

T/SAIAS

上海市人工智能行业协会团体标准

T/SAIAS XXXX—2025

人形机器人操作任务数据集数据格式要求

Requirement for humanoid robot operation task dataset data format

（征求意见稿）

在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上。

2023 – XX – XX 发布

2024 – XX – XX 实施

上海市人工智能协会 发布

目 次

前言 II

1 范围 3

2 规范性引用文件 3

3 术语和定义 3

4 缩略语 4

5 一般要求 4

6 数据集描述 4

7 数据集总体框架 6

8 数据格式要求 6

附录 A （资料性） 数据集描述文件示例..... 10

附录 B （资料性） 数据存储格式示例..... 23

附录 C （资料性） 数据集转换 RLDS 规范示例..... 24

参考文献 26

前 言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由上海市人工智能行业协会提出。

本文件由上海市人工智能行业协会归口。

本文件起草单位：

本文件主要起草人：

本文件为首次发布。

人形机器人操作任务数据集数据格式要求

1 范围

本文件规定了人形机器人操作任务数据集的一般要求、数据集描述、数据集总体框架、数据格式要求。

本文件适用于人形机器人操作任务数据集。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

ISO 8000-2 数据质量—第2部分：术语(Dataquality-part2:Vocabulary)
ISO/IEC 14977:1996 信息技术 语法元语言 扩展的BNF

3 术语和定义

ISO 8000-2界定的以及下列术语和定义适用于本文件。

3.1

人形机器人 humanoid robot

具有躯干、头部和部分或全部肢体，外观和动作与人类相似，实现感知、认知、决策、执行等能力的机器人。

[来源：GB/T 12643—202X，4.15.5，有修改]

3.2

操作任务 operation task

由机器人全身肢体关节协调执行的场景任务。

示例：抓取、捡拾、拖拽、推拉、开闭、踢、位移等。

3.3

数据 data

信息的可再解释的形式化表示，以适用于通信、解释或处理。

注：数据可以由人工或自动的方式加工、处理。

[来源：GB/T 5271.1—2000，01.01.02]

3.4

数据集 dataset

具有一定主题，可标识并能被计算机系统处理的数据集合。

[来源：ISO 19115-1：2014，4.3]

3.5

元数据 metadata

定义和描述其他数据的数据。

[来源：GB/T 18391.1—2009，3.2.16]

3.6

主数据 master data

数据集中与目标任务相关的场景、任务、机器人本体等数据部分，用于支撑主要的分析、建模和决策任务。

示例：家政厨房场景生活照料服务（冲泡咖啡）人形机器人数据集包含环境数据（视觉数据如厨房布局、听觉数据

如操作指令、环境声等)、机器人构型、机器人本体数据(运动学状态数据、动力学状态数据)、场景任务数据(机器人执行冲泡咖啡的任务步骤)等。

3.7

数据类型 datatype

一些可区分的值的集合,这种区别由这些值的性质以及对这些值的运算所表征。

[来源:GB/T 18221—2000, 4.11]

3.8

数据标识符 data identifier; ID

注册机构内一个管理项的唯一标识符。

[来源:GB/T 18391.1—2009, 3.1.10]

4 缩略语

下列缩略语适用于本文件:

CSV: 逗号分隔值 (comma separated values)

JSON: JS键值对数据 (JavaScript object notation);

Parquet: 支持嵌套结构的列式存储格式

XML: 可扩展标记语言 (Extensible Markup Language)

5 一般要求

5.1 人形机器人操作任务数据集(以下简称“数据集”)来源应可追溯,标识符应具有规范性、唯一性。

5.2 数据集应支持多种通用数据交换格式,提供版本号及获取路径。

5.3 数据集应使用符合 ISO/IEC 14977:1996 规定的形式文法进行定义。

5.4 数据集的数据语义编码要求包括:

注:数据语义编码是指一种将数据所蕴含的语义内容,通过特定的规则和方式,转化为计算机能够识别和处理的特定符号或数字系统。

a) 文本文件应使用 UTF-8 编码。

b) 键名使用下划线命名法,首字母小写,单词间以下划线分隔。对同一类型的数据应使用相同键名,并在语义文档中进行定义。

c) 对于不同类型的数据,键名应具有唯一性;可通过添加命名空间前缀或 URI 前缀进行区分。在键名前应使用领域本体中定义类名称作为前缀。

d) 键名应具备自解释性,优先采用国际通用、文化中立的术语,并提供多语言翻译。

e) 嵌套数据结构中的键名应准确反映层级位置,并在语义映射中定义对应的概念与关系。

f) 不应将编程语言或数据格式的保留字用作键名,扩展字段时可添加后缀并在语义文档中注册新概念。提供语义映射文件并保持与数据格式定义同步更新。

5.5 数据集主数据包括但不限于:

——环境感知主数据;

——本体状态主数据;

——场景任务主数据。

5.6 数据集组件包括但不限于:

——数据文件:图像、深度图、点云、音频等类型的具体数据;

——数据集描述文件:整体描述文件与片段描述文件,符合第 6 章的要求。

6 数据集描述

6.1 数据集整体描述文件的内容宜包括但不限于:

a) 数据集名称:

- 1) 采用英文字符、数字、符号等组合方式进行命名,语义清晰,避免冗长或复杂表述;
- 2) 包含关键词反映数据集的主要特征、数据类型或研究领域等;

- 3) 包含时间、地点、对象等关键信息,若使用缩写,附加后缀标注进行说明;
- 4) 在同一项目或组织中应遵循统一的命名规则,避免使用不常用缩写,若使用则附加后缀标注进行说明;
- 5) 在名称中明确版本号或日期;
- 6) 避免使用特殊字符(如空格、异形符号等);
- 7) 具备可扩展性,

示例: RobotVisionDataset_202403_USA_Humanoid_v1.0 (人形机器人视觉数据集, 2024 年 3 月, 美国采集, 版本 1.0)。

- b) 数据集内容信息: 包括数据集中包含的片段数量描述、数据类型描述(如视觉、听觉等数据及对应数值格式等)、机器人和传感器配置信息等;
- c) 数据集用途信息: 包括数据集的应用场景、任务目的、数据集的适用范围和限制描述等;
- d) 元数据取值范围建议;
- e) 数据集来源信息: 包括数据收集环境(温度、湿度、光照等)、方式、时间和地点、提供数据提供者等相关信息;
- f) 数据集版本号与存储信息。

6.2 片段描述文件的内容可参考表 1。

表1 片段描述文件内容

名称(推荐)	字段名(推荐)	数据类型	选项	说明
场景ID	scene_id	string	可选	唯一应用场景标识
场景名称	scene_name	string	可选	机器人应用场景名称
场景描述	scene_description	string	可选	描述机器人应用场景信息(环境类型、环境地图、物体布局、尺寸、主要特征等)
任务ID	task_id	string	可选	唯一任务标识
时间戳	Timestamp	见表5	必选	数据采集的绝对时间
任务名称	task_name	string	可选	机器人执行具体任务名称, 如搬运、浇花、叠衣服等
任务描述	task_description	string	必选	描述机器人执行具体任务特征(目标、约束条件等)
任务操作步骤	task_instruction	string	可选	描述机器人执行具体任务的过程或步骤, 包括短序列(不超过3步(含3步)) 和长序列(超过3步)
技能描述	skill_sequence	string	可选	描述执行任务时所需技能, 通常是单一动作
操作模式	operation_mode	string	可选	描述机器人操作模式, 如遥控操作、跟随运动或算法生成轨迹等
机器人任务状态数据	robot_task_status_data	见表9、10	可选	1) 机器人当前任务起始位置与姿态信息: 位置信息如机器人在工厂中执行任务时的空间坐标(X、Y、Z坐标); 姿态/关节状态信息包括各个关节的角度、速度、加速度等; 2) 机器人起始运动状态。
环境及感知数据	environmental_and_perception_data	见表2、3、4、6、7、8	可选	1) 起始环境信息: 包括布局、工作区域坐标、障碍物位置以及环境参数(如温度、光照)等; 2) 传感器数据: 包括视觉、听觉、触觉、味觉、嗅觉数据; 3) 环境动态数据: 可能影响任务的动态因素, 如人员流动或设备运行状态等。

表 1 片段描述文件内容（续）

名称（推荐）	字段名（推荐）	数据类型	选项	说明
控制指令与反馈	control_commands_and_feedback	string	可选	1) 下达给机器人的具体控制指令（如“抓取”、“移动至指定位置”、“放置”等），可采用文字描述或编码指令； 2) 在任务执行过程中若出现故障，描述相关错误码、报警信息及故障等。
执行状态	execution_status	string	可选	机器人执行结果状态信息，描述操作成功/失败/部分完成等。

6.3 数据集描述文件示例见附录 A。

7 数据集总体框架

人形机器人操作任务数据集总体框架见图1。

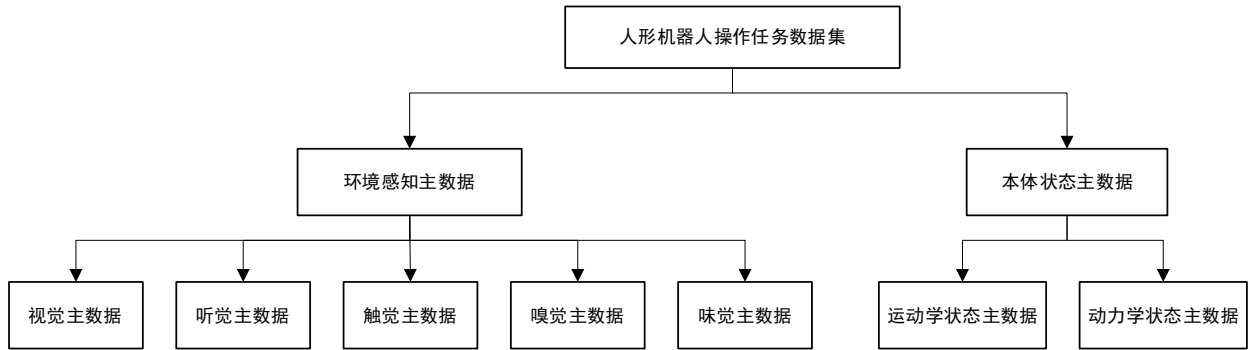


图1 人形机器人操作任务数据集总体架构

8 数据格式要求

8.1 元数据格式

元数据格式见表2～表11。

注：与表2～表11中规定的数据单位不一致时需特别注明。

表2 视觉元数据

中文名称	数据标识符	数据类型	描述	单位
2维图像	image	uint8	外部环境图像，像素值格式为RGB8，一个字节8位（取值范围是0～255）。	/
深度图	image_with_depth	float32	外部环境深度图，像素值表示距离，至少保留2位小数。	米, m
点云	point_cloud	float32	外部环境点云，点云数据中每个点的坐标通常表示为(x, y, z)，其中 x, y和z分别是点在三个正交方向上的坐标值，坐标值表示相对于坐标系原点的平移距离，至少保留2位小数。	米, m

表3 听觉元数据

中文名称	数据标识符	数据类型	描述	单位
指令音	audio	uint8	指令声音数据。	/
环境音	audio	uint8	外部环境背景声音数据。	/

表4 附加音频元数据

中文名称	数据标识符	数据类型	描述	单位
采样格式	sample_format	string	音频采样格式，描述音频数据存储方式。	/
通道数	channels	uint8	音频文件的声道数。	个
采样率	sample_rate	uint32	描述一秒内对模拟声音信号进行数字化采样的次数。	赫兹，Hz
比特率	bitrate	uint32	每秒传输的二进制数据量，描述压缩音频的质量。	比特/秒，bit/s
编码格式	coding_format	string	用于压缩和解压缩音频数据的算法。 编码格式可使用 WAV，MP3	/

表5 时间戳

中文名称	数据标识符	数据类型	描述
原始时间戳	raw_timestamp	int64	传感器本地时钟记录的原始Unix时间戳，记录数据采集的绝对时间
时间同步源	raw_time_source	string	时间同步来源（如“系统时钟”、“GPS”、“NTP”）传感器
对齐方法	alignment_method	string	时间戳对齐算法

表6 触觉元数据

中文名称	数据标识符	数据类型	描述	单位
压力	pressure	float64	标量，单位面积上受到的垂直力。	帕斯卡，Pa
接触力	contact_force	float64	矢量，两个物体相互作用时受到的以沿任何方向作用力，由作用在物体上的三个正交分量[Fx, Fy, Fz]组成。	牛顿，N
纹理	surface_texture	float64	表面触感描述（如“光滑”、“粗糙”）。	/
接触点	touchpoints	float64	受力表面关键点坐标（x, y, z）	米，m
硬度	hardness	float64	表面抗压能力。	兆帕斯卡，Mpa
温度	temperature	float64	外部环境温度数据。	摄氏度，℃
湿度	humidity	float64	外部环境湿度数据。	相对湿度百分比，%RH

表7 嗅觉元数据

中文名称	数据标识符	数据类型	描述	单位
气体分类	gas_type	string	气体类型（如二氧化碳、甲烷等）。	/
气体浓度	gas_concentration	float64	外部环境气体浓度数据。	ppm、ppb、mg/m³等
有害气体	is_hazardous	bool	是否为有害气体（true/false）。	/
特殊气体	is_special	bool	是否为特殊气体（如惰性气体）。	/

表8 味觉元数据

中文名称	数据标识符	数据类型	描述	单位
化学成分	chemical_components	string	化合物列表（如NaCl、H₂O）。	/
化合物含量	compound_content	float64	外部环境化合物含量数据。	毫克/升，mg/L

表9 运动学状态元数据

中文名称	数据标识符	数据类型	描述	单位
绝对位置和姿态	absolute_pose	float64	世界坐标系下的位置和姿态数据,由[x, y, z, q_x, q_y, q_z, q_w]组成	位置部分[x, y, z]单位为米, m; 姿态部分[q_x, q_y, q_z, q_w]无物理单位
相对位置和姿态	relative_pose	float64	相对位置初始坐标系的位姿数据, 由[x, y, z, q_x, q_y, q_z, q_w]组成	位置部分[x, y, z]单位为米, m; 姿态部分[q_x, q_y, q_z, q_w]无物理单位
线速度	linear_velocity	float64	笛卡尔空间位移线速度数据, 由[vx, vy, vz]组成(默认世界坐标系)	米每秒, m/s
角速度	angular_velocity	float64	角速度数据, 由[ωx, ωy, ωz]组成	弧度每秒, rad/s
关节位置	joint_position	float64	本体各个关节角度信息, 由[j1, j2, j3, j4, j5, j6...]组成	弧度, rad
关节速度	joint_velocity	float64	本体各个关节速度信息, 由[v1, v2, v3, v4, v5, v6...]组成	弧度每秒, rad/s
关节加速度	joint_acceleration	float64	本体各个关节加速度信息, 由[a1, a2, a3, a4, a5, a6...]组成	弧度每秒平方, rad/s ²

表10 动力学状态元数据

中文名称	数据标识符	数据类型	描述	单位
关节力	joint_force	float64	本体各个关节力信息, 由[f1, f2, f3, f4, f5, f6...]组成	牛顿, N
关节力矩	joint_torque	float64	本体各个关节力矩信息, 由[m1, m2, m3, m4, m5, m6...]组成	牛顿·米, N·m
末端六维力	end_effector_wrench	float64	末端执行器的六维力/力矩信息, 格式为[Fx, Fy, Fz, Mx, My, Mz]	末端三轴力[Fx, Fy, Fz]单位为牛顿, N; 末端三轴力矩[Mx, My, Mz]单位为牛顿·米, N·m

表11 传感器标定信息

中文名称	数据标识符	数据类型	描述
传感器设备序列号及名称	sensor_serial_name	string	传感器设备的唯一标识符和名称(如“SN-001:Realsense_D435”),用于区分不同设备。
传感器的类型及数量	sensor_type_count	string	传感器类型与数量,键为类型(如camera/IMU/LiDAR等),值为子类型和数量(如{"camera": {"type": "RGB-D", "count": 2}})。
传感器采集频率	sensor_sampling_rate	float32	传感器采集数据的频率,单位为赫兹(Hz),如相机帧率、IMU数据采集频率等。
传感器内参	intrinsic_params	dict (JSON格式)	传感器的内参,如用于相机标定和图像矫正,包含: 焦距(fx, fy); 主点坐标(cx, cy); 畸变系数(k1, k2, p1, p2, k3)。
传感器外参	extrinsic_params	dict (JSON格式)	传感器的外参,表示传感器在世界坐标系中的位置和姿态,如相机包括: 平移向量(t_x, t_y, t_z) 旋转四元数(q_w, q_x, q_y, q_z)或旋转矩阵(3×3)。

8.2 本体坐标系设置

8.2.1 宜采用采集的第一帧画面中人形机器人站立姿势下肩部中心点作为坐标系原点(其他传感器同坐标原点),见图2,对于定制人形机器人应提供详细本体坐标位置说明。坐标轴定义如下:

- X轴方向:机器人正前方,即机器人前进方向;
- Y轴方向:垂直于X轴,指向机器人左侧(左手所在的坐标系);

——Z 轴方向：垂直向上，与 X 轴和 Y 轴构成右手定则关系。

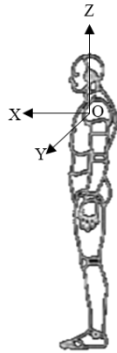


图2 人形机器人坐标系示例

8.2.2 若涉及与其他坐标系的转换，应提供明确的转换公式或方法说明。

8.3 数据集存储格式

8.3.1 人形机器人操作任务数据集在流转过程中应通过计算机（软件）验证是否符合规范。

8.3.2 人形机器人操作任务数据集存储格式见附录 B，内容包括数据集名称、主数据目录，主数据数据格式（parquet）、元数据目录、元数据数据格式（JSON）等。

8.4 数据集共享格式

人形机器人操作任务数据集应支持转换机器学习生态实现所需的格式，如tensorflow等。附录C给出了数据集转换RLDS规范的示例。

附录 A
(资料性)
数据集描述文件示例

A.1 数据集整体描述文件示例

A.1.1 示例1如下:

```
dataset name // 数据集的名称 (6.1中a)
ContactDualArm-2027-US-Lab-v1.2
size_categories: // 数据集大小
- n>1T
task_categories: // 数据集任务分类
- other
language: // 数据集使用语言
- en
tags: // 数据集标签
- real-world
- dual-arm
- Robotics manipulation

# Data Volume // 数据集内容信息 (6.1中b)
- **Trajectory Segments**:Over 3 million operational trajectory segments

# Data Types
-Visual Data: RGB images, depth images, video data.
-Auditory Data: Voice commands, ambient sounds.
-Tactile Data: Force sensor data.
-Kinematic Data: Joint angles, positions, velocities.
-Dynamic Data: Joint torques, accelerations.
-Formats: Image data in PNG format, video data in MP4 format, numerical data in
float32 type.

# Robot and Sensor Configuration
-Robot Type: Humanoid robot with dual-arm manipulation capabilities.
-Sensors: Visual sensors (cameras), auditory sensors (microphones), tactile sensors
(force sensors), kinematic sensors (IMU).
-Sensor Parameters: Camera intrinsic and extrinsic parameters, IMU sampling frequency,
etc.

# Application Scenarios // 数据集用途信息 (6.1中c)
Robot manipulation tasks in a laboratory setting, such as object grasping and
placement.

# Purpose
Supports the learning and optimization of robot manipulation skills, including
contact-rich operations, long-horizon planning, and multi-robot collaboration.
```

```

# Scope and Limitations
- **License**: CC BY-NC-SA 4.0 ([Details](https://creativecommons.org/licenses/by-nc-sa/4.0/))
- **Environmental Sensitivity**:
  - Requires lighting normalization below 600 Lux
  - Temperature compensation needed beyond 25°C

# Metadata Value Range Recommendations // 元数据取值建议 (6.1中d)
- Timestamps: Nanosecond precision, unit: ns.
- Joint Angles: In radians, ranging from 0 to  $2\pi$ .
- Position Data: In meters, precise to three decimal places.
- Forces and Torques: In Newtons (N) and Newton-meters (N·m), precise to two decimal places.

# Provenance Information // 数据集来源信息 (6.1中e)
## Collection Environment
- Temperature: 20°C -25°C
- Humidity: 40%-60%

## Acquisition Details
- **Location**: Robotics Lab, United States
- **Period**: 2027 (6-month collection cycle)
- **Data Provider**: Robotics Research Group

# Version and Access // 数据集版本与存储信息 (6.1中f)
## Version Control
- Current Version: v1.2 (2024-12-30)
- Changelog: Added 3 new industrial assembly scenarios

## Storage & Access
- **Repository**: [AWS S3](s3://dataset-bucket/ContactDualArm2027)
- **Mirror**: [Academic Cloud Storage](https://academic.cloud/dataset/ContactDualArm2027)

# Key Features // 数据集关键特征
- **3 million+** trajectories from 300 robots.
- **200+ real-world scenarios** across 6 target domains.
- **Cutting-edge hardware**: visual tactile sensors / 6-DoF dexterous hand / mobile dual-arm robots
- **Tasks involving**:
  - Contact-rich manipulation
  - Long-horizon planning
  - Multi-robot collaboration

# Table of Contents // 提供文档的目录结构
- [Key Features ](#key-features-)
- [News ](#news-)
- [TODO List ](#todo-list-)
- [Get started ](#get-started-)

```

- [Download the Dataset] (#download-the-dataset)
- [Dataset Structure] (#dataset-structure)
- [Explanation of Proprioceptive State] (#explanation-of-proprioceptive-state)
- [Dataset Preprocessing] (#dataset-preprocessing)
- [License and Citation] (#license-and-citation)

Get started // 使用指导

Download the Dataset

To download the full dataset, you can use the following code. If you encounter any issues, please refer to the official XXX documentation.

Dataset Preprocessing

Our project relies solely on the [XXXX] (https://), please follow their [installation instructions] (https://).

Dataset Structure

// 数据集结构

Folder hierarchy

// 文件夹层级说明

```

data
├── task_info
│   ├── task_327.json
│   └── ...
├── observations
│   ├── 327 # This represents the task id.
│   │   ├── 648642 # This represents the episode id.
│   │   └── depth # This is a folder containing depth information saved in
PNG format.
│   │       └── videos # This is a folder containing videos from all camera
perspectives.
│   │           └── ...
│   └── ...
├── parameters
│   ├── 327
│   │   ├── 648642
│   │   │   ├── camera
│   │   │   └── ...
│   └── 352
│       ├── 648544
│       └── camera # This contains all the cameras' intrinsic and extrinsic
parameters.
└── ...
├── proprio_stats
│   ├── 327[task_id]
│   │   ├── 648642[episode_id]
│   │   └── proprio_stats.h5 # This file contains all the robot's
proprioceptive information.
│   └── 648649
│       └── proprio_stats.h5

```



```

|-- state
  |-- effector
    |-- force
    |-- position
  |-- end
    |-- angular
    |-- orientation
    |-- position
    |-- velocity
    |-- wrench
  |-- head
    |-- effort
    |-- position
    |-- velocity
  |-- joint
    |-- current_value
    |-- effort
    |-- position
    |-- velocity
  |-- robot
    |-- orientation
    |-- orientation_drift
    |-- position
    |-- position_drift
  |-- waist
    |-- effort
    |-- position
    |-- velocity
|-- action
  |-- effector
    |-- force
    |-- index
    |-- position
  |-- end
    |-- orientation
    |-- position
  |-- head
    |-- effort
    |-- position
    |-- velocity
  |-- joint
    |-- effort
    |-- index
    |-- position
    |-- velocity
  |-- robot
    |-- index
    |-- orientation
    |-- position

```

```

    |-- velocity
|-- waist
    |-- effort
    |-- position
    |-- velocity

### Value shapes and ranges                                     // 数据形状与量程说明
| Group | Shape | Meaning |
| --- | :--- | :--- |
| /timestamp | [N] | timestamp in nanoseconds |
| /state/effector/position (dexhand) | [N, 12] | left `[:, :6]`, right `[:, 6:]`,
joint angle in rad |
| /state/end/orientation | [N, 2, 4] | left `[:, 0, :]`, right `[:, 1, :]`, flange
quaternion with xyzw |
| /state/end/position | [N, 2, 3] | left `[:, 0, :]`, right `[:, 1, :]`, flange xyz
in meters |
| /state/head/position | [N, 2] | yaw `[:, 0]`, pitch `[:, 1]`, rad |
| /state/joint/current_value | [N, 14] | left arm `[:, :7]`, right arm `[:, 7:]` |
| /state/joint/position | [N, 14] | left arm `[:, :7]`, right arm `[:, 7:]`, rad |
| /state/robot/orientation | [N, 4] | quaternion in xyzw, yaw only |
| /state/robot/position | [N, 3] | xyz position, where z is always 0 in meters |
| /state/waist/position | [N, 2] | pitch `[:, 0]` in rad, lift `[:, 1]` in meters |
| /action/*/index | [M] | actions indexes refer to when the control source is
actually sending signals |
| /action/effector/position (gripper) | [N, 2] | left `[:, 0]`, right `[:, 1]`, 0 for
full open and 1 for full close |
| /action/effector/position (dexhand) | [N, 12] | same as /state/effector/position
| /action/effector/index | [M_1] | index when the control source for end effector is
sending control signals |
| /action/end/orientation | [N, 2, 4] | same as /state/end/orientation |
| /action/end/position | [N, 2, 3] | same as /state/end/position |
| /action/end/index | [M_2] | same as other indexes |
| /action/head/position | [N, 2] | same as /state/head/position |
| /action/head/index | [M_3] | same as other indexes |
| /action/joint/position | [N, 14] | same as /state/joint/position |
| /action/joint/index | [M_4] | same as other indexes |
| /action/robot/velocity | [N, 2] | vel along x axis `[:, 0]`, yaw rate `[:, 1]` |
| /action/robot/index | [M_5] | same as other indexes |
| /action/waist/position | [N, 2] | same as /state/waist/position |
| /action/waist/index | [M_6] | same as other indexes |

```

A.1.2 示例2如下:

```

{
    "codebase_version": "v2.0",           // 代码库版本
    "robot_type": "gr1t1",                // 机器人型号
    "total_episodes": 100,                 // 总片段数量
    "total_frames": 44876,                 // 总数据帧数
    "total_tasks": 16,                     // 涵盖任务类型数
    "total_videos": 100,                   // 视频文件总数

```



```

        "total_chunks": 1, // 数据分块总数
        "chunks_size": 1000, // 每个分块包含1000个情节
        "fps": 30, // 数据采集帧率（帧/秒）
        "splits": {
            "train": "0:100"
        },
        "data_path": "data/chunk-{episode_chunk:03d}/episode_{episode_index:06d}.parquet",
// 状态数据存储路径
        "video_path": "videos/chunk-{episode_chunk:03d}/{video_key}/episode_{episode_index:06d}.mp4", // 视频存储路径
        "features": { // 顶部摄像头观测
            "observation.images.top": {
                "dtype": "video", // 视频数据类型
                "copy": true, // 数据拷贝存储
                "shape": [ // [通道数, 高度, 宽度]
                    3,
                    800,
                    1280
                ],
                "names": [
                    "channel",
                    "height",
                    "width"
                ],
                "info": { // 视频编码参数
                    "video.fps": 30.0, // 帧率
                    "video.height": 800, // 画面高度
                    "video.width": 1280, // 画面宽度
                    "video.channels": 3, // RGB三通道
                    "video.codec": "h264", // H. 264编码
                    "video.pix_fmt": "yuv420p", // 像素格式
                    "video.is_depth_map": false, // 非深度图
                    "has_audio": false // 无音频
                }
            },
            "observation.state": { // 机器人关节状态观测
                "dtype": "float32",
                "shape": [
                    32
                ],
                "names": [ // 关节名称列表
                    "waist_yaw_joint", // 腰部偏航关节
                    "waist_pitch_joint", // 腰部俯仰关节
                    "waist_roll_joint", // 腰部横滚关节
                    "head_yaw_joint", // 头部偏航
                    "head_roll_joint", // 头部横滚
                    "head_pitch_joint", // 头部俯仰
                    "left_shoulder_pitch_joint", // 左肩俯仰
                    "left_shoulder_roll_joint", // 左肩横滚

```

```

        "left_shoulder_yaw_joint",    // 左肩偏航
        "left_elbow_pitch_joint",    // 左肘俯仰
        "left_wrist_yaw_joint",      // 左腕偏航
        "left_wrist_roll_joint",     // 左腕横滚
        "left_wrist_pitch_joint",    // 左腕俯仰
        "right_shoulder_pitch_joint", // 右肩俯仰
        "right_shoulder_roll_joint",  // 右肩横滚
        "right_shoulder_yaw_joint",   // 右肩偏航
        "right_elbow_pitch_joint",    // 右肘俯仰
        "right_wrist_yaw_joint",      // 右腕偏航
        "right_wrist_roll_joint",     // 右腕横滚
        "right_wrist_pitch_joint",    // 右腕俯仰
        "L_pinky_proximal_joint",     // 左手小指近端关节
        "L_ring_proximal_joint",      // 左手无名指近端
        "L_middle_proximal_joint",    // 左手中指近端
        "L_index_proximal_joint",     // 左手食指近端
        "L_thumb_proximal_pitch_joint", // 左手拇指俯仰
        "L_thumb_proximal_yaw_joint",  // 左手拇指偏航
        "R_pinky_proximal_joint",     // 右手小指近端关节
        "R_ring_proximal_joint",      // 右手无名指近端
        "R_middle_proximal_joint",    // 右手中指近端
        "R_index_proximal_joint",     // 右手食指近端
        "R_thumb_proximal_pitch_joint", // 右手拇指俯仰
        "R_thumb_proximal_yaw_joint"  // 右手拇指偏航
    ]
},
"action": {                                // 动作控制指令
    "dtype": "float32",
    "shape": [
        32
    ],
    "names": [
        "waist_yaw_joint",
        "waist_pitch_joint",
        "waist_roll_joint",
        "head_yaw_joint",
        "head_roll_joint",
        "head_pitch_joint",
        "left_shoulder_pitch_joint",
        "left_shoulder_roll_joint",
        "left_shoulder_yaw_joint",
        "left_elbow_pitch_joint",
        "left_wrist_yaw_joint",
        "left_wrist_roll_joint",
        "left_wrist_pitch_joint",
        "right_shoulder_pitch_joint",
        "right_shoulder_roll_joint",
        "right_shoulder_yaw_joint",
        "right_elbow_pitch_joint",
    ]
}

```

```

        "right_wrist_yaw_joint",
        "right_wrist_roll_joint",
        "right_wrist_pitch_joint",
        "L_pinky_proximal_joint",
        "L_ring_proximal_joint",
        "L_middle_proximal_joint",
        "L_index_proximal_joint",
        "L_thumb_proximal_pitch_joint",
        "L_thumb_proximal_yaw_joint",
        "R_pinky_proximal_joint",
        "R_ring_proximal_joint",
        "R_middle_proximal_joint",
        "R_index_proximal_joint",
        "R_thumb_proximal_pitch_joint",
        "R_thumb_proximal_yaw_joint"
    ]
},
"observation.state.pose": {           // 末端执行器位姿观测
    "dtype": "float32",
    "shape": [
        36
    ],
    "names": [
        "left_hand_x",           // 左手三维坐标
        "left_hand_y",
        "left_hand_z",
        "left_hand_ortho6d_1",   // 左手6D旋转表示
        "left_hand_ortho6d_2",
        "left_hand_ortho6d_3",
        "left_hand_ortho6d_4",
        "left_hand_ortho6d_5",
        "left_hand_ortho6d_6",
        "right_hand_x",          // 右手三维坐标
        "right_hand_y",
        "right_hand_z",
        "right_hand_ortho6d_1",  // 右手6D旋转表示
        "right_hand_ortho6d_2",
        "right_hand_ortho6d_3",
        "right_hand_ortho6d_4",
        "right_hand_ortho6d_5",
        "right_hand_ortho6d_6",
        "head_ortho6d_1",        // 头部6D姿态
        "head_ortho6d_2",
        "head_ortho6d_3",
        "head_ortho6d_4",
        "head_ortho6d_5",
        "head_ortho6d_6",
        "L_pinky_proximal_joint", // 手部关节
        "L_ring_proximal_joint",

```

```

        "L_middle_proximal_joint",
        "L_index_proximal_joint",
        "L_thumb_proximal_pitch_joint",
        "L_thumb_proximal_yaw_joint",
        "R_pinky_proximal_joint",
        "R_ring_proximal_joint",
        "R_middle_proximal_joint",
        "R_index_proximal_joint",
        "R_thumb_proximal_pitch_joint",
        "R_thumb_proximal_yaw_joint"
    ]
},
"action.pose": {                                     // 位姿控制指令
    "dtype": "float32",
    "shape": [
        36
    ],
    "names": [
        "left_hand_x",
        "left_hand_y",
        "left_hand_z",
        "left_hand_ortho6d_1",
        "left_hand_ortho6d_2",
        "left_hand_ortho6d_3",
        "left_hand_ortho6d_4",
        "left_hand_ortho6d_5",
        "left_hand_ortho6d_6",
        "right_hand_x",
        "right_hand_y",
        "right_hand_z",
        "right_hand_ortho6d_1",
        "right_hand_ortho6d_2",
        "right_hand_ortho6d_3",
        "right_hand_ortho6d_4",
        "right_hand_ortho6d_5",
        "right_hand_ortho6d_6",
        "head_ortho6d_1",
        "head_ortho6d_2",
        "head_ortho6d_3",
        "head_ortho6d_4",
        "head_ortho6d_5",
        "head_ortho6d_6",
        "L_pinky_proximal_joint",
        "L_ring_proximal_joint",
        "L_middle_proximal_joint",
        "L_index_proximal_joint",
        "L_thumb_proximal_pitch_joint",
        "L_thumb_proximal_yaw_joint",
        "R_pinky_proximal_joint",

```

```

        "R_ring_proximal_joint",
        "R_middle_proximal_joint",
        "R_index_proximal_joint",
        "R_thumb_proximal_pitch_joint",
        "R_thumb_proximal_yaw_joint"
    ]
},
"timestamp": {                                // 时序戳（秒）
    "dtype": "float32",
    "shape": [
        1
    ],
    "names": null
},
"frame_index": {                              // 帧序列号
    "dtype": "int64",
    "shape": [
        1
    ],
    "names": null
},
"episode_index": {                           // 片段编号
    "dtype": "int64",
    "shape": [
        1
    ],
    "names": null
},
"index": {                                    // 全局索引
    "dtype": "int64",
    "shape": [
        1
    ],
    "names": null
},
"task_index": {                              // 任务类型编号
    "dtype": "int64",
    "shape": [
        1
    ],
    "names": null
},
"next.done": {                               // 片段终止标志
    "dtype": "bool",
    "shape": [
        1
    ],
    "names": null
}
}

```

```

    }
}

```

A.2 片段描述文件示例

```

[
  {
    "episode_id": 111,                // 片段唯一标识
    "task_id": "001",                // 任务类型编号
    "task_name": "Hanging clothes with a hanger", // 任务名称
    "task_description": "A robot stands beside the bed or sofa, with clothes and
a hanger placed on the bed or sofa.", // 任务特征描述
    "timestamp": "2023-07-20T14:30:00Z", // 时间戳
    "init_scene_text": "A robot stands beside the bed or sofa, with clothes and a
hanger placed on the bed or sofa.", // 场景描述
    "environmental_and_perception_data": { // 环境感知数据
      "layout": {
        "work_area": "2.5x1.8m area", // 工作区域尺寸
        "obstacles": ["bed", "sofa"], // 障碍物列表
        "coordinates": { // 关键物体坐标
          "hanger": [1.2, 0.8, 0.6],
          "clothes": [1.0, 0.7, 0.5]
        }
      },
      "sensors": { // 传感器配置
        "camera": {
          "type": "RGBD",
          "resolution": "1280x800"
        },
        "force_torque": "OnRobot SGT-50"
      },
      "lighting": 350 // 光照强度
    },
    "label_info": {
      "action_config": [ // 动作序列配置
        {
          "start_frame": 0, // 动作起始帧
          "end_frame": 302, // 结束帧
          "skill": "Pick" // 技能类型
        },
        "control_command": "Move to hang at (x=1.2,y=0.8,z=2.0)", // 控制
        "robot_task_status_data": { // 机器人状态
          "position": [1.0, 0.7, 0.5],
          "joint_angles": [45, 30, 15, 0, 20, ...], // 关节角度
          "gripper_force": 0.5 // 抓取力度
        }
      },
      ...
    ]
  }
]

```

指令

```

        "start_frame": 1913,
        "end_frame": 2417,
        "skill": "Hang",
        "control_command": "End effector moves to (x=1.2,y=0.8,z=2.0)",
        "environmental_dynamics": { // 环境动态
            "human_distance": 2.3 // 人员距离
        }
    },
    ],
    "key_frame": [ // 关键帧标注
        {
            "frame": 302, // 首次抓取完成帧
            "sensor_data": { // 传感器数据
                "rgb_image": "frame_0302.jpg",
                "force_peak": 0.52
            }
        }
    ]
},

"execution_status": { // 执行状态
    "result": "success", // 总体结果
    "accuracy": 98.7, // 悬挂位置精度
    "wrinkle_level": 0.2 // 衣物褶皱系数
}

}

]

```

附录 B (资料性) 数据存储格式示例

```

RobotDataset_202403_kitchen_fxk_v1.0           // 数据集名称
├── data                                           // 主数据存储文件夹
│   └── 356                                       // 分块编号
│       ├── episode_000000.parquet               // 主数据格式
│       ├── episode_000001.parquet
│       └── ....
├── meta                                           // 元数据管理文件夹
│   ├── episodes.jsonl                           // 元数据格式
│   ├── info.json                               // 数据集整体描述文件信息
│   └── tasks.jsonl                             // 任务定义文件
├── videos
│   └── 356
│       ├── observation_image_back_left_fisheye_color
│       │   ├── episode_000000.mp4
│       │   └── .....
│       ├── observation_image_back_right_fisheye_color
│       │   ├── episode_000000.mp4
│       │   ├── episode_000001.mp4
│       │   └── .....
│       ├── observation_image_hand_left_color
│       │   ├── episode_000000.mp4
│       │   └── .....
│       ├── observation_image_hand_right_color
│       │   ├── episode_000000.mp4
│       │   └── .....
│       ├── observation_image_head_center_fisheye_color
│       │   ├── episode_000000.mp4
│       │   └── .....
│       ├── observation_image_head_color
│       │   ├── episode_000000.mp4
│       │   └── .....
│       ├── observation_image_head_left_fisheye_color
│       │   ├── episode_000000.mp4
│       │   └── .....
│       └── observation_image_head_right_fisheye_color
│           ├── episode_000000.mp4
│           └── .....
XX directories, XX files                         // 总计XX个目录，XX个文件

```


附录 C (资料性) 数据集转换 RLDS 规范示例

C.1 转换 RLDS 应遵循 tensorflow 数据集库，转换后的 RLDS 应遵循主数据固有格式要求为：

```
Python
{'action': {'open_gripper': {tf.Tensor: shape=(), dtype=bool, numpy=True}}} # action
是一个TensorFlow张量对象
{'is_first': {tf.Tensor: shape=(), dtype=bool, numpy=True}}
{'is_last': {tf.Tensor: shape=(), dtype=bool, numpy=False}}
{'is_terminal': {tf.Tensor: shape=(), dtype=bool, numpy=False}}
{'observation': {'image': {tf.Tensor: shape=(480, 640, 3), dtype=uint8,
numpy=array([], dtype=uint8)}, 'natural_language_embedding': {tf.Tensor: shape=(512,),
dtype=float32, numpy=array([], dtype=float32)}, 'natural_language_instruction':
{tf.Tensor: shape=(), dtype=string, numpy=b''}, 'state': {tf.Tensor: shape=(7,),
dtype=float32, numpy=array([], dtype=float32)}}}
{'reward': {tf.Tensor: shape=(), dtype=float64, numpy=0.0}}
{'discount': {tf.Tensor: shape=(), dtype=float64, numpy=0.0}} # reward的折扣因子
```

C.2 转换后的 RLDS 的元数据应遵循元数据固有格式要求为：

```
JSON
{
    "citation": "",
    "description": "",
    "fileFormat": "",
    "location": {"urls": [""]},
    "moduleName": "", "name": "",
    "releaseNotes": {"0.1.0": ""},
    "splits": [{"filepathTemplate": "", "name": "", "numBytes": "", "shardLengths":
[]}]
    ],
    "version": "0.1.0"
}
```

C.3 转换后的 RLDS 的特征字典数据应遵循的特征字典数据固有格式要求为：

```
JSON
{
    "pythonClassName": "",
    "featuresDict": {
        "features": {
            "episode_metadata": {"pythonClassName": "", "featuresDict": {"features":
{"file_path": {"pythonClassName": "", "text": {}, "description": ""}}}},
            "steps": {"pythonClassName": "", "sequence": {"feature":
{"pythonClassName": "", // step特征表示强化学习过程中的一个步骤，该步骤步骤包含了智能体
的动作、观察、奖励等信息
            "featuresDict": {
                "features": {
```


参 考 文 献

- [1] GB/T 5271.1—2000 信息技术词汇 第1部分：基本术语
 - [2] GB/T 12643—202X 机器人与机器人装备 词汇
 - [3] GB/T 15237.1-2000 术语工作 词汇 第1部分：理论与应用
 - [4] GB/T 18221—2000 信息技术 程序设计语言、环境与系统软件接口 独立于语言的数据类型
 - [5] GB/T 18391.1—2009 信息技术 实现元数据注册系统（MDR）内容一致性的规程 第1部分：
数据元
 - [6] GB/T 43340.1—2023 财经新闻信息 第1部分：元数据
 - [7] ISO 8000-2: 2022 Data quality — Part 2: Vocabulary
 - [8] ISO 19115-1: 2014 Geographic information — Metadata Part 1: Fundamentals
-