

Instituto Tecnológico y de Estudios Superiores de Monterrey

Inteligencia Artificial Avanzada para la Ciencia de Datos II (Gpo 501)



**Tecnológico
de Monterrey**

Momento de Retroalimentación: Reto Evaluación

Equipo 5

Jorge Eduardo De León Reyna - A00829759

David Esquer Ramos - A01114940

Francisco Mestizo Hernández - A01731549

Adrián Emmanuel Faz Mercado - A01570770

Noviembre 15, 2023

Análisis de la literatura

En esta sección del documento se presentan cuatro artículos científicos donde se habla de problemas similares al de tomar asistencia o participación de una clase, cómo se resolvieron utilizando inteligencia artificial y las ventajas y desventajas de hacerlo así.

An attention-based CNN for automatic whole-body postural assessment (1):

En este trabajo, se desarrolla un enfoque de evaluación postural totalmente automático mediante una red neuronal convolucional basada en la atención. La red aprende a identificar regiones cruciales del cuerpo y extraer automáticamente características para estimar el riesgo postural. La evaluación del método se realiza utilizando el conjunto de datos pH36M, creado a partir de Human3.6M, y se utiliza el criterio Rapid Entire Body Assessment (REBA). Para medir el rendimiento, se emplean el error absoluto medio (MAE) y el coeficiente Kappa de Cohen en lugar de la precisión, dado que esta última no es fiable en distribuciones desequilibradas de riesgos. Los resultados experimentales muestran un rendimiento prometedor, con un valor de Kappa de 0.73, indicando una correlación sustancial entre las predicciones del modelo y las evaluaciones humanas, abordando de manera efectiva la distribución desigual de riesgos.

A Convolutional Neural Network (CNN) Based Approach for the Recognition and Evaluation of Classroom Teaching Behavior

En el artículo, los autores proponen un modelo basado en redes neuronales convolucionales para identificar y evaluar comportamientos de enseñanza en el aula. Utilizan datos característicos del comportamiento docente obtenidos mediante Máquinas de Vectores de Soporte (SVM) y los procesan con CNN. Además, aplican un análisis jerárquico para calcular el peso de los indicadores de comportamiento docente y verificar su consistencia. Para evaluar el modelo, se enfocan en la precisión de la identificación del comportamiento en el aula y el error en la evaluación. Sin embargo, una limitación destacada es que la identificación y evaluación no pueden realizarse al mismo tiempo, y la eficacia del modelo depende en gran medida de la calidad de los datos de entrada para la SVM y la CNN.

Student attendance with face recognition (LBPH or CNN): Systematic literature review

En este artículo se habla sobre diferentes algoritmos que existen en la literatura para hacer la detección de la cara de una persona. Se menciona un algoritmo tradicional llamado

LBPH y uno utilizando Machine Learning llamado CNN. Además, de compararlos con otros algoritmos, el objetivo principal del paper fue decidir cuál de los dos es mejor para la detección de caras, de acuerdo a los resultados de accuracy encontrados en aproximadamente.

No se especifica porque se compara en los modelos el accuracy pero se entiende que lo que se busca es que el modelo sepa distinguir perfectamente a cada persona. Se encontraron resultados que las CNN son significativamente mejores, además que se desempeñan bien en escenarios con poca iluminación, con personas con lentes o con accesorios en la cabeza. La desventaja es que necesitan muchos datos para ser entrenadas, a comparación del LPBH. Igualmente se menciona que se necesitan aproximadamente 15 imágenes por usuario en diferentes condiciones para lograr hacer una detección con accuracy alto.

YOLO Series for Human Hand Action Detection and Classification from Egocentric Videos

En este artículo, se aborda la detección y clasificación de manos mediante redes YOLO en el contexto de aplicaciones de estimación de pose tridimensional y reconocimiento de actividades manuales. Se realiza un estudio detallado de las arquitecturas YOLO (versión 1 a 7), evaluando la eficiencia en conjuntos de datos de visión egocéntrica (EV) como FPHAB, HOI4D y RehabHand. La evaluación se basa en métricas como el valor de Intersección sobre Unión (IOU) para la detección de poses de manos, con un umbral TheshIOU. Además, se emplean medidas de precisión (P), sensibilidad (R), F1-Score (F1), y el mAP.5 (mean Average Precision) para la clasificación de acciones de mano. Los resultados indican que la red YOLOv7 y sus variantes destacan en la detección y clasificación de manos, logrando una precisión del 95% o más en los conjuntos de datos mencionados, con velocidades de procesamiento específicas para resoluciones de imagen definidas.

Evaluación del modelo utilizado

Una vez que se hizo la investigación de la literatura, se decidió probar el modelo utilizando el accuracy, ya que es el que más se utilizó, pero también se evaluarán otras métricas como el F1 score, el recall y el precision.

Las pruebas que se pueden realizar del modelo es grabar videos donde haya diferente cantidad de personas y que cada uno haga o no haga participaciones (levantar la mano). Para

evaluar el desempeño se pasarán estos videos por el sistema y se registrarán los FP, FN y VP. A continuación se muestra una tabla con el set de pruebas y los resultados obtenido:

Descripción	Participaciones reales	Participaciones detectadas	VP	FP	FN
Una persona cerca fuera de frame	3	5	3	2	0
Una persona cerca con manos semi levantadas pero no participaciones, por ejemplo tocándose la cara	3	4	3	1	0
Una persona sin hacer nada	0	0	0	0	0
Un video sin personas	0	0	0	0	0
Dos personas, una cerca y una lejos. Algunas participaciones son simultáneas	5 (3 y 2)	7 (4 y 3)	5	2	0
Dos personas lejos. Una persona deja la mano levantada mucho tiempo	5 (3 y 2)	6 (3 y 3)	5	1	0
Una persona con otra persona caminando en la escena	3	6	3	2	1

Ahora, basados en estos resultados, se pueden calcular las métricas que se mencionaron anteriormente.

Precision: 0.7037037037037037

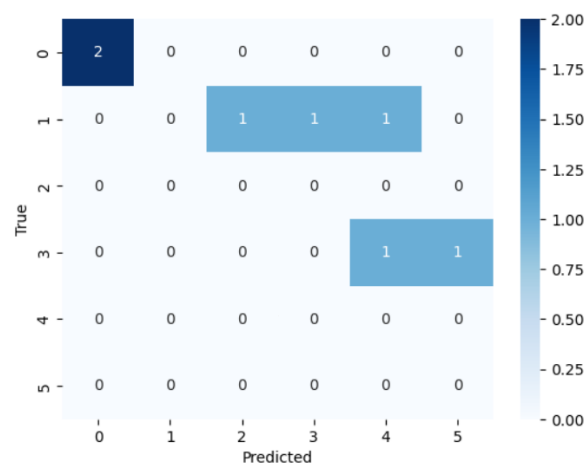
Recall: 0.95

F1 Score: 0.8085106382978724

Accuracy: 57.14285714285714%

Los resultados que se obtienen, indican que el modelo no lo hace tan mal en casos generales, pero tiende a fallar. El 57% de las participaciones fueron detectadas correctamente y el resto no. Es importante mencionar que estas mediciones se hicieron comparando si las participaciones totales de los videos fueron correctas o no.

Para verificar dónde está el origen de los resultados, se analizará la siguiente matriz de confusión.



La matriz marca la cantidad de participaciones que tenía que haber y las que el modelo contó. Por lo tanto, el hecho de que haya registros sobre la diagonal principal, indica que el modelo suele hacer predicciones de manos levantadas donde no las hay, pero no falla en detectar participaciones. Es decir, los alumnos nunca tendrán menos participaciones de las que deberían, pero sí pueden tener números inflados.

Esto se sustenta con el cálculo del error medio y del error porcentual. Aproximadamente el modelo detecta 2.7 participaciones de más.

Mean Squared Error: 2.7142857142857144

El 37% de las participaciones que asigna el sistema son erróneas.

Average Percentage Error: 37.142857142857146%

Conclusiones

El modelo hace buenas predicciones para las participaciones de los alumnos y no les cuenta de menos. Pero parece que es muy sensible, porque tiende a detectar participaciones donde no las hay. Incluso este comportamiento se ve cuando se marca más de una participación, aunque solo se haya levantado la mano una vez.

Esto quiere decir que el modelo se tiene que hacer más preciso al detectar participaciones y que no las detecte de más. Esto se puede lograr cambiando la forma en que se hace la detección al levantar la mano.

Por otro lado, el modelo es excelente en la detección de caras. Todas las personas en las pruebas de este documento fueron detectadas correctamente y se puede mejorar aún más disminuyendo la sensibilidad, para que no confunda caras similares

Bibliografía

Bhargavi, P. et al. (2022). *Face Recognition Attendance System for Online Classes*. Sardar Patel Institute of Technology: India.

<https://www.scopus.com/record/display.uri?origin=recordpage&zone=relatedDocuments&eid=2-s2.0-85146342954&noHighlight=false&sort=plf-f&src=s&sid=a87d2b9be098d82bb01af4345663e0e4&sot=b&sdt=b&sl=47&s=TITLE-ABS-KEY%28attendance+taking+neural+network%29&relpos=0>

Budiman, A. et al. (2023). *Student attendance with face recognition (LBPH or CNN): Systematic literature review*. Bina Nusantara University: Indonesia.

https://www.sciencedirect.com/science/article/pii/S187705092202186X?ref=pdf_download&fr=RR-2&rr=8264c5179d6749e9

Hung-Cuong, N. (2023). *YOLO Series for Human Hand Action Detection and Classification from Egocentric Videos*. Czestochowa University of Technology: Poland.

<https://www.scopus.com/record/display.uri?eid=2-s2.0-85151215953&origin=resultslist&sort=plf-f&src=s&sid=37590ee2bbeec52f39fc5ddb825398f5&sot=b&sdt=b&s=TITLE-ABS-KEY%28human+pose+AND+detection+AND+yolo%29&sl=29&sessionSearchId=37590ee2bbeec52f39fc5ddb825398f5>

Zewei, D. et al. (2023). *An attention-based CNN for automatic whole-body postural assessment*. University of Wollongong: Australia.
<https://www.scopus.com/record/display.uri?eid=2-s2.0-85175579705&origin=resultslist&sort=plf-f&src=s&sid=37590ee2bbeec52f39fc5ddb825398f5&sot=b&sdt=b&s=TITLE-ABS-KEY%28pose+estimation%29&sl=29&sessionSearchId=37590ee2bbeec52f39fc5ddb825398f5>