



Tecnológico de Monterrey

Instituto Tecnológico y de Estudios Superiores de Monterrey

Momento de Retroalimentación: Reto Evaluación

TC3007C.501 Inteligencia Artificial Avanzada para la Ciencia de Datos II

Profesores:

Iván Mauricio Amaya Contreras

Blanca Rosa Ruiz Hernández

Félix Ricardo Botello Urrutia

Edgar Covantes Osuna

Felipe Castillo Rendón

Hugo Terashima Marín

Equipo 2

Integrantes:

Luis Ángel Guzmán Iribe – A01741757

Julian Lawrence Gil Soares – A00832272

Alberto H Orozco Ramos – A00831719

8 de Noviembre de 2023

1. Realicen una búsqueda de literatura reciente acerca de las métricas utilizadas por otros autores para evaluar el desempeño de un modelo, en problemas similares (puede ser la identificación de rostro, reconocimiento de participación, estimación de pose, etc.). Se recomienda que realicen la búsqueda a través de Scopus (<https://www.scopus.com/home.uri>Links to an external site.)
2. Seleccionen por lo menos cuatro (4) artículos y realicen un resumen (un párrafo corto por artículo) donde mencionan lo que los autores hicieron, la métrica que utilizaron, y alguna debilidad de la solución propuesta por los autores

Durante nuestra investigación encontramos que para los modelos de reconocimiento de objetos se utiliza la precisión como la principal métrica de evaluación. Específicamente utilizamos una comparación de los verdaderos positivos y verdaderos negativos contra los falsos positivos y falsos negativos. En nuestra investigación encontramos la siguiente ecuación: $Precisión = \frac{TP}{TP+FP}$ $Recall = \frac{TP}{TP+FN}$ $F-1 score = \frac{2*recall*precisión}{recall+precisión}$ Estas ecuaciones nos permiten evaluar el desempeño de nuestro modelo a través de medir cuántas veces nuestro modelo atina en su reconocimiento.

Otra metodología de evaluación de desempeño que encontramos que se utiliza como un estándar es el llamado intersection over union (IOU). La manera en la que se utiliza se puede describir con la siguiente ecuación: $IoU = \frac{Area\ of\ Union}{Area\ of\ Intersection}$. Este método evalúa la ubicación y tamaño de la caja que la computadora pone encima del objeto que se detectó. El área de unión se refiere al área total del área de el objeto real dentro de la imagen y la caja de la precisión que el algoritmo arroja, el área de intercesión se refiere a cuánta área del objeto real está dentro de la caja de previsión. Su división forma un rango de 0 a 1 que representa básicamente que tan certero es la predicción arrojada y este número se puede utilizar para evaluar el rendimiento del algoritmo.

En un artículo que trata acerca del uso de computer vision y modelos de deep learning para la detección de parpadeos voluntarios en pacientes que padecen de esclerosis lateral amiotrófica. El artículo trata con la implementación de un modelo que funciona en diferentes capas o fases, de manera similar al nuestro, con la primera fase siendo la detección del ojo, la cual emplea la técnica de IOU descrita anteriormente para el entrenamiento del modelo detección de posición del ojo. Posteriormente, se realizan diferentes pruebas de video para evaluar la capacidad del modelo, en las que se define un parpadeo como la superposición del IOU en el sobre la muestra en el dataset, es decir, cuando se detecta una modificación en el área de interés en una determinada cantidad de tiempo. Una vez contabilizados los parpadeos se buscan las superposiciones en el tiempo con la muestra real, y se obtiene la matriz de confusión que da a conocer los resultados positivos y negativos, verdaderos y falsos, que a su vez permiten conocer las métricas de desempeño descritas anteriormente.

Otro artículo (De Lima Medeiros et al., 2022) que se revisó habla sobre el desarrollo de aplicaciones web que implementan deep learning, su respectivo deployment, enfoque de desarrollo y al final se mencionan algunos ejemplos. Resumiendo un poco de lo que trata el artículo, en primera instancia se abordan las aplicaciones web que implementan modelos de Deep Learning del lado del cliente en los navegadores. El autor comparte su opinión diciendo que este tipo aplicaciones poseen muchas ventajas como amplia búsqueda, decremento en costos para servidores y privacidad de los datos de los usuarios. También,

hay secciones del documento dónde se discute el porqué los navegadores, el lenguaje de programación JavaScript y la librería de TensorFlow.js son ideales para este conjunto, así como las restricciones y consideraciones involucradas. Igualmente, se describen 4 enfoques para la creación y optimización de modelos de Deep Learning que mejor se ajusten para aplicaciones web por el lado del frontend. Dichos enfoques son: *reusar*, *personalizar*, *convertir* y *optimizar* modelos existentes, y resaltar sus ventajas y desventajas. Finalmente, en el artículo contiene una sección donde incluye algunos proyectos de desarrollo web que hacen uso de Deep Learning, tales como “*Education playground Teachable Machine*”, “*Visualization playground GAN Lab*” o “*Development playground Milo*”, sin embargo, existen algunas en particular que poseen elementos que utilizamos en el reto tales como detección de pose y gestos. Algunos ejemplos que se mencionan son “*Rehabilitation and monitoring*”, “*Physical activity coaching*”, y “*Gesture tracking*”. Dichos ejemplos utilizan principalmente el modelo de pose

“*PoseNet*” para estimar la postura y los gestos de las personas en diferentes escenarios, ya sea para rehabilitación, monitoreo, evaluación de riesgos, detección de caídas, baile y entrenamiento físico; dichos modelos hacen uso de dos cámaras o la integración de voz para mejorar la interacción y precisión de los resultados.

En cuanto a la relación entre los modelos propuestos por el autor y nuestro modelo existen diversos puntos a considerar. Por ejemplo, nosotros si hacemos uso de servicios en la nube para implementar los modelos de Deep Learning pre entrenados que utilizamos para la toma de asistencia y participación de los alumnos. Otra diferencia es que el autor menciona el uso de Tensorflow.js para la simplificación durante la implementación de los modelos, cosa que no contemplamos desde un inicio con el fin de evitar recurrir a servidores de la nube para ejecutar los modelos y elevar el costo computacional. Algunas de las métricas que utilizan este tipo de modelos son:

- *Focal Loss Function*: Función de pérdida computarizada para manejar la pérdida de datos por cada ejemplo analizado y reajusta los pesos del modelo acorde a su dificultad, la cual se ajusta al uso de su entrenamiento.
- *Mean Average Precision* y *IoU*: Estas métricas utilizadas para evaluar la exactitud del modelo miden el porcentaje de las predicciones correctas por cada una de los valores a predecir que son desconocidos por el modelo. Por otro lado, IoU evalúa la superposición entre los valores predichos y las cajas delimitadoras de verdad.

3. Seleccionen métricas (o indicadores de desempeño) a medir adecuadas al reto y justifiquen su selección.










Debido a que utilizamos modelos pre entrenados, buscaremos evaluar el desempeño de uno de los modelos, en nuestro caso será el modelo de reconocimiento facial, realizando pruebas específicas para medir su efectividad reconociendo a cada uno de los integrantes del equipo en distintas condiciones.




4. Evalúen los modelos usando las métricas (o indicadores de desempeño).




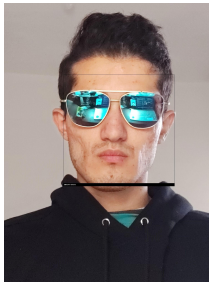




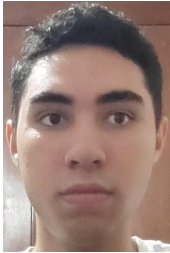






Se evaluará el modelo de forma sencilla, marcando un reconocimiento facial exitoso, es decir, que reconoció el rostro de la persona y colocó un recuadro con su nombre en el rostro de la persona, o bien, que falló en reconocer a la persona. Esto se hace debido a que como no llevamos a cabo el entrenamiento y los modelos son pre entrenados, la mejor forma de evaluar el rendimiento del modelo de reconocimiento facial sería verificar si reconocer a cierta persona o no. Para evaluarlo y asegurarnos de que efectivamente encuentra a la

persona correcta, se cargaran todas las caras codificadas de cada uno de los integrantes del equipo en cada prueba, y posteriormente se evaluará el desempeño del modelo en base a los escenarios, condiciones o métricas (fotos en donde alguno de los integrantes aparezca) establecidos en la siguiente tabla:

- 5. Generen comparaciones entre los modelos con las métricas (o indicadores de desempeño).
- 6. Grafiquen y visualicen los resultados obtenidos.

Condiciones Evaluadas	Alberto	Julian	Luis
Rostro	 <p>Imagen de referencia</p>	 <p>Imagen de referencia</p>	 <p>Imagen de referencia</p>
	 <p>Imagen de prueba</p>	 <p>Imagen de prueba</p>	 <p>Imagen de prueba</p>
	 <p>Reconocimiento exitoso</p>	 <p>Reconocimiento exitoso</p>	 <p>Reconocimiento exitoso</p>

Utilizando una gorra	 <p>Imagen de referencia</p>	 <p>Imagen de referencia</p>	 <p>Imagen de referencia</p>
	 <p>Imagen de prueba</p>	 <p>Imagen de prueba</p>	 <p>Imagen de prueba</p>
	 <p>Reconocimiento exitoso</p>	 <p>Reconocimiento exitoso</p>	 <p>Reconocimiento exitoso</p>
Utilizando lentes	 <p>Imagen de referencia</p>	 <p>Imagen de referencia</p>	 <p>Imagen de referencia</p>

	 <p>Imagen de prueba</p>	 <p>Imagen de prueba</p>	 <p>Imagen de prueba</p>
	 <p>Reconocimiento exitoso</p>	 <p>Reconocimiento exitoso</p>	 <p>Reconocimiento exitoso</p>
	 <p>Imagen de referencia</p>	 <p>Imagen de referencia</p>	 <p>Imagen de referencia</p>
	 <p>Imagen de prueba</p>	 <p>Imagen de prueba</p>	 <p>Imagen de prueba</p>
Utilizando una hoodie	 <p>Imagen de prueba</p>	 <p>Reconocimiento exitoso</p>	 <p>Reconocimiento exitoso</p>

	Reconocimiento exitoso		
--	------------------------	--	--

7. Utilicen los elementos anteriores para poder hacer una interpretación clara y de calidad

Como podemos observar en la tabla, el modelo de reconocimiento facial es muy preciso y certero en los casos de prueba que definimos para evaluarlo. Debido a cuestiones de tiempo no nos fue posible probar con más casos, pero del conjunto que logramos registrar, existen algunos casos en los cuáles la persona fue sorprendentemente reconocida. El modelo logra reconocer personas a pesar de que llevan lentes (más sorprendente lo hace con lentes de sol), también si tienen puesta una gorra o bien una hoodie, esto nos deja como conclusión que el modelo fue entrando exhaustivamente para lograr este nivel de precisión. A pesar de esto, no descartamos que en algunas ocasiones no sea posible reconocer a las personas debido a las condiciones de iluminación, la calidad de las imágenes tanto de entrada como de prueba, el ángulo de la cara con respecto a la cámara, la distancia, entre otras variables que pueden afectar al rendimiento del modelo, aún así y con varias dificultades, logra arrojar resultados impecables..

Referencias

- Farkhod, A., Abdusalomov, A. B., Mukhiddinov, M., & Cho, Y.-I. (2022). Development of Real-Time Landmark-Based Emotion Recognition CNN for Masked Faces. *Sensors* (14248220), 22(22), 8704. <https://0-doi-org.biblioteca-ils.tec.mx/10.3390/s22228704>
- Goh, H.-A., Ho, C.-K., & Abas, F. S. (2023). Front-end deep learning web apps development and deployment: a review. *Applied Intelligence*, 53(1), 15923-15945.
- Jeune, P. L., & Mokraoui, A. (2023). Rethinking Intersection Over Union for Small Object Detection in Few-Shot Regime.
- De Lima Medeiros, P. A., Da Silva, G. V. S., Fernandes, F., Sánchez-Gendriz, I., De Castro Lins, H. W., Da Silva Barros, D. M., Nagem, D. a. P., & De Medeiros Valentim, R. A. (2022). Efficient machine learning approach for volunteer eye-blink detection in real-time using webcam. *Expert Systems With Applications*, 188, 116073. <https://doi.org/10.1016/j.eswa.2021.116073>