Actividad 2 - Teacheable Machine: Pose Estimation

Diego Iván Perea Montealegre (2185751) diego.perea@uao.edu.co

Carlos Iván Osorio Moreno carlos_ivan.osorio@uao.edu.co

Samir Hassan Ordoñez samir.hassan@uao.edu.co

Luis Fernando Pareja Bernal luis.pareja@uao.edu.co

Gabriel Angel Jeannot Viaña gabriel.jeannot@uao.edu.co

Facultad de Ingeniería, Universidad Autónoma de Occidente

Cali, Valle del Cauca

Las técnicas de inteligencia artificial más utilizadas en la aplicación de detección de pose incluyen redes neuronales, cálculo de recursos, aprendizaje y estudio de datos. Algoritmos pre-entrenados también se usan para estimar la pose.

Los algoritmos pre-entrenados son herramientas de gran poder creadas por algunos de los más grandes actores en el mundo de la Inteligencia Artificial. Estos algoritmos están disponibles de manera gratuita y son fáciles de usar, diseñados para que puedas obtener valor de ellos con la mayor facilidad posible. No tienes que entrenarlos con grandes cantidades de datos ni tomar decisiones sobre la arquitectura a utilizar, ya que todo eso ha sido hecho previamente. La mayoría de los algoritmos pre-entrenados tienen un enfoque genérico, lo que significa que cada usuario puede adaptarlos a sus necesidades específicas. Sin embargo, la mayoría de ellos pueden funcionar sin algún tipo de adaptación inmediatamente después de descargarlos. Se podría decir que son como "plug-and-play", lo que significa que se pueden utilizar inmediatamente después de descargarlos. Hay una familia de algoritmos pre-entrenados que son increíblemente útiles para un gran número de aplicaciones relacionadas con la posición corporal: La Estimación de la pose.

La Estimación de la pose es un conjunto de técnicas de visión por computadora que se utilizan para detectar figuras humanas en imágenes y videos, y mostrar las articulaciones clave de esas figuras. El resultado final es una representación detallada de las articulaciones y puntos clave de la posición corporal, lo que permite detectar la posición de una persona en tiempo real. [1]

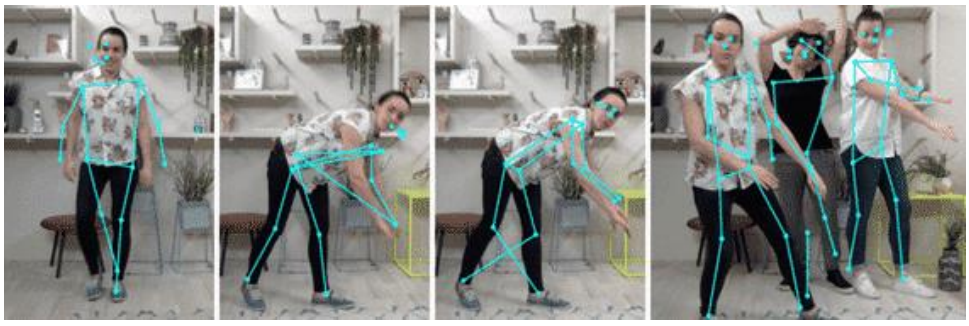
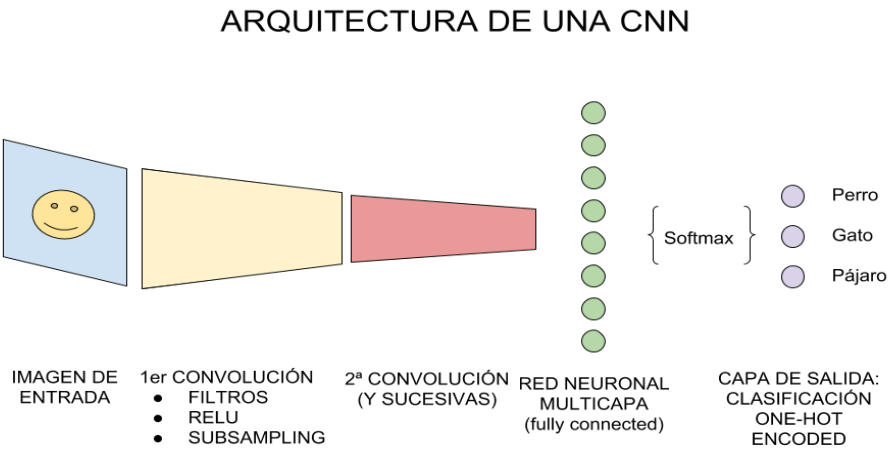


Figura 1. Ejemplo de detección de pose

La detección de pose puede implementarse utilizando diferentes enfoques, pero en general, las técnicas modernas se basan en el uso de redes neuronales convolucionales (CNN) profundas. Las CNN son una clase de redes neuronales que se especializan en procesar datos con estructura espacial, como las

imágenes. La estructura típica de una red neuronal para la detección de pose es un modelo de red neuronal basado en CNN, que se entrena en grandes conjuntos de datos de imágenes etiquetadas con información de pose. El modelo de red neuronal utiliza la información visual de la imagen para inferir la posición de las articulaciones humanas. Una arquitectura común utilizada en la detección de pose es el llamado modelo de red neuronal de una etapa (Single-Stage Neural Network), como Hourglass o OpenPose, que utiliza una sola red neuronal para estimar la posición de las articulaciones en la imagen. Otro enfoque común es la utilización de modelos de dos etapas (Two-Stage Models), como Mask R-CNN o Faster R-CNN, que primero detectan la región en la que se encuentra una persona y luego estiman la pose dentro de esa región.



En este caso el algoritmo que usa Teachable Machine es MobileNetV1, este algoritmo está especialmente diseñado para dispositivos Edge.

MobileNet usa primordialmente convoluciones separables en profundidad (Depthwise Separable Convolutions) en lugar de las convoluciones estándar usadas en arquitecturas anteriores para la construcción modelos más ligeros. Las MobileNets introducen dos nuevos hiperparámetros globales (width multiplier and resolution multiplier) esto les permite a los desarrolladores de modelos intercambiar latencia o precisión por velocidad y reducir tamaño dependiendo de los requerimientos.

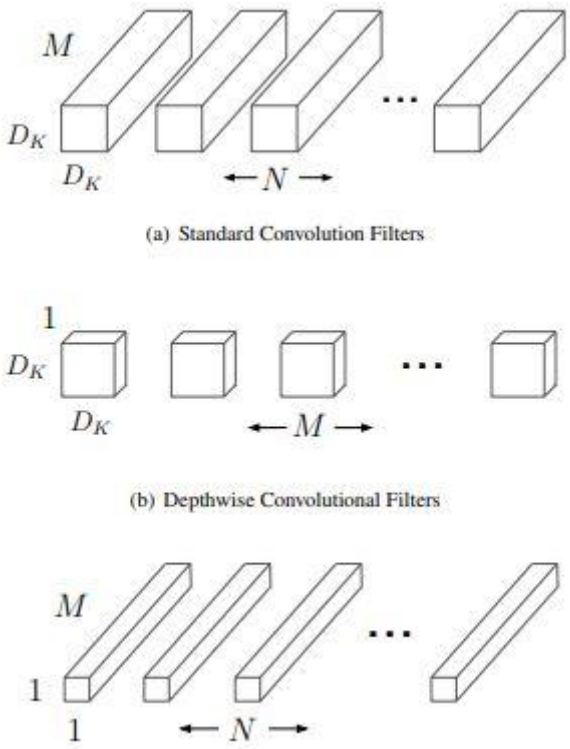


Figura 2. Depthwise convolutional filters comparsion

La arquitectura de las MobileNets consta de 28 capas (si contamos las convoluciones de profundidad y de punto como capas separadas). Una MobileNet estándar cuenta con 4.2 millones de parámetros los cuales pueden ser reducidos posteriormente afinando el hiperparámetro del multiplicador de ancho. Las imágenes que recibe son del tamaño 224 x 224 x 3.

En general, la detección de pose es una tarea compleja que requiere una gran cantidad de datos etiquetados y poder de procesamiento para entrenar y ejecutar los modelos de red neuronal. Sin embargo, las técnicas de detección de pose modernas han logrado un alto grado de precisión y son capaces de funcionar en tiempo real en algunas aplicaciones.

Ejemplos prácticos

Detección de pose de orientación de la cabeza y cuerpo: Usando la plataforma de Teacheable Machine [4] se realizó el modelaje de proyecto de posturas que con este se puede detectar la postura del cuerpo. Se decidió que la detección de pose se enfocaría en la cabeza y cuerpo debido a que puede resolver algunas problemáticas de las cuales son:

1. Control de sistemas: la aplicación puede ser utilizada para controlar un sistema, como una pantalla o un robot, en el que la dirección de la cabeza se utiliza para controlar la dirección del sistema.
2. Análisis de la atención: la detección de la orientación de la cabeza puede ser utilizada como una métrica para medir la atención de una persona hacia un objeto o una tarea en particular.
3. Asistencia médica: la aplicación puede ser utilizada como una herramienta para monitorear la postura de una persona y detectar si están experimentando dolor en el cuello o la cabeza debido a una postura inadecuada.
4. Seguridad: la aplicación puede ser utilizada en sistemas de seguridad para detectar la dirección de la cabeza de una persona, lo que podría ser útil en la prevención de accidentes o en la detección de posibles situaciones de peligro.

Por lo que la detección de la orientación de la cabeza y el cuerpo puede ser útil en cualquier aplicación en la que la dirección de la cabeza de una persona sea importante para el control, la atención o la seguridad.

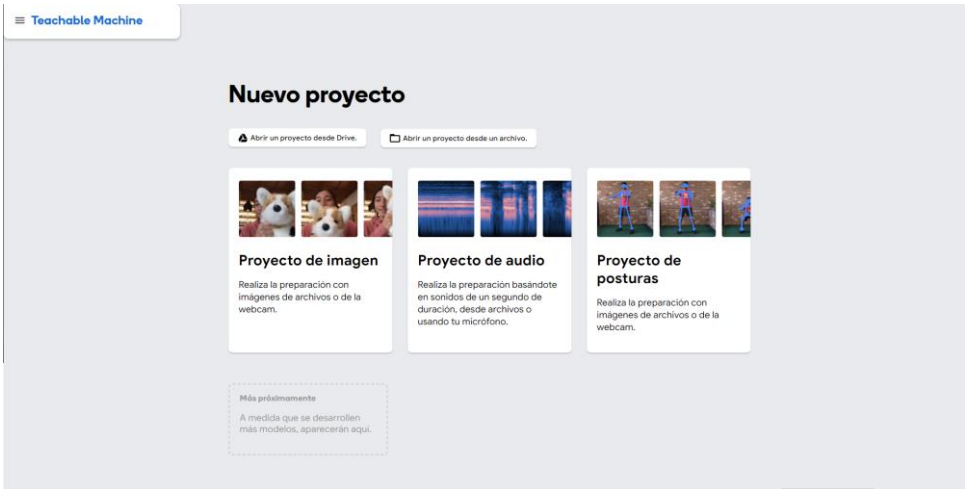


Figura 1 .Selección de proyecto a realizar en la página de Teacheable Machine

Ya seleccionado el proyecto a efectuar se realiza la toma de muestras dándole click al botón “Webcam”

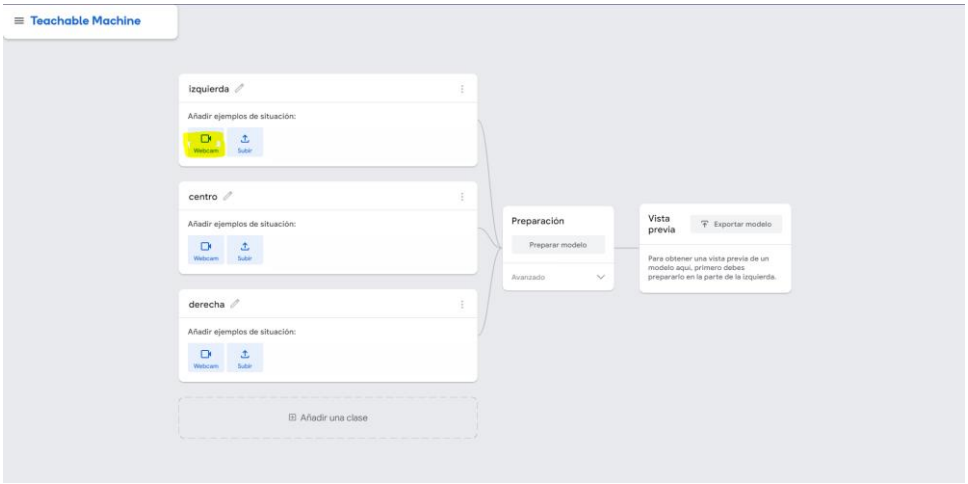


Figura 2. Inicio de toma de muestras de detección de pose

Con la toma de muestras de la primera clase nombrada “izquierda” almacenará muestras de orientación de cabeza y cuerpo del lado izquierdo, en la segunda clase nombrada “centro” almacenará muestras de orientación de cabeza y cuerpo del lado centro y la tercera clase nombrada “derecha” almacenará muestras de orientación de cabeza y cuerpo del lado derecho

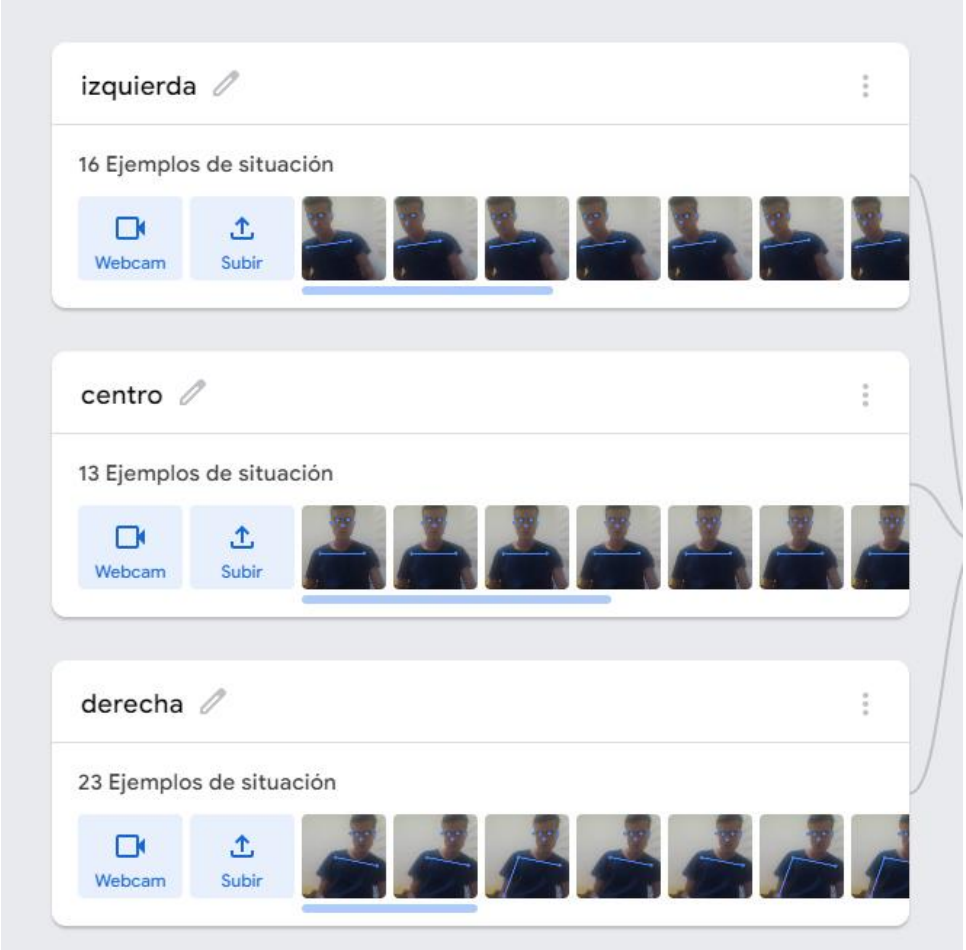


Figura 3. Toma de muestras totales de detección de pose de orientación de la cabeza y del cuerpo

Para entrenar el modelo se selecciona en el botón “Preparar modelo” y en las funciones avanzadas se deja por defecto.

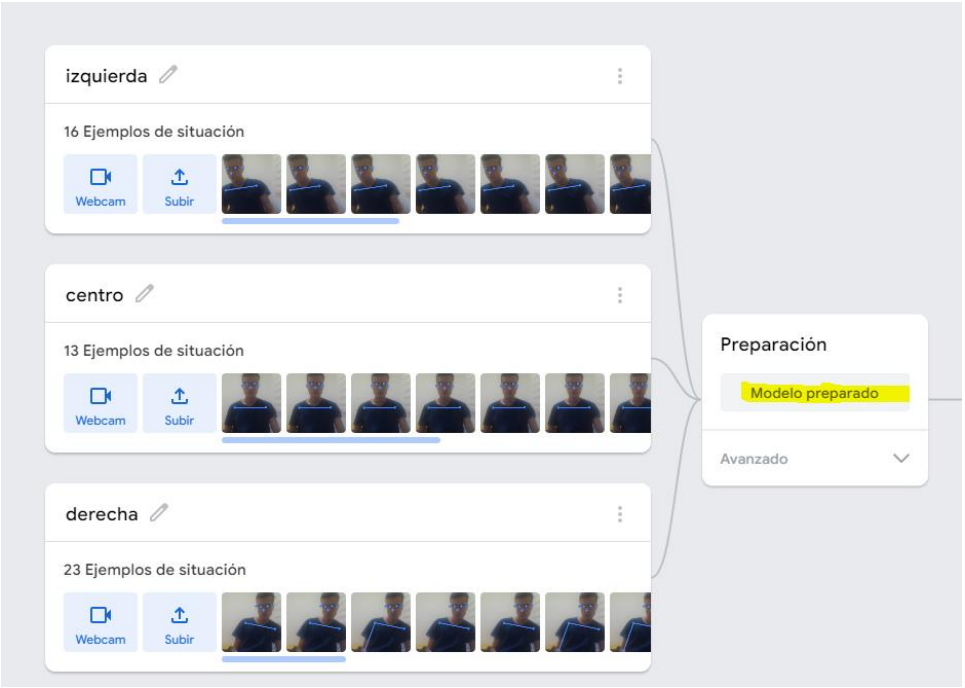


Figura 4. Entrenamiento del modelo de detección de pose de orientación de la cabeza y del cuerpo

Ahora se comprueba la funcionalidad del modelo en el cual es satisfactoria, por lo que detecta la pose frente la diferente orientación de la cabeza y del cuerpo, tal como en las clases de muestras. Esta detección de cada una de las clases está dada en porcentajes para determinar la clasificación de pose a la clase.

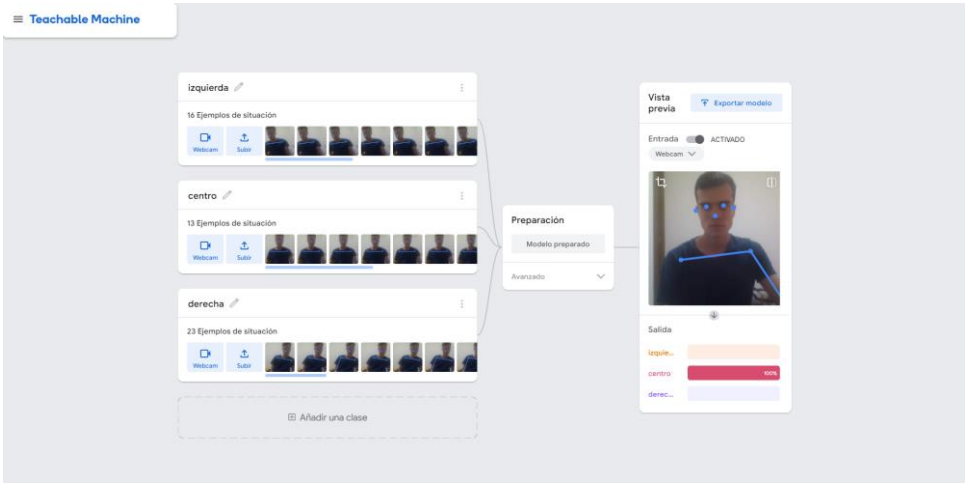


Figura 5. Funcionamiento del modelo de detección de pose de orientación de la cabeza y del cuerpo

Para exportar el modelo se da click en el botón “Exportar modelo” en donde aparece una pestaña para descargarlo y un fragmento de código de JavaScript para usar el modelo.

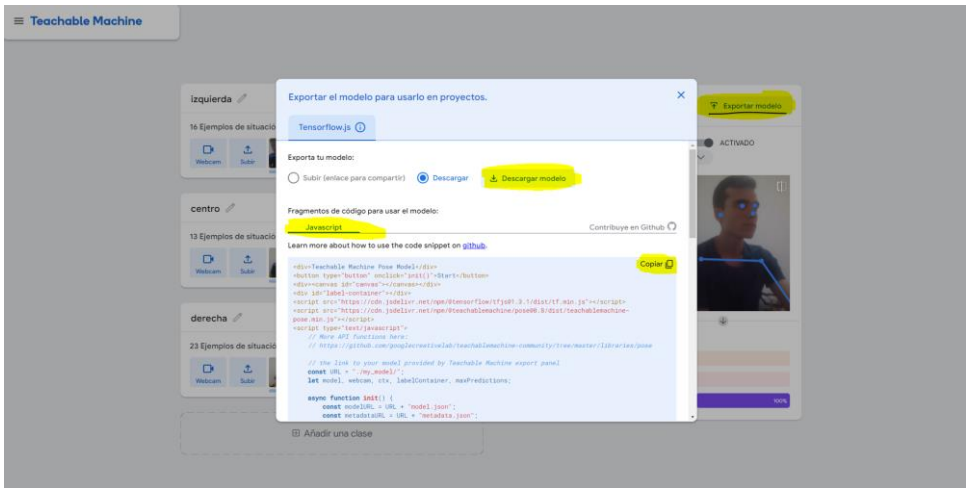


Figura 6. Exportación del modelo de detección de pose de orientación de la cabeza y del cuerpo

Para que se visualice en una interfaz el modelo y su aplicabilidad se debe de crear una carpeta, que en este caso se llamó “ethics_activity2”, y descomprimido el modelo descargado como se muestra en la figura 6 se debe cambiar el nombre de la carpeta “my_model” y este debe agregar dentro de la carpeta nueva creada. Dentro de la carpeta de crea un archivo HTML, en este caso se nombró “pose_lado.html”. Para crear la interfaz en este archivo HTML se realiza el código HTML básico para que funcione como se muestra en la figura 8.

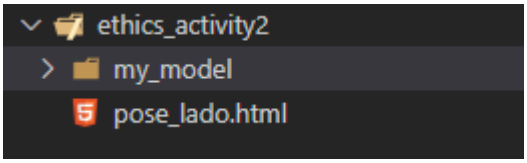


Figura 7. Creación de carpeta, archivo y modificación de carpeta del modelo

```
<!DOCTYPE html>
<html lang="en">
<head>
    <meta charset="UTF-8">
    <meta http-equiv="X-UA-Compatible" content="IE=edge">
    <meta name="viewport" content="width=device-width, initial-scale=1.0">
    <title>Document</title>
</head>
<body>


</body>
</html>
```

Figura 8. Creación de archivo HTML

Dentro del archivo HTML entre el apartado de body se inserta el código JavaScript previamente copiado cuando se exportó el modelo, tal y como se muestra en la figura 6.

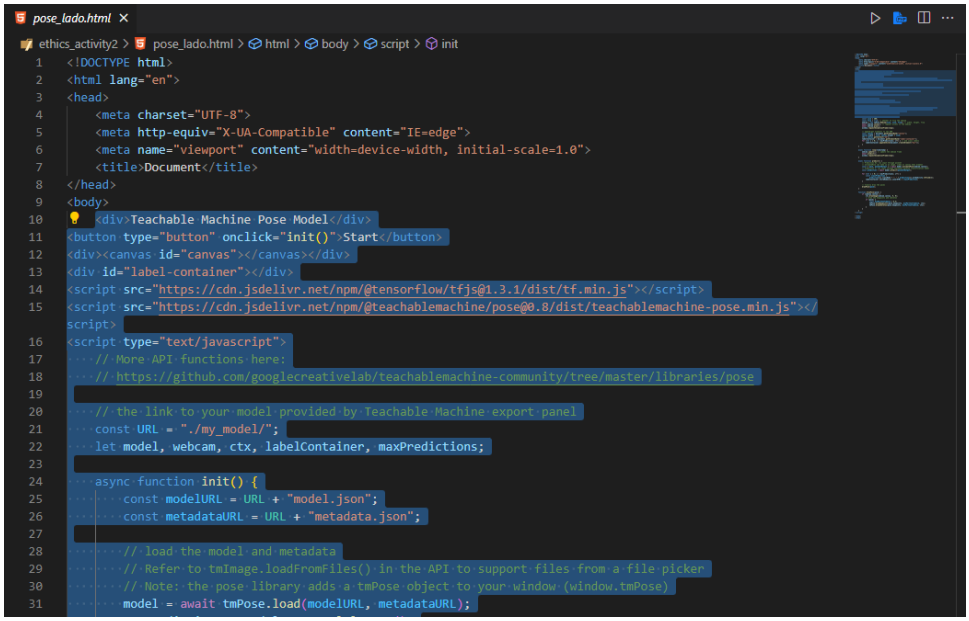


Figura 9. Archivo HTML completo “paso_lado.html”

Una vez realizado esto, se procede activar el servidor local o servidor a preferencia y se abre el puerto a desear. En este caso con el editor de texto VSCode. Este proceso se realiza de manera rápida debido a que se ejecuta con la extensión Live Server (ver figura 10).

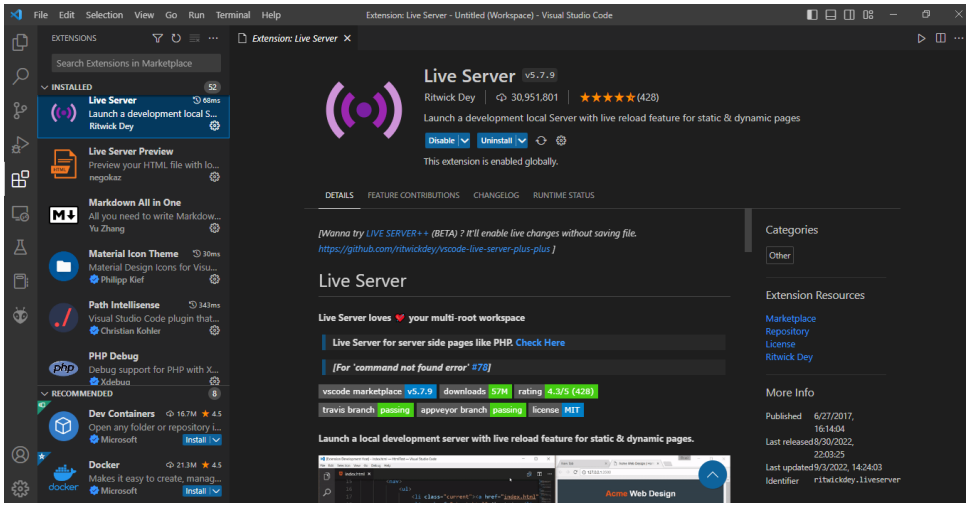


Figura 10 Extensión live server

Con darle click derecho al archivo HTML y seleccionado “Open Live Server”, se activa el servidor local y el puerto dado. En este caso es 127.0.0.1:5500, que es una dirección IP local que se utiliza para acceder a un servidor web en el mismo dispositivo en el que se está ejecutando el navegador web. El 127.0.0.1 es la dirección IP local de la máquina, también conocida como "localhost". El 5500 se refiere al puerto en el que el servidor web está escuchando las solicitudes entrantes.

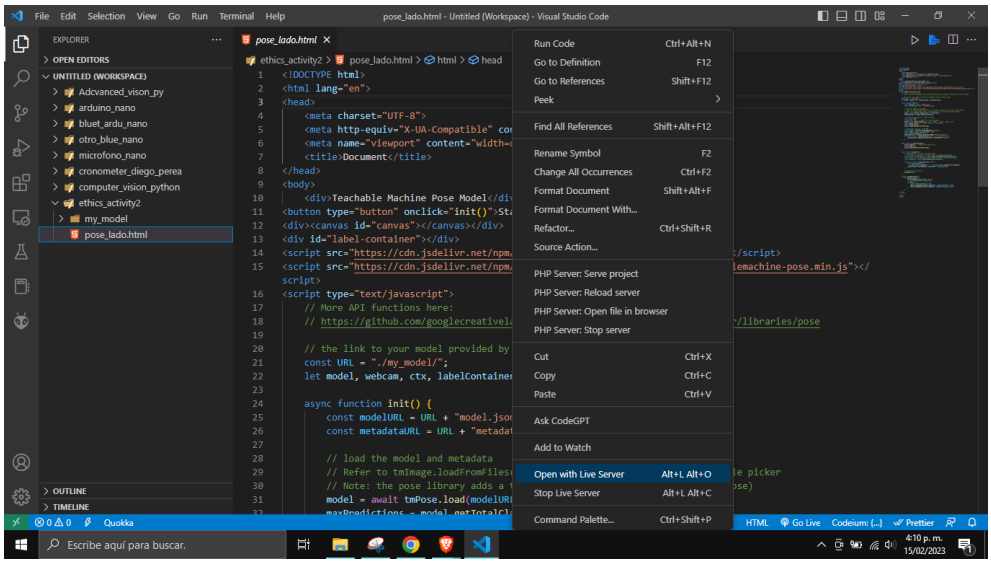


Figura 11. Activación del servidor

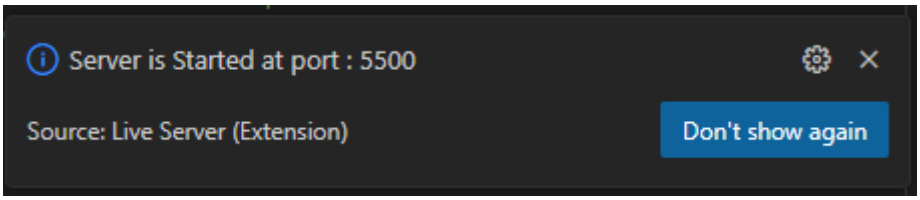


Figura 12. Visualización de activación del servidor

Lo primero que se visualizará será esta página en dónde, para la activación del modelo, se hace click al botón “Start” , y darle “Permitir” a la pestaña emergente para que de acceso a la cámara web.

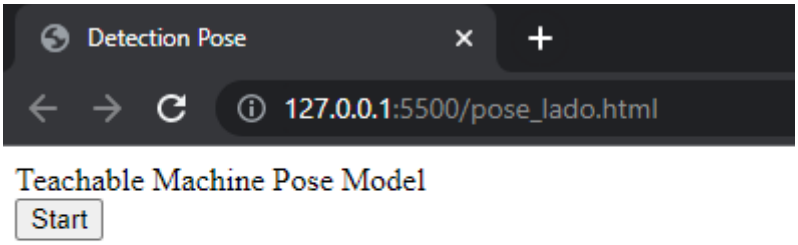


Figura 13. Aplicación web del modelo de detección de pose de orientación de la cabeza y del cuerpo

La detección de la pose está dada en porcentajes, donde 0.00 equivale a 0% y 1.00 equivale a 100% coincidente. Desde la cámara web visualizada se puede ver la pose de captación momentánea que está realizando el modelo. A continuación, se muestra la efectividad del modelo:

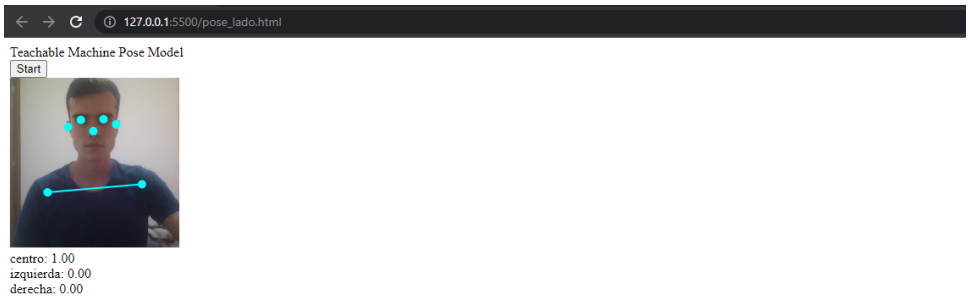


Figura 14. Clasificación de clase centro en detección de pose de orientación de la cabeza y del cuerpo

Figura 15. Clasificación de clase izquierda en detección de pose de orientación de la cabeza y del cuerpo

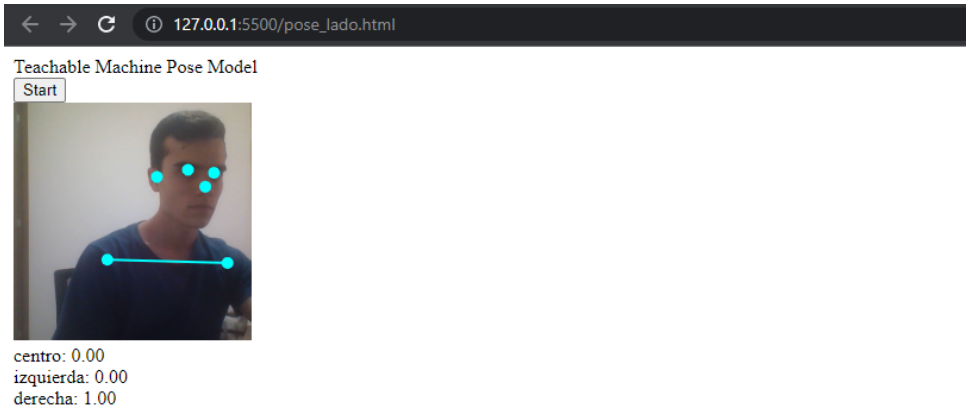


Figura 16. Clasificación de clase derecha en detección de pose de orientación de la cabeza y del cuerpo

La interfaz gráfica de la aplicación web se puede mejorar mediante la implementación de diferentes estilos, y se puede aplicar este modelo de detección de pose de orientación de la cabeza y del cuerpo hacia otras aplicaciones como anteriormente se afirmó, dando a entender la buena práctica del modelo y de las muestras tomadas.

Conclusiones

Sin duda, las capacidades de acción que se desglosan de los distintos sistemas de inteligencia artificial que actualmente están en desarrollo son exponencialmente grandes, y una de sus ramas, la estimación de poses, resulta útil para distintos campos que requieren de procesamiento en tiempo real del comportamiento humano en cuanto a sus movimientos, tal y como se evidencia actualmente en campos como la salud y bienestar mediante el monitoreo de la actividad física de las personas, el deporte para medir rendimiento y técnica de los atletas, la robótica para la interacción humano-robot y el mundo del entretenimiento como los videojuegos, la realidad virtual y aumentada.

El actual uso de algoritmos pre-entrenados para la estimación de la pose, junto con las técnicas de redes neuronales convolucionales profundas, permite detectar la posición corporal con precisión y en tiempo real. La aplicación de la detección de pose sigue siendo un campo de investigación en evolución, y se espera que las técnicas y los modelos sigan mejorando y evolucionando para satisfacer las necesidades de diversos sectores en el futuro. Se espera y desea que el avance tecnológico en esta área fomente el bienestar y desarrollo de los individuos desde una variedad de campos inmensa.

Referencias

[1]*Estimación de la pose con deep learning* (2022) *ENIIT*. Available at: https://eniit.es/estimacion-de-la-pose-con-deep-learning/ (Accessed: February 12, 2023).

[2]*MobileNetV1 Architecture*. Available at: MobileNet V1 Architecture (opengenus.org) (Accessed: February 15, 2023).

[3]*Introducing the next generation of on-device vision models: mobileNetV3* at: MobileNetV3 introducing (Accessed: February 14, 2023).

[4] Google. [Online]. Available: https://teachablemachine.withgoogle.com/. [Accessed: 15-Feb-2023]