

多模态数据集数据标注方案

1 语义标注方案 (Semantic Annotation Strategy)

为了赋予原始传感器数据以行为语义（如“行走”、“跌倒”），我们需要对对齐后的数据进行时间轴标注。针对本项目特点，我们设计了以下三种标注方案，并推荐采用方案一作为基准。

1.1 方案一：基于 ELAN 的时间轴标注（推荐）

ELAN 是多模态研究领域最标准的标注工具，适用于精确到毫秒的行为分割。

- **操作流程：**
 1. 将对齐后的主视频（如 Camera 1）导入 ELAN。
 2. 在时间轴上划分行为区间（例如：05.2s – 08.4s 为 “Walking”）。
 3. 导出标注文件（CSV/TXT），利用 Python 脚本将标签映射至主数据集的每一帧。
- **优势：**精度高，人工校验成本低，适合行为识别任务。
- **适用性：**本项目首选方案，因雷达数据难以直观解读，需依赖同步视频进行“所见即所得”的标注。

1.2 方案二：基于 CVAT 的细粒度标注

CVAT (Computer Vision Annotation Tool) 适用于需要空间位置信息的任务。

- **操作流程：**在视频画面上绘制检测框 (Bounding Box) 或骨骼关键点 (Keypoints)。
- **优势：**提供空间语义 (Target Localization)，可支持更复杂的姿态估计任务。
- **劣势：**工作量巨大，对于仅需行为分类的任务略显冗余。

1.3 方案三：AI 辅助自动化标注

利用大模型（如 GPT-4o 或 Gemini Pro 1.5）的视频理解能力进行预标注。

- **操作流程：**将视频片段输入多模态大模型，提示其生成“时间戳-动作”列表（JSON 格式），再由人工进行微调。
- **优势：**极大提高标注效率，适合处理海量数据。
- **劣势：**对细微动作（如“微小晃动”）的捕捉可能不如人工精确。

2 最终数据集结构 (Final Dataset Structure)

经过时空对齐与数据融合，最终生成的 `Xiaomi_HRI_Master_Dataset.csv` 采用了以主相机帧率为基准的统一格式。

2.1 数据字段定义

数据集每一行代表一个时间步 (Time Step)，包含以下关键信息：

Table 1: 主数据集字段说明

字段类别	字段名	物理含义与说明
时序基准	Frame_ID	主相机 (Camera 1) 的视频帧号，作为全表主键。
	Timestamp	统一的时间戳 (秒)，所有传感器均已对此对齐。
雷达数据	Radar_X, Y, Z	毫米波雷达检测到的人体目标在物理空间中的真实坐标 (米)。
视觉数据	C1_U, C1_V	目标在相机 1 图像平面的投影坐标 (像素)。
	C2_U, C2_V	目标在相机 2 图像平面的投影坐标 (像素)。
	...	(包含 C3, C4 等所有机位数据)
惯性数据	IMU_Acc_X/Y/Z	被试者佩戴 IMU 的三轴加速度 (m/s^2)。
	IMU_Gyro_X/Y/Z	被试者佩戴 IMU 的三轴角速度 (rad/s)。
语义标签	Action_Label	该时刻对应的人体行为类别 (如 Walking, Falling)。

2.2 数据样本示例

以下展示了数据集的前 3 行样本数据 (数值仅为示例)：

Frame	Time	Radar_X	Radar_Y	C1_U	C1_V	Acc_X	Acc_Y	Label
100	3.33	1.52	4.20	500	800	0.02	9.81	Walking
101	3.36	1.55	4.21	505	802	0.05	9.79	Walking
102	3.39	1.58	4.22	510	805	0.03	9.80	Walking

Table 2: 融合后的多模态数据样本片段