



INSTITUTO TECNOLÓGICO Y DE ESTUDIOS SUPERIORES DE
MONTERREY

Escuela de Ingeniería y Ciencias

Ingeniería en Ciencia de Datos y Matemáticas

Momento de Retroalimentación: Reto Modelo y Refinamiento

INTELIGENCIA ARTIFICIAL AVANZADA PARA LA CIENCIA DE
DATOS II

Federico Medina García Corral A01721441

Michelle Yareni Morales Ramón A01552627

Paola Sofía Reyes Mancheno A00831314

Maria Fernanda Torres Alcubilla A01285041

Supervisado por:

Iván Mauricio Amaya Contreras, Ph.D.

Edgar Covantes Osuna, Ph.D.

Hugo Terashima Marín, Ph.D.

Blanca R. Ruiz Hernández, Ph. D.

Christian Mendoza Buenrostro, Ph. D.

Prof. Felipe Castillo Rendón

Prof. Félix Ricardo Botello Urrutia

Monterrey, Nuevo León. Fecha, 8 de noviembre de 2023

1. Benchmark

Con el objetivo de establecer una manera de medir y comparar nuestros modelos, decidimos implementar un benchmark, el cual está conformado por dos pruebas:

- Funcionamiento del modelo cuando tiene los registros de la(s) persona(s) y rostros completos.
- Funcionamiento del modelo cuando tiene los registros de la(s) persona(s) y se muestra solamente una porción del rostro.
- Funcionamiento del modelo cuando una persona no se encuentra registrada, esto con el objetivo de mostrar el comportamiento del modelo cuando no se tiene información del rostro.

2. Modelo inicial

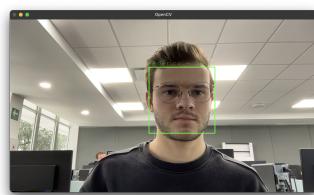
Para la situación problema en la que se está trabajando, se comenzó aplicando un modelo de reconocimiento facial de Haar Cascade. Este es un modelo clasificador preentrenado que se utiliza para reconocer patrones visuales en imágenes. Como tal, Haar Cascade no es una técnica de Machine Learning, pues su enfoque es más clásico. Este funciona separando a las imágenes en trazas blancas y negras según características de la foto, y posteriormente se divide al set de entrenamiento en positivo y negativo para la clasificación [1].

Al momento de incorporar este modelo, se observó que cumplía con la función de reconocer los rostros de una persona; desafortunadamente este se basa en características muy específicas de cada persona. Por ejemplo, si en el entrenamiento la persona está utilizando algún accesorio como lentes, a la hora de ejecutar el modelo este no podrá identificar de manera correcta a la persona. Hecho que no permitía cumplir con uno de los objetivos de la situación problema. De igual manera, al ser un modelo “antiguo” y casi obsoleto, se veía reflejado en un desempeño bastante lento y cortado.

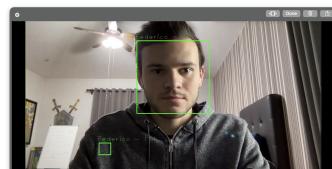
Para la evaluación del benchmark, se realizaron diversas pruebas en distintos escenarios. Para el registro de “Federico” se utilizó la foto 1 y las restantes del equipo, este modelo fue utilizado en la imagen 2a, donde se muestra que una predicción correcta. Por otro lado, usando el mismo modelo, en la imagen 2b se realizó una prueba con distinta iluminación y aunque el reconocimiento es correcto, observamos que la identificación de rostros no es tan precisa, ya que de manera errónea detecta un rostro en la prenda.



Figura 1: Foto utilizada para entrenar modelo de Haar Cascade



(a) Mismos accesorios



(b) Identificación facial errónea

Figura 2: Reconocimiento de rostro con el modelo Haar Cascade

La prueba de rostro parcialmente cubierto se observa en la imagen 3, en esta el modelo logra tanto la detección como reconocimiento correcto.

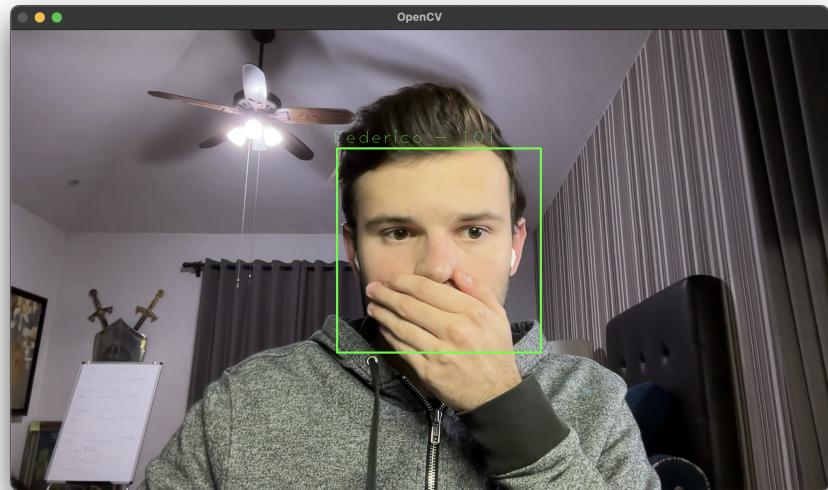


Figura 3: Predicción correcta con rostro parcialmente cubierto

Por último, en la imagen 4, se observa la predicción del modelo entrenado con fotos del resto del equipo (excluyendo a “Federico”), podemos observar que el modelo fuerza la predicción a tener un resultado conocido, por lo que en este modelo no existe el caso de “unknown”.

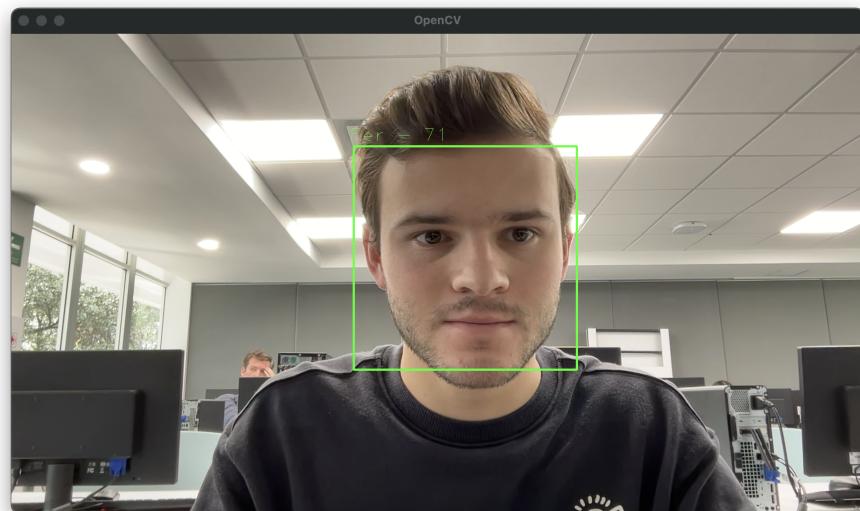


Figura 4: Reconocimiento erróneo de rostro con el modelo Haar Cascade

Tanto el error de identificación observado en la imagen 2b y el etiquetado obligado en los rostros

no registrados observado en la imagen 4 no se alinean con los objetivos y contexto de la situación planteada, por lo que se decide que este modelo no es viable.

3. Nueva solución

La nueva solución, con la que se estará trabajando, es el modelo de reconocimiento facial de la librería dlib, el cual cuenta con una precisión del 99.38 % en el benchmark *Labeled Faces in the Wild*. Para entrenar el modelo con nuestro dataset, solo es necesario una foto de cada persona que se quiera identificar. Además, considerando la escalabilidad del proyecto, si se cuenta con una computadora con múltiples CPU cores, la librería te permite trabajar en paralelo, lo que acelera el reconocimiento facial al procesar más imágenes en el mismo tiempo. [2] Cabe recalcar que para el entrenamiento del nuevo modelo se utilizarán los mismos datos de prueba como para el Haar Cascade, esto con el objetivo de mantener bajo el mismo ambiente ambos modelos y evaluar cuál es el mejor para la situación problema que se busca solucionar.

3.1. Evaluación y ajuste

El único parámetro que se permite ajustar es el parámetro *tolerancia*, el cual refleja que tan estricto es el modelo al momento de comparar rostros. Mientras menor sea este, más estricto es el modelo.

Primero, se probó con un valor alto de 0.8, en la imagen prueba 5a, observamos que cuando el rostro está despejado se clasifica correctamente, como es en el caso de “Federico”, sin embargo, si el rostro es parcial, el modelo varía en su clasificación con las personas similares registradas, en el ejemplo se observa el resultado de “Sofi” cuando debería ser “Fer”.

Por otro lado, en el escenario donde el rostro no se encuentre registrado, se tiene mayor probabilidad de ajuste al rostro más similar y la posibilidad de “unknown” es menor, en este caso se puede observar que en la imagen prueba 5b se clasifican a ambas personas como “Federico”.

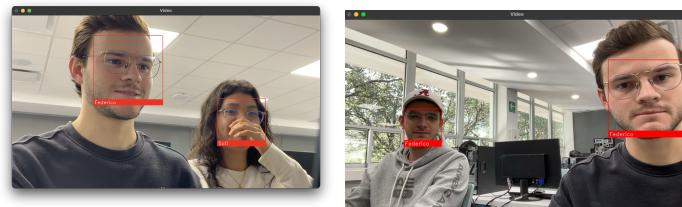
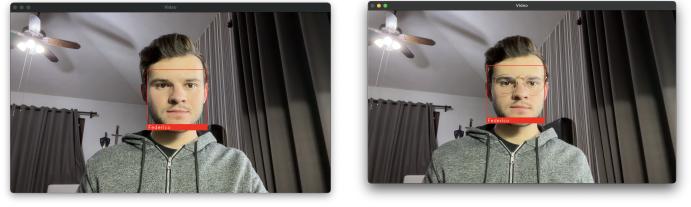


Figura 5: Error en el modelo face_recognition con parámetro de *tolerance* = 0.8

Debido a esto, se decidió disminuir este valor a 0.6, en la imagen prueba 6a, se observa que igualmente se da un reconocimiento correcto cuando el rostro se encuentra despejado, sin embargo, a diferencia del anterior, cuando el rostro es parcial se clasifica como “unknown” en lugar de un registro, lo que representa una mejora al momento de considerar la toma de asistencia.

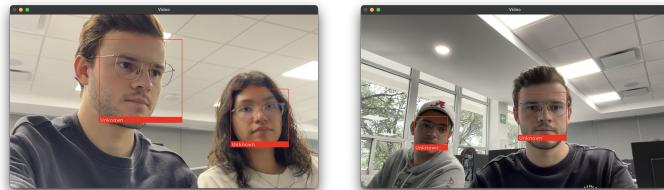
Por otro lado, en la imagen 6b se plantea el escenario cuando el rostro no se encuentra registrado, donde a diferencia del modelo anterior, se clasifica como “unknown”.



(a) Reconocimiento rostro sin
obstáculo (b) Reconocimiento de rostro con
obstáculo

Figura 6: Modelo `face_recognition` con parámetro de $tolerance = 0.6$ con obstáculo

Por último, se decidió probar con un valor de 0.4, como podemos observar en la imagen prueba 7a, se observa que aún con rostros despejados no predice correctamente y lo clasifica como “unknown”. De igual manera, en la imagen 7b podemos observar que se da esta misma clasificación cuando el rostro no se encuentra registrado.



(a) Reconocimiento rostros regis-
trados (b) Reconocimiento de rostro desco-
nocido

Figura 7: Error en el modelo `face_recognition` con parámetro de $tolerance = 0.4$

Es por esto, que considerando las pruebas con los distintos valores en el parámetro de tolerancia, se decide usar un valor de tolerancia de 0.6.

4. Comparación de modelos

Finalmente, para comparar modelos, se utilizaron diferentes poses al igual que ángulos y objetos en los rostros para identificar cuál modelo es el mejor para la problemática que se busca solucionar. En cuanto a reconocimiento de cara, evidentemente el mejor modelo entre el Haar Cascade y el `face_recognition` es evidentemente el último debido a que identifica correctamente a las personas al utilizar aprendizaje profundo y ver más allá de las áreas grises de las imágenes. Por otro lado, al mover los valores de $tolerance$, observamos que el valor que mejor se ajusta a nuestros datos de entrenamiento es definitivamente el de 0.6 ya que este identifica claramente a todas las personas que están registradas en el dataset al igual que a aquellas que no están registradas, al clasificarlas como “unknown”, ya que si tomamos el valor de 0.8 el modelo hace lo posible por etiquetar a todos aunque no estén registrados, minimizando a los “unknown”. Y finalmente, el tener una tolerancia muy pequeña, hace que los rostros registrados difícilmente sean clasificados como personas, sino

que los etiqueta como desconocidos, por lo que el modelo funciona mucho mejor con un *tolerance* = 0.6.

Como métrica de medición, se decidió utilizar todas las instancias en donde cada modelo con su respectivo parámetro tomaba los valores incorrectos de las personas, sin embargo, no valió la pena hacer un análisis de comparación ya que difícilmente los modelos captaban correctamente a las personas o las etiquetaba de manera correcta, mientras que el modelo con el parámetro igual a 0.6 simplemente clasificaba a las personas de manera correcta en casi todos los intentos, siendo alrededor de 19 de 20 veces que las clasificaba correctamente. Con esto en mente, podemos concluir que el mejor modelo para este tipo de situación es el *face_recognition* con una tolerancia cercana a 0.6 o inclusive a 0.5 si se busca ser muy específico con el registro de los alumnos en la clase que se llevará a cabo por el socio formador, con la única desventaja de que entre más sensible sea el modelo, más difícilmente podrá identificar correctamente a los alumnos al guardar una distancia de la cámara o al estar en un ángulo distinto a la foto de prueba.

Referencias

- [1] Samiksha Malhotra et al. «Comparison between attendance system implemented through haar cascade classifier and face recognition library». En: (feb. de 2021). URL: <https://iopscience.iop.org/article/10.1088/1757-899X/1022/1/012045/pdf>.
- [2] *Face Recognition*. URL: <https://face-recognition.readthedocs.io/en/latest/readme.html>.