rostros artificiales (Deepfakes) 14
4 TRABAJO REALIZADO 15
4.1 Descripción del Resultado Final 15
4.2 Trabajo computacional 18
5 VALIDACIÓN DEL TRABAJO 19
5.1 Metodología de prueba 19
5.2 Validación de los resultados del trabajo 19
5.2.1 Pruebas de validación de modelo 19
5.2.2 Pruebas finales 23
5.3 Evaluación del plan de trabajo 30
6 DISCUSIÓN 30
7 CONCLUSIONES 31
8 AGRADECIMIENTOS 32
9 REFERENCIAS 33
10 APENDICES 35

1 INTRODUCCIÓN

En la era actual, caracterizada por avances tecnológicos acelerados y la creciente necesidad de garantizar la seguridad en diversos entornos, el reconocimiento facial ha emergido como una herramienta fundamental. En este contexto, la presente investigación se enfoca en un sistema de seguridad innovador desarrollado para abordar los desafíos contemporáneos en la gestión de identidades. Este sistema, basado en tres modelos de redes convolucionales preentrenados que operan de manera conjunta, constituye un avance significativo en la eficiencia y la adaptabilidad de los sistemas de reconocimiento facial. El núcleo de esta investigación reside en la capacidad del sistema para identificar personas presentes en una base de datos predeterminada, al tiempo que distingue entre individuos conocidos y desconocidos. Además, su habilidad para discernir entre rostros reales y artificiales añade una capa adicional de seguridad, evitando posibles intentos de falsificación de identidad. La implementación del sistema se ha llevado a cabo en una interfaz amigable y de fácil uso.

La elección de este tema se fundamenta en la creciente importancia del reconocimiento facial en el ámbito de la seguridad. En un mundo donde la precisión en la identificación personal es esencial, surge la necesidad imperativa de sistemas avanzados que no solo sean eficientes y precisos, sino también adaptables a diversas condiciones y entornos. Este trabajo no solo busca proporcionar una solución técnica, sino que también aborda las implicaciones éticas asociadas con el uso de tecnologías de reconocimiento facial. La relevancia de esta investigación se enfoca en la mejora sustancial de la seguridad en lugares con una población delimitada, como hoteles, universidades o instituciones gubernamentales, donde la gestión de identidades es esencial. Sin embargo, es crucial considerar y abordar los posibles impactos negativos, como el riesgo de discriminación y el acceso no autorizado a datos personales. Este estudio se posiciona como una contribución integral que no solo promueve el avance técnico, sino que también reflexiona sobre las responsabilidades éticas asociadas con las tecnologías de reconocimiento facial en un mundo cada vez más interconectado y dependiente de la información.

2 MARCO TEÓRICO, CONCEPTUAL E HISTÓRICO

2.1 Marco Teórico

2.1.1 Deep Learning (Aprendizaje profundo)

En la última década, el campo de la inteligencia artificial ha experimentado una metamorfosis fundamental con la irrupción del deep learning. El artículo que dio paso al deep learning fue "Learning representations by back-propagating errors" (1986), de David Rumelhart, Geoffrey Hinton y Ronald Williams [1]. Este artículo presentó el algoritmo de retro-propagación, un método para entrenar redes neuronales artificiales. La retro-propagación permitió a las redes neuronales aprender a representar datos complejos, abriendo así el camino a una serie de avances en el campo del deep learning. Esta rama del machine learning, también conocida como aprendizaje profundo, se basa en la estructura y función de las redes neuronales artificiales con el objetivo de modelar y resolver problemas complejos. A diferencia de los métodos tradicionales de machine learning, el deep learning implica modelos con múltiples capas, llamadas capas profundas, que forman una arquitectura jerárquica. Las redes neuronales profundas, compuestas por capas interconectadas de neuronas, tienen la capacidad de aprender representaciones jerárquicas de datos complejos. Este dominio ha experimentado un crecimiento vertiginoso, impulsado por avances teóricos y tecnológicos. Estos progresos han permitido a las redes neuronales profundas alcanzar un rendimiento excepcional en tareas que van desde el reconocimiento de imágenes hasta la traducción de idiomas.

2.1.2 Reconocimiento Facial

El reconocimiento facial es una tecnología de biometría que se utiliza para identificar o verificar la identidad de una persona a través de sus rasgos faciales únicos. Esta tecnología utiliza algoritmos para analizar patrones faciales, como la forma de los ojos, la nariz, la boca y otros elementos distintivos, y luego compara esta información con una base de datos de rostros previamente registrados. En el ámbito del reconocimiento facial, el deep learning ha provocado una transformación revolucionaria. Las redes neuronales profundas han propiciado el desarrollo de sistemas de reconocimiento facial significativamente más precisos y eficientes que sus contrapartes tradicionales.

Hitos como "Deep Residual Learning for Image Recognition" (2015), de Kaiming He, Xiangyu Zhang, Shaoqing Ren y Jian Sun [2], presentaron la arquitectura ResNet, que superó a AlexNet en el ILSVRC 2015, demostrando la capacidad de entrenar

redes profundas incluso con numerosas capas. Otro avance clave es "Attention Is All You Need" (2017), de Vaswani et al [3]. Esta investigación introdujo la arquitectura de red neuronal de atención, destacando su capacidad para mejorar el rendimiento en diversas tareas, incluyendo el reconocimiento facial, al permitir que la red se centre en regiones específicas de la entrada. En el ámbito específico de reconocimiento facial, "FaceNet: A Unified Embedding for Face Recognition and Clustering" (2015), de Florian Schroff, Dmitry Kalenichenko y James Philbin [4], presentó FaceNet, una red neuronal convolucional optimizada para generar representaciones vectoriales de rostros. Esta innovación ha sido esencial para tareas como identificación y verificación facial. Además, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification" (2014) [5], de Yilun Wang, Andrew Zisserman y Tony Jebara, presentó DeepFace, un sistema de reconocimiento facial basado en redes neuronales profundas que logró un rendimiento comparable al de los humanos en la tarea de verificación facial. Estos logros representan solo una fracción de los hitos en deep learning que han influido significativamente en el reconocimiento facial, destacando la continua evolución de este campo y anticipando mejoras continuas en la eficiencia de los sistemas de reconocimiento facial en los años por venir.

2.1.3 Transfer Learning (Transferencia de Aprendizaje)

El aprendizaje por transferencia o transfer learning es un paradigma en el aprendizaje automático (Machine Learning) que se basa en la idea de transferir conocimiento adquirido en una tarea fuente a una tarea objetivo-relacionada. Este enfoque se ha vuelto fundamental en el campo de la inteligencia artificial, ya que permite aprovechar el aprendizaje previo en un dominio específico para mejorar el rendimiento en una tarea relacionada, incluso cuando los conjuntos de datos son limitados o inexistentes para la tarea objetivo. El concepto de transferencia de conocimiento no es nuevo y ha sido explorado en diversas disciplinas, pero su aplicación en el aprendizaje automático ha ganado popularidad en los últimos años.

El artículo seminal "A Survey of Transfer Learning" de Sinno Jialin Pan y Qiang Yang (2010) [6], proporciona una visión general exhaustiva de las técnicas y enfoques de transferencia de conocimiento en ese momento. Este artículo destaca la importancia de la transferencia de conocimiento en diversos dominios y presenta una taxonomía que clasifica los métodos de transferencia en diferentes categorías, como la transferencia inductiva, la transferencia transductiva y la transferencia multi-instantánea. Otro artículo influyente en el campo es "A Survey on Deep

Transfer Learning and Beyond" de Fuchao Yu, Xianchao Xiu, y Yunhui Li (2022) [7]. Este artículo aborda específicamente el aprendizaje por transferencia en el contexto de las redes neuronales profundas. Proporciona una revisión detallada de los enfoques de transferencia en redes neuronales y explora cómo se pueden transferir pesos preentrenados de una tarea a otra para mejorar el rendimiento. También discute los desafíos y oportunidades en este tipo de enfoque.

En el ámbito de la visión por computadora, "How transferable are features in deep neural networks?" de Yosinski et al. (2014) [8], es un artículo significativo. Explora cómo las capas de características aprendidas en una tarea pueden ser reutilizadas en una red neuronal convolucional (CNN) para mejorar el rendimiento en una tarea relacionada. Este trabajo destaca la importancia de la selección de capas y la adaptación de características para lograr una transferencia efectiva en tareas de reconocimiento de imágenes. Más recientemente, "Universal Language Model Fine-tuning for Text Classification" de Jeremy Howard y Sebastian Ruder (2018) [9], ha contribuido al campo del procesamiento del lenguaje natural (NLP). Este trabajo propone el uso de un modelo de lenguaje preentrenado, como el ULMFiT, y muestra cómo puede adaptarse eficientemente a tareas específicas con conjuntos de datos limitados, demostrando así la efectividad del aprendizaje por transferencia en NLP. En resumen, el aprendizaje por transferencia ha evolucionado significativamente a lo largo de los años, y la literatura en este campo proporciona un marco sólido para comprender los fundamentos teóricos y las aplicaciones prácticas de este enfoque en diversas áreas del aprendizaje automático.

2.2 Marco Conceptual

2.2.1 Detección de Rostros

La detección de rostros es una tecnología de visión por computadora que identifica y localiza caras humanas en imágenes o videos. A lo largo de su evolución, ha experimentado notables avances. En las décadas de 1970 y 1980, los primeros intentos se centraron en algoritmos para identificar patrones faciales, pero las limitaciones tecnológicas restringieron el progreso. En la década de 1990, con mejoras en el procesamiento y algoritmos, se utilizaron en aplicaciones como el enfoque automático de cámaras. El cambio significativo llegó en la década de 2000 con el auge de la tecnología digital. El aprendizaje profundo y las redes neuronales convolucionales se volvieron fundamentales. Algoritmos como Viola-Jones (2001)

[10], y desarrollos en aprendizaje profundo, como MTCNN [11] y SSD [12], impulsaron la eficiencia.

MTCNN: Multi-task Cascaded Convolutional Networks

El artículo Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks, publicado en 2016 por Kaipeng Zhang, Shaoting Zhang, Xiaogang Wang y Jian Sun [11], presenta un método de detección y alineación de rostros basado en redes neuronales convolucionales (CNN). El método, denominado MTCNN (Multi-Task Cascaded Convolutional Neural Networks), se basa en una arquitectura en cascada de tres etapas, cada una de las cuales utiliza una CNN diferente para realizar una tarea específica.

La primera etapa, denominada P-Net, se encarga de generar una gran cantidad de regiones candidatas de posible presencia de un rostro. La P-Net es una red pequeña y rápida que se ejecuta en paralelo en una GPU. La segunda etapa, denominada R-Net, se encarga de clasificar las regiones candidatas de la primera etapa como rostros o no rostros. La R-Net es una red más grande y compleja que la P-Net, y utiliza una función de pérdida de clasificación para aprender a distinguir entre rostros y no rostros. La tercera etapa, denominada O-Net, se encarga de refinar las regiones candidatas de la segunda etapa y de estimar la posición de los cinco puntos de referencia faciales (ojos, nariz y boca). La O-Net es la red más grande y compleja de las tres, y utiliza una función de pérdida de regresión para aprender a estimar la posición de los puntos de referencia faciales. El método MTCNN ha demostrado ser muy eficaz en la detección y alineación de rostros. En el conjunto de datos FDDB, MTCNN alcanza un rendimiento de 99,5% en precisión y 0,2% en falsos positivos.

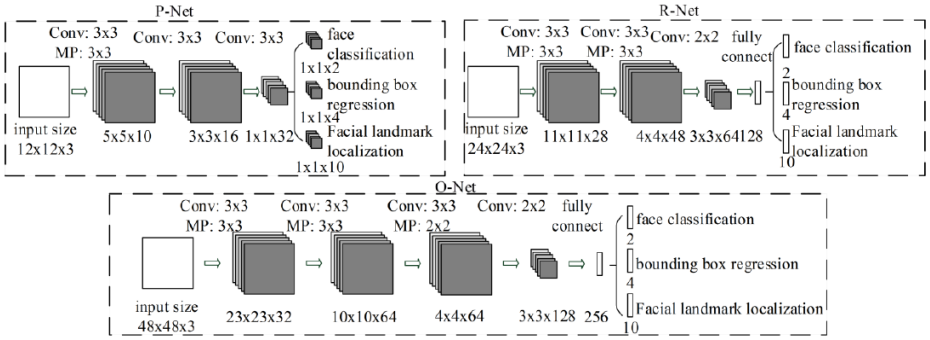


Figura 1 Arquitectura MTCNN tomado de [11]

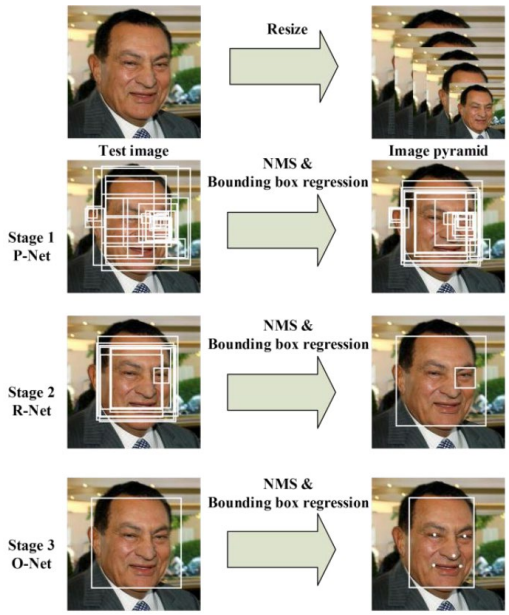


Figura 2 Funcionamiento MTCNN tomado de [11]

2.2.2 Clasificación de Rostros

La clasificación de rostros es una rama de la visión por computadora y el procesamiento de imágenes que se centra en identificar y categorizar rostros humanos en imágenes o vídeos. Su objetivo principal es asignar etiquetas o clasificaciones a los rostros según ciertas características o criterios, como la identificación de individuos, el género, la edad, las emociones, entre otros. Esta tarea se ha vuelto esencial en una amplia gama de aplicaciones, desde la seguridad y vigilancia hasta la interacción humano-computadora y la organización de contenido multimedia.

La clasificación de rostros ha experimentado avances significativos en múltiples frentes, especialmente en la última década. Inicialmente, los métodos tradicionales se basaban en características geométricas y estadísticas, pero el advenimiento del aprendizaje automático impulsó una revolución en la precisión y la generalización de los modelos. Las redes neuronales convolucionales (CNN) se convirtieron en una herramienta fundamental al permitir el aprendizaje de jerarquías de características, superando las limitaciones de los enfoques manuales. Además, la introducción de descriptores locales como Histogramas de Gradientes Orientados (HOG) y Descriptores Binarios Locales (LBP) mejoró la capacidad de los modelos

para capturar información discriminativa, especialmente en situaciones donde la iluminación y las poses variaban considerablemente. El surgimiento de arquitecturas de redes pre-entrenadas, como VGGFace [13] y FaceNet [4], facilitó la transferencia de conocimientos y aceleró el desarrollo de modelos más precisos y eficientes. En el ámbito de la representación de rostros, la adopción de embeddings revolucionó la forma en que los sistemas de clasificación procesan y comparan rostros. Estos vectores numéricos compactos capturan características fundamentales de manera más eficiente, permitiendo una comparación más precisa y una clasificación robusta.

VGG-Face

VGG-Face es presentado en el artículo "Deep Face Recognition" de Parkhi et al. [13], la extracción de características se lleva a cabo mediante una Convolutional Neural Network (CNN) de dos capas. La primera capa, la capa de convolución, aplica filtros convolucionales a la imagen de entrada, matrices numéricas diseñadas para identificar patrones en las imágenes. La segunda capa, la capa de pooling, reduce la complejidad computacional al disminuir el tamaño de la representación de características. El modelo CNN se entrena en un extenso conjunto de datos de 2.6 millones de imágenes de 2,600 personas, etiquetadas con la identidad correspondiente. La formación de la CNN utiliza el algoritmo de retropropagación, que ajusta los pesos de los filtros convolucionales para minimizar la discrepancia entre las predicciones del modelo y las etiquetas reales de las imágenes de entrenamiento.

En la fase de clasificación, la capa de salida de la CNN produce embeddings faciales, representaciones vectoriales que capturan las características invariantes a factores como la iluminación, la pose y la expresión facial. Estos embeddings faciales se emplean para entrenar un clasificador de apoyo vectorial, un algoritmo de aprendizaje automático que separa los datos de diferentes clases al encontrar puntos de apoyo. Para clasificar nuevas imágenes, se extraen las características faciales de la imagen utilizando la CNN, y luego estas características se utilizan para generar embeddings faciales. Estos embeddings se introducen en el clasificador de apoyo vectorial, que asigna la nueva imagen a una de las identidades conocidas. Este enfoque garantiza una representación robusta de las características faciales, permitiendo una clasificación eficaz incluso en condiciones variables.

Este método supera a los enfoques tradicionales al ser más robusto ante variaciones en las condiciones y al poder aprender características discriminativas a

partir de conjuntos de datos más extensos. El modelo CNN propuesto consta de dos capas: una capa de convolución y una capa de pooling. El enfoque se evaluó en conjuntos de datos como LFW [14] y YTF [15], logrando tasas de error de identificación del 1.6% y 3.8%, respectivamente. Estos resultados comparables al estado del arte, pero con mayor eficiencia computacional, validaron la eficacia del método.

layer	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
type	input	conv	relu	conv	relu	mpool	conv	relu	conv	relu	mpool	conv	relu	conv	relu	conv	relu	mpool	conv
name	–	conv1_1	relu1_1	conv1_2	relu1_2	pool1	conv2_1	relu2_1	conv2_2	relu2_2	pool2	conv3_1	relu3_1	conv3_2	relu3_2	conv3_3	relu3_3	pool3	conv4_1
support	–	3	1	3	1	2	3	1	3	1	2	3	1	3	1	3	1	2	3
filt dim	–	3	–	64	–	–	64	–	128	–	–	128	–	256	–	256	–	–	256
num filts	–	64	–	64	–	–	128	–	128	–	–	256	–	256	–	256	–	–	512
stride	–	1	1	1	1	2	1	1	1	1	2	1	1	1	1	1	1	2	1
pad	–	1	0	1	0	0	1	0	1	0	0	1	0	1	0	1	0	0	1

layer	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
type	relu	conv	relu	conv	relu	mpool	conv	relu	conv	relu	conv	relu	mpool	conv	relu	conv	relu	conv	softmx
name	relu4_1	conv4_2	relu4_2	conv4_3	relu4_3	pool4	conv5_1	relu5_1	conv5_2	relu5_2	conv5_3	relu5_3	pool5	fc6	relu6	fc7	relu7	fc8	prob
support	1	3	1	3	1	2	3	1	3	1	3	1	2	7	1	1	1	1	1
filt dim	–	512	–	512	–	–	512	–	512	–	512	–	–	512	–	4096	–	4096	–
num filts	–	512	–	512	–	–	512	–	512	–	512	–	–	4096	–	4096	–	2622	–
stride	1	1	1	1	1	2	1	1	1	1	1	1	2	1	1	1	1	1	1
pad	0	1	0	1	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0

Figura 3 Arquitectura VGGFace tomado de [13]

2.2.3 Detección de Rostros Artificiales (Deepfakes)

La detección de rostros artificiales, en el contexto de los deepfakes, se refiere al proceso de identificar manipulaciones digitales en medios multimedia, especialmente videos e imágenes, generadas por algoritmos de aprendizaje profundo. Este fenómeno, surgido a principios de la década de 2010, ha evolucionado rápidamente gracias a la sofisticación de las técnicas de generación de deepfakes. En respuesta a las preocupaciones sobre la proliferación de contenido engañoso, la investigación ha avanzado en el desarrollo de algoritmos específicos de inteligencia artificial para detectar estas manipulaciones. A medida que los generadores de deepfakes y las tecnologías de detección continúan su competencia, la colaboración entre la comunidad de investigación, la industria y los legisladores se ha vuelto esencial para abordar este desafío en constante cambio, dando lugar a una mejora constante en ambas áreas y la promulgación de normativas destinadas a mitigar los riesgos asociados con la manipulación digital.

En este enfoque, se han desarrollado algoritmos de aprendizaje profundo específicamente diseñados para identificar estas creaciones digitales engañosas. Estos algoritmos, en muchos casos, se nutren de conjuntos de datos extensos y diversos que contienen tanto deepfakes como contenido auténtico. La efectividad de estas técnicas de detección radica en la capacidad de los modelos para identificar patrones sutiles y anomalías en la generación de deepfakes. Factores

como inconsistencias en el parpadeo, movimientos faciales irregulares y errores en la sincronización de audio y video se convierten en puntos clave de análisis. Así, la Detección Basada en IA no solo se ha convertido en un componente esencial en la lucha contra la manipulación digital, sino que también ha demostrado ser una herramienta crucial para mantener la integridad y la autenticidad en el contenido multimedia.

MesoNet: a Compact Facial Video Forgery Detection Network

El artículo "MesoNet: a Compact Facial Video Forgery Detection Network" [16], presenta una red neuronal llamada MesoNet, diseñada para detectar falsificaciones faciales en videos de manera eficiente. MesoNet es compacta y adecuada para dispositivos con recursos limitados, como dispositivos móviles. La arquitectura de MesoNet consta de tres capas principales: extracción de características, fusión de características y clasificación. La capa de extracción de características utiliza una red neuronal convolucional profunda para extraer características espaciales, temporales y de textura de los videos faciales. La capa de fusión de características combina estas características mediante concatenación, permitiendo aprender relaciones entre ellas. La capa de clasificación utiliza un clasificador de regresión logística.

La evaluación de MesoNet en un conjunto de datos mostró una precisión del 98,5% en la detección de falsificaciones faciales. El artículo detalla la arquitectura, datos de entrenamiento y evaluación, y compara MesoNet con otros métodos, destacando su compactibilidad y eficiencia. MesoNet tiene aplicaciones potenciales en la verificación de identidad, detección de propaganda y protección de la privacidad. Puede ayudar a prevenir la falsificación de identidad, la difusión de propaganda y la manipulación de videos faciales. En resumen, MesoNet es una herramienta prometedora para abordar problemas de seguridad y privacidad relacionados con videos faciales manipulados.

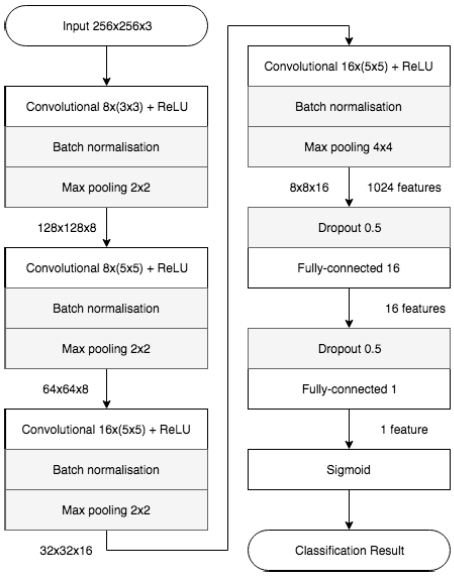


Figura 4 Arquitectura MesoNet tomado de [16]

2.3 Marco Histórico

2.3.1 Antecedentes internos

En el año 2019 el trabajo desarrollado por Juan Felipe Chávez Fonseca en su proyecto de grado tuvo como objetivo la implementación de un sistema de identificación personal mediante CNN, el cual permitiera brindar apoyo en actividades de seguridad. Sin embargo, el objetivo no se cumplió totalmente, ya que únicamente se implementó el sistema de identificación de rostros, el cual realizaba su función con más de 98% de precisión, al ser testeada en la base de datos LFW, pero este no realizaba la función de clasificar rostros dada una base de datos [17].

2.3.2 Antecedentes externos

En el estado del arte actual en el campo de reconocimiento facial, se ha demostrado que las técnicas de aprendizaje profundo han superado significativamente a los métodos tradicionales. Algunos ejemplos son la red neuronal convolucional (CNN), que se ha utilizado para detectar y reconocer caras en imágenes, y las redes adversarias generativas (GAN), que se han utilizado para sintetizar imágenes de caras realistas. También se han desarrollado técnicas de codificación de características como la red neuronal de codificación de caras (FaceNet), que permite la identificación precisa de personas en imágenes. Además, se han desarrollado redes de atención que permiten a los sistemas de

reconocimiento facial enfocarse en las partes importantes de una imagen y, por lo tanto, mejorar la precisión del reconocimiento facial. Sin embargo, todavía existen desafíos importantes que deben superarse, como la variabilidad en la iluminación, la expresión facial y el envejecimiento, para que los sistemas de reconocimiento facial sean más robustos y precisos en situaciones reales [18].

Por otro lado, con respecto a sistemas de seguridad basados en reconocimiento personal, empresas como Amazon, Paravisión, Cognitec, entre otros, se han empeñado en diseñar sistemas que permiten identificar características personales mediante detección de rostros, tales como género y edad. Además, han diseñado sistemas en tiempo real que permiten detectar suplantaciones de identidad o número de apariciones en un sitio especifico [19].

3 DEFINICION Y ESPECIFICACION DEL TRABAJO

3.1 Definición

En el contexto de la implementación de un sistema de seguridad basado en reconocimiento facial, a partir de una base de datos, con tres modelos principales (detección y clasificación de rostros, e identificación de deepfakes), el problema se centra en establecer un marco efectivo y ético para la seguridad en entornos específicos con una población definida. Este sistema se diseñó para operar en lugares que cuentan con una base de datos predefinida, limitando su aplicabilidad a entornos donde se pueda gestionar y mantener dicha base de datos.

La implementación del sistema de reconocimiento facial busca optimizar la eficiencia de la seguridad en entornos específicos, reduciendo la dependencia de recursos humanos para la vigilancia. Aunque el sistema puede contribuir rápidamente a la detección eficiente de amenazas y mejorar la seguridad comunitaria, es esencial abordar preocupaciones sobre la privacidad y la posible discriminación, lo que requiere la implementación de salvaguardias éticas. La crítica preocupación ética en torno al reconocimiento facial subraya la necesidad de establecer políticas y prácticas claras para garantizar un uso ético y prevenir abusos, como la vigilancia indiscriminada. Además, la implementación debe cumplir con todas las leyes y regulaciones legales relacionadas con la privacidad en la ubicación específica de despliegue, y el sistema debe protegerse contra manipulaciones o ataques para garantizar la integridad y confiabilidad de los resultados. Aunque no hay un impacto directo en la manufactura, se deben considerar aspectos de calidad y actualización del hardware, y la implementación sostenible implica la actualización continua de

modelos y la consideración de tecnologías más eficientes desde el punto de vista energético.

3.2 Especificaciones

3.2.1 Detección de rostros

La elección del modelo de detección de rostros debes ser seleccionado por su eficiencia y precisión en la detección de rostros, garantizando un tiempo de respuesta aceptable y una tasa de detección confiable. Las siguientes especificaciones se proponen para un contexto de máximo 20 rostros en un frame.

Parámetro	Restricciones	Niveles de Satisfacción
Eficiencia en la detección	Tiempo de respuesta < 100ms	Aceptable: 100ms Deseado: 50ms
Precisión	Tasa de detección de rostros correcta > 95%	Aceptable: 90% Deseado: 99%

Tabla 1 Especificaciones modelo detección de rostros

3.2.2 Clasificación de rostros

El modelo de clasificación de rostros se debe seleccionar por su capacidad para clasificar rostros con alta precisión y su resistencia a variaciones en la iluminación y la pose, lo que lo hace adecuado para entornos del mundo real.

Parámetro	Restricciones	Niveles de Satisfacción
Exactitud de clasificación	Precisión de clasificación > 90%	Aceptable: 90% Deseado: 97%
Manejo de variabilidad	Tolerancia a cambios de iluminación y pose del individuo	Aceptable: 80% Deseado: 95%

Tabla 2 Especificaciones modelo clasificación de rostros

3.2.3 Detección de rostros artificiales (Deepfakes)

El modelo de detección de detección de rostros artificiales se incorpora para identificar deepfakes con alta sensibilidad y especificidad, ofreciendo una capa adicional de seguridad al sistema.

Parámetro	Restricciones	Niveles de Satisfacción
Sensibilidad a deepfakes	Sensibilidad > 92%	Aceptable: 92% Deseado: 99%
Especificidad	Especificidad > 95%	Aceptable: 95% Deseado: 99%

Tabla 3 Especificaciones modelo de detección de deepfakes

4 TRABAJO REALIZADO

El trabajo realizado consiste en un sistema de seguridad basado en reconocimiento facial [20], el cual está compuesto por 3 modelos de redes convolucionales profundas preentrenados y unidos a través de transfer learning. A través de una interfaz, el usuario puede identificar personas previamente almacenadas en una base de datos mediante una cámara. Además, el sistema permite verificar a personas desconocidas o que no forman parte de la base de datos, mientras se lleva a cabo una verificación de la realidad de las personas detectadas. Este sistema fue implementado en el lenguaje de programación Python, elegido debido a su popularidad en temas de inteligencia artificial, como machine learning o deep learning. Además, este lenguaje cuenta con herramientas que facilitan el diseño del sistema, como librerías de visión por computadora o modelos de redes profundas. El sistema diseñado funciona de manera lineal, es decir, que el proceso se realiza de forma secuencial, uno tras otro. El proceso de funcionamiento consta de tres etapas esenciales: preprocesamiento, procesamiento y observación. Dicho proceso se puede observar en la Figura 6, la cual describe el proceso de un fotograma. Sin embargo, para una mejor comprensión, se presenta cada etapa de funcionamiento detallada.

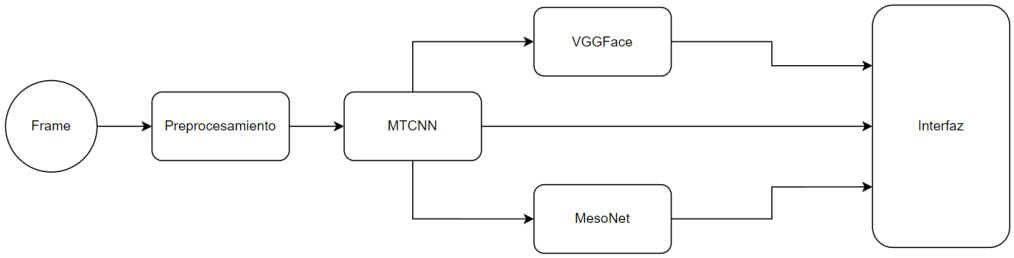


Figura 5 Diagrama de bloques funcionamiento del sistema

4.1 Descripción del Resultado Final

Etapa de preprocesamiento

La primera etapa del sistema inicia haciendo uso de la interfaz, en la cual se selecciona la base de datos sobre la cual se quiere hacer el reconocimiento. Esta base de datos debe encontrarse en una carpeta, la cual contenga subcarpetas con el nombre asociado a cada identidad, y en ellas se encuentren almacenadas las fotos en diferentes ángulos del rostro de cada una de ellas. Posterior a ello, inicia una etapa de preparación en la cual se verifica que se hayan creado previamente las representaciones vectoriales de cada una de las identidades de la base de datos (embeddings). Si este proceso no ha sido realizado, el modelo de clasificación VGGFace, haciendo uso de una de sus múltiples funcionalidades, genera un archivo en el cual se almacenan cada una de estas representaciones vectoriales. Este archivo será de gran utilidad a la hora de realizar el proceso de clasificación de un rostro, ya que gracias a él se realizarán operaciones vectoriales de una forma más fácil, reduciendo costos computacionales y permitiendo identificar un rostro de manera más eficiente. Cabe resaltar, que cada vez que se agreguen nuevas fotos o personas a la base de datos, este se deberá borrar, y se creará de nuevo haciendo uso del botón “Actualizar Base de datos” en la interfaz. Una vez creado el archivo de representación vectorial, se inicia la captura de video, donde cada fotograma es modificado aplicando diferentes correcciones de iluminación y un re-escalado que reduzca el tiempo de procesamiento y permita el correcto funcionamiento de la siguiente etapa.

Etapa de procesamiento

En esta etapa, se lleva a cabo un proceso para cada fotograma capturado en la etapa de preprocesamiento. Al finalizar la etapa anterior, el fotograma resultante ingresa inicialmente al modelo de detección de rostros, MTCNN. Este procesa el fotograma y obtiene las coordenadas de cada rostro detectado, así como información sobre el ancho y la altura del recuadro que enmarca cada rostro. Se modifica el fotograma original dibujando recuadros en cada rostro identificado con la información obtenida. Además, utilizando los datos proporcionados por MTCNN sobre el fotograma, se realiza un recorte para cada rostro, dando continuidad al proceso de detección de rostros artificiales y clasificación.

Antes de ingresar al modelo de detección de deepfakes, MesoNet, el rostro se re-escala debido a que el modelo cuenta con una capa de entrada de dimensiones diferentes a las del recorte del rostro. Una vez modificadas las dimensiones del recorte del rostro, este ingresa a MesoNet, que establece la probabilidad de que el rostro sea real, asignando un número entre 0 y 1, donde 0 indica la probabilidad de un rostro artificial y 1 la probabilidad de uno real. Similar al modelo de detección de rostros, el resultado de MesoNet se representa en el fotograma original dibujando, en la esquina

superior derecha del recuadro de cada rostro, una "R" para rostros reales y una "F" si se consideran falsos o deepfakes. Es importante señalar que el umbral elegido para determinar si un rostro es real, en condiciones donde las personas generalmente se encuentran de 1 a 6 metros de la cámara, oscila entre 0.2 y 0.5, dependiendo de la iluminación del entorno.

Retomando el proceso, para la etapa de clasificación del rostro se utilizan las dimensiones del recorte original de cada rostro para ser procesadas por VGGFace. Este modelo crea una representación vectorial del rostro identificado y se compara este vector con todos los vectores de la base de datos. La comparación se realiza mediante la distancia coseno, definida por:

$$dis_{cosine} = 1 - sim_{cosine}$$

Donde sim_{cosine} es la similaridad del coseno definida por:

$$sim_{cosine} = \frac{a^T b}{\sqrt{a^T a}\sqrt{b^T b}}$$

En esta fórmula, a es la representación vectorial del rostro identificado y b es la representación vectorial del rostro de la base de datos. Al finalizar la comparación, el modelo permite identificar los rostros más parecidos en la base de datos, es decir, aquellos con menor distancia coseno. Sin embargo, para realizar una clasificación precisa y evitar falsos positivos, se determina un umbral mínimo para determinar si un rostro pertenece a una identidad de la base de datos. El umbral determinado para las condiciones previamente definidas se encuentra en el rango de 0.1 a 0.3, dependiendo de las condiciones lumínicas y la precisión que se le quiera dar al modelo. Finalmente, si un rostro está clasificado por encima del umbral, significa que es un desconocido; mientras que, si uno o más rostros están por debajo del umbral, la identidad asignada al rostro será aquella con menor distancia coseno. De esta manera, y mediante un subproceso adicional, se obtiene el nombre de la identidad o, en caso contrario, se asigna "Desconocido", el cual se dibuja en la parte superior izquierda del recuadro.

Etapa de observación

En la fase final, se unen los resultados de todo el sistema en la interfaz. El fotograma modificado por la etapa anterior se re-escala para ser visualizado en la interfaz. Además, esta última presenta una lista observable de los nombres de las personas almacenadas en la base de datos, la cual cambia de color si el sistema reconoce una

identidad registrada. Este diseño se concibió con el propósito de facilitar al usuario la observación de una persona específica en situaciones donde puedan detectarse numerosas personas, dificultando la distinción entre desconocidos y registrados. Finalmente, la interfaz posee la capacidad de detener el reconocimiento en cualquier momento, con el objetivo de evitar un gasto computacional excesivo en situaciones donde no sea completamente necesario llevar a cabo el reconocimiento.

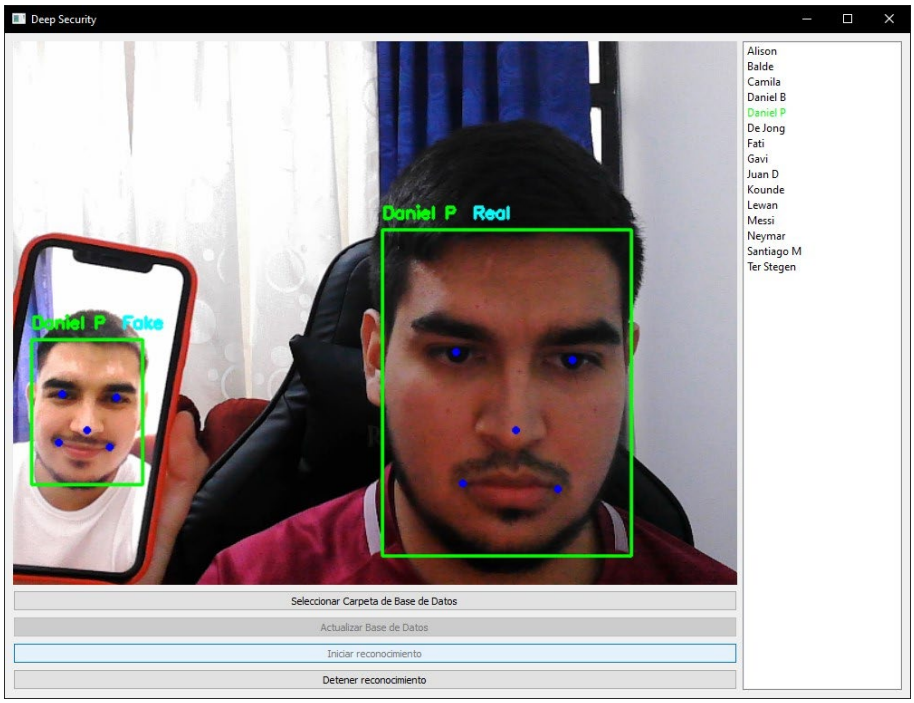


Figura 6 Interfaz del sistema

4.2 Trabajo computacional

El sistema de seguridad basado en reconocimiento facial desarrollado implica el uso de modelos matemáticos claves basados en redes neuronales profundas, como el clasificador VGGFace para generar representaciones vectoriales, el detector de rostros MTCNN para ubicar rostros, y el detector de deepfakes MesoNet para evaluar la autenticidad. Estos se eligieron debido a que, individualmente, cumplen con las especificaciones definidas, y presentan un gran desempeño computacionalmente logrando grandiosos resultados en múltiples situaciones. Además, la compatibilidad con Python fue un factor determinante a la hora de elegir dichos modelos. Por otro lado, los resultados se presentan de una forma gráfica a través de una interfaz que muestra el fotograma modificado, en donde se observará la salida de cada uno de los modelos como el recuadro donde se encuentran los rostros, la identidad de cada uno

y si este es real o no. Finalmente, las condiciones y dominio de validez del sistema incluyen la actualización periódica de la base de datos, la adaptabilidad del umbral de detección de deepfakes según la iluminación y distancia, y la capacidad de ajustar el umbral de clasificación de rostros para garantizar una precisión óptima y evitar falsos positivos. Un aspecto importante que es necesario recalcar y no mencionado anteriormente es la dependencia de una GPU o tarjeta gráfica, así como la necesidad de un ambiente con buena iluminación, las cuales, como se verá en secciones posteriores, permitirá que el sistema funcione de una forma correcta y eficiente.

5 VALIDACIÓN DEL TRABAJO

5.1 Metodología de prueba

Para evaluar el correcto funcionamiento del sistema, se efectuaron dos tipos de pruebas. El primero de ellos consistió en pruebas de validación de modelo, las cuales se realizaron posteriormente a la implementación de cada etapa individual. En otras palabras, cada modelo, después de su implementación, fue sometido a una evaluación de precisión y rendimiento computacional considerando tiempos de procesamiento, contrastando el uso de una GPU, para comprobar su correcto funcionamiento. Dichas pruebas se realizaron haciendo uso de una tarjeta gráfica NVIDIA RTX 2060 de 6GB, e imágenes capturadas de uso propio, y de 3 bases de datos, LFW (modelo de clasificación), Wider Face (modelo de detección y clasificación) [21], y Face2Face (modelo de deepfakes) [22]. El otro tipo de pruebas son las pruebas conjuntas o pruebas finales, las cuales se llevaron a cabo con el sistema completo (todos los modelos simultáneamente). Estas pruebas tuvieron como objetivo verificar el correcto funcionamiento del sistema e identificar tasas de error y acierto, teniendo en cuenta métricas como la precisión o la tasa de falsos positivos. Es importante mencionar que las pruebas finales se ejecutaron en diferentes contextos computacionales y de visión. Finalmente, la base de datos creada para aquellas pruebas en las que se evalué clasificación de rostros cuenta con un aproximado de 40 identidades.

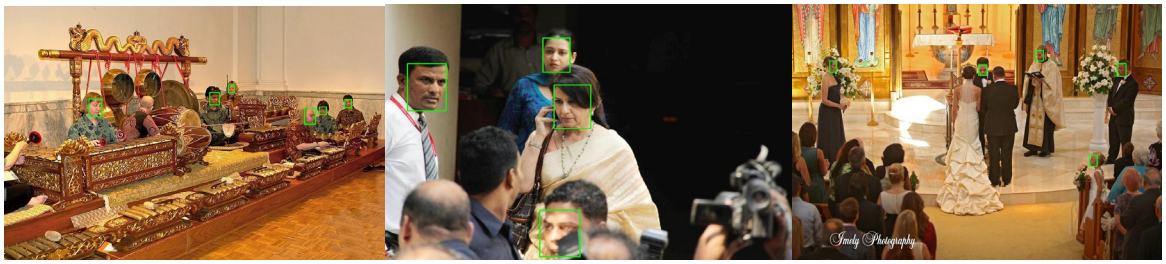
5.2 Validación de los resultados del trabajo

5.2.1 Pruebas de validación de modelo

Estas pruebas se dividieron por cada etapa del sistema implementado. Cabe resaltar, que, a pesar de haber mencionado la individualidad de cada modelo en su prueba, los modelos de detección de deepfakes, y clasificación dependen de la etapa de detección de rostros por lo cual, se considerará en la prueba de precisión únicamente los resultados medidos por el modelo, mientras que en la prueba de

rendimiento en tiempo real se evaluará el tiempo conjunto con la etapa de detección.

Etapa de detección de rostros



Figuras 7-8-9 Imágenes de prueba modelo clasificación de rostros

# Prueba	Número de rostros por imagen	Precisión
0	10	60%
1	6	100%
2	5	100%
3	4	100%
4	3	100%
5	5	100%
6	17	94%
7	15	93%
8	1	100%
9	7	100%
10	16	100%
Resultados Promedio		95%

Tabla 1 Resultados prueba con imágenes Wider Face

# *Prueba*	Número de rostros por toma	Tiempo de procesamiento (ms) Con GPU	Sin GPU
0	10	78	312
1	14	109	436
2	8	62	249
3	1	8	31
4	9	70	280
5	6	47	187
6	13	101	405
7	4	31	125
8	12	93	374
9	11	86	343
10	15	117	467
Resultados Promedio		73	292

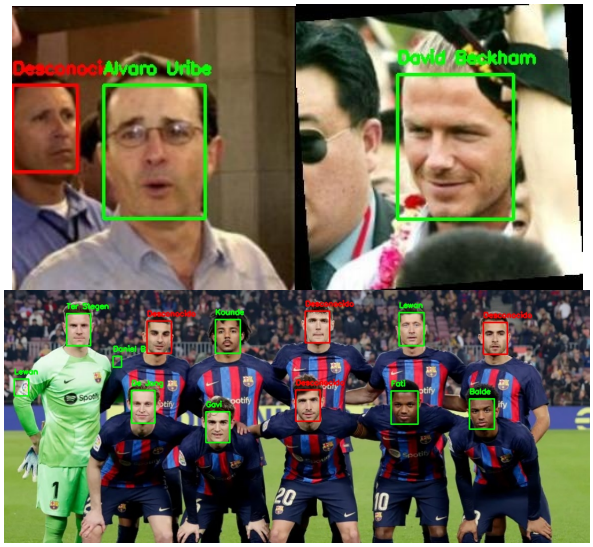
Tabla 2 Resultados prueba en tiempo real modelo de detección de rostros

Los resultados obtenidos por las pruebas de precisión y rendimiento para el modelo de detección de rostros MTCNN expuestos en las Tablas 1 y 2 respectivamente son satisfactorios, ya que comprobamos que el modelo funciona correctamente, y de una forma precisa, llegando a obtener en 10 pruebas una precisión promedio del 95% en imágenes que cuentan con menos de 20 rostros a la vez. Así mismo, el rendimiento del modelo en su implementación en tiempo real es el esperado, obteniendo un mejor rendimiento con el uso dedicado de una GPU, presentando tiempos de respuesta bajos. Además, en las Figuras 6, 7, 8 y 9 se

observa que los recortes de cada rostro se hacen de una forma coherente. Finalmente, se observa que el tiempo de procesamiento es proporcional al número de rostros detectados, es decir que aquellos fotogramas que cuentan con un mayor número de rostros observables en una imagen tardarán mayor tiempo en ser procesados.

Etapa de clasificación de rostros

Con respecto a la etapa de clasificación, el modelo VGGFace cuenta con una precisión promedio aproximada del 98% en las pruebas realizadas (Tabla 3), por lo que su funcionamiento es correcto. Por otro lado, el tiempo de procesamiento incrementó con respecto al rendimiento del modelo de detección de rostros un 40% sin usar GPU, esto debido a que se están realizando dos procesos, el de la etapa de detección de rostros y de clasificación ellos. Sin embargo, a pesar de este incremento observamos que al usar GPU el incremento se reduce a un 10% (Tabla 4), evidenciando una vez más la importancia de esta característica para un mejor rendimiento del sistema.



Figuras 10-11-12 Imágenes de prueba modelo clasificación de rostros

# Prueba	Número de rostros por imagen	Precisión
0	2	100%
1	14	92%
2	5	100%
3	11	100%
4	3	100%
5	1	100%
6	17	94%
7	15	93%
8	1	100%
9	7	100%
10	1	100%
	Resultados Promedio	98%

Tabla 3 Resultados prueba con imágenes Wider Face y LFW

# *Prueba*	Número de rostros por toma	Tiempo de procesamiento (ms) Con GPU	Sin GPU
0	12	116	699
1	1	10	58
2	15	145	873
3	2	19	116
4	6	58	349
5	15	145	873
6	8	77	466
7	14	135	815
8	11	106	640
9	4	39	233
10	5	48	291
	Resultados Promedio	82	492

Tabla 4 Resultados prueba en tiempo real modelo de clasificación de rostros

Etapa de detección de deepfakes

Figuras 13-14-15-16 Imágenes de prueba modelo detección de deepfakes

# Prueba	Número de rostros por imagen	Precisión
0	1	100%
1	7	85%
2	8	100%
3	5	100%
4	1	100%
5	2	100%
6	4	100%
7	9	93%
8	1	100%
9	10	100%
10	12	83%
Resultados Promedio		96%

Tabla 5 Resultados prueba con imágenes Face2Face

# *Prueba*	Número de rostros por toma	Tiempo de procesamiento (ms) Con GPU	Sin GPU
0	10	78	503
1	2	16	101
2	11	86	553
3	8	62	402
4	5	39	252
5	7	54	352
6	1	8	50
7	15	117	755
8	3	23	151
9	2	16	101
10	9	70	453
Resultados Promedio		52	334

Tabla 6 Resultados prueba en tiempo real modelo de detección de deepfakes

Finalmente, el rendimiento observado en las dos pruebas realizadas por el modelo de detección de rostros artificiales MesoNet en las Tablas 5 y 6 se puede concluir que su rendimiento es óptimo. Fue interesante observar la reducción de tiempo de preprocesamiento al usar GPU con respecto a la primera etapa. Esto sucede debido a que las pruebas fueron realizadas con Face2Face una base de datos especial para detectar deepfakes, la cual cuenta con videos en los que el número de personas que aparecen o son observables es bajo, por lo que la etapa de detección de rostros reduce su tiempo de procesamiento. Además, el modelo MesoNet no cuenta con una arquitectura tan compleja o profunda, por lo que esto también contribuirá a tener tiempos de procesamiento bajos. Finalmente, su precisión promedio es muy alta, por lo que podemos afirmar que la elección de este modelo fue la correcta.

5.2.2 Pruebas finales

Para probar el sistema completo se realizaron pruebas en contextos espaciales diferentes, tales como distancia entre la cámara y las personas identificadas, calidad de imagen y algunas modificaciones de iluminación. A continuación, se presentan resultados obtenidos en el procesamiento de video haciendo uso del modelo completo. Cada prueba registrada consta de un video de 15 a 30 segundos en el que aparece un numero de personas promedio. Estos videos son recortes de

[23]. Los siguientes resultados se encuentran consignados en una tabla donde se mide la tasa de falsos positivos y negativos, su precisión y su rendimiento computacional entiempo de procesamiento por fotograma. Es necesario mencionar que la prueba 4 siempre presentará una modificación en su iluminación, empeorando las condiciones y agregándole dificultad al sistema.

En primera instancia, se realizó una prueba en videos de calidad baja (480p) en un rango de 1 a 6 metros de distancia entre los identificados y la cámara, donde ninguna persona de la base de datos apareciera en este, con el fin de evaluar el desempeño del sistema a la hora de identificar "Desconocidos" y medir taza de Falsos Positivos. Para ello, inicialmente se definieron los umbrales de clasificación y de deepfakes en 0.25 y 0.5.

					Tiempo de procesamiento por Frame (ms)	
# Prueba	Número de rostros	Falsos Negativos	Falsos Positivos	Precisión	Con GPU	Sin GPU
0	4	0%	50%	50%	35	235
1	2	0%	0%	100%	18	117
2	7	0%	0%	100%	62	411
3	11	0%	54%	46%	97	646
4	5	20%	20%	60%	44	294
5	10	0%	50%	50%	88	587
6	1	0%	0%	100%	9	59
7	15	0%	53%	47%	132	881
8	3	0%	0%	100%	26	176
9	2	0%	0%	100%	18	117
10	9	0%	0%	100%	79	528
Resultados Promedio		1,8%	20.6%	78%	55	368

Tabla 7 Resultados prueba con umbrales 0.25 y 0.5

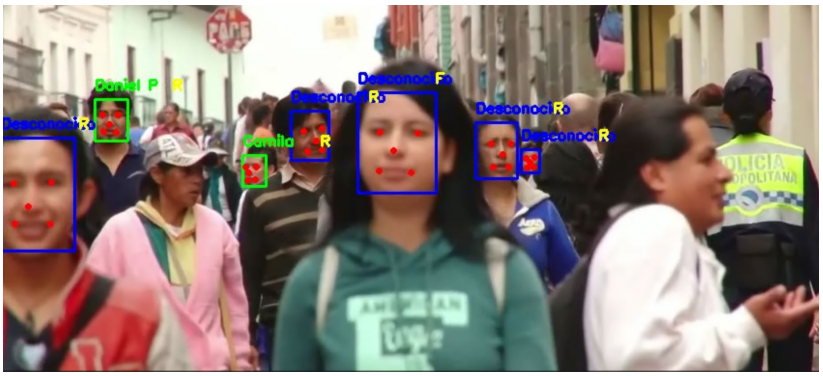


Figura 17 Captura de video umbrales 0.25 y 0.5

Como se puede observar en la Figura 17 y comprobar en la Tabla 7 los resultados no son satisfactorios, obteniendo una tasa muy alta de falsos positivos, y una relativamente baja precisión. Este problema se solucionara con un ajuste en los limites de los umbrales. Para ello se realizan otras pruebas preeliminares que consisten en procesar videos similares diferentes al set de prueba con diferentes umbrales y elegir aquel con mejor precisión y menor tasa de falsos positivos y negativos. Una vez terminado el re-entrenamiento, se eligieron los valores de 0.15 y 0.23, con los cuales se probo en el mismo set de prueba, obteniendo mejores resultados, observados en la Figura 18 y la tabla 8. Dichos resultados evidencian un cambio significativo en la precisión y una disminución sustancial de los falsos positivos. Es importante recalcar que a pesar de no contar con la calidad adecuada (720p) el sistema funciona bien, exceptuando el caso de baja iluminación el cual por razones obvias no obtiene resultados optimos.

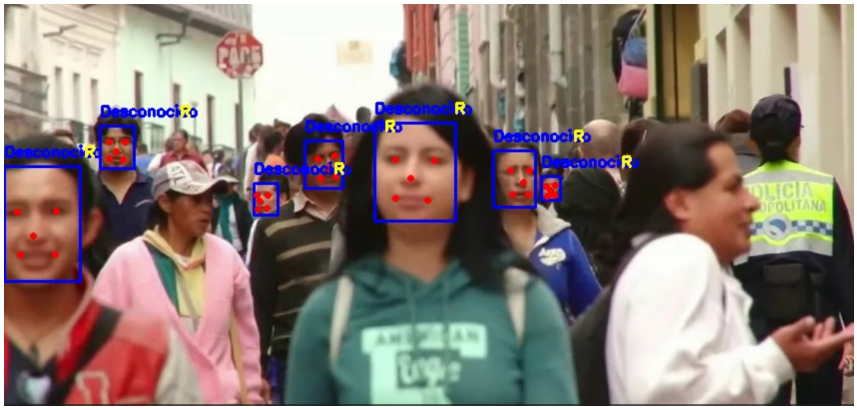


Figura 18 Captura de video umbrales 0.15 y 0.23

					Tiempo de procesamiento por Frame (ms)	
# Prueba	Número de rostros	Falsos Negativos	Falsos Positivos	Precisión	Con GPU	Sin GPU
0	4	0%	0%	100%	35	235
1	2	0%	0%	100%	18	117
2	7	0%	0%	100%	62	411
3	11	0%	0%	100%	97	646
4	5	20%	20%	60%	44	294
5	10	0%	0%	100%	88	587
6	1	0%	0%	100%	9	59
7	15	0%	7%	93%	132	881
8	3	0%	0%	100%	26	176
9	2	0%	0%	100%	18	117
10	9	0%	0%	100%	79	528
Resultados Promedio		1,8%	2,5%	96%	55	368

Tabla 8 Resultados prueba con umbrales 0.15 y 0.23

Figura 19 Prueba 4 mala iluminación

Una vez definidos los umbrales y comprobado el correcto funcionamiento, se realizaron pruebas en 3 contextos diferentes, definidos así:

- Contexto 1: Distancia de camara 1-6m, calidad 1080p, 40 personas en base de datos, con apariciones de personas en la base de datos a lo largo del video.
- Contexto 2: Distancia de camara 10-15m, calidad 720p, 40 personas en base de datos, con apariciones de personas en la base de datos a lo largo del video.
- Contexto 3: Caracteristicas similares a contexto 2, con apariciones de personas en la base de datos a lo largo del video, pero sin estar registradas en ella.

Los dos primeros contextos permitirán evaluar el desempeño en situaciones reales, con una base de datos grande donde deverá distinguir entre desconocidos y conocidos, mientras que el contexto 3 presenta una prueba con una dificultad mayor donde debera unicamente evitar falsos positivos y negativos.

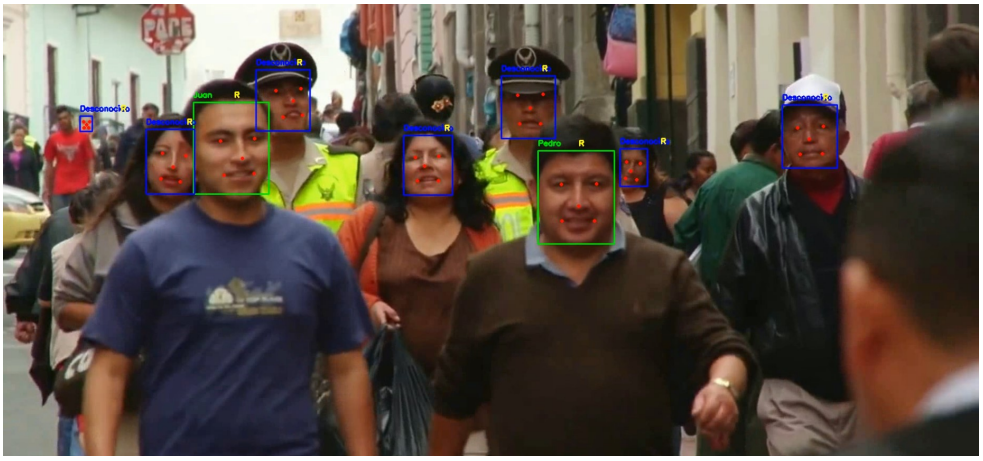


Figura 20 Captura de video contexto 1

					Tiempo de procesamiento por Frame (ms)	
# Prueba	Número de rostros	Falsos Negativos	Falsos Positivos	Precisión	Con GPU	Sin GPU
0	3	0%	0%	100%	34	193
1	11	0%	10%	90%	123	707
2	8	0%	0%	100%	90	514
3	6	0%	0%	100%	67	386
4	10	20%	0%	80%	112	643
5	4	0%	0%	100%	45	257
6	2	0%	0%	100%	22	129
7	15	0%	7%	93%	168	965
8	3	0%	0%	100%	34	193
9	5	0%	0%	100%	56	322
10	9	0%	12%	88%	101	579
Resultados Promedio		1,8%	2,6%	96%	77	444

Tabla 9 Resultados prueba contexto 1

Los resultados de las pruebas en el contexto 1 fueron satisfactorios, ya que sus valores en las métricas fueron similares a los de la calibración. Esto quiere decir que el modelo es robusto a diferentes situaciones de resolución del video. Por otro lado, existe un incremento del 25% en tiempo de procesamiento, con respecto a las pruebas anteriores, causado por el modelo de clasificación, debido la existencia de personas registradas en la base de datos. Finalmente, la prueba de iluminación en este contexto no ha cambiado, obteniendo baja precisión con respecto a las demás pruebas.

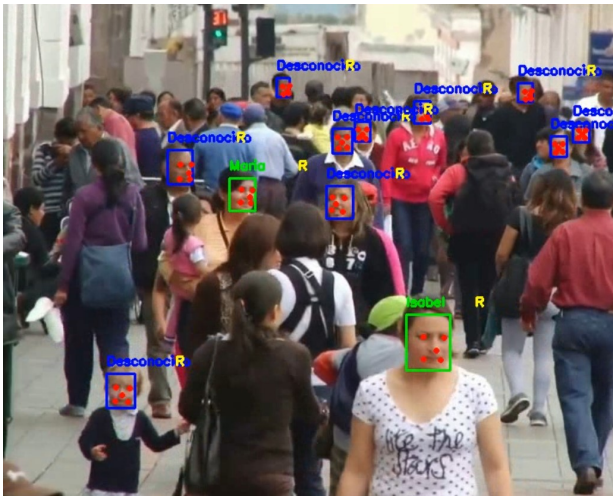


Figura 21 Captura de video contexto 2

					Tiempo de procesamiento por Frame (ms)	
# Prueba	Número de rostros	Falsos Negativos	Falsos Positivos	Precisión	Con GPU	Sin GPU
0	8	0%	0%	100%	77	498
1	14	0%	0%	100%	135	871
2	5	0%	0%	100%	48	311
3	11	0%	5%	90%	106	684
4	3	33%	33%	33%	29	187
5	5	0%	0%	100%	48	311
6	17	0%	0%	100%	164	1058
7	15	0%	7%	93%	145	933
8	1	0%	0%	100%	10	62
9	7	0%	0%	100%	67	435
10	16	0%	6%	94%	154	995
Resultados Promedio		3,0%	4,6%	92%	89	577

Tabla 10 Resultados prueba contexto 2

Los resultados observados en la Figura 21 y la tabla 10 reafirman el correcto funcionamiento del sistema. A pesar de que la precisión promedio de las pruebas disminuyo 4%, y el tiempo de procesamiento incremento, los resultados son apropiados debido a que este es un contexto poco habitual, con una gran distancia entre las personas y la cámara, y una relativamente baja resolución (720p). De esta forma, ratifica la robustez del sistema a cambios de distancia. Además, es apropiado mencionar el incremento de las tasas de falsos positivos y negativos, la cual es causa de la aparición de rostros fuera del margen de funcionamiento establecido (10-15m).

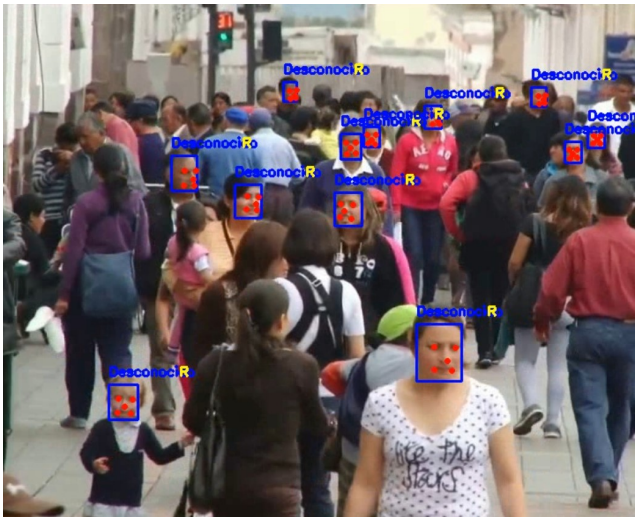


Figura 22 Captura de video contexto 3

					Tiempo de procesamiento por Frame (ms)	
# Prueba	Número de rostros	Falsos Negativos	Falsos Positivos	Precisión	Con GPU	Sin GPU
0	8	0%	0%	100%	71	474
1	14	8%	0%	92%	125	830
2	5	0%	0%	100%	45	297
3	11	0%	5%	90%	98	652
4	3	33%	0%	66%	27	178
5	5	0%	0%	100%	45	297
6	17	6%	0%	94%	151	1008
7	15	0%	7%	93%	134	890
8	1	0%	0%	100%	9	59
9	7	0%	0%	100%	62	415
10	16	6%	6%	88%	143	949
Resultados Promedio		4,8%	1,6%	93%	83	550

Tabla 11 Resultados contexto 3

Finalmente, las pruebas en el contexto 3 cuentan con resultados similares al contexto 2. En este punto es necesario mencionar la precisión de la unión de los tres modelos que conforman el sistema, y resaltar su eficiencia computacional gracias al uso de una GPU. Sin embargo, al finalizar todas las pruebas comprobamos que el sistema en situaciones de baja iluminación o iluminación extrema experimenta imprecisión, por lo que esta será una restricción al momento de este ser implementado.

5.3 Evaluación del plan de trabajo

La ejecución del plan de trabajo propuesto en la propuesta de tesis se llevó a cabo de manera eficiente, reflejando una planificación cuidadosa y la consecución exitosa de cada etapa del proyecto. La fase inicial de investigación sentó las bases teóricas con una revisión exhaustiva de técnicas de preprocesamiento de imágenes y modelos de deep learning. La configuración del entorno de trabajo, incluyendo la instalación de software y la integración de hardware, se completó en una semana, facilitando la transición a las fases de implementación.

La implementación del modelo MTCNN y la elección del modelo VGGFace transcurrieron según el cronograma, destacando una eficiencia en el desarrollo y pruebas específicas. La integración de las etapas de detección y clasificación se ejecutó en el tiempo planificado, demostrando una aplicación exitosa de transfer learning. La implementación de la etapa de seguridad con MesoNet, crucial para la detección de rostros artificiales, se extendió a 3 semanas, reflejando la dedicación necesaria para asegurar la eficacia y precisión del modelo. El desarrollo de la interfaz se realizó según lo planeado en 2 semanas, considerando el diseño amigable y funcional, así como la implementación de funciones esenciales.

Las pruebas finales se llevaron a cabo durante el período asignado de 2 semanas, evaluando el rendimiento del sistema en diversas condiciones y documentando los resultados obtenidos. La fase final de elaboración del informe se completó en 1 semana, permitiendo una compilación detallada de los resultados y la preparación para la presentación final del proyecto. En resumen, la evaluación del plan de trabajo refleja una ejecución coherente y eficiente, subrayando la importancia de una planificación cuidadosa y la flexibilidad en la duración de algunas etapas para garantizar la calidad y eficacia del sistema de reconocimiento facial desarrollado.

6 DISCUSIÓN

El trabajo realizado en esta investigación ha culminado con el desarrollo de un sistema de reconocimiento facial basado en técnicas avanzadas de machine learning, deep learning y transfer learning. El objetivo general de diseñar un sistema altamente eficiente y preciso centrado en el reconocimiento e identificación de rostros para entornos de seguridad en espacios con una población definida ha sido alcanzado en gran medida. Sin embargo, las pruebas en tiempo real también revelaron variaciones en el rendimiento según la presencia o ausencia de una GPU, lo que sugiere la importancia de considerar la infraestructura computacional al implementar el sistema. Esta restricción podría afectar la accesibilidad del sistema en entornos que no cuenten con esta capacidad de procesamiento. De la misma forma, las condiciones lumínicas

del entorno en el cual se ejecutará el sistema corren con gran importancia a la hora de obtener un resultado satisfactorio, de esta forma, se podría considerar una limitación que restringiría el número de espacios, o momentos en los que este puede funcionar apropiadamente.

En términos de los objetivos iniciales del proyecto, el sistema ha logrado cumplir con la mayoría de ellos de manera satisfactoria. La detección de rostros, la clasificación y la evaluación de la autenticidad han sido abordadas de manera eficiente. Sin embargo, a pesar de los logros alcanzados, hay áreas que pueden ser objeto de mejoras y trabajo futuro. La dependencia de una GPU podría mitigarse mediante optimizaciones y adaptaciones para el uso en entornos con recursos limitados. La automatización de la actualización de la base de datos y la mejora continua de los algoritmos de procesamiento de imágenes podrían considerarse para hacer el sistema más robusto y autónomo. Además, explorar técnicas de aumento de datos y el uso de arquitecturas más avanzadas podría elevar aún más la precisión del sistema.

En resumen, este trabajo ha logrado diseñar un sistema de reconocimiento facial avanzado que cumple con los objetivos establecidos. Sin embargo, las limitaciones identificadas subrayan la importancia de la adaptabilidad y la mejora continua para garantizar la eficacia del sistema en diversas condiciones y entornos. El análisis cualitativo y cuantitativo de los resultados proporciona una base sólida para futuras investigaciones y mejoras en el campo del reconocimiento facial.

7 CONCLUSIONES

En conclusión, el trabajo realizado ha culminado con el desarrollo exitoso de un sistema de reconocimiento facial basado en técnicas avanzadas de machine learning y deep learning. El sistema logra una detección precisa de rostros, clasificación efectiva y evaluación de la autenticidad, proporcionando una solución integral para entornos de seguridad. La implementación de modelos como VGGFace, MTCNN y MesoNet ha demostrado ser fundamental para el rendimiento del sistema, obteniendo resultados consistentes en diversas pruebas y contextos computacionales.

Por otro lado, este trabajo aporta contribuciones originales que abarcan varios aspectos significativos en el ámbito del reconocimiento facial. En primer lugar, se destaca el diseño e implementación exitosos de un sistema altamente eficiente y preciso, marcando un avance fundamental en la tecnología de reconocimiento facial. Además, la integración de técnicas avanzadas de machine learning y transfer learning ha permitido abordar de manera efectiva la detección de rostros, clasificación e identificación de deepfakes, aspecto crucial para mejorar la autenticidad del sistema.

Asimismo, se ha logrado un hito en la usabilidad del sistema mediante el desarrollo de una interfaz visual intuitiva, facilitando la observación y verificación de resultados de manera accesible. Por último, el trabajo se distingue por la exhaustiva evaluación del desempeño del sistema, llevada a cabo a través de pruebas de validación de modelos y pruebas finales en tiempo real, asegurando la robustez y confiabilidad del sistema en diversas situaciones y entornos. Estas contribuciones conjuntas consolidan este trabajo como una referencia valiosa en el campo del reconocimiento facial avanzado.

En términos de consecuencias globales, el sistema puede tener un impacto positivo en la mejora de la seguridad en espacios públicos y privados. Económicamente, la implementación del sistema podría traducirse en ahorros significativos en comparación con soluciones de seguridad tradicionales. Asimismo, el proyecto ha implicado costos asociados con el desarrollo de software, la adquisición de recursos computacionales y la realización de pruebas. Sin embargo, en comparación con sistemas de seguridad convencionales, la practicidad y eficiencia del sistema de reconocimiento facial podrían resultar en una inversión más rentable a largo plazo. La adaptabilidad del sistema a diferentes entornos y su capacidad para mejorar la seguridad podrían hacerlo atractivo en el mercado de la ingeniería de seguridad, por lo que se sugiere una evaluación continua de costos y beneficios para garantizar la sostenibilidad y la practicidad del sistema en el mercado.

8 AGRADECIMIENTOS

En primer lugar, quiero expresar mi profunda gratitud a Dios, fuente inagotable de fortaleza y guía a lo largo de este arduo pero gratificante camino académico.

A mis padres, Carlos Ramiro Pantoja Paredes y Edilma Magaly Bernal, les debo todo. Su amor incondicional, apoyo constante y sacrificios han sido la fuerza impulsora detrás de cada logro. Su sabiduría y aliento han sido mi ancla en las tormentas y mi faro en la oscuridad. Este logro también es suyo.

Mi familia y amigos, los cuales han sido mi red de seguridad emocional. Su paciencia, comprensión y aliento me han sostenido en momentos difíciles. Agradezco la alegría que compartimos y el respaldo inquebrantable que me brindaron en cada paso de este viaje.

Mi profundo agradecimiento se extiende hacia mi asesor de tesis, Fernando Enrique Lozano Martínez, cuya orientación experta y apoyo constante fueron fundamentales en la realización de este trabajo. Sus conocimientos, paciencia y dedicación a la excelencia académica han dejado una huella imborrable en mi formación.

A mis queridos compañeros de carrera, agradezco la camaradería, los intercambios de ideas y el apoyo mutuo. Juntos, enfrentamos desafíos y celebramos éxitos, creando recuerdos que atesoraré siempre.

A mis compañeros de seminario, les agradezco por enriquecer nuestras discusiones académicas y por el estímulo constante. Sus perspectivas únicas han contribuido significativamente a mi crecimiento intelectual.

En conjunto, cada persona mencionada ha dejado una marca indeleble en mi trayectoria académica y personal. Este logro es el resultado de un esfuerzo colectivo y del amor, apoyo y contribuciones de aquellos que me rodean.

9 REFERENCIAS

[1] D. E. Rumelhart, G. E. Hinton, y R. J. Williams, "Learning representations by back-propagating errors," Nature, vol. 323, pp. 533-536, 1986. [En línea]. Disponible en: https://api.semanticscholar.org/CorpusID:205001834

[2] K. He, X. Zhang, S. Ren y J. Sun, "Deep residual learning for image recognition," en Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770-778, 2016.

[3] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser y I. Polosukhin, "Attention is all you need," en Advances in Neural Information Processing Systems, vol. 30, 2017.

[4] 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), Boston, MA, USA, 2015, pp. 815-823, doi: 10.1109/CVPR.2015.7298682.

[5] Y. Taigman, M. Yang, M. 'A. Ranzato y L. Wolf, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification," en Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1701-1708, 2014.

[6] S. J. Pan y Q. Yang, "A survey on transfer learning," IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 10, pp. 1345-1359, 2009.

[7] F. Yu, X. Xiu, and Y. Li, “A Survey on Deep Transfer Learning and Beyond,” Mathematics, vol. 10, no. 19, p. 3619, Oct. 2022, doi: 10.3390/math10193619.

[8] J. Yosinski, J. Clune, Y. Bengio y H. Lipson, "How transferable are features in deep neural networks?" en Advances in Neural Information Processing Systems, vol. 27, 2014.

[9] J. Howard y S. Ruder, "Universal language model fine-tuning for text classification," arXiv preprint arXiv:1801.06146, 2018.

[10] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, vol. 1, pp. I-511-I-518 vol.1, 2001.

[11] J. Xiang y G. Zhu, "Joint face detection and facial expression recognition with MTCNN," en 2017 4th International Conference on Information Science and Control Engineering (ICISCE), pp. 424-427, 2017. IEEE.

[12] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu y A. C. Berg, "SSD: Single Shot Multibox Detector," en Computer Vision--ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11--14, 2016, Proceedings, Part I, pp. 21-37, 2016. Springer.

[13] O. Parkhi, A. Vedaldi y A. Zisserman, "Deep face recognition," en BMVC 2015 - Proceedings of the British Machine Vision Conference 2015, British Machine Vision Association, 2015.

[14] G. B. Huang, M. Ramesh, T. Berg y E. Learned-Miller, "Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments," Informe Técnico, University of Massachusetts, Amherst, No. 07-49, octubre de 2007.

[15] Lior Wolf, Tal Hassner and Itay Maoz Face Recognition in Unconstrained Videos with Matched Background Similarity. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2011.

[16] D. Afchar, V. Nozick, J. Yamagishi e I. Echizen, "Mesonet: A Compact Facial Video Forgery Detection Network," en 2018 IEEE International Workshop on Information Forensics and Security (WIFS), pp. 1-7, 2018. IEEE.

[17] Chaves Fonseca, J. (2019). Diseño de un sistema de apoyo en seguridad basado en reconocimiento de múltiples rostros e identificación de usuarios. Universidad de los Andes.

[18] M. T. H. Fuad et al., "Recent Advances in Deep Learning Techniques for Face Recognition," in IEEE Access, vol. 9, pp. 99112-99142, 2021, doi: 10.1109/ACCESS.2021.3096136.

[19] Okeke, F. (2022). Facial recognition: Top software vendors. TechRepublic. https://www.techrepublic.com/article/facial-recognition-software/

[20] J. D. Pantoja Bernal, “DeepSecurity.zip,” Google Drive, https://drive.google.com/file/d/13HDNk1jxrJBtxPgicBWq8lPfbUcgtRVb/view?usp=drive_link (accessed Nov. 22, 2023).

[21] S. Yang, P. Luo, C. C. Loy y X. Tang, "WIDER FACE: A Face Detection Benchmark," en IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[22] Saahil Sood, «Face2Face Video Dataset». Zenodo, abr. 10, 2022. doi: 10.5281/zenodo.6430891.

[23] Productores Independientes, “Gente Caminando 2,” YouTube, https://www.youtube.com/watch?v=g0FmqhivlMg&t=15s&pp=ygUPZ2VudGUgY2FtaW5hbmRv (accessed Nov. 22, 2023).

10 APENDICES

Propuesta inicial del proyecto.

UNIVERSIDAD DE LOS ANDES
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA Y ELECTRÓNICA

PRESENTACIÓN DE PROPUESTA DE PROYECTO DE GRADO

SEMESTRE: 2023-10
FECHA: 24/05/2023

PROYECTO O TESIS DE GRADO PARA OPTAR EL TÍTULO DE:
INGENIERO ELECTRÓNICO
ESTUDIANTE: JOSÉ DANIEL PANTOJA BERNAL CÓDIGO: 202014653

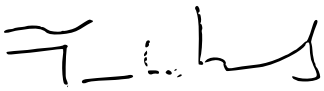
TÍTULO DE LA TESIS O PROYECTO:

SISTEMA DE SEGURIDAD A PARTIR DE UNA BASE DE DATOS MEDIANTE MACHINE LEARNING, BASADO EN RECONOCIMIENTO DE ROSTROS

DECLARACIÓN:
Soy consciente que cualquier tipo de fraude en esta Tesis es considerado como una falta grave en la Universidad. Al firmar, entregar y presentar esta propuesta de Tesis o Proyecto de Grado, doy expreso testimonio de que esta propuesta fue desarrollada de acuerdo con las normas establecidas por la Universidad. Del mismo modo, aseguro que no participé en ningún tipo de fraude y que en el trabajo se expresan debidamente los conceptos o ideas que son tomadas de otras fuentes.

Soy consciente de que el trabajo que realizaré incluirá ideas y conceptos del autor y el Asesor y podrá incluir material de cursos o trabajos anteriores realizados en la Universidad y por lo tanto, daré el crédito correspondiente y utilizaré este material de acuerdo con las normas de derechos de autor. Así mismo, no haré publicaciones, informes, artículos o presentaciones en congresos, seminarios o conferencias sin la revisión o autorización expresa del Asesor, quien representará en este caso a la Universidad.

Firma: José Daniel Pantoja Bernal
Código: 202014653
CC: 1.004.770.997



Vo.Bo. ASESOR (Firma)
Nombre: Fernando Enrique Lozano Martínez

1. JUSTIFICACIÓN Y DESCRIPCIÓN

Con el fin de adquirir y profundizar conocimientos en el área de inteligencia artificial, aplicar conocimientos obtenidos en el pregrado del programa de ingeniería electrónica y ayudar con el desarrollo de nuevas tecnologías que brinden una mayor seguridad a la sociedad, se propone este proyecto cuyo objetivo es diseñar un sistema de reconocimiento personal a partir de una base de datos mediante Machine Learning. El sistema utilizará técnicas de procesamiento de imágenes, las cuales, mediante aprendizaje automático y aprendizaje transferido ayudarán a diseñar e implementar un sistema basado en reconocimiento de rostros mediante una red neuronal profunda, la cual a través de redes neuronales convolucionales (CNN) permitirá extraer características particulares de cada rostro, las cuales por medio de una capa de clasificación serán asociadas a una identidad personal, la cual se almacenará previamente en una base de datos.

2. MARCO TEÓRICO

Antecedentes Externos:

En el estado del arte actual en el campo de reconocimiento facial, se ha demostrado que las técnicas de aprendizaje profundo han superado significativamente a los métodos tradicionales. Algunos ejemplos son la red neuronal convolucional (CNN), que se ha utilizado para detectar y reconocer caras en imágenes, y las redes adversarias generativas (GAN), que se han utilizado para sintetizar imágenes de caras realistas. También se han desarrollado técnicas de codificación de características como la red neuronal de codificación de caras (FaceNet), que permite la identificación precisa de personas en imágenes. Además, se han desarrollado redes de atención que permiten a los sistemas de reconocimiento facial enfocarse en las partes importantes de una imagen y, por lo tanto, mejorar la precisión del reconocimiento facial. Sin embargo, todavía existen desafíos importantes que deben superarse, como la variabilidad en la iluminación, la expresión facial y el envejecimiento, para que los sistemas de reconocimiento facial sean más robustos y precisos en situaciones reales [1].

Por otro lado, con respecto a sistemas de seguridad basados en reconocimiento personal, empresas como Amazon, Paravisión, Cognitec, entre otros, se han empeñado en diseñar sistemas que permiten identificar características personales mediante detección de rostros, tales como género y edad. Además, han diseñado sistemas en tiempo real que permiten detectar suplantaciones de identidad o número de apariciones en un sitio especifico [2].

Antecedentes locales:

En el año 2019 el trabajo desarrollado por Juan Felipe Chávez Fonseca en su proyecto de grado(https://repositorio.uniandes.edu.co/flexpaper/handle/1992/44747/u830933.pdf?sequence=1&isAllowed=y#page=1), tuvo como objetivo la implementación de un sistema de identificación personal mediante CNN, el cual permitiera brindar apoyo en actividades de seguridad. Sin embargo, el objetivo no se cumplió totalmente, ya que únicamente se implementó el sistema de identificación de rostros, el cual realizaba su función con más de 99% de precisión, al ser testeada en la base de datos LFW, pero este no realizaba la función de clasificar rostros dada una base de datos [3].

3. CARACTERIZACIÓN DEL PROYECTO

Objetivos Principales:

- Diseñar un sistema de reconocimiento personal, basado en reconocimiento e identificación de rostros, eficiente y preciso mediante el uso de técnicas de aprendizaje automático y transferido.
- Implementar un algoritmo de procesamiento de imágenes para preprocesar las imágenes y videos que se utilizarán en la base de datos.
- Evaluar el desempeño del sistema de reconocimiento personal utilizando diferentes métricas de evaluación.

Objetivos Específicos:

- Investigar y seleccionar los algoritmos de aprendizaje automático más adecuados para el problema de reconocimiento personal.
- Diseñar una base de datos de imágenes y videos para entrenar y evaluar el sistema de reconocimiento personal.
- Implementar técnicas de preprocesamiento de imágenes, como la eliminación de ruido y la normalización de la iluminación, para mejorar la calidad de las imágenes y videos.
- Implementar, mediante aprendizaje automático y transferido, una red neuronal que permita reconocer rostros.
- Diseñar un clasificador que permita reconocer identidades a partir de una base de datos
- Implementar el sistema de reconocimiento personal diseñado.
- Entrenar el sistema de reconocimiento personal utilizando la base de datos diseñada.
- Evaluar el rendimiento del sistema utilizando diferentes métricas de evaluación, como la precisión y la tasa de falsos positivos.
- Analizar los resultados obtenidos y proponer mejoras para el sistema de reconocimiento personal.

Alcances:

- El sistema se ejecutará en una aplicación de computador diseñada para su fácil uso.
- El sistema estará optimizado para funcionar en cámaras de resolución mayor a 720p, 1MP.
- Se espera que el sistema pueda ser usado en computadores con características de procesamiento relativamente bajas.

4. CONTEXTO DEL PROYECTO Y TRATAMIENTOS

Metodología:

- Se llevará a cabo una revisión exhaustiva de la literatura científica relacionada con técnicas de reconocimiento personal mediante aprendizaje automático y transferido.
- Se seleccionará una base de datos de imágenes y videos adecuada para el entrenamiento y evaluación del sistema.
- Se aplicarán técnicas de preprocesamiento de imágenes para mejorar la calidad de las imágenes y videos, y así permitir el uso de estos en el sistema.
- Se elegirá la arquitectura de la red neuronal a usar para el reconocimiento de rostros.
- Se diseñará el clasificador, que permitirá clasificar los rostros teniendo en cuenta una base de datos dada.

- Se entrenará el sistema de reconocimiento personal mediante técnicas de Machine Learning utilizando la base de datos seleccionada.
- Se evaluará el rendimiento del sistema utilizando diferentes métricas de evaluación.
- Se analizarán los resultados obtenidos y se propondrán mejoras para el sistema de reconocimiento personal.

Compromisos:

- El sistema desarrollado será eficiente y preciso.
- Las imágenes o videos que se usarán para el entrenamiento o prueba del sistema serán exentos de derechos de autor o de autoría propia.
- Cada fuente utilizada para el desarrollo del proyecto será correctamente referenciada.
- Se cumplirán los horarios consensuados con el Asesor para la revisión del desarrollo del proyecto.

Posibles restricciones en el desarrollo del proyecto:

- El número de datos de entrenamiento o prueba del sistema pueden ser limitados, por lo que la elección de estos debe ser adecuada.
- De igual forma, los recursos computacionales son limitados, por lo que se deberá tener en cuenta factores como tiempos de entrenamiento del sistema.
- El tiempo de desarrollo del proyecto se espera que sea aproximadamente de 4 meses, tiempo en el cual se espera finalizar este proyecto.

5. CRONOGRAMA

Semana 1-2:

- Revisión bibliográfica y definición del problema.
- Selección de la base de datos adecuada para el entrenamiento y evaluación del sistema.

Semana 3-4:

- Preprocesamiento de imágenes y videos: eliminación de ruido, normalización de iluminación y otras técnicas de mejora de calidad de las imágenes y videos.

Semana 5-8:

- Implementación de algoritmos de aprendizaje automático y transferido para el entrenamiento del sistema de reconocimiento personal.
- Implementación del clasificador mediante una base de datos.

Semana 9-10:

- Evaluación del sistema utilizando diferentes métricas de evaluación, como la precisión y la tasa de falsos positivos.
- Análisis de resultados preliminares y propuestas de mejoras al sistema.

Semana 11-12:

- Implementación de mejoras al sistema de reconocimiento personal basadas en los análisis de los resultados preliminares.

Semana 13-14:

- Evaluación final del sistema mejorado utilizando la base de datos completa.
- Análisis de resultados y discusión de conclusiones.
-

Semana 15-16:

- Preparación del informe final sobre el proyecto.
- Revisión y corrección del informe final.

6. BIBLIOGRAFÍA

[1] M. T. H. Fuad et al., "Recent Advances in Deep Learning Techniques for Face Recognition," in IEEE Access, vol. 9, pp. 99112-99142, 2021, doi: 10.1109/ACCESS.2021.3096136.

[2] Okeke, F. (2022). Facial recognition: Top software vendors. TechRepublic. htps://www.techrepublic.com/article/facial-recognition-software/

[3] Chaves Fonseca, J. (2019). Diseño de un sistema de apoyo en seguridad basado en reconocimiento de múltiples rostros e identificación de usuarios. Universidad de los Andes.

7. ANEXOS

a. Formato resumen de Solicitud Proyecto Especial
b. Formato de sesión de derechos de Autor (autorización de uso a nombre de la Universidad de los Andes).

Universidad de los Andes
Facultad de Ingeniería

FORMATO DE SOLICITUD

CURSO TUTORIAL*	
PROYECTO ESPECIAL**	**X**

* Curso de naturaleza teórica dirigido por un profesor tutor. Válido como materia del área de profundización.
** Curso de naturaleza práctica dirigida por un profesor asesor. Válido como materia del área de profundización del estudiante o del área de vinculación del profesor asesor.

Estos cursos deben solicitarse ante un interés particular del estudiante o asesor, en temas para los cuales no existe oferta similar en la planeación de cursos por parte del departamento.

Nota: la calificación de este curso es cuantitativa.

Estudiante: José Daniel Pantoja Bernal Código: 202014653
Programa: Ingeniería Electrónica Semestre: 7
Email: j.pantojab@uniandes.edu.co

Título del Curso: Sistema de seguridad a partir de una base de datos mediante Machine Learning, basado en reconocimiento de rostros.

1. Descripción

   Con el fin de adquirir y profundizar conocimientos en el área de inteligencia artificial, aplicar conocimientos obtenidos en el pregrado del programa de ingeniería electrónica y ayudar con el desarrollo de nuevas tecnologías que brinden una mayor seguridad a la sociedad, se propone este proyecto cuyo objetivo es diseñar un sistema de reconocimiento personal a partir de una base de datos mediante Machine Learning. El sistema utilizará técnicas de procesamiento de imágenes, las cuales, mediante aprendizaje automático y aprendizaje transferido ayudarán a diseñar e implementar un sistema basado en reconocimiento de rostros mediante una red neuronal profunda, la cual a través de redes neuronales convolucionales (CNN) permitirá extraer características particulares de cada rostro, las cuales por medio de una capa de clasificación serán asociadas a una identidad personal, la cual se almacenará previamente en una base de datos.

2. Marco Teórico

   Antecedentes Externos:

   En el estado del arte actual en el campo de reconocimiento facial, se ha demostrado que las técnicas de aprendizaje profundo han superado significativamente a los métodos tradicionales.

Algunos ejemplos son la red neuronal convolucional (CNN), que se ha utilizado para detectar y reconocer caras en imágenes, y las redes adversarias generativas (GAN), que se han utilizado para sintetizar imágenes de caras realistas. También se han desarrollado técnicas de codificación de características como la red neuronal de codificación de caras (FaceNet), que permite la identificación precisa de personas en imágenes. Además, se han desarrollado redes de atención que permiten a los sistemas de reconocimiento facial enfocarse en las partes importantes de una imagen y, por lo tanto, mejorar la precisión del reconocimiento facial. Sin embargo, todavía existen desafíos importantes que deben superarse, como la variabilidad en la iluminación, la expresión facial y el envejecimiento, para que los sistemas de reconocimiento facial sean más robustos y precisos en situaciones reales [1].

Por otro lado, con respecto a sistemas de seguridad basados en reconocimiento personal, empresas como Amazon, Paravisión, Cognitec, entre otros, se han empeñado en diseñar sistemas que permiten identificar características personales mediante detección de rostros, tales como género y edad. Además, han diseñado sistemas en tiempo real que permiten detectar suplantaciones de identidad o número de apariciones en un sitio especifico [2].

Antecedentes locales:

En el año 2019 el trabajo desarrollado por Juan Felipe Chávez Fonseca en su proyecto de grado (https://repositorio.uniandes.edu.co/flexpaper/handle/1992/44747/u830933.pdf?sequence=1&isAllowed=y#page=1), tuvo como objetivo la implementación de un sistema de identificación personal mediante CNN, el cual permitiera brindar apoyo en actividades de seguridad. Sin embargo, el objetivo no se cumplió totalmente, ya que únicamente se implementó el sistema de identificación de rostros, el cual realizaba su función con más de 99% de precisión, al ser testeada en la base de datos LFW, pero este no realizaba la función de clasificar rostros dada una base de datos [3].

3. Objetivos

Objetivos Principales:

- Diseñar un sistema de reconocimiento personal, basado en reconocimiento e identificación de rostros, eficiente y preciso mediante el uso de técnicas de aprendizaje automático y transferido.
- Implementar un algoritmo de procesamiento de imágenes para preprocesar las imágenes y videos que se utilizarán en la base de datos.
- Evaluar el desempeño del sistema de reconocimiento personal utilizando diferentes métricas de evaluación.

Objetivos Específicos:

- Investigar y seleccionar los algoritmos de aprendizaje automático más adecuados para el problema de reconocimiento personal.
- Diseñar una base de datos de imágenes y videos para entrenar y evaluar el sistema de reconocimiento personal.
- Implementar técnicas de preprocesamiento de imágenes, como la eliminación de ruido y la normalización de la iluminación, para mejorar la calidad de las imágenes y videos.
- Implementar, mediante aprendizaje automático y transferido, una red neuronal que permita reconocer rostros.
- Diseñar un clasificador que permita reconocer identidades a partir de una base de datos
- Implementar el sistema de reconocimiento personal diseñado.
- Entrenar el sistema de reconocimiento personal utilizando la base de datos diseñada.
- Evaluar el rendimiento del sistema utilizando diferentes métricas de evaluación, como la precisión y la tasa de falsos positivos.
- Analizar los resultados obtenidos y proponer mejoras para el sistema de reconocimiento personal.

Alcances:

- El sistema se ejecutará en una aplicación de computador diseñada para su fácil uso.
- El sistema estará optimizado para funcionar en cámaras de resolución mayor a 720p, 1MP.
- Se espera que el sistema pueda ser usado en computadores con características de procesamiento relativamente bajas.

4. Observaciones

Metodología:

- Se llevará a cabo una revisión exhaustiva de la literatura científica relacionada con técnicas de reconocimiento personal mediante aprendizaje automático y transferido.
- Se seleccionará una base de datos de imágenes y videos adecuada para el entrenamiento y evaluación del sistema.
- Se aplicarán técnicas de preprocesamiento de imágenes para mejorar la calidad de las imágenes y videos, y así permitir el uso de estos en el sistema.
- Se elegirá la arquitectura de la red neuronal a usar para el reconocimiento de rostros.

- Se diseñará el clasificador, que permitirá clasificar los rostros teniendo en cuenta una base de datos dada.
- Se entrenará el sistema de reconocimiento personal mediante técnicas de Machine Learning utilizando la base de datos seleccionada.
- Se evaluará el rendimiento del sistema utilizando diferentes métricas de evaluación.
- Se analizarán los resultados obtenidos y se propondrán mejoras para el sistema de reconocimiento personal.

Compromisos:

- El sistema desarrollado será eficiente y preciso.
- Las imágenes o videos que se usarán para el entrenamiento o prueba del sistema serán exentos de derechos de autor o de autoría propia.
- Cada fuente utilizada para el desarrollo del proyecto será correctamente referenciada.
- Se cumplirán los horarios consensuados con el Asesor para la revisión del desarrollo del proyecto.

Posibles restricciones en el desarrollo del proyecto:

- El número de datos de entrenamiento o prueba del sistema pueden ser limitados, por lo que la elección de estos debe ser adecuada.
- De igual forma, los recursos computacionales son limitados, por lo que se deberá tener en cuenta factores como tiempos de entrenamiento del sistema.
- El tiempo de desarrollo del proyecto se espera que sea aproximadamente de 4 meses, tiempo en el cual se espera finalizar este proyecto.

5. Cronograma

Semana 1-2:

- Revisión bibliográfica y definición del problema.
- Selección de la base de datos adecuada para el entrenamiento y evaluación del sistema.

Semana 3-4:

- Preprocesamiento de imágenes y videos: eliminación de ruido, normalización de iluminación y otras técnicas de mejora de calidad de las imágenes y videos.

Semana 5-8:

- Implementación de algoritmos de aprendizaje automático y transferido para el entrenamiento del sistema de reconocimiento personal.
- Implementación del clasificador mediante una base de datos.

Universidad de los Andes
Facultad de Ingeniería

Semana 9-10:

- Evaluación del sistema utilizando diferentes métricas de evaluación, como la precisión y la tasa de falsos positivos.
- Análisis de resultados preliminares y propuestas de mejoras al sistema.

Semana 11-12:

- Implementación de mejoras al sistema de reconocimiento personal basadas en los análisis de los resultados preliminares.

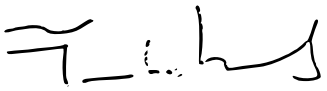
Semana 13-14:

- Evaluación final del sistema mejorado utilizando la base de datos completa.
- Análisis de resultados y discusión de conclusiones.

Semana 15-16:

- Preparación del informe final sobre el proyecto.
- Revisión y corrección del informe final.

Firma estudiante: José Daniel Pantoja Bernal



Firma asesor: Fernando Enrique Lozano Martínez

Para ser llenado por secretaría de coordinación:

Fecha de recepción: _____

Recibido por: _____

Departamento de Ingeniería Eléctrica y Electrónica
Carrera 1 Este No. 19A-40 , Bogotá – Colombia | Tel: (57-1) 3 394999 Ext: 2830 Fax (57-1) 3 324316

Referencias:

[1] M. T. H. Fuad et al., "Recent Advances in Deep Learning Techniques for Face Recognition," in IEEE Access, vol. 9, pp. 99112-99142, 2021, doi: 10.1109/ACCESS.2021.3096136.

[2] Okeke, F. (2022). Facial recognition: Top software vendors. TechRepublic. htps://www.techrepublic.com/article/facial-recognition-software/

[3] Chaves Fonseca, J. (2019). Diseño de un sistema de apoyo en seguridad basado en reconocimiento de múltiples rostros e identificación de usuarios. Universidad de los Andes.

TRABAJO DE GRADO
AUTORIZACIÓN DE SU USO A FAVOR DE LA UNIVERSIDAD DE LOS ANDES

Yo José Daniel Pantoja Bernal, mayor de edad, vecino de Bogotá D.C., identificado con la Cédula de Ciudadanía N° 1.004.770.997 de Guaitarilla-Nariño, actuando en nombre propio, en mi calidad de autor del trabajo de tesis, monografía o trabajo de grado denominado: Sistema de seguridad a partir de una base de datos mediante Machine Learning, basado en reconocimiento de rostros, haré entrega del ejemplar respectivo y de sus anexos del ser el caso, en formato digital o electrónico (CD-ROM) y autorizo a LA UNIVERSIDAD DE LOS ANDES, para que en los términos establecidos en la Ley 23 de 1982, Ley 44 de 1993, Decisión Andina 351 de 1993, Decreto 460 de 1995 y demás normas generales sobre la materia, utilice y use en todas sus formas, los derechos patrimoniales de reproducción, comunicación pública, transformación y distribución (alquiler, préstamo público e importación) que me corresponden como creador de la obra objeto del documento. PARÁGRAFO: La presente autorización se hace extensiva no sólo a las facultades y derechos de uso sobre la obra en formato o soporte material, sino también para formato virtual, electrónico, digital, óptico, usos en red, internet, extranet, intranet, etc., y en general para cualquier formato conocido o por conocer.

EL AUTOR - ESTUDIANTES, manifiesta que la obra objeto de la presente autorización es original y la realizará sin violar o usurpar derechos de autor de terceros, por lo tanto, la obra es de su exclusiva autoría y tiene la titularidad sobre la misma. PARÁGRAFO: En caso de presentarse cualquier reclamación o acción por parte de un tercero en cuanto a los derechos de autor sobre la obra en cuestión, EL ESTUDIANTE - AUTOR, asumirá toda la responsabilidad, y saldrá en defensa de los derechos aquí autorizados; para todos los efectos la Universidad actúa como un tercero de buena fe.

EL AUTOR - ESTUDIANTE.

(Firma) .....

Nombre José Daniel Pantoja Bernal

C.C. N° 1.004.770.997 de Guaitarilla-Nariño