



SECCIÓN CIENCIAS TÉCNICAS

Artículo original de investigación

Algoritmos basados en aprendizaje profundo para aplicaciones reales de reconocimiento de rostros

Heydi Méndez-Vázquez ^{1*} <https://orcid.org/0000-0002-7834-1791>

Yoanna Martínez-Díaz ¹ <https://orcid.org/0000-0002-1591-5989>

Annette Morales González-Quevedo ¹ <https://orcid.org/0000-0003-2716-3144>

Fabiola Becerra-Riera ¹ <https://orcid.org/0000-0002-0639-259X>

Nelson Méndez-Llanes ¹ <https://orcid.org/0000-0003-4003-1496>

¹ Centro de Aplicaciones de Tecnologías de Avanzada. La Habana, Cuba

* Autor para la correspondencia: hmendez@cenatav.co.cu

Editor

Lisset González Navarro
Academia de Ciencias de Cuba.
La Habana, Cuba

Traductor

Darwin A. Arduengo García
Academia de Ciencias de Cuba.
La Habana, Cuba

RESUMEN

Introducción: El reconocimiento automático de rostros se ha mantenido por varios años como un tema de investigación activo en el área de visión por computadoras. Es una de las técnicas biométricas más utilizadas, pero su eficacia en sistemas prácticos se ve afectada ante condiciones no controladas de captura como son la baja resolución, el uso de mascarilla o el reconocimiento de retratos hablados en una investigación criminal. **Objetivos:** Se plantea como objetivo de esta investigación desarrollar un conjunto de algoritmos para el reconocimiento automático de rostros basados en aprendizaje profundo. **Métodos:** La investigación está sustentada en los avances más recientes de las técnicas de aprendizaje profundo, que han sido adecuadas y contextualizadas para ser utilizadas en aplicaciones prácticas de reconocimiento de rostros. **Resultados:** En particular, se proponen métodos novedosos para las tareas de: a) reconocimiento automático de rostros en video de manera eficiente, b) reconocimiento de rostros en videos de baja resolución, c) selección de cuadros y regiones para un reconocimiento de rostro más efectivo, d) clasificación de atributos débiles del rostro, e) reconocimiento automático de rostros con mascarillas, y f) reconocimiento de retratos hablados. **Conclusiones:** Los algoritmos desarrollados logran aumentar la eficacia en las diferentes tareas enunciadas, en comparación con algoritmos similares del estado actual de la literatura y han mostrado ser efectivos en diferentes entornos prácticos.

Palabras clave: reconocimiento automático de rostros; aprendizaje profundo

Deep learning-based methods for real face recognition applications

ABSTRACT

Introduction: Automatic face recognition has remained an active research topic in the field of computer vision for several years. It is one of the most widely used biometric techniques,

but its effectiveness in practical systems is affected by uncontrolled capture conditions such as low resolution, the use of masks, or the recognition of facial sketches in a criminal investigation. **Objective:** To develop a set of algorithms for automatic face recognition based on deep learning techniques, in order to be used in real-world applications in uncontrolled environments. **Methods:** The research is supported by the latest advancements in deep learning techniques, which have been adapted and contextualized for practical face recognition applications for security and law-enforcement. **Results:** In particular, novel methods are proposed for the tasks of: a) efficient automatic face recognition in videos, b) face recognition in low-resolution videos, c) selection of frames and regions for more effective face recognition, d) classification of soft face attributes, e) automatic face recognition with masked faces, and f) recognition of facial sketches. **Conclusions:** The developed algorithms achieve increased effectiveness in the aforementioned tasks compared to similar algorithms in the state-of-the-art and have shown to be effective in different practical environments.

Keywords: automatic face recognition; deep learning

INTRODUCCIÓN

El reconocimiento de rostros es uno de los campos de investigación más activos en visión por computadoras y biometría. ⁽¹⁾ Recientemente han existido avances significativos en esta temática, gracias a la ayuda de algoritmos de aprendizaje profundo y de mayor disponibilidad de conjuntos de datos. ⁽²⁾ Sistemas que incorporan estos algoritmos se utilizan cada vez por los organismos encargados de hacer cumplir la ley, volviéndose imprescindibles para la investigación criminal y otras aplicaciones civiles. Sin embargo, la mayoría de las imágenes de rostro disponibles para entrenar los métodos basados en aprendizaje profundo son capturadas en condiciones controladas (alta resolución, buena iluminación, posición frontal, etc.) y esto provoca que, en entornos no controlados la eficacia de los sistemas de reconocimiento de rostros existentes, se vea reducida significativamente. ⁽³⁾

La video protección es uno de los escenarios no controlados de mayor importancia de cara a la seguridad interna, siendo utilizado tanto por entidades gubernamentales como civiles, en la mayoría de los países del mundo. Esto hace que la eficacia del reconocimiento facial sea un factor de gran importancia, unido a la necesidad de sistemas que permitan un procesamiento eficiente de los cuadros de los videos. ^(4,5) En este tipo de aplicaciones, a la variabilidad intraclase dada por las condiciones no controladas de captura, se le suma la complejidad de la baja resolución a la que usualmente se captan las imágenes por la distancia de los sujetos en estos entornos, que provoca pérdidas en las características identificativas de los rostros. ⁽⁶⁾ De igual manera, el uso de mascarillas, que se volvió obligatorio durante la pandemia de COVID-19, es una oclusión que afecta significativamente el rendimiento de sistemas de reconocimiento facial. ⁽⁷⁾

Otro escenario de vital importancia para el esclarecimiento de hechos delictivos, es el reconocimiento automático de retratos hablados. ⁽⁸⁾ Hay muchos casos en los que la policía no puede capturar fotografías de un sospechoso, pero hay algún testigo o víctima que ayuda a hacer un retrato hablado. Este retrato hablado es solo una aproximación a la identidad de la persona, pero puede ser utilizado para buscar en una base de datos de fotografías y así reducir la lista de posibles sospechosos.

Teniendo en cuenta los aspectos mencionados se plantea como objetivo de este trabajo desarrollar un conjunto de algoritmos para el reconocimiento automático de rostros basados en aprendizaje profundo, que permitan aumentar la eficacia de su uso en sistemas reales para la seguridad y el orden interior.

MÉTODOS

Las principales contribuciones de este trabajo van encaminadas a desarrollar un grupo de métodos para tareas relacionadas con el reconocimiento de rostros basado en aprendizaje profundo, que logren obtener altos niveles de eficiencia y eficacia comparados con algoritmos similares del estado actual de la literatura.

En los últimos años, las redes neuronales convolucionales profundas (CNN) han demostrado resultados impresionantes en diferentes áreas de la visión por computadoras y en particular han alcanzado un rendimiento superior o similar a los humanos en el reconocimiento de rostros. ⁽⁹⁾ Estas redes son difíciles de entrenar debido a la gran cantidad de hiperparámetros que deben ser ajustados. Para encontrar la mejor configuración de una red para un problema específico, se requiere una gran capacidad computacional. Sin embargo, una vez que se entrena el modelo se puede utilizar para otras

tareas sin necesidad de volver a entrenarlo. Cuando no se tienen suficientes recursos para entrenar una red desde 0, una opción comúnmente utilizada es utilizar modelos preentrenados y ajustarlos.

En esta investigación se realizó un análisis del comportamiento de diferentes representaciones obtenidas de una red preentrenada en tareas relacionadas con el reconocimiento de rostros, con el objetivo de determinar la forma más efectiva de lograr un aprendizaje profundo con un poder discriminativo mayor sin necesidad de entrenamiento adicional.⁽¹⁰⁾ Este análisis permitió determinar que el mejor rendimiento de una red preentrenada se obtiene utilizando un vector de agrupación promedio de la última capa convolucional como descriptor. También se propuso aprender una métrica sobre estos rasgos para problemas específicos de reconocimiento de rostros en entornos no controlados. A partir de este estudio inicial, se sentaron las bases para el desarrollo de los algoritmos propuestos.

Redes profundas ligeras para el reconocimiento de rostros

Las redes profundas son cada vez más complejas y consumen muchos recursos, lo que dificulta su implementación en aplicaciones en tiempo real o en dispositivos con recursos limitados. Por lo tanto, la investigación se centra en diseñar redes profundas eficientes sin perder precisión en el reconocimiento. Uno de los modelos propuestos es el ShuffleFaceNet, que mejora el reconocimiento de rostros utilizando una capa de convolución global en profundidad y la función de activación PReLU.⁽¹¹⁾ Los resultados experimentales muestran que este modelo es más eficiente y compacto que el ShuffleNet original. Además, se realizó una comparación exhaustiva con otras arquitecturas de redes ligeras para el reconocimiento facial, evaluando su rendimiento, tiempo de inferencia y requisitos computacionales. Teniendo en cuenta este resultado, el resto de la investigación se basó principalmente en arquitecturas de redes ligeras.

Reconocimiento de rostros en video de baja resolución

El reconocimiento de rostros en imágenes o videos de baja resolución es un desafío importante en aplicaciones de videoprotección. En la literatura, se han propuesto 2 enfoques principales para abordar este problema: aprender características invariantes a la resolución y restaurar imágenes de baja resolución. En esta investigación se profundizó en ambos enfoques. Por un lado, se propuso una representación de disimilitud utilizando características convolucionales profundas para encontrar un espacio común más discriminativo para

imágenes de baja y alta resolución.⁽¹²⁾ Este enfoque demostró ser efectivo en comparación con otros métodos actuales, pero se basa en redes complejas y costosas.

Por otro lado, se evaluaron métodos recientes de restauración de imágenes de baja resolución.⁽¹³⁾ Estos métodos utilizan redes convolucionales y redes adversarias generativas de aprendizaje profundo para recuperar imágenes de alta calidad a partir de imágenes de baja resolución con degradaciones desconocidas. Sin embargo, la mayoría de estos métodos han sido evaluados en conjuntos de imágenes sintéticas y no se ha evaluado su eficacia en imágenes nativas de baja resolución. En este trabajo, se evaluaron los métodos de última generación GPEN, GFP-GAN y CodeFormer para generar imágenes de alta resolución y se analizó su impacto en el reconocimiento facial.

La investigación se centró entonces en encontrar una solución de bajo costo computacional para su implementación en aplicaciones prácticas de videoprotección. Los modelos ligeros, como se vio han demostrado un alto nivel de precisión en tareas de identificación y verificación facial, pero su eficacia en escenarios específicos y desafiantes como el reconocimiento de rostros de baja resolución no ha sido ampliamente evaluada. En este trabajo, se propone evaluar modelos faciales ligeros en imágenes de baja resolución capturadas en escenarios reales de videoprotección.⁽¹³⁾ Se utilizaron 2 modelos ligeros recientes, MobileFaceNet y ShuffleFaceNet, que son extremadamente eficientes y han demostrado un alto rendimiento en diferentes tareas de reconocimiento facial. Además, se utilizó estos mismos modelos para evaluar el impacto de 3 métodos de restauración ciega seleccionados en el reconocimiento de imágenes de rostros de baja resolución capturadas en escenarios reales.

Selección y agregación de la información presente en una secuencia de videos

En la literatura, se han propuesto diferentes enfoques para abordar las variaciones en el reconocimiento facial en entornos no controlados, como son la pose, la iluminación y las expresiones faciales. Los 2 enfoques principales son el basado en la calidad y el basado en la agregación. El enfoque basado en la calidad selecciona 1 subconjunto de cuadros del video que son más representativos de la identidad de la persona, utilizando criterios como la pose facial, la apertura de la boca o los ojos y la resolución de la imagen. En cambio, el enfoque basado en la agregación combina información de todos los cuadros del video para generar una única representación del rostro, asignando un peso a cada cuadro en función de su relevancia. En el trabajo se proponen métodos para ambos enfoques.

Enfoque basado en calidad

La calidad de la imagen facial es un factor importante en la precisión de los sistemas de reconocimiento biométrico. Se han propuesto muchos métodos para evaluarla basados en diversas métricas. Uno de los enfoques más utilizados es seleccionar un conjunto de cuadros del video teniendo en cuenta su calidad. ⁽¹⁴⁾ En el presente trabajo se analizó y propuso heurísticas para seleccionar eficientemente los cuadros de la cara con mayor valor identificativo. ⁽¹⁵⁾ Se propusieron medidas de calidad para evaluar la borrosidad, la pose y la expresión de los ojos y la boca, basados en un detector de puntos característicos del rostro con baja demanda computacional. Basado en estas medidas, se propone un método que analiza la imagen facial de manera local y permite determinar las regiones faciales con información más útil para el reconocimiento. ⁽¹⁶⁾ De esta forma se extraen rasgos de las redes de reconocimiento solo de las regiones determinadas con calidad y estos son los utilizados para representar el rostro. En el caso de los videos se obtiene la representación local para el cuadro del video y los rasgos correspondientes a las regiones seleccionadas con buena calidad son promediados a través de los diferentes cuadros de la secuencia.

Enfoque basado en agregación

El enfoque basado en la agregación ofrece un paso en el que los descriptores extraídos de cada cuadro se combinan para generar una representación facial única del video. La media ponderada es el operador más comúnmente utilizado para la agregación, sin embargo, existen otros operadores más flexibles como la media ponderada ordenada (OWA), que puede comportarse como el máximo, el mínimo, la media, la mediana y estadísticas de orden k , y utiliza información posicional en el proceso de agregación.

Se propuso un nuevo esquema de agregación adaptativa basado en operadores OWA en contraste con la media ponderada. ⁽¹⁷⁾ Se desarrollaron 2 implementaciones, OWANet y WOWANet para evaluar su idoneidad en el reconocimiento facial en video. El esquema de agregación propuesto consta de un módulo de codificación que extrae los vectores de características de los cuadros de entrada y un módulo de agregación que combina los vectores en uno solo. La agregación se realiza mediante un bloque de atención para obtener un vector de relevancia, ordenar las características según la relevancia y agregarlas utilizando pesos obtenidos mediante un cuantificador lingüístico.

Clasificación de atributos débiles

En los últimos años ha habido un aumento en el uso de atributos visuales descriptibles para enriquecer las represen-

taciones faciales existentes. Estos atributos incluyen aspectos relacionados con el entorno y rasgos biométricos débiles que proporcionan información demográfica y características físicas del rostro. Aunque no identifican a una persona específica, estos atributos han demostrado mejorar la precisión en tareas como la localización, el reconocimiento y la reidentificación de individuos. ⁽¹⁸⁾ Uno de los objetivos de este trabajo es desarrollar un método eficaz para estimar atributos demográficos a partir de secuencias de rostros en entornos no controlados de videoprotección. A través de la selección de información relevante y la combinación de diferentes estrategias de fusión, se busca superar las limitaciones actuales y mejorar la precisión de la predicción final de atributos demográficos en videos.

El enfoque propuesto consiste en identificar los cuadros de mejor calidad al considerar 12 factores que afectan las imágenes de rostros, como pose, la iluminación, la oclusión, la baja resolución, el desenfoque, la boca abierta, los ojos cerrados, la dirección no frontal de la mirada, el balance de color, el rostro no centrado, los ojos rojos y el fondo no uniforme. ^(19,20) Para evaluar estos factores se utilizan 12 medidas automáticas que determinan el poder identificativo y la calidad global de imágenes de rostros. Debido a que determinada calidad de imágenes puede ser útil para clasificar género, pero no para estimar edad (y viceversa), se propone usar un clasificador para aprender la relevancia de cada medida para la estimación de género y edad, respectivamente. Después de evaluar la calidad de los cuadros, se seleccionan los mejores para la estimación de género o edad y se representan de acuerdo a la siguiente estrategia: cada imagen facial se divide en 10 regiones que se definen utilizando los puntos característicos del rostro; las regiones definidas permiten una cobertura completa del rostro y se asignan rasgos visuales utilizando capas intermedias de CNN preentrenadas; estos rasgos se reducen a un único vector representativo que facilita la estimación de los atributos. ⁽²¹⁾ En resumen, se utiliza una estrategia que selecciona los mejores cuadros, divide la imagen facial en regiones y utiliza rasgos visuales de CNN preentrenadas para estimar los atributos específicos.

La novedad de esta propuesta radica en la codificación local del rostro usando rasgos de capas intermedias de CNN para estimar el género y la edad. Se utilizaron 2 redes basadas en el modelo ligero MobileNet, para desarrollar un método de clasificación que funcionara eficazmente en computadoras con recursos limitados de memoria y procesamiento.

Reconocimiento de rostros con mascarillas

Con el surgimiento de la pandemia de COVID-19 el uso de mascarillas se convirtió en una medida sanitaria clave para

prevenir la propagación del virus. Sin embargo, esto afectó negativamente la eficacia de los sistemas de reconocimiento facial utilizados en el control de acceso, autenticación y búsquedas operativas, ya que no podían reconocer rostros con mascarillas. Ante este problema los científicos comenzaron a buscar soluciones, como el reentrenamiento de modelos existentes utilizando imágenes de rostros con mascarillas sintéticas, o centrarse en la región periocular del rostro como una alternativa para superar el efecto de las mascarillas. ⁽²²⁾

En este trabajo se estudió y analizó la efectividad de modelos ligeros de reconocimiento facial para abordar el problema de la presencia de mascarillas. ⁽²³⁾ Se consideraron 2 enfoques: incluir imágenes de rostros con mascarillas en el proceso de aprendizaje de estos modelos, y aplicar estos modelos en el reconocimiento de imágenes periorcarias. Se evaluaron los modelos MobileFaceNet, ShuffleFaceNet y VarGFaceNet para determinar su eficacia en esta tarea.

Reconocimiento de retratos hablados

El reconocimiento de retratos hablados es una aplicación relevante para organismos encargados de hacer cumplir la ley. En estos casos se utilizan descripciones verbales o compuestos de sospechosos para buscar en una base de datos de fotografías policiales y reducir la lista de posibles sospechosos. Los retratos hablados pueden ser dibujados a mano o compuestos utilizando *software* especializado. Sin embargo, hay diferencias significativas entre los retratos hablados y las imágenes reales de rostros debido a la dificultad de describir un rostro de memoria y las distorsiones al combinar componentes faciales. Se han propuesto algoritmos para abordar el reconocimiento de retratos hablados compuestos utilizando aprendizaje profundo, pero requieren una gran cantidad de muestras de entrenamiento. En este trabajo, se propone explorar el uso de la información local utilizando rasgos de redes neuronales profundas, para el reconocimiento de retratos hablados. ⁽²⁴⁾ Adicionalmente se propone utilizar enfoque de aprendizaje de métrica que puede ayudar a extender el uso de un modelo entrenado en un dominio general a otro dominio específico

RESULTADOS

Redes ligeras

Para mostrar la efectividad de las redes ligeras de aprendizaje profundo en el reconocimiento de rostros, se evaluaron 5 modelos ligeros y eficientes en 3 bases de datos utilizadas ampliamente para el reconocimiento de rostros: CALFW, IJB-C y CPLFW. Los modelos ligeros evaluados obtienen resultados comparables y en algunos casos, hasta mejores que el resto de los métodos muy profundos reportados en la literatura.

Entre los modelos ligeros, el modelo MobileFaceNet alcanza las mayores tasas de verificación, superior al 96 % en la base de datos IJB-C, seguido de la VarGFaceNet. Además, se realizaron experimentos en bases de datos de imágenes y videos de grandes volúmenes de datos, comparando los resultados obtenidos por los modelos ligeros originales y sus correspondientes versiones diseñadas para el reconocimiento de rostros. Estas evaluaciones mostraron que las modificaciones introducidas en las redes ligeras para el caso específico del reconocimiento de rostros, permiten en todos los casos, incrementar los valores de eficacia, sobre todo en el reconocimiento a gran escala. Por otra parte, aunque las arquitecturas ligeras originales tienen un consumo de memoria ligeramente menor y un menor número de parámetros que sus correspondientes versiones de rostro, sus tiempos de inferencia son mayores. Por ejemplo, en los casos de MobileNetV1 y MobileNetV2, los tiempos de ejecución de la CPU se reducen por un factor de 2 en la versión adaptada para rostros, lo cual es de suma importancia para aplicaciones en tiempo real.

Baja resolución

Los modelos ligeros ShuffleFaceNet y MobileFaceNet (ajustados o reentrenados para este problema), fueron evaluados para el reconocimiento de imágenes nativas de baja resolución. Adicionalmente, se evaluaron los modelos escogidos para la restauración ciega (GPEN, GFP-GAN y CodeFormer). En este último caso, las imágenes de baja resolución fueron convertidas a imágenes de alta resolución que se utilizaron como entrada a las redes ligeras adaptadas al reconocimiento de rostros sin ningún reentrenamiento. Los resultados muestran que aprender un espacio común de características invariantes a partir de imágenes de baja resolución, ofrece las mayores tasas de reconocimiento, mientras que el uso de la restauración ciega degrada el desempeño de los modelos de reconocimiento facial. Esto refleja que los métodos de restauración ciega entrenados en imágenes de baja resolución sintetizadas artificialmente no generalizan bien a imágenes nativas de baja resolución. La falta de pares de imágenes nativas de alta y baja resolución para el aprendizaje de estos métodos, hace que las distribuciones de datos de entrenamiento y prueba sean muy diferentes, lo que implica la necesidad adicional de adaptación del dominio. Además, en los escenarios de videoprotección, las imágenes capturadas tienen otras variaciones adicionales como oclusiones, cambios de pose e iluminación, que añaden complejidad al proceso de restauración.

Selección y agregación de cuadros

Se evalúan el método basado en la calidad local de los rostros y el método propuesto basado en la agregación en

bases de datos de videos con variaciones significativas en la expresión facial, la iluminación, la pose, la resolución y el fondo. Los experimentos realizados muestran que los mejores resultados se obtienen con la propuesta basada en extraer los rasgos profundos asociados a las regiones del rostro con mayor calidad.

Clasificación de atributos débiles

Se evaluaron los clasificadores de género y edad en 2 bases de datos de video que contienen las etiquetas de estos atributos. En las comparaciones se incluyen los resultados de combinar el procedimiento de selección de los n ($n = 10$) mejores cuadros con el enfoque propuesto de extracción de rasgos del modelo MobileFaceNet a partir de las regiones del rostro y con un enfoque de clasificación de principio a fin, tanto para los propios modelos MobileFaceNet, como para otras 2 redes del estado del arte reconocidos por su efectividad en estas tareas. En los experimentos realizados se puede observar cómo la propuesta de selección de cuadros de videos y la representación local por regiones obtiene mejores resultados de manera general en todos los casos. La clasificación de estos atributos se utiliza con el objetivo de mejorar o filtrar la búsqueda de rostros en videos.

Reconocimiento de rostros con mascarillas

Para los experimentos de reconocimiento de rostros con mascarillas, los modelos ligeros de reconocimiento de rostros seleccionados: VarGFaceNet, ShuffleFaceNet y MobileFaceNet fueron reentrenados de forma independiente tanto en las imágenes con mascarillas sintéticas como en las periorculares creadas a partir de la base de datos CASIA-WebFace. Los modelos fueron evaluados en los conjuntos de datos de prueba generados con mascarillas a partir de las bases de datos LFW, AgeDB-30 y CALFW. En los experimentos se demuestra que el uso de la información periorcular del rostro introduce mejoras en la verificación con respecto a los modelos existentes para el reconocimiento de rostros sin mascarillas. No obstante, los modelos ligeros reentrenados con imágenes con mascarillas alcanzan los mejores resultados.

Reconocimiento de retratos hablados

El método propuesto para el reconocimiento de retratos hablados es evaluado y comparado con otros métodos de la literatura en 2 de las bases de datos más utilizadas para este propósito: PRIP-VSGC y UoM-SGFS. Se evalúa el método de extracción de rasgos local propuesto, utilizando 2 modelos profundos: DEEPS (entrenado con retratos hablados) y el modelo genérico para el reconocimiento de rostros Dlib. Se comparan los resultados con los modelos originales y con los modelos ligeros utilizados en el resto del trabajo. Los experi-

mentos demuestran que la propuesta utilizando ambos modelos supera al resto de los métodos de la literatura, incluido el modelo DEEPS original, entrenado para el reconocimiento de retratos hablados. En el caso del uso directo de los modelos ligeros, a pesar de no estar entrenados para esta tarea, permiten alcanzar resultados bastante cercanos a los modelos entrenados específicamente para reconocer retratos hablados. Como trabajo futuro de esta investigación queda pendiente utilizar para los retratos hablados la estrategia de extraer rasgos de manera local utilizando modelos ligeros entrenados para el reconocimiento de rostros en general.

DISCUSIÓN

Las arquitecturas de redes ligeras utilizadas en este trabajo para diferentes tareas relacionadas con el reconocimiento de rostros, exhiben mejoras significativas en comparación con las redes complejas tradicionales. Especialmente su velocidad de inferencia y los requisitos de almacenamiento, las hace muy adecuadas para sistemas con recursos de cómputo limitados. En las diversas tareas evaluadas como son el reconocimiento de rostros en video, de baja resolución, con mascarillas y de retratos hablados se evidencia el potencial de las redes ligeras con respecto a modelos más complejos, ofreciendo un balance adecuado entre eficacia y eficiencia para su uso en aplicaciones prácticas. Otro aspecto importante en los experimentos realizados fue comprobar la importancia de ajustar (o reentrenar) los modelos a los problemas en específico a abordar. Analizar y modificar las arquitecturas de las redes neuronales convolucionales en dependencia del problema a abordar hace más efectivo su uso en la práctica.

Conclusiones

En el presente trabajo se presentan un grupo de métodos basados en aprendizaje profundo para tareas relacionadas con el reconocimiento de rostros. Primeramente, se profundizó en las CNN existentes en la literatura para el reconocimiento de rostros con el objetivo de determinar la mejor manera de aplicar modelos preentrenados. Además, se propuso una nueva red ligera y basado en esta se adaptaron otros modelos ligeros de la literatura para permitir mantener los altos niveles de eficacia, utilizando muy bajos recursos computacionales en comparación con otros modelos profundos. Basado en el uso de los modelos profundos ligeros se obtuvieron métodos novedosos para mejorar el reconocimiento de rostros en video. En específico se adaptaron las redes para el caso de la muy baja resolución y de la presencia de mascarillas faciales, se propusieron métodos para la selección y agregación de la información de un rostro presente en un video y métodos para la clasificación de atributos débiles, que en su conjunto per-

miten obtener altos niveles de eficacia comparados con otros métodos reportados en la literatura y al ser evaluados en escenarios reales. Finalmente se propone un método basado en el uso de rasgos de redes profundas de manera local que permite la búsqueda de retratos hablados con altos niveles de eficacia comparado con sus referentes en la literatura. Los métodos propuestos han sido introducidos en aplicaciones prácticas cubanas.

REFERENCIAS BIBLIOGRÁFICAS

- Taskiran M, Kahraman N, Erdem CE. Face recognition: past, present and future (a review). *Digital Signal Processing*. 2020 Nov; 1;106:102809.
- Wang M, Deng W. Deep face recognition: a survey. *Neurocomputing*. 2021 Mar 14;429:215-44.
- Anwarul S, Dahiya S. A comprehensive review on face recognition methods and factors affecting facial recognition accuracy. In: *Proceedings of ICRIC 2019: Recent Innovations in Computing*. Cham: Springer; 2020. p. 495-514.
- Bashbaghi S, Granger E, Sabourin R, Parchami M. Deep learning architectures for face recognition in video surveillance. In: *Deep learning in object detection and recognition*. Singapore: Springer; 2019. p. 133-54.
- Liu W, Zhou L, Chen J. Face recognition based on lightweight convolutional neural networks. *Information*. 2021 Apr 28;12(5):191.
- Li P, Prieto L, Mery D, Flynn PJ. On low-resolution face recognition in the wild: comparisons and new techniques. *IEEE Transactions on Information Forensics and Security*. 2019 Jan 3;14(8):2000-12.
- Alzubi A, Albalas F, Al-Hadhrani T, Younis LB, Bashayreh A. Masked face recognition using deep learning: a review. *Electronics*. 2021 Oct 31;10(21):2666.
- Jacquet M, Champod C. Automated face recognition in forensic science: review and perspectives. *Forensic Science International*. 2020 Feb 1;307:110124.
- Parkhi O, Vedaldi A, Zisserman A. Deep face recognition. In: *Proceedings of the British Machine Vision Conference (BMVC) 2015*; 2015 Sep 7-10; Swansea, UK. Swansea: British Machine Vision Association; 2015. p. 41.1-41.12.
- López-Avila L, Plasencia-Calaña Y, Martínez-Díaz Y, Méndez-Vázquez H. On the use of pre-trained neural networks for different face recognition tasks. In: *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications: 22nd Iberoamerican Congress, CIARP 2017*; 2017 Nov 7-10; Valparaíso, Chile. Cham: Springer; 2018. p. 356-64.
- Martínez-Díaz Y, Luevano LS, Méndez-Vázquez H, Nicolás-Díaz M, Chang L, González-Mendoza M. Shufflefacenet: a lightweight face architecture for efficient and highly-accurate face recognition. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*; 2019 Oct 27-Nov 2; Seoul, Korea. Piscataway: IEEE; 2019. p. 0-0.
- Hernández-Durán M, Plasencia-Calaña Y, Méndez-Vázquez H. Low-resolution face recognition with deep convolutional features in the dissimilarity space. In: *Progress in Artificial Intelligence and Pattern Recognition: 6th International Workshop, IWAIPR 2018*; 2018 Sep 24-26; Havana, Cuba. Cham: Springer; 2018. p. 95-103.
- Martínez-Díaz Y, Méndez-Vázquez H, Luevano LS. Effectiveness of blind face restoration to boost face recognition performance at low-resolution images. In: *Progress in Artificial Intelligence and Pattern Recognition: 8th International Workshop on Artificial Intelligence and Pattern Recognition (IWAIPR 2023)*; 2023 Sep 26-28; Varadero, Cuba. Cham: Springer; 2023. p. 455-67.
- Bahroun S, Abed R, Zagrouba E. KS-FQA: keyframe selection based on face quality assessment for efficient face recognition in video. *IET Image Processing*. 2021 Jan;15(1):77-90.
- Méndez-Llanes N, Castillo-Rosado K, Méndez-Vázquez H, Kheilat-Kihel S, Tistarelli M. Face recognition on mobile devices based on frames selection. In: *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications: 24th Iberoamerican Congress, CIARP 2019*; 2019 Oct 28-31; Havana, Cuba. Cham: Springer; 2019. p. 316-25.
- Méndez-Llanes N, Castillo-Rosado K, Méndez-Vázquez H, Tistarelli M. Quality-based representation for unconstrained face recognition. In: *2020 25th International Conference on Pattern Recognition (ICPR)*; 2021 Jan 10-15; Milan, Italy. IEEE; 2021. p. 6494-500.
- Rivero-Hernández J, Morales-Gonzalez A, Denis LG, Méndez-Vázquez H. Ordered weighted aggregation networks for video face recognition. *Pattern Recognition Letters*. 2021 Jun 1;146:237-43.
- Becerra-Riera F, Morales-González A, Méndez-Vázquez H. A survey on facial soft biometrics for video surveillance and forensic applications. *Artificial Intelligence Review*. 2019 Aug;52(2):1155-87.
- Becerra-Riera F, Morales-Gonzalez A, Méndez-Vázquez H, Dugelay JL. Attribute-based quality assessment for demographic estimation in face videos. In: *2020 25th International Conference on Pattern Recognition (ICPR)*; 2021 Jan 10-15; Milan, Italy. IEEE; 2021. p. 5875-82.
- Becerra-Riera F, Morales-González A, Méndez-Vázquez H, Dugelay JL. Demographic attribute estimation in face videos combining local information and quality assessment. *Machine Vision and Applications*. 2022 Mar;33(2):26.
- Becerra-Riera F, Morales-González A, Méndez-Vázquez H. Exploring local deep representations for facial gender classification in videos. In: *Progress in Artificial Intelligence and Pattern Recognition: 6th International Workshop, IWAIPR 2018*; 2018 Sep 24-26; Havana, Cuba. Cham: Springer; 2018. p. 104-12.
- Deng J, Guo J, An X, Zhu Z, Zafeiriou S. Masked face recognition challenge: the insightface track report. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*; 2021 Oct 11-17; Virtual. IEEE; 2021. p. 1437-44.
- Martínez-Díaz Y, Méndez-Vázquez H, Luevano LS, Nicolás-Díaz M, Chang L, González-Mendoza M. Towards accurate and lightweight masked face recognition: an experimental evaluation. *IEEE Access*. 2021 Dec 13;10:7341-53.
- Méndez-Vázquez H, Becerra-Riera F, Morales-González A, López-Avila L, Tistarelli M. Local deep features for composite face sketch recognition. In: *2019 7th International Workshop on Biometrics and Forensics (IWBF)*; 2019 May 2-3; Cancun, Mexico. IEEE; 2019. p. 1-6.

Recibido: 16/06/2025

Aprobado: 30/06/2024

Agradecimientos

Los autores agradecen a todos aquellos que colaboraron de una forma u otra en la investigación presentada en este trabajo: Luis S. Luevano, Jacinto Rivero, Leyanis López, Yenisel Plasencia, Mairelys Hernández, Miguel Nicolás-Díaz, Katy Castillo, Leonardo Chang, Miguel González-Mendoza, Jean-Luc Dugelay, Lester Guerra, Massimo Tistarelli.

Conflictos de intereses

Los autores declaran que no tienen conflictos de intereses económicos o relaciones personales que pudieran haber influido en el trabajo reportado en este documento.

Contribuciones de los autores

- Conceptualización: Heydi Méndez
- Curación de datos: Yoanna Martínez, Fabiola Becerra, Nelson Méndez
- Análisis formal: Heydi Méndez, Yoanna Martínez, Annette Morales
- Adquisición de fondos: Heydi Méndez
- Investigación: Heydi Méndez, Yoanna Martínez, Annette Morales, Fabiola Becerra
- Metodologías: Heydi Méndez, Annette Morales
- Administración de proyecto: Heydi Méndez
- **Software:** Yoanna Martínez, Fabiola Becerra, Nelson Méndez
- Supervisión: Heydi Méndez, Annette Morales

- Validación: Yoanna Martínez, Fabiola Becerra, Nelson Méndez
- Visualización: Heydi Méndez, Yoanna Martínez, Annette Morales
- Redacción-borrador original: Heydi Méndez, Yoanna Martínez, Annette Morales
- Redacción-revisión y edición: Heydi Méndez

Financiamientos

No se utilizó financiamiento específico para realizar la investigación presentada.

Cómo citar este artículo

Méndez-Vázquez H, Martínez-Díaz Y, Morales González-Quevedo A, Becerra-Riera F, Méndez-Llanes N. Algoritmos basados en aprendizaje profundo para aplicaciones reales de reconocimiento de rostros. An Acad Cienc Cuba [Internet] 2025 [citado en día, mes y año];15(2):e3094. Disponible en: <http://www.revistaccuba.cu/index.php/revacc/article/view/3094>

El artículo se difunde en acceso abierto según los términos de una licencia Creative Commons de Atribución/Reconocimiento-NoComercial 4.0 Internacional (CC BY-NC-SA 4.0), que le atribuye la libertad de copiar, compartir, distribuir, exhibir o implementar sin permiso, salvo con las siguientes condiciones: reconocer a sus autores (atribución), indicar los cambios que haya realizado y no usar el material con fines comerciales (no comercial).

© Los autores, 2025.

