



Inteligencia artificial avanzada para la ciencia de datos II (Gpo 501)

Innovación en la Gestión de Asistencias y Participaciones en Aulas Multitudinarias: Un Enfoque Pionero en Detección de Rostros y Reconocimiento de Postura con YOLOv8 y Deep Learning

Autores:

Juan Pablo Castañeda Serrano | A01752030
Aldo Daniel Villaseñor Fierro | A01637907
Francisco Castorena Salazar | A00827756
José Alfredo García Rodríguez | A00830952

Profesores de Reto:

Ivan Mauricio Amaya Contreras
Edgar Covantes Osuna
Hugo Terashima Marìn

Noviembre 2023

Índice

1. FaceTime - Deep learning based face recognition attendance system [1]	2
2. Performance Evaluation and Comparison of Software for Face Recognition, Based on Dlib and Opencv Library [2]	2
3. FAREC - CNN based efficient face recognition technique using Dlib [3]	2
4. Face image recognition based on convolutional neural network [4]	3
5. Métricas Seleccionadas	3
6. Pruebas y comparaciones	4
6.1. Pruebas sobre el Algoritmo de Participación	4
6.1.1. Prueba Video 1 (Fondo Estable)	4
6.2. Pruebas para verificar el funcionamiento del Sistema Integrado	4
7. Resultados	5
7.1. Sobre Pruebas de Reconocimiento Facial	5
7.1.1. VGG15 - face	6
7.1.2. Meta - DeepFace	6
7.1.3. Comparación y selección	7
7.2. Sobre Pruebas de Algoritmo de Participación	7
8. Conclusiones	8
9. Anexos	9

1. FaceTime - Deep learning based face recognition attendance system [1]

En este documento se propone un nuevo sistema de asistencia basado en reconocimiento facial mediante aprendizaje profundo, aprovechando los avances recientes en las redes neuronales convolucionales profundas (CNNs) para tareas de detección y reconocimiento facial. El objetivo principal de esta investigación fue aplicar de manera práctica estos enfoques de aprendizaje profundo de vanguardia en tareas de reconocimiento facial. Dado que las CNNs logran los mejores resultados con conjuntos de datos más grandes, lo cual no es común en entornos de producción, el desafío principal fue aplicar estos métodos en conjuntos de datos más pequeños. Se propone un nuevo enfoque para la ampliación de imágenes en tareas de reconocimiento facial. La precisión general fue del 95.02 % en un conjunto de datos reducido de imágenes originales de empleados en un entorno de tiempo real. Se sugiere que el modelo de reconocimiento facial propuesto podría integrarse en otro sistema con o sin algunas modificaciones menores como componente principal o de apoyo para fines de monitoreo y otra de sus debilidades sería la necesidad de que el sistema esté constantemente detectando y monitoreando en busca de los rostros de los empleados, considerando que en 3 meses fueron menos de 300 entradas/salidas por cada uno así que esto no es necesario.

$$\text{Precisión} = \frac{\text{Verdaderos positivos}}{\text{Verdaderos positivos} + \text{Falsos positivos}} \times 100 \%$$

2. Performance Evaluation and Comparison of Software for Face Recognition, Based on Dlib and Opencv Library [2]

El artículo ofrece una visión general y un análisis de la complejidad temporal de los algoritmos de visión por computadora para el reconocimiento facial. Se centra en comparar dos bibliotecas populares de visión por computadora, OpenCV y dlib, explorando sus características, analizando sus ventajas y desventajas, y determinando en qué situaciones cada una es más adecuada. Se desarrollaron tecnologías de visión por computadora para el reconocimiento facial y se realizó una investigación exhaustiva de estas dos bibliotecas. Se analizan sus características, se evalúan sus pros y contras, y se presentan ejemplos de aplicaciones de reconocimiento facial basadas en histogramas orientados a gradientes para la detección facial, estimación de puntos de referencia para la orientación facial y redes neuronales convolucionales profundas para la comparación con rostros conocidos. El artículo resume el concepto de reconocimiento facial, describiendo la base científica para ello y la construcción de un sistema completo de reconocimiento. Se formulan los principios básicos de los programas para el reconocimiento facial y se realiza un análisis comparativo de la productividad de ambas bibliotecas en relación con el tiempo de ejecución y el número de iteraciones de los algoritmos aplicados. Se crearon dos aplicaciones simples de reconocimiento facial utilizando estas bibliotecas y se comparó su rendimiento. En la evaluación del modelo, se midió el tiempo que los algoritmos de reconocimiento facial tardaban en identificar a las personas, observándose que tanto con OpenCV como con dlib el tiempo de reconocimiento era menor a 2 segundos, siendo menos de 1 segundo en el caso de OpenCV y siendo el que tuvo mejor rendimiento de los 2 y cito 'the OpenCV library is more productive, has better performance for face detection and detection' [2].

3. FAREC - CNN based efficient face recognition technique using Dlib [3]

El artículo destaca la creciente atención que ha recibido el reconocimiento facial en las últimas décadas, tanto en la investigación como en los mercados comerciales. Propone una técnica eficiente para un sistema de reconocimiento facial basado en Deep Learning utilizando Redes Neuronales Convolucionales (CNN) con alineación facial Dlib. Describe detalladamente el proceso involucrado en el reconocimiento facial, como la

alineación y extracción de características faciales. Además, resalta la importancia de la alineación facial, ya que la precisión y la Tasa de Falsas Aceptaciones (FAR) se observan utilizando la técnica propuesta. El análisis computacional muestra un rendimiento superior a otros enfoques de vanguardia. El estudio se realizó en el conjunto de datos del Face Recognition Grand Challenge (FRGC), logrando una precisión del 96 % con una FAR de 0.1.

Se evaluaron la Tasa de Aceptación Verdadera (TAR) y la Tasa de Falsas Aceptaciones (FAR) en el conjunto de datos FRGC para el reconocimiento facial. La TAR se utiliza para medir la aceptación o reconocimiento correcto de rostros y es el porcentaje de veces que un sistema reconoce correctamente la identidad verdadera de una persona. La FAR es la medida de la probabilidad de que el sistema de reconocimiento acepte incorrectamente a una persona no autorizada y es la relación entre el número de aceptaciones falsas y el número total de intentos de reconocimiento.

$$FAR = \frac{\text{Número de falsas aceptaciones}}{\text{Número total de intentos de verificación}} \times 100 \%$$

4. Face image recognition based on convolutional neural network [4]

El texto presenta el sistema FaceNet, que aborda los desafíos en la implementación eficiente y a gran escala de la verificación y reconocimiento facial, a pesar de los avances recientes en el campo del reconocimiento facial. FaceNet aprende directamente una representación de imágenes faciales en un espacio euclidiano compacto donde las distancias reflejan la similitud facial. Este espacio permite la implementación sencilla de tareas como reconocimiento, verificación y agrupación facial utilizando incrustaciones de FaceNet como vectores de características. Se emplea una red convolucional profunda para optimizar directamente la incrustación, a diferencia de enfoques previos que usaban una capa intermedia como cuello de botella. El entrenamiento se realiza con tripletes de parches faciales alineados y relacionados/no relacionados, generados mediante un método de minería de tripletes en línea novedoso. La ventaja de este enfoque es una eficiencia representacional mucho mayor, logrando un rendimiento de reconocimiento facial de vanguardia utilizando solo 128 bytes por rostro. En el conjunto de datos Labeled Faces in the Wild (LFW), el sistema alcanza una precisión récord del 99.63 %, y en YouTube Faces DB logra un 95.12 %. Se reduce la tasa de error en un 30 % en comparación con los mejores resultados publicados en ambos conjuntos de datos. La clasificación de precisión obtenida es de $98.87\% \pm 0.15$ con el método de corte fijo y un asombroso $99.63\% \pm 0.09$ con alineación adicional. Este rendimiento supera en más de siete veces el error reportado para DeepFace y en un 30 % al estado del arte previo reportado para DeepId2+. Incluso el modelo más pequeño, NN3, alcanza un rendimiento que no es estadísticamente diferente.

$$\text{Precisión} = \frac{\text{Verdaderos positivos}}{\text{Verdaderos positivos} + \text{Falsos positivos}} \times 100 \%$$

5. Métricas Seleccionadas

A lo largo del desarrollo del reto tuvimos varios desafíos que se nos fueron presentando y es por eso que tuvimos que probar entre diferentes soluciones o maneras de acercarnos a la problemática para poder resolver satisfactoriamente el reto, específicamente para el caso de face recognition fue necesario probar varias alternativas de entre las cuales seleccionamos la mejor tomando en cuenta 3 factores principalmente:

1. El tiempo de procesamiento.
2. La proporción de personas detectadas correctamente (Recall).

3. Consistencia del algoritmo en condiciones no favorables (iluminación, ángulo, etc.) (accuracy / exactitud).

Entonces para evaluar nuestro algoritmo de reconocimiento de caras decidimos hacer algo similar a lo que se hace con los algoritmos de clasificación e interpretar los resultados con una matriz de confusión y un reporte de clasificación. Además revisando en la literatura decidimos agregar algunas métricas extras, como:

Tomando las métricas más importantes del reporte de clasificación específicamente para nuestro problema:

Exactitud (Accuracy)

$$\text{Accuracy} = \frac{\text{Número de predicciones correctas}}{\text{Número total de muestras}}$$

Recall (Sensibilidad o Exhaustividad):

$$\text{Recall} = \frac{\text{Verdaderos positivos}}{\text{Verdaderos positivos} + \text{Falsos negativos}}$$

Puntuación F1:

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

6. Pruebas y comparaciones

6.1. Pruebas sobre el Algoritmo de Participación

A lo largo del proyecto, se realizaron varios videos de prueba, de varias personas en el aula levantando la mano para detectar sus participaciones.

Se realizó la prueba principal con un video de duración de aproximadamente 3 minutos, el video tenía 5 personas, el video simulaba el levantamiento de las manos de las personas, este video fue realizado en un salón de clases donde el fondo no cambiaba, véase figura 1 y se tenía buena iluminación.

6.1.1. Prueba Video 1 (Fondo Estable)

- Número de Personas: 5
- Duración del Video: 2 minutos, 44 segundos.
- Tiempo de Procesamiento en backend: 4 minutos, 30 segundos.

De observaciones sobre el video se tiene que era una ventaja tener un fondo estable sin interferencias externas, se obtuvo un mejor rendimiento general.

6.2. Pruebas para verificar el funcionamiento del Sistema Integrado

En esta sección se busca ver que las distintas partes que conforman nuestro sistema de registro de asistencias y participaciones funcionen bien de forma conjunta. Para esto se verificará que el flujo de datos es correcto, con esto nos referimos a que la interfaz sea capaz de, registrar un nuevo alumno, grabar un video de asistencias, grabar un video de clases, que estos sean almacenados correctamente en base de datos y que después sean procesados en nuestro backend, para que finalmente los registros sobre los alumnos y las clases



Figura 1: Imagen sobre el Video 1 de Prueba, se observa que el fondo es estable.

se actualizen y también se muestren las distintas gráficas e indicadores en la sección de Estadísticas de la UI actualizadas.

A continuación se mostrarán los elementos que serán evaluados para saber si la interfaz funciona de forma correcta.

- Log in correcto a la interfaz web, indicador de los datos de estudiantes siendo cargados a la interfaz.
- Generación correcta de gráficas y estadísticas al ser actualizada la base de datos.
- Proceso de grabar y procesar videos de asistencias llamando al backend desde el servidor en la interfaz de usuario.

7. Resultados

7.1. Sobre Pruebas de Reconocimiento Facial

face_recognition **Matriz de confusión**

	Alfredo	Aldo	Juan	Pablo	Paco	Unknown
Alfredo	9	1	0	0	0	0
Aldo	1	8	1	0	0	0
Juan	0	0	9	1	0	0
Pablo	0	1	0	8	1	0
Paco	0	1	0	0	9	0
Unknown	0	1	2	0	1	6

Tiempo: 2 segundos.

Reporte de clasificación

Clase	Precisión	Recall	F1-score	Soporte
Alfredo	0.90	0.90	0.90	10
Aldo	0.67	0.80	0.73	10
Juan	0.75	0.90	0.82	10
Pablo	0.89	0.80	0.84	10
Paco	0.82	0.90	0.86	10
Unknown	1.00	0.60	0.75	10
Exactitud	0.82 (60 muestras)			
Promedio	0.84	0.82	0.82	60

Cuadro 1: Métricas de clasificación

7.1.1. VGG15 - face

Matriz de confusión

	Alfredo	Aldo	Juan	Pablo	Paco	Unknown
Alfredo	10	0	0	0	0	0
Aldo	0	9	1	0	0	0
Juan	0	0	9	1	0	0
Pablo	1	0	0	8	1	0
Paco	0	1	0	0	9	0
Unknown	1	1	2	0	1	5

Tiempo: 5 segundos

Reporte de clasificación

Clase	Precisión	Recall	F1-score	Soporte
Alfredo	0.83	1.00	0.91	10
Aldo	0.82	0.90	0.86	10
Juan	0.75	0.90	0.82	10
Pablo	0.89	0.80	0.84	10
Paco	0.82	0.90	0.86	10
Desconocido	1.00	0.50	0.67	10
Exactitud	0.83 (60 muestras)			
Promedio	0.85	0.83	0.83	60

Cuadro 2: Métricas de clasificación

7.1.2. Meta - DeepFace

Matriz de confusión

	Alfredo	Aldo	Juan	Pablo	Paco	Unknown
Alfredo	10	0	0	0	0	0
Aldo	0	9	1	0	0	0
Juan	0	0	9	1	0	0
Pablo	1	0	0	10	0	0
Paco	0	1	0	0	9	0
Unknown	1	1	0	1	0	8

Tiempo 11 segundos.

Reporte de clasificación

Clase	Precisión	Recall	F1-score	Soporte
Alfredo	0.83	1.00	0.91	10
Aldo	0.82	0.90	0.86	10
Juan	0.90	0.90	0.90	10
Pablo	0.83	0.91	0.87	11
Paco	1.00	0.90	0.95	10
Desconocido	1.00	0.73	0.84	11
Exactitud	0.89 (62 muestras)			
Promedio	0.90	0.89	0.89	62

Cuadro 3: Reporte de clasificación

7.1.3. Comparación y selección

Modelo Promedio F1-score	Tiempo (seg)	Exactitud	Promedio Precisión	Promedio Recall
Modelo 1 (face_recon) 0.82	2	0.82	0.84	0.82
Modelo 2 (VGG15 - face) 0.83	5	0.83	0.85	0.83
Modelo 3 (Meta - DeepFace) 0.89	11	0.89	0.90	0.89

Cuadro 4: Comparación de métricas de rendimiento de modelos

Podemos ver en los resultados de nuestras evaluaciones que los modelos tuvieron rendimientos bastante similares entre ellos en 4, sin embargo, un aspecto clave en nuestra solución del reto es que planeamos que el procesamiento sea lo mas ligero y por consecuencia barato que se pueda, esto para que pueda ser implementado sin impactar en el bolsillo del cliente.

Es por eso que tomando en cuenta que el algoritmo que toma mucho menos tiempo que los demás, además de obtener unas métricas buenísimas considerando que no eran las mejores condiciones, cosa contraria a un salón de clases en donde el ambiente por lo regular siempre estará en buenas condiciones. Considerando todo lo anterior decidimos decantarnos por el algoritmo de face_recognition que nos brinda de todo lo que necesitamos a un bajo coste computacional.

7.2. Sobre Pruebas de Algoritmo de Participación

A continuación se muestra la matriz de confusión correspondiente al video de prueba.

		Real		Total
		Positive	Negative	
Predicha	Positive	62	0	62
	Negative	16	0	16
Total		78	0	

En este caso, no se presentan verdaderos negativos debido a que serían demasiados por frame de vídeo, nuestro algoritmo no registra cuando un alumno no está participando, solo cuando este participa, sin embargo podemos darnos una muy buena idea del desempeño tan solo mirando los verdaderos positivos y los Falsos negativos.

Teniendo en cuenta la métrica de accuracy

$$Accuracy = \frac{VP + VN}{VP + VP + FN + TN}$$

Vemos que el accuracy en la prueba fue del 79 %.

Teniendo en cuenta la métrica de precisión

$$precision = \frac{VP}{VP + FP}$$

Se observa que se obtuvo una precisión de 1, por lo que esta métrica no es muy útil en nuestro caso.

8. Conclusiones

La evaluación del modelo de detección de poses revela un buen rendimiento global con un accuracy del 79 % en el caso de prueba. Este indicador sugiere que el modelo realiza predicciones precisas en la mayoría de los casos. Sin embargo, se observa una tasa significativa de falsos negativos, indicando que algunas participaciones no se detectan de manera óptima, una consideración crítica en un contexto educativo donde cada participación es valiosa.

Como sugerencias de mejora, se propone experimentar con la modificación del umbral de detección para equilibrar la sensibilidad y precisión del modelo, lo que podría reducir la tasa de falsos negativos. Otra oportunidad para mejorar el rendimiento del algoritmo se encuentra en implementar condiciones basadas en la distancia de un alumno detectado como `Unknown` un alumno que si fue detectado en frames pasado, ya que hay ciertas condiciones en las que el algoritmo falla por errores en el reconocimiento facial, pero que con arreglos en el código podrían solucionarse esos errores aprovechándonos de los supuestos de los que partimos del código y del correcto funcionamiento de los modelos en frames anteriores. Además, explorar modelos alternativos de detección de poses puede proporcionar nuevas perspectivas y mejorar la capacidad del sistema para capturar poses de manera más efectiva.

La limitación de recursos computacionales en las computadoras personales de los integrantes del proyecto ha impactado en el tiempo de procesamiento, especialmente al procesar vídeos de participación. Se sugiere buscar recursos computacionales más sofisticados para implementar algoritmos de detección de poses más robustos y mejorar las velocidades de procesamiento, especialmente si se ejecuta el backend localmente.

Referencias

- [1] M. Arsenovic, S. Sladojevic, A. Anderla, and D. Stefanovic, "Facetime — deep learning based face recognition attendance system," in *2017 IEEE 15th International Symposium on Intelligent Systems and Informatics (SISY)*, 2017, pp. 000 053–000 058.
- [2] N. Boyko, O. Basystiuk, and N. Shakhovska, "Performance evaluation and comparison of software for face recognition, based on dlib and opencv library," in *2018 IEEE Second International Conference on Data Stream Mining Processing (DSMP)*, 2018, pp. 478–482.

- [3] S. Sharma, K. Shanmugasundaram, and S. K. Ramasamy, “Farec — cnn based efficient face recognition technique using dlib,” in *2016 International Conference on Advanced Communication Control and Computing Technologies (ICACCCT)*, 2016, pp. 192–195.
- [4] G. Lou and H. Shi, “Face image recognition based on convolutional neural network,” *China Communications*, vol. 17, no. 2, pp. 117–124, 2020.

9. Anexos

Repositorio de github: <https://github.com/a01752030/PortafolioRetoFC>

*En caso de requerir acceso, mandar correo a cualquier integrantes del equipo A00830952@tec.mx