

caption

大学难度问题

数学推理

图表文档阅读

general visual questioning answering

video understanding

能力类别与数据集示例

能力类别	数据集示例
图像描述生成	COCO-Cap, Flickr30K, NoCaps, TextCaps
图文检索	COCO-Cap, Flickr30K, RefCOCO, RefCOCO+, RefCOCOG
图表/文档问答	AI2D, ChartQA, DocVQA, InfoVQA
自然场景问答	GQA, VizWiz-VQA, TextVQA, OK-VQA, ScienceQA, MathVista
综合能力评测	MMBench, MME, MMVET
特定领域任务	POPE, ScienceQA, MathVista

图像描述生成

- 能力：生成自然语言描述，评估视觉到语言的生成能力。
- 数据集：COCO-Cap, Flickr30K, NoCaps, TextCaps。

图文检索（跨模态对齐）

- 能力：图像和文本模态之间的检索与对齐能力。
- 数据集：COCO-Cap, Flickr30K, RefCOCO, RefCOCO+, RefCOCOG。

图表/文档问答

- 能力：从结构化或半结构化数据中进行信息提取与推理。
- 数据集：AI2D, ChartQA, DocVQA, InfoVQA。

自然场景问答

- 能力：回答与自然场景相关的问题，评估视觉内容理解与语言推理能力。
- 数据集：GQA, VizWiz-VQA, TextVQA, OK-VQA, ScienceQA, MathVista。

综合能力评测

- 能力：测试模型在生成、推理、对齐等多方面的能力。
- 数据集：MMBench, MME, MMVET。

特定领域任务

- **能力**：在领域特定任务中的应用能力。
- **数据集**：POPE, ScienceQA, MathVista。