Mediapipe是Google在2019年开发并提出的一款开源的跨平台多媒体处理框架，用于构建基于机器学习的应用程序，特别是涉及到计算机视觉、音频处理、姿势估计等领域。Mediapipe实际上是一个集成的机器学习视觉算法的工具库(包含人脸检测、人脸关键点、手势识别、头像分割和姿态识别等各种模型)，该库提供了一系列预训练模型和工具，可以帮助开发者快速构建和部署计算机视觉和音频处理应用。它提供了一套易于使用的API，可以用于实时姿势估计、手势识别、面部识别、目标跟踪等任务。

关于Mediapipe的3d坐标

<https://blog.csdn.net/kalakalabala/article/details/135272822>

<https://github.com/PINTO0309/PINTO_model_zoo?tab=readme-ov-file>

2D/3D Human/Animal Pose Estimation

<https://github.com/facebookresearch/VideoPose3D>

<https://github.com/Daniil-Osokin/lightweight-human-pose-estimation-3d-demo.pytorch>

<https://github.com/zhenhuat/STCFormer?tab=readme-ov-file>