**人体姿态识别项目汇报文档**

**制作人：王奕力**

**专业：计算机科学与技术**

**学号：1211007011**

1. **概述**
2. **定义**

人体姿态识别（Human Pose Estimation）是计算机视觉的核心任务，旨在通过分析图像或视频数据，‌精确识别和定位人体关键骨骼点的空间位置‌（如头部、肩膀、肘部、手腕、臀部、膝盖、脚踝等），并‌构建描述人体姿势的骨架结构‌。其核心步骤是检测图像中的人体，然后对这些关键点进行定位（通常用坐标表示）。技术主要分为：

‌2D姿态识别‌：在图像平面上定位关键点坐标 (x, y)。

‌3D姿态识别‌：估算关键点在三维空间中的位置 (x, y, z)，包含深度信息。

该技术依赖深度学习（尤其是卷积神经网络和目标检测技术），处理遮挡、复杂背景和姿态多样性是主要挑战。

1. **当前行业现状**

国外人体姿态识别研究起步较早，技术体系成熟。美国、欧洲等地区以‌学术机构（如CMU、斯坦福）与科技巨头（Google、Meta、微软）‌为主导，聚焦‌3D姿态估计、实时多目标跟踪‌及‌弱监督学习‌方向。代表性成果包括OpenPose（CMU）、AlphaPose（MPII）等开源框架。近年来，‌Transformer架构与多模态融合‌成为热点，例如Meta的DINOv2通过自监督学习提升泛化能力。挑战在于‌复杂场景下的鲁棒性‌（如遮挡、光照变化），部分研究结合‌神经辐射场（NeRF）‌优化三维重建精度。产业应用集中在‌医疗康复、元宇宙交互‌等领域，苹果Vision Pro已集成高精度手势追踪技术。‌

国内研究由‌高校（清华、北大、浙大）与企业（百度、商汤、旷视）‌共同推动，侧重‌轻量化部署与垂直场景落地‌。百度PaddlePaddle、商汤SensePose等框架在2D实时检测领域达到国际水平。研究热点包括‌基于国产芯片的端侧推理优化‌（如华为昇腾）和‌遮挡姿态补全算法‌（如上海交大的Occlusion-Net）。政策支持加速技术转化，例如冬奥会中应用的AI裁判系统。但‌基础理论创新‌仍依赖国外开源模型，3D姿态数据稀缺制约发展。未来方向聚焦‌低成本传感器融合‌（如RGB-D摄像头）与‌工业质检、智慧养老‌等本土化场景。

1. **组内研究成果**

在本学期中，我们小组首先了解了本领域的基本概念、定义、行业历史、工作原理、研究方向和应用可能，对这门学科有了比较基础的了解。然后我们对该系统的工作原理进行了重点学习和研究，原定计划选择openpose作为我们的软件，但实践之后发现它有诸多不足，于是我们改用了mediapipe作为软件，解决了openpose时期的诸多问题，例如硬件要求过高、帧率较低、处理速度较慢等，最终成功在设备上实现了人体的姿态识别。但我们小组在实践过程中对于如何界定动作是否标准或者是否应该计数等问题有不同的看法，且并未能取得统一，所以在这方面的实验成果并不是十分理想。

1. **总结**

在本学期的学习过程中，我们收获颇丰，不仅学习了许多人工智能相关的前沿知识，还训练和培养了动手能力。随着项目的逐渐深入，我在实操过程中对理论知识有了更深刻的理解和认识，并逐渐将理论和实践结合起来。同时，在实验过程中，我们也遇到了许多问题，例如数据集的选择，动作类型的定义，以及单人动作识别到多人动作识别的转换问题，我们不仅组内进行了数次研讨，同时还定期向老师汇报和交流，最终大部分问题得到了解决。此项目的研究，使我学习了大量知识并加以应用，相信在未来它们会对我的学习之路大有裨益。