

Utilización de Algoritmo de Agrupación de Múltiples Vistas para la Detección de Rostros

Pedro Luis González Roa

Abstract—En los últimos años se ha demostrado la complejidad en la realización de algoritmos de detección y reconocimiento de rostros, ya que estos comparten un alto grado de similitud estructural. De esta manera, se requieren de la utilización *redes neuronales convolucionales* (CNN por sus siglas en inglés) para la extracción de características de la imagen. Las cuales requieren de una base de datos de gran tamaño para el entrenamiento exitoso de estas. Por lo que se utilizan se utilizan CNN's previamente entrenadas para la extracción de las características las cuales definirán el grupo ó cara a la que pertenecen. Se propone la utilización de múltiples *redes neuronales convolucionales* en conjunto, combinando las características que extraen cada una de ellas para la parte interior y la exterior de la cara.

I. ANTECEDENTES

I-A. Agrupamiento de imágenes

El reconocimiento facial se encuentra directamente relacionado con uno de los problemas que han recibido mucha atención en las últimas tres décadas. El problema de agrupamiento de imágenes, también conocido como *IC* por sus siglas en inglés, se centra en la búsqueda y comparación de características similares entre imágenes. En otras palabras, *IC* consiste subconjuntos de objetos de una misma naturaleza y separarlos de los objetos con características diferentes. [1]

I-B. Combinación de CNN previamente entrenados

En previas investigaciones diferentes modelos de *deep clustering algorithm* han demostrado un buen desempeño para pequeñas imágenes. Por otro lado, para obtener resultados exitosos en la agrupación de imágenes de mayor complejidad (cómo imágenes de objetos con estructuras complejas) es necesario la utilización de CNNs ya entrenadas en un paso previo para la extracción de características específicas. Existen una variedad de estos modelos, los cuales tienen un mismo objetivo pero aportan una perspectiva diferente. Para casos específicos un modelo puede tener mejor desempeño en la obtención de características definitivas, pero en otros existe la posibilidad de lo contrario. [1]

Al igual que en diferentes temas en la era de *Big Data*, existen diferentes perspectivas del mismo objeto que se pueden utilizar para complementarse mutuamente. Por lo cual, el algoritmo de *Agrupación de vistas múltiples* (*Multi-view clustering - MVC*) ha ganado popularidad por su éxito en modelos que obtienen información del mismo objeto desde diferentes fuentes. [3] Guérin en su artículo *Combining pretrained CNN feature extractors to enhance clustering of complex natural images* utiliza diferentes modelos CNN cómo las múltiples perspectivas, para obtener un mejor rendimiento en la agrupación de imágenes. Utilizando dicha

arquitectura se demostró que se reduce el riesgo de elegir una red CNN que no obtenga las mejores características para la discriminación de las imágenes. [1]

I-C. Modelo IE-CNN

Previamente se mencionó que el reconocimiento facial se encuentra relacionado a *IC*. Esto es porque también es importante la detección de los rasgos que diferencian a una persona de otra; donde por la estructura general de nuestros rostros no es una tarea sencilla. Por lo que investigaciones y propuestas exitosas en este ámbito han utilizado una gran cantidad de imágenes de los mismos rostros. An-Ping Song en su artículo *Similar Face Recognition Using the IE-CNN Model* remarca que en estas investigaciones se ha enfocado a la cara como un solo objeto. Cuando nuestras cabezas se descomponen en dos áreas:

- La interna: que se compone por los ojos, nariz y boca.
- La externa: que se compone por la cabeza, la barbilla y las orejas.

El modelo *IE-CNN* de Song se enfoca en separar dichas áreas y utilizarlas cómo diferentes partes de una misma entrada de datos antes de las estructuras convolucionales, obteniendo mejores resultados. [2]

II. PLANTEAMIENTO DEL PROBLEMA

La utilización de redes neuronales profundas ha obtenido un buen resultado para la detección de rostros. Aunque un atributo importante para el éxito de estas ha sido el enorme tamaño de las bases de datos utilizadas para el entrenamiento. Esto es por la similitudes estructurales que se tienen entre los diferentes rostros humanos, exigiendo una gran cantidad de muestras para que un modelo CNN convencional sea capaz de aprender sobre las características necesarias para discernir entre ellas.

La obtención de dicha cantidad de datos requiere de un personal extenso y acceso a mucha información, la cual siempre es pública. An-Ping Song en su investigación describe la gran diferencia entre los repositorios públicos y los repositorios privados de rostros humanos para entrenamiento en la tabla *Table 1*. [2] Al no ser tan viable la recolección de estos datos, en ocasiones es necesario buscar alternativas para eficientizar los modelos alrededor de la cantidad de información obtenida.

III. CONTRIBUCIONES ESPERADAS

Se propone un modelo basado en los resultados de las investigaciones previamente mencionadas. Esta arquitectura

TABLE I
TAMAÑO DE DIFERENTES REPOSITORIOS DE IMÁGENES DE ROSTROS

Repositorio	Identidades	Imágenes
LFW	5,749	13,233
WDRef	2,995	13,233
CelebFaces	10,177	202,599
CASIA-WebFace	10,575	494,414
Facebook	4,030	4.4M
Google (Privado)	8M	200M

buscará utilizar diferentes perspectivas de modelos *CNN* previamente entrenados sobre las dos diferentes áreas propuestas por Song, utilizando agrupamiento de diferentes vistas. Buscando aprovechar al máximo la cantidad de información en los dataset públicos, y tener el mismo desempeño, o incluso una mejora, a comparación de otros modelos.

El desempeño del modelo se mide de acuerdo al porcentaje de rostros correctamente agrupados. Es decir que los grupos de las imágenes correspondan al mismo rostro.

REFERENCES

- [1] Joris Guérin, Stéphane Thiery, Eric Nyiri, Olivier Gibaru, and Byron Boots. Combining pretrained cnn feature extractors to enhance clustering of complex natural images. *Neurocomputing*, 423:551–571, 1 2021.
- [2] An Ping Song, Qian Hu, Xue Hai Ding, Xin Yi Di, and Zi Heng Song. Similar face recognition using the ie-cnn model. *IEEE Access*, 8:45244–45253, 2020.
- [3] Yan Yang and Hao Wang. Multi-view clustering: A survey. *Big Data Mining and Analytics*, 1:83–107, 6 2018.