

多模型协同的科普挂图资源生成智能体系统

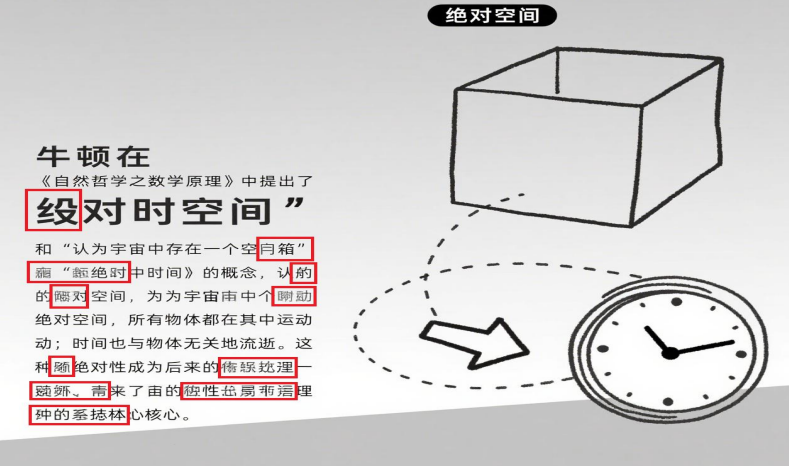
关贝, 李峻羽, 张芸鹏, 孙哲, 王永吉
集成创新中心

联系人: 关贝, 15011316325, guanbei@iscas.ac.cn

系统简介

科普挂图作为面向科学知识传播的多模态资源, 通常包含密集的文字说明和相应的图像、图表、插图等元素。其传统制作流程长, 涵盖主题界定、文案撰写、图像设计与排版、审核修改等环节, 既需保证内容之权威性与通俗性, 又要求较高的图像制作技术, 审核返工成本亦高。为此, 该系统工具提出基于多模型协同的端到端科普挂图全自动生成智能体, 有效克服当前文生图大模型在中文长文本精准渲染上的局限, 显著促进科普资源的高效生成与广泛传播。

牛顿的绝对时空观



文生图大模型存在中文长文本准确渲染缺陷

功能指标

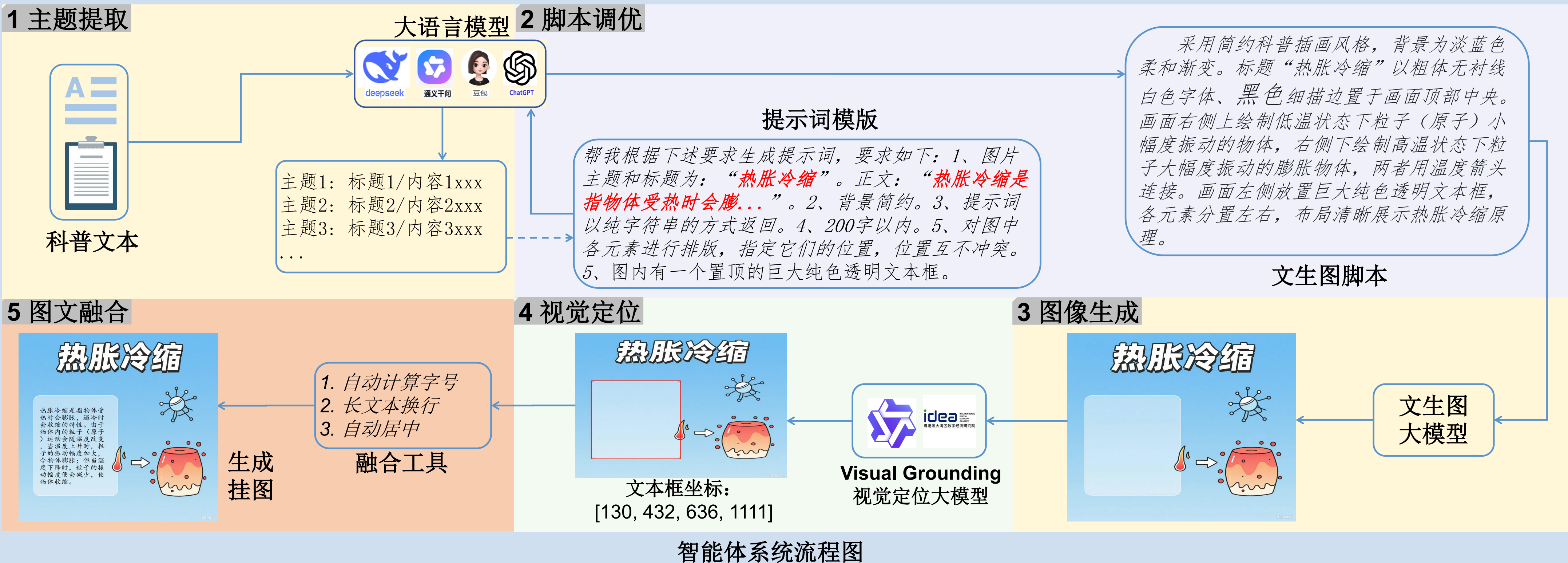
- 中文长文本图像生成: 以图文融合策略, 弥合多模态大模型在中文长文本准确渲染上的缺陷。
- 图文内容一致性: 依托大模型提示词重写实现文生图脚本调优, 确保图像与文本内容高度一致。
- 端到端全自动生成: 基于科普文本自动高效产出科普挂图, 支持后验校验与迭代优化, 全流程无需人工干预。

创新点

- 提出一种多模型协同的中文长文本图像生成方法, 通过字形-语义-布局三重引擎联动, 精准渲染任意长度的密集中文内容, 解决现有文生图大模型在中文长文本上的失真问题。
- 提出一种基于智能体的端到端全自动生成科普挂图资源的新范式, 实现“一次输入-秒级出图-立即可用”的全流程无人化生产, 成倍提升科普资源产生效率, 为科普事业注入持续动力。

技术细节

- 主题提取: 基于大语言模型对输入的科普长文本进行主题解析, 生成同系列异主题科普语料。
- 脚本调优: 依托预制提示词规则模板, 调用大模型生成文生图脚本, 保障图文语义一致。
- 图像生成: 基于文生图多模态大模型生成高分辨率、结构合理的科普图像, 并预留文本框区域。
- 视觉定位: 调用具备强几何能力的视觉定位大模型精确识别文本框并给出精准坐标。
- 图文融合: 用户脚本自适应计算文本字号等布局参数, 完成图文融合并自动评估融合质量。



成果展示

[天文] 绝对时空

牛顿的绝对时空观

[医药] 麻苧止咳糖浆介绍

麻苧止咳糖浆的主要成分是什么?

麻苧止咳糖浆的功能与主治是什么?

[科技] 机器人“触摸”: 植物识别的创新探索

机器人通过电极触感知植物

传统监测设备的局限性

[物理] 热胀冷缩的科学原理

热胀冷缩

原子

温度

受人类皮肤启发的机器人结构

机器学习用于植物识别

不同科普领域挂图资源自动生成效果展示

应用领域

- 该科普挂图生成智能体框架可广泛应用于以下场景与行业, 显著提升知识传播效率、降低制作门槛, 并催生新的服务模式:
- 科技馆、博物馆与主题展览: 基于已有文本科普资源批量输出科普挂图, 依托热点事件24h内上线“特展挂图”等;
 - 基础教育与课外科普: 生成主题海报、实验流程挂图、科学家故事墙等;
 - 政府、社区、公共卫生与应急科普: 用于突发传染病、防灾减灾宣传栏及农业技术挂图、气象防灾指南等;
 - 企业员工培训: 将技术手册即时转化为可视化培训材料等;
 - 长文本图像合成训练数据等。