

PIXEL-REASONER 数据格式

Using *tyslides*

尹超

2026-02-10

<https://github.com/TIGER-AI-Lab/Pixel-Reasoner>

CONTENTS

1. 数据输入格式与输出

数据输入格式与输出

- Pixel-Reasoner 项目主要处理三种类型的数据：
 1. RL / Eval 数据 (parquet 格式)
 2. SFT 数据 (消息列表 format)
 3. 最终模型输入 (chat message 格式)
- 每种数据类型有独特的格式，要求图像/视频路径与其他字段的一致性。

- 数据格式: parquet
- 主要字段:
 - ▶ question: 题目文本
 - ▶ answer: 答案 (通常为 \boxed{} 格式)
 - ▶ image: 图像路径
 - ▶ is_video: 是否为视频题目
 - ▶ qid: 唯一 ID

- 示例：

```
{  
    "qid": "train-xxx-123",  
    "question": "What is written on the orange box?\nchoices:\nA: ...  
    \nB: ...",  
    "answer": ["\\boxed{A}"],  
    "image": ["/abs/path/to/images/0.jpg"],  
    "is_video": false  
}
```

- 数据格式: JSON 格式, 包含消息列表 (message_list)
- 主要字段:
 - ▶ message_list: 一系列对话消息
 - ▶ qid: 唯一 ID
- 每个消息包含:
 - ▶ role: 用户、助手、系统
 - ▶ content: 图像或文本内容

- 示例：

```
{  
  "qid": "0",  
  "message_list": [  
    {  
      "role": "system",  
      "content": [  
        {"text": "You are a helpful assistant...", "image": null, "video":  
null}  
      ]  
    },  
    {  
      "role": "user",  
      "content": [  
        {"text": "What is written on the orange box?\nchoices:\nA: ...",  
"image": null, "video": null},  
      ]  
    }  
  ]  
}
```

```
    {"text": null, "image": "images/0-0.jpg", "video": null}
  ]
},
{
  "role": "assistant",
  "content": [
    {"text": "Step-by-step reasoning... \boxed{A}", "image": null,
"video": null}
  ]
}
}
```

- 模型输入：将数据格式化为 chat message 格式
- 将 question、image（单图或视频帧）、is_video 转换为消息列表
- 示例：

```
[  
  {"role": "user", "content": [{"text": "What is written on the orange box?  
\nchoices:\nA: ...", "image": null, "video": null}, {"text": null, "image":  
"images/0-0.jpg", "video": null}]}]  

```

- 工具调用格式：每次工具调用返回一个 JSON，格式如下：

```
<tool_call>
{"name": "crop_image_normalized", "arguments": {"bbox_2d": [0.12, 0.30, 0.55, 0.62], "target_image": 1}}
</tool_call>
```

- 可用于裁剪图像、选取视频帧等操作。

- 最终答案：在推理过程中返回最终答案，格式为 `\boxed{}`。
- 示例：

```
{  
  "response": "\boxed{A}"  
}
```

- 逐样本输出：每个样本在评估过程中会生成一个 JSONL 文件，记录评估结果。
- 示例：

```
{"qids": "vstar-001", "question": "What is written on the orange  
box?", "response": "...  
\boxed{B}", "match": true, "round0_correctness": 0, "round1_correctness": 1,  
"num_toolcalls": 2}
```

- 聚合指标输出：评估汇总结果，以 JSON 格式存储。
- 示例：

```
{"benchname": "Visual  
Reasoning", "pass1": 0.85, "ncorrect": 50, "ntotal": 100, "modelpath": "model_v1"}
```