

DISEÑO DE UN FRAMEWORK PARA LA EXTRACCIÓN Y ANÁLISIS DE PUNTOS CLAVE BASADO EN DEEP LEARNING PARA LA MEJORA DE LA TÉCNICA EMPLEADA EN ACTIVIDADES FÍSICAS

Autor: Hernández Bas, Ignacio.

Director: Clemente Verdú, Álvaro.

Entidad Colaboradora: BigML, Inc

RESUMEN DEL PROYECTO

Las ramas del *Deep Learning* y *Pose Detection* se están desarrollando con rapidez, pero a pesar de esto su aprovechamiento es muy complejo. Para facilitar su uso, se ha desarrollado una aplicación que permite la detección y análisis de la técnica realizada por el usuario durante la ejecución de diferentes actividades físicas. A partir de la creación de un *dataset* y la evaluación del sistema sobre este conjunto de datos, se han obtenido resultados satisfactorios que muestran el gran potencial que posee la detección de poses en el ámbito del deporte.

Palabras clave: *Deep Learning, Pose Detection, keypoints*

1. Introducción

Durante los últimos años, la rama del *Deep Learning* propia del aprendizaje computacional (*Machine Learning*) ha aumentado en popularidad considerablemente gracias al constante desarrollo de las tecnologías y a la aparición de nuevas técnicas e investigaciones. Deep Learning se define como el conjunto de técnicas las cuales se basan en algoritmos que tratan de replicar la estructura y funcionamiento del cerebro humano a partir de la creación de Redes Neuronales. Entre las aplicaciones más populares se encuentra su uso para el desarrollo de automóviles que circulen de forma autónoma, la creación de modelos de lenguaje y de Procesamiento de Lenguaje Natural (NLP) empleadas por herramientas tales como ChatGPT [1] o para el procesamiento y clasificación de imágenes.

Del avance en técnicas de procesamiento de imágenes, surge la necesidad de la clasificación de eventos, tales como los movimientos o poses realizados por un individuo. La rama encargada de llevar a cabo estas detecciones recibe el nombre de *Pose Detection*, empleando técnicas donde los ordenadores tratan de replicar la forma en la que el ser humano asimila la información a través de los ojos. *Pose Detection* se centra en el diseño y creación de modelos que permitan localizar los puntos clave o *keypoints* de la entidad que está realizando la acción a analizar (por ejemplo, un humano). Estos puntos clave hacen referencia a las partes del cuerpo detectadas por el modelo de detección de pose, el cual devuelve las coordenadas en el espacio de cada uno de estos puntos en un instante de tiempo determinado.

Una vez se obtienen dichas coordenadas del modelo, otros sistemas pueden utilizar estos resultados y combinarlos con otros tipos de información, tal como el contexto o actividad que se está realizando, para obtener una visión más enriquecida de la acción empleada por el usuario.

A pesar de la existencia de aplicaciones que ya incorporan técnicas de detección de pose tales como la detección de posibles errores en la realización de posturas de yoga [2] o el análisis del swing empleado por un golfista [3], estas se limitan a realizar análisis sencillos que aportan una información limitada al usuario y en su mayoría, estos están solamente enfocados al análisis de un rango específico de acciones.

2. Definición del proyecto

Este proyecto define la creación de un sistema el cual permita acercar al usuario múltiples de las funcionalidades y herramientas relativas a la detección y análisis de poses en el ámbito del deporte y del ejercicio físico. Por lo tanto, se va a diseñar una aplicación de detección y análisis de actividades físicas, en la cual se le facilite al usuario la información relativa a sus movimientos realizados y los detalles para llevar a cabo las correcciones necesarias en el caso de que no los esté ejecutando de manera correcta en función del tipo de ejercicio que haya realizado.

Este proyecto esta principalmente enfocado en la detección de ejercicios físicos que un usuario podría realizar ya sea en su hogar o en el gimnasio y con la posibilidad de que pueda emplear diferentes objetos tales como pesas o barras para llevar a cabo su actividad.

3. Descripción del sistema

El sistema creado busca reducir el nivel de comprensión necesario para poder hacer uso de los beneficios de la detección de poses a partir de la creación de una herramienta de gran usabilidad que permita detectar con eficacia y precisión los eventos que suceden durante la realización de una actividad física. Para poder llevar a cabo esta función, el sistema se divide en tres bloques fundamentales, los cuales se muestran representados en la figura 1.

- **Aplicación Web:** gracias a la interfaz web creada en se permite al usuario interactuar con el sistema permitiendo que este pueda cargar sus respectivos vídeos para su procesado, análisis y representación de resultados.
- **Procesado y detección de poses:** sección del sistema encargada de realizar la ingesta del vídeo introducido por el usuario, llevando a cabo la detección de la pose del usuario en cada una de las imágenes (*frames*) que componen el vídeo. Dicha detección será realizada a partir del modelo de Pose Detection de código abierto *MoveNet* [4].
- **Análisis y obtención de eventos:** una vez ya se obtienen todas las imágenes del vídeo procesado con todas sus poses detectadas, se realiza el análisis de cada una de estas poses en función del tipo de ejercicio físico que el usuario ha ejecutado. Cada tipo de ejercicio supone que se deben de aplicar una serie de reglas determinadas para alcanzar la pose objetivo que se considera como correcta. Esto se realiza a partir de la agrupación de una serie de reglas generales, las cuales engloban el movimiento y posición de un conjunto de ciertas partes del cuerpo como las piernas, brazos o el torso.

Gracias a esto se dispone de un sistema adaptable que permite incorporar nuevas actividades físicas sin la necesidad de generar reglas específicas en función de la actividad que se desee incorporar a la aplicación. Tras el análisis, se lleva a cabo la síntesis de los resultados en eventos únicos que aporten información de utilidad al usuario donde se definen las acciones que se han realizado en el vídeo.

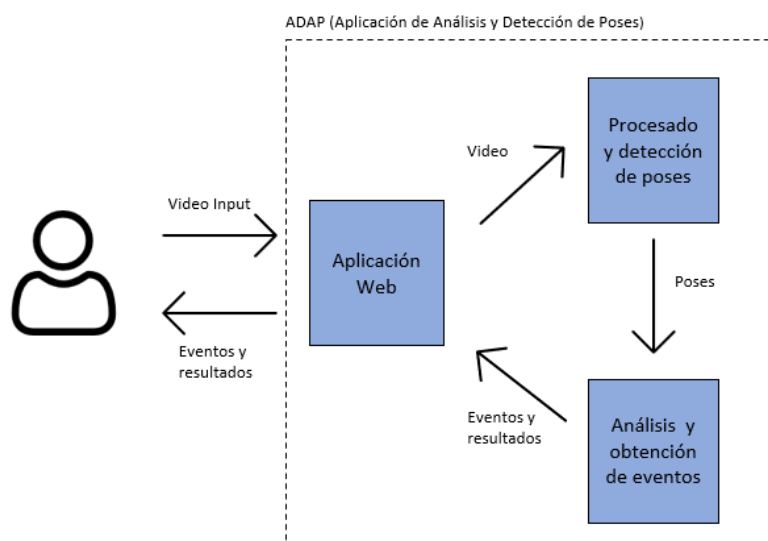


Figura 1 Secciones principales del sistema

4. Resultados

Tras la implementación del sistema, es importante evaluar el comportamiento del sistema, determinando si las predicciones realizadas por la sección de análisis se están realizando de forma correcta.



Figura 2 Pose objetivo alcanzada



Figura 3 Pose objetivo no alcanzada

Las figuras 2 y 3 muestran ejemplos de las imágenes resultantes del proceso de análisis y detección de un vídeo. Estas representaciones sirven como indicadores para el usuario sobre el rendimiento obtenido en la actividad. Las diferentes partes del cuerpo se resaltan en color verde o magenta, según si se encuentran dentro o fuera del rango objetivo de la pose que se debe alcanzar en función de la actividad.

Mediante la creación de un *dataset* donde se ha llevado a cabo el etiquetado de múltiples actividades físicas diferentes, se ha podido medir cuanto de preciso es el sistema a la hora de llevar a cabo las detecciones de las diferentes actividades físicas.

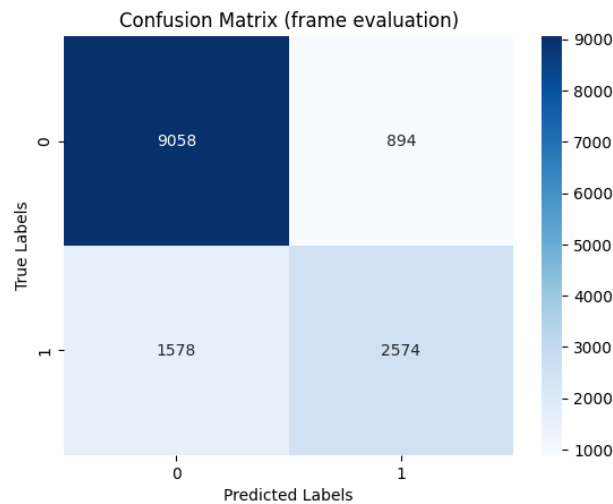


Figura 4 Matriz de confusión resultado de la predicción

Métrica	Valor
Accuracy	0.82
Precision	0.74
Recall	0.61
F1 Score	0.67

Tabla 1 Métricas de evaluación obtenidas

La figura 4 muestra la matriz de confusión resultado de la comparación entre todas las imágenes etiquetadas y las imágenes predichas por el sistema. Los resultados del *accuracy* muestran que más del 80% de los *frames* etiquetados se predicen correctamente, otras métricas como la *precision* indica que el 74% de las veces que se detecta una pose sin errores, esta también se había calificado como correcta.

Por otro lado, el *recall* supera el 60%, lo cual refleja que el sistema por lo general detecta correctamente los frames con una pose correcta, pero existen casos que el sistema no detecta. Ante estos resultados se puede afirmar que el funcionamiento del sistema es satisfactorio, pero presenta un posible margen de mejora en sus predicciones.

5. Conclusiones

Gracias a la creación de este proyecto, se ha podido presentar al usuario las utilidades que presenta la detección de poses a través del deporte, produciendo que se abra una nueva puerta a una innovadora forma de hacer ejercicio.

Los resultados obtenidos corroboran el buen funcionamiento del sistema, pero es de gran importancia mencionar que la cantidad de datos que se poseen y la variabilidad entre estos es determinante a la hora de llevar a cabo una evaluación. A mayor número de datos y situaciones que se poseen, más información se puede obtener de la evaluación de los datos predichos, obteniendo mejores conclusiones acerca del funcionamiento del sistema. Por lo que se recomienda tomar el *dataset* establecido como un punto de partida útil que permite concretar la precisión actual que presenta el sistema.

Respecto a los posibles trabajos futuros, sería interesante la composición de un dataset especializado en el marco de trabajo del proyecto, el cual presente una extensión mucho mayor y concentre numerosas actividades físicas, realizadas por una gran cantidad de personas y desde un mayor número de perspectivas distintas que las que presenta el conjunto de datos actual.

Por otro lado, se recomienda la investigación de métodos de *Pose Detection* que incorporen múltiples cámaras a la hora de realizar la grabación de la actividad. A pesar de que MoveNet es un modelo que aporta buenos resultados con el uso de una sola cámara, un análisis en tres dimensiones de la acción podría suponer una mejora en la detección de los puntos clave del usuario y por lo tanto obtener una mayor precisión en el análisis de las actividades físicas.

6. Referencias

- [1] ABDULLAH, Malak; MADAIN, Alia; JARARWEH, Yaser. ChatGPT: Fundamentals, applications and social impacts. En *2022 Ninth International Conference on Social Networks Analysis, Management and Security (SNAMS)*. IEEE, 2022. p. 1-8.
- [2] Ajay Chaudhari et al. YOG-GURU: REAL-TIME YOGA POSE CORRECTION SYSTEM USING DEEP LEARNING METHODS.
[URL:https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9509937](https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9509937).
- [3] Marc Marais and Dane Brown. Golf Swing Sequencing Using Computer Vision.
URL: https://link.springer.com/chapter/10.1007/978-3-031-04881-4_28
- [4] Google. MoveNet: modelo de detección de pose ultrarrápido y preciso. URL <https://www.tensorflow.org/hub/tutorials/movenet>