

3D Human Pose Estimation (HPE) in video

I introduced pose estimation to my 7-year-old nephew so that he could enjoy his somersault moves and try new things for fun.

Human Pose Estimation is a computer vision-based technology that identifies and classifies specific points on the human body. 33 points represent our limbs and joints to compute the angle of flexion, and measure, human pose well.

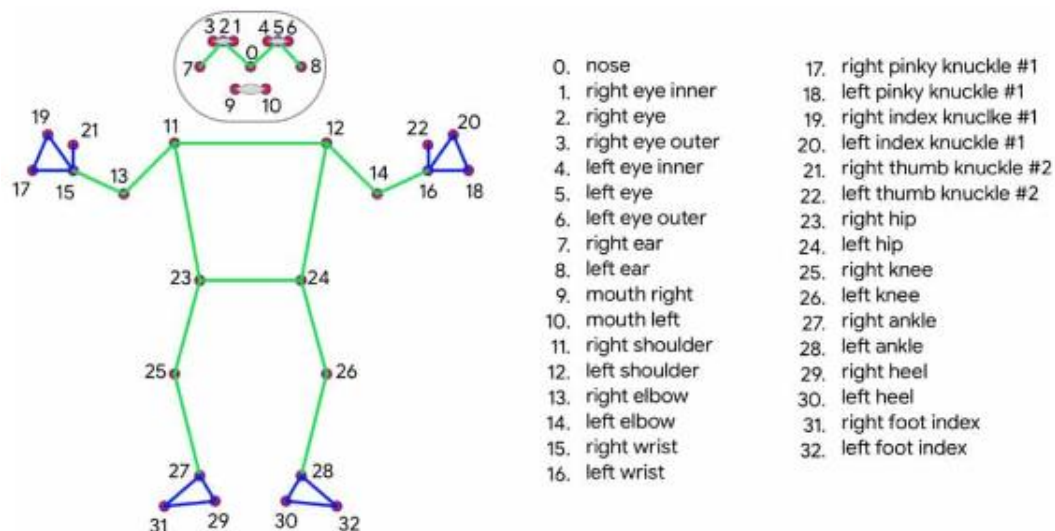
3D human poses in video(s) can be effectively estimated with a fully convolutional model based on dilated temporal convolutions over 2D key point detection followed by 3D pose estimation.

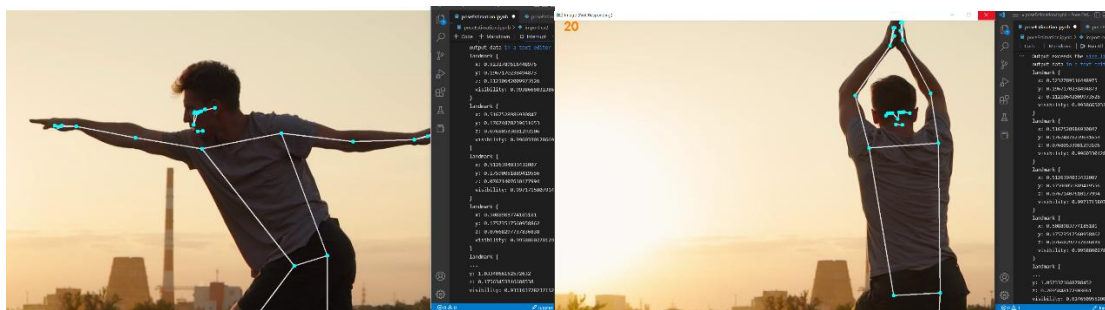
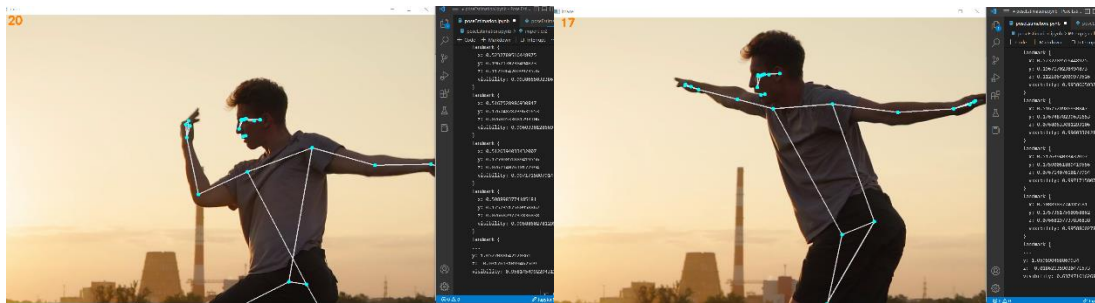
Keypoint estimation can be utilized to interpret movements to 3D models in the Metaverse such as Augmented Reality (AR), Virtual Reality (VR), Mixed Reality (MR) is used in association with webcam/smart phone cameras. It is used in sports industry, healthcare, Yoga, and business.

It is beneficial for physiotherapists, trainers, or artists to know the right angle of a joint in a specific work out. It aids in analysing pose/movement if it is correctly performed such as person squatting, running, bending, jumping in association with machine learning technology, is applicable to myriad fields.

Most of HPE methods are based on recording an RGB image with the optical sensor to detect body parts and over all poses. It is used for fitness, rehabilitation, augmented reality applications, and surveillance. The skeleton-based model is used in human pose estimation this is because it consists of a set of joints like ankles, knees, shoulders, elbows, wrists, and limb orientations comprising the skeletal structure of a human body.

I have used OpenCV and MediaPipe is well-balanced framework for pose estimation. It uses BlazePose 33 landmark topology. It works in two stages – **detection and tracking**. As detection is not performed in each frame, MediaPipe is able to **perform inference faster**.





Comparison of MediaPipe vs YOLOv7

YOLOv7	MediaPipe
Detection runs for all frames	Detection runs once followed by tracker
YOLOv7 detects multi-person detection framework	MediaPipe is a single person pose estimation framework
YOLOv7 supports both CPU and GPU	MediaPipe supports only CPU
YOLO performs better in low light condition than MediaPipe	
When it comes to occlusion YOLOv7 is wins	MediaPipe is comparatively good at detecting far-away objects
It is better at estimating fast movements given the high resolution	