

Comparación de fenotipos faciales mediante el uso de redes neuronales convolucionales

Comparing facial phenotypes using convolutional neural networks

Francho Melendez, Sergio González
Fenomatch S.L. , Zaragoza 50001, España

RESUMEN

El análisis automático de los rasgos fenotípicos faciales a partir de imágenes abre la posibilidad de utilizar el análisis facial para incrementar las posibilidades de que los descendientes nacidos de donación de gametos se parezcan a sus progenitores. En el presente trabajo, se presentan los experimentos en el uso de las tecnologías de deep learning en el ámbito del análisis facial para avanzar en este objetivo. Entre ellos, se presenta la creación de una base de datos de caras para la evaluación del parecido facial, el uso de la segmentación facial, y la detección de distancias fenotípicas tridimensionales.

Palabras clave: *Análisis facial, Inteligencia Artificial, Generative Adversarial Networks, Selección embrionaria, Time-lapse, Reproducción Asistida, Fenomatch.*

SUMMARY

Automatic analysis of facial phenotypic features from images opens up the possibility of using facial analysis to increase the chances that offspring born from gamete donation will resemble their parents. In this paper, experiments in the use of deep learning technologies in the field of facial analysis to advance this goal are presented, including the creation of a database of faces for the evaluation of facial resemblance, the use of facial segmentation, and the detection of three-dimensional phenotypic distances.

Key words: *Facial analysis, Artificial Intelligence, Generative Adversarial Networks, Embryo selection, Time-lapse, Assisted Reproduction, Fenomatch*

Correspondencia: Luis José Arenaz Villalba
Fenomatch SL
Calle San Juan y San Pedro 7 local 1
Zaragoza 50.001
España
SOLICITUD REIMPRESIÓN: Email: fertilidad@editorialmedica.com

INTRODUCCIÓN

El estudio de los rasgos fenotípicos faciales y la heredabilidad de los mismos ha sido estudiado extensivamente desde mitad del siglo XX. Los resultados de estas investigaciones junto con el desarrollo de tecnologías de análisis facial permiten aplicaciones específicas en la reproducción asistida que, analizando a donantes de gametos y pacientes, se propone aumentar las posibilidades de que progenitores y descendientes compartan rasgos fenotípicos faciales.

El estudio de la heredabilidad de los rasgos faciales así como la evaluación del parecido entre dos personas es compleja tanto desde el punto de vista genético como perceptual. Desde el punto de vista genético, normalmente se utilizan bases de datos de gemelos idénticos en combinación con modelos de caracterización matemática facial a partir de fotografías, buscando así correlaciones comunes entre gemelos. Desde el punto de vista perceptual, tradicionalmente se ha intentado crear modelos que reprodujeran los criterios expresados por personas y recogidos cuidadosamente en experimentos. A este respecto, con la incursión de la *machine learning* en el análisis de imágenes, se ha comprobado que las redes neuronales convolucionales son efectivas en modelar ciertos aspectos de la percepción visual. Simultáneamente, los algoritmos de *matching* facial pueden entrenarse específicamente para dicho propósito.

En el presente trabajo se presentan nuestros avances en un nuevo algoritmo de *matching* facial dedicado a dotar de mayor peso a ciertas partes del rostro para permitir una mayor configurabilidad del algoritmo. Esto lleva a mejorar aún más la capacidad de proveer al usuario de mayor control sobre el algoritmo de selección de donante.

MATERIAL Y MÉTODO

nuevo algoritmo de matching facial

Se ha creado un nuevo algoritmo que calcula un índice de idoneidad entre dos fotografías de dos personas, las cuales, en el uso típico del algoritmo, pertenecerán al receptor y a un donante cuya similitud se quiere evaluar. Dicho índice es una media ponderada configurable de diferentes índices de similitud calculados independientemente. El proceso empieza detectando la sección de la fotografía correspondiente a la cara. Una vez identificada, computa un modelo tridimensional de la misma que se utiliza para calcular diferentes distancias fenotípicas. También detecta las diferentes secciones como la nariz o la boca y calcula el parecido perceptual general así como de forma sectorizada. Por último, se genera un índice de *matching* fenotípico con una red neuronal entrenada para esta tarea específicamente. A continuación se presentan los detalles de cada uno de estos procesos.

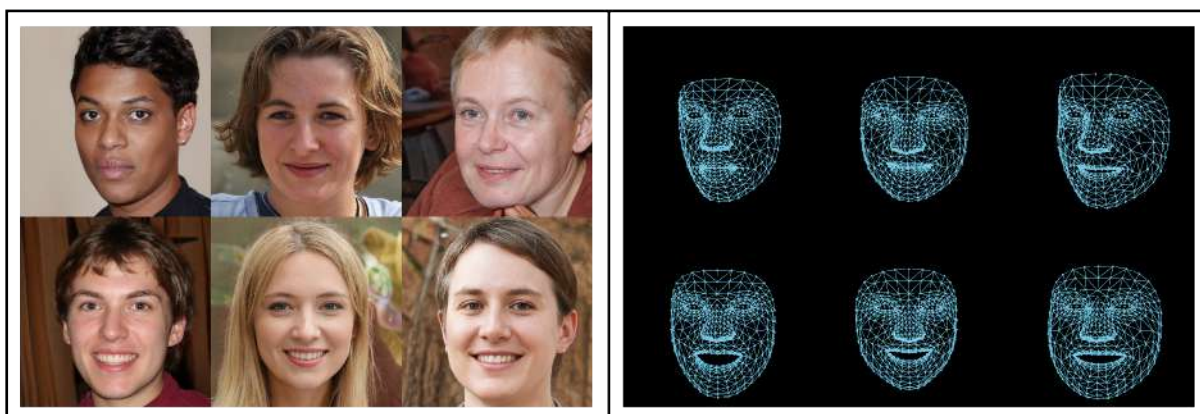
A. Extracción de puntos tridimensionales

El algoritmo de extracción de puntos (1) detecta los contornos de la cara y deforma un modelo tridimensional predefinido optimizando la posición de los puntos para que concuerden con la imagen. De este proceso se extraen 468 puntos de referencia, los cuales se utilizan para calcular diferentes distancias fenotípicas que han sido estudiadas en la literatura como heredables (2), como las distancias nasión-basión, nasión-silla o anchura de la nariz.

En el ejemplo de la figura 1, se pueden observar diferentes caras y los modelos tridimensionales detectados correspondientes. La principal ventaja de utilizar puntos de referencia

FIGURA 1

Extracción de puntos y estimación de pose facial en tres dimensiones



tridimensionales sobre los bidimensionales es que se trata de una manera más efectiva las desviaciones sobre la pose de cabeza en la imagen con respecto a la pose puramente frontal así como las deformaciones relacionadas con la lente utilizada para tomar la fotografía. Todo esto hace por tanto que las distancias medidas sean más precisas. Una vez calculadas las distancias fenotípicas, se crea un vector con todas ellas y se calcula la distancia entre los vectores de dos imágenes, dando como resultado el *índice de distancias fenotípicas*.

B. Segmentación facial

Utilizando una red neuronal enfocada a la detección de las diferentes partes de la cara, se pueden etiquetar cada uno de los píxeles de la fotografía a una de las secciones definidas: ojos, cejas, pelo, nariz, boca, cuello, cuerpo y fondo. El objetivo es poder medir el parecido entre dos fotografías con una granularidad mayor, pudiendo así ajustar los pesos a diferentes áreas faciales.

Este proceso utiliza una red neuronal de segmentación (3) entrenada en la base de datos *CelebFaces Attributes* que contiene más de 10.000 fotografías de caras con los atributos faciales segmentados. En la Figura 2 se pueden observar los resultados de este proceso al aplicarlo sobre las imágenes de la Figura 1.

C. Parecido Perceptual

Para comparar las fotografías de ambas caras, se utiliza una métrica de similitud perceptual basada en redes neuronales convolucionales. Se ha demostrado que esta métrica reproduce con sorprendente exactitud las evaluaciones de similitud entre dos fotografías por parte de personas (4). Se utiliza esta métrica tanto para toda la cara, como para las diferentes secciones extraídas con la segmentación facial. Se distinguen tres secciones diferenciadas: ojos y cejas, nariz, y boca. El algoritmo proporciona información sobre la localización de las diferentes partes del rostro, diferen-

ciando entre lado derecho e izquierdo, dientes, labio superior e inferior, por lo que se agrupan los elementos pertenecientes a cada sección creando una única imagen. Cada sección da lugar a un índice numérico de similitud, los cuales pueden ser ponderados según las preferencias del usuario.

D. Distancia Facial Aprendida

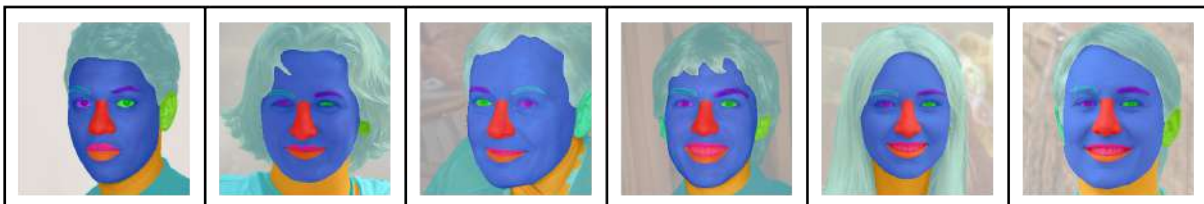
En última instancia, se utiliza una red neuronal que crea una *embedding* de cada fotografía y calcula la distancia entre ambas. Este *embedding* o representación matemática multidimensional de la fotografía representa una serie de características faciales de forma que la distancia entre dos representaciones en dicho espacio representa la probabilidad de esas dos fotografías a pertenecer a la misma familia. Dicha representación matemática se genera por medio del paso de la imagen a través de una red neuronal que realiza una serie de operaciones matemáticas cuyos parámetros ha aprendido a través de un proceso de optimización. La utilización de este *embedding* permite a la red neuronal calcular una distancia que es invariable a la posición de la cara o la iluminación y que está específicamente dedicada a la representación de caras. La arquitectura utilizada es similar a las utilizadas para el reconocimiento facial, pero ha sido optimizada para la tarea en cuestión, que no es determinar si las fotografías son de la misma persona sino si comparten fenotipos faciales. En este sentido, ignora aspectos como el sexo, que sería determinante en el caso de reconocimiento facial.

UNA BASE DE DATOS PARA EVALUACIÓN DE FENOTIPOS FACIALES

En este estudio se presenta una base de datos creada para la evaluación del análisis facial. Para este propósito, se utiliza un algoritmo de generación de caras que permite la creación de retratos fotorrealistas (5). Esto permite generar un número ilimitado de caras, modificarlas y compartirlas sin las

FIGURA 2

Segmentación de cara en secciones codificadas con diferentes colores, ojos, cejas, nariz, y boca, aplicada a las fotografías de la Figura 1



limitaciones de las leyes de protección de datos. Este algoritmo, además, permite mezclar los parámetros de la representación vectorial de la imagen pudiendo combinar los parámetros de dos imágenes y creando así variaciones de la imagen original.

Experimentando con diferentes niveles de mezcla de los parámetros de estilo, se ha creado una base de datos con caras automáticamente clasificadas en cuatro niveles de similitud:

- **Nivel 0:** Los individuos conservan los rasgos faciales de ojos, nariz, y boca prácticamente inalterados. Conserva también el color de pelo y piel. Cambia la pose de la cara y el peinado y puede variar la estructura de la cabeza. Consideramos este nivel cómo misma persona o hermanos gemelos.

- **Nivel 1:** Hermanos o hermanos gemelos. La boca y la nariz se mantienen. La estructura de la cabeza, forma de las cejas y mandíbula puede variar notablemente.

- **Nivel 2:** Familia. Comparten etnia, color de ojos y pelo. Estructura de ojos y boca es muy similar pero varía la nariz, la forma de las cejas. La forma de la cabeza, peinado y mandíbula pueden variar. La iluminación general de la imagen se conserva.

- **Nivel 3:** Comparte color de piel, de pelo y de ojos. Puede apreciarse algún parecido, pero siempre muy mezclado con los rasgos de otra persona. Las personas podrían ser familia, pero con una influencia muy fuerte de otra persona. Figura 3

Se han generado 50 imágenes, 25 de mujeres y 25 de hombres, de diferentes edades y etnias. Para cada nivel, se combina una imagen con las 49 restantes. El total es de 9654 imágenes. Algunas de las combinaciones de parámetros producen imágenes poco realistas o son demasiado parecidas y no proporcionan valor a la base de datos. Después de hacer una criba manual, el conjunto final de imágenes contiene 8000 imágenes de caras.

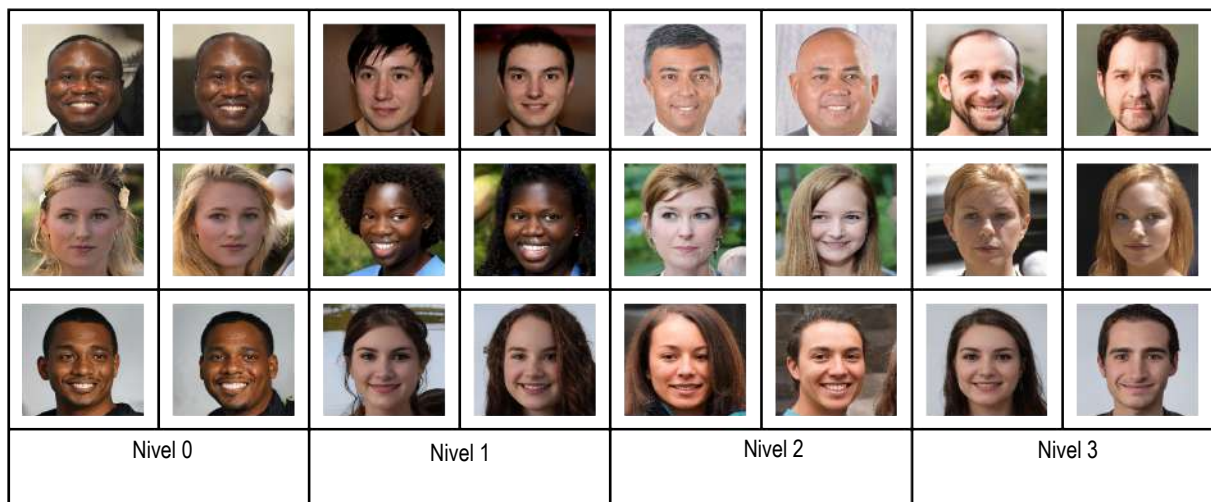
El objetivo de generar estas imágenes es tener un data set extenso y controlable. Pese a que las imágenes generadas no pueden expresar una relación familiar o genética real, sí que proporcionan un orden en el nivel de parecido. Esto proporciona una herramienta útil para evaluar diferentes iteraciones del algoritmo de forma controlada. También es posible utilizar esta base de datos para validar los algoritmos con respecto al parecido perceptual comparando los resultados con validaciones realizadas por personas. Recoger un número de imágenes similar de personas reales con los correspondientes permisos de uso conllevaría un gran esfuerzo tanto en tiempo como en recursos.

TRABAJO FUTURO

Tras la presentación de los nuevos cambios propuestos para el algoritmo de *matching* facial teniendo en cuenta rasgos heredables, así como la presentación de un dataset basado

FIGURA 3

Niveles de similitud dentro de la base de datos propuesta para el análisis facial



en diferentes niveles de similitud, es necesario un mayor estudio acerca de la influencia del factor genético sobre los rasgos faciales.

La generación de datos sintéticos es un útil primer paso, sin embargo el método actual, no incluye datos reales sobre parentesco. El uso de bases de datos reales de gemelos idénticos o familiares con diferentes niveles de parecido y relación familiar, permitiría la inclusión de dichos parámetros en el proceso generativo y ayudaría a mejorar el proceso de comparación facial y su heredabilidad en la descendencia.

CONCLUSIÓN

Los rasgos faciales y su heredabilidad son factores de gran peso para la descendencia. En tratamientos de reproducción asistida que involucran donantes, la selección de donante se realiza teniendo en cuenta aspectos fenotípicos. Así pues, la elección de un donante que comparta rasgos faciales con la persona receptora puede permitir incrementar el parecido facial de la descendencia resultante. La utilización de algoritmos de deep learning basados en redes neuronales convolucionales, permite realizar un estudio más profundo acerca de la similitud entre dos individuos.

Por una parte, la incorporación del *matching* facial basado en el análisis de puntos tridimensionales permite mejorar el algoritmo, evitando, por ejemplo, errores de posicionamiento del rostro y consiguiendo distancias faciales más precisas, resultando en una mejor comparación facial.

Por otra parte, el uso de la segmentación durante el análisis de esta imagen, permite dar más peso en la comparación a

ciertas fracciones de la cara que se conocen como heredables, dando como resultado un índice de comparación mucho más personalizado.

Así pues, con estas nuevas modificaciones, se puede realizar una mejora en el *matching* facial, lo que permite así disminuir tanto la incertidumbre como el error humano durante el proceso de selección de donante.

BIBLIOGRAFÍA

1. **Yury Kartynnik Artsiom Ablavatski Ivan Grishchenko Matthias Grundmann**, CVPR Workshop on Computer Vision for Augmented and Virtual Reality 2019, IEEE, Long Beach, CA
2. **Dimosthenis Tsagkrasoulis, Pirro Hysi, Tim Spector & Giovanni Montana**, Heritability maps of human face morphology through large-scale automated three-dimensional phenotyping. Scientific reports, 2017.
3. **Yu, Changqian, et al.** "Bisenet: Bilateral segmentation network for real-time semantic segmentation." Proceedings of the European conference on computer vision (ECCV). 2018.
4. **R. Zhang and P. Isola and A. A. Efros and E. Shechtman and O. Wang**, The Unreasonable Effectiveness of Deep Features as a Perceptual Metric, 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018.
5. **Karras, Tero and Laine, Samuli and Aila, Timo** A style-based generator architecture for generative adversarial networks, , Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019.