项目编号：

广东技术师范大学
大学生创新创业训练计划项目申报书

（创新训练项目）

项目名称：依托家电实训中心软件平台制作改良文生图模型符号质量的数据集

项目负责人姓名、学号：罗煜斌 2023044743109

项目负责人所在院系：电子与信息学院

项目负责人所学专业：网络工程

本校指导教师姓名、工号：欧阳明俊 2306003

校外指导教师姓名、单位：无

申请日期：2024.03.28

一、基本情况

<table>
<tr><td>项目
名称</td><td colspan="7">依托家电实训中心软件平台制作改良文生图模型符号质量的数据集</td></tr>
<tr><td>所属
学科</td><td colspan="7">0809 计算机类</td></tr>
<tr><td>申请
金额</td><td colspan="2">8000 元</td><td colspan="2">起止年月</td><td colspan="3">2024 年 4 月至 2025 年 4 月</td></tr>
<tr><td>负责人
姓名</td><td>罗煜斌</td><td>性别</td><td>男</td><td>民族</td><td>汉</td><td>出生
年月</td><td>2004 年 08
月</td></tr>
<tr><td>负责人
学号</td><td colspan="3">2023044743109</td><td>电话/
邮箱</td><td colspan="3">13510829172
2185494983@qq.com</td></tr>
<tr><td>第一指导
教师姓名</td><td colspan="3">欧阳明俊</td><td>电话/
邮箱</td><td colspan="3">13316257131
ouymj@gpnu.edu.cn</td></tr>
<tr><td colspan="2">负责人曾经参与
科研的情况</td><td colspan="6">本项目负责人罗煜斌学习过编程,较熟练地掌握C语言和python，可根据项目要求运用所学知识并在其基础上进行运用创新。对神经网络、人工智能、深度学习、数据集都有所了解。具有较强的创新能力和说干就干的行为准则，对本项目在思路和行动上有较大的帮助。</td></tr>
<tr><td colspan="2">指导教师承担
科研课题情况</td><td colspan="6">(1) 2023 年-至今，企业横向，基于 IMU 的作业单兵室内定位系统研究，项目负责人。
(2) 2022 年-至今，企业横向，电网低成本 RTK 板卡研发，项目实际负责人。
(3) 2022 年-2023 年，企业横向，次纳秒量级授时接收机研制，项目实际负责人。
(4) 2019 年-2022 年，北斗重大专项，基于北斗三号***传输技术研究，骨干。
(5) 2019 年-2022 年，北斗重大专项，北斗与 5G 通信导航融合演示验证，骨干。</td></tr>
<tr><td colspan="2">指导教师对本项
目的支持情况</td><td colspan="6">在该项目中，欧阳明俊老师给予了我们如下支持：
（1）为项目团队的研究背景和研究设计提供方向性的指导，在难点和重点部分进行把关，帮助项目团队解决项目的难点；
（2）为项目团队的实践创造条件，根据项目团队的实际需求，可以为项目团队协调实地考察方面的工作；
（3）为项目团队的研究提供较为前沿的学术观点的指导
（4）依托学院实验室资源，为项目团队的实验提供设备上的支持</td></tr>
<tr><td rowspan="2">团队所有成员（按序填写）</td><td>姓名</td><td>学号</td><td>学院（校外学生填学校名称）</td><td>学历</td><td>专业</td><td colspan="2">分工</td></tr>
<tr><td>罗煜斌</td><td>2023044
743109</td><td>电子与信息学院</td><td>本科</td><td>网络工程</td><td colspan="2">撰写项目计划书；设计场景，收集数据，制作数据集。</td></tr>
</table>

	张凯	2023044743122	电子与信息学院	本科	网络工程	撰写项目计划书；负责主要的构建自动生成数据的工作流程。
	陈炜逸	2023044743110	电子与信息学院	本科	网络工程	撰写项目计划书；总结各实验的报告，制作PPT
	林镇州	2023044743116	电子与信息学院	本科	网络工程	撰写项目计划书；做组织工作；跟指导老师进行沟通；收集文献和项目有关的资料。
	廖思齐	2023044743112	电子与信息学院	本科	网络工程	撰写项目计划书；开展项目的阶段性总结会议。
指导教师（按序填写）	姓名	工号	所在单位	学历	职务/职称	分工
	欧阳明俊	2306003	电子与信息学院	博士	校聘副教授	指导项目的进行。

二、立项依据（可加页）

（一）研究目的

近年来，基于扩散模型的文生图，文生视频神经网络表现出令人印象深刻的生成能力，但它们在符号系统的生成上尚有不足。符号系统包括了文字和几何图形，是一系列拥有严格的形状和排序规则的标识。文生视频模型基于扩散网络结构，以用户输入的提示词为初始权重，经过一系列非线性变换后得到连续图像，输出的图像在时空上具有连续性。符号系统的标识在时空上具有连续性，逻辑性，和合理性，而现有扩散模型生成的符号只具有连续性，缺乏了逻辑和合理性，具体表现为破碎凌乱的文字，或不符合几何规律的数学图形。

文生视频模型 Sora—车道线破碎且毫无逻辑

当前，针对文生图，文生视频的模型工业化，民用化是热门趋势。此类模型在文化传播领域，新媒体领域有广阔应用前景。但是其在符号系统生成的问题限制了其应用场景。

- **学院基础：**

学院现有省级教学团队、省级科研团队 2 个，国家级精品课程 1 门、省级精品课程 3 门，拥有省部级科研平台 6 个、省部级教学平台 6 个；拥有 16 个校外专业、教育实习实训基地；拥有教育部-中兴 ICT 产教融合基地、广东省电子信息工程应用型人才培养示范基地、广东省通信工程应用型人才培养示范基地、“电子与信息工程”教育部职教师资培养培训基地、“卓越网络工程师”广东省大学生校外实践教学基地、“电子与信息工程”广东省实验示范中心等 6 个实践基地。拥有广东省知识产权大数据重点实验室、广东省智能信息处理与嵌入式人工智能工程中心、广东高校未来网络工程技术研究中心、广东高校移动信息工程技术研究中心、广东省粤港制造云国际合作基地、广东省民族发展大数据工程技术研究中心、广东省普通高校国际暨港澳台合作创新平台等省部级科研平台；以及研究生联合培养基地 10 个，其中 4 个省级，6 个院级。

家电实训中心软件平台：

2023 年，为顺应数字经济对“数字工匠”、“数字工匠之师”、“信息技术应用创新”的人才需求和本校面向“数字经济”涉及到的新一代电子信息技术领域培养高素质“技术+师范”信息技术应用创新人才的要求，本院落地建设了广东智能家电产业集群科产教融合实训中心，以信创技术构建“数字场景”之家，打造“场景式”教学培训产品体系，实现“产学研转创用”。

目前，广东智能家电产业集群科产教融合实训中心相关软件平台和模块已部署上线，形成了以智慧家电实训教学与科研科创开发交互平台，运动健康实验教学与科研科创开发交互平台，智慧康养实验教学与科研科创开发交互平台，智能机器人助理实验教学与科研科创开发交互平台，实训中心 AI 教学与科研科创平台，实训中心实验教学与科研科创大数据系统，实训中心源码管理与调用平台，实训中心教学与科研后台业务管理系统为核心，全面服务于

教学，科研和科创的软件服务平台。基于广东智能家电产业集群科产教融合实训中心实际应用场景，围绕上述软件服务平台，电子与信息学院将在以下方面持续发力，切实依托家电实训中心软件服务平台，推动教学、科研、科创高质量发展。

在科研方面，电子与信息学院拟依托实训中心软件平台，开发数字孪生应用平台，构建空间模型和设备控制模型。开展基于人工智能算法、GET3、RGBD 相机、三维条纹投影重构、机器视觉等领域的数字孪生建模技术等领域的相关研究，承接相关科研项目并发表高水平论文，以推动家电产业的智能化、高效化和可持续发展。

1.数字孪生应用平台开发

与思林杰科技公司、卡特加特公司合作，共同研发 1 个数字孪生应用平台，该平台将集成孪生场景建模、设备控制、数据分析等功能，为家电产品的设计、测试、优化提供全方位支持。共同确定平台的功能需求和技术架构，并进行系统设计和开发。构建一个高精度、高仿真度的空间模型，用于模拟家电产品的实际运行环境。开发 10 个设备控制模型，实现对家电产品的虚拟控制和仿真测试。

2.基于人工智能算法的数字孪生建模研究

研究如何运用人工智能算法对家电产品进行智能化建模，提高模型的精度和仿真度。探索深度学习、机器学习等算法在数字孪生建模中的应用，提升模型的自适应性和预测能力。基于 GET3、RGBD 相机和三维条纹投影重构的数字孪生技术研究研究 GET3 技术在数字孪生中的应用，实现对家电产品的高精度测量和建模。利用 RGBD 相机和三维条纹投影重构技术，获取家电产品的三维信息，为建模提供丰富的数据支持。研究机器视觉技术在数字孪生建模中的应用，实现对家电产品的自动化识别和建模。探索基于机器视觉的模型优化方法，提高模型的准确性和可靠性。

3.预期成果

成功开发出一个功能完善的数字孪生应用平台，为家电产品的设计、测试、优化提供有力支持。构建出高精度、高仿真度的空间模型和设备控制模型，为家电产品的虚拟仿真和测试提供可靠基础。在数字孪生建模技术领域取得一系列创新成果，形成具有自主知识产权的核心技术。培养一支掌握数字孪生技术的科研团队，为未来的科技创新和人才培养奠定坚实基础。

广东省智能信息处理与嵌入式人工智能工程技术研究中心：

本中心的前身是广东省现代职业教育信息化及应用服务工程技术研究中心，根据学校的发展规划和电子与信息学院的学科发展要求，结合数字经济这个大时代背景，为了充分发挥云平台的应用价值，本中心不再仅仅只是将其运用在职业教育信息化这个领域，本中心还计划将云平台应用到智慧医疗和工业互联网这两个领域。为了突出本工程中心的任务与工作重心的转变，本中心拟更名为广东省智能信息处理与嵌入式人工智能工程技术研究中心。本工程中心拥有一支学历高、业务能力强、科研经验丰富的技术研发与成果推广的团队。现有专职研发开发人员 32 人，其中高级职称 8 人、中级职称 11 人、获得博士学位者 22 人、获得硕士学位者 9 人。中心组成员在职业教育、云计算、移动互联网、信息安全、大数据处理、软件开发和系统集成等领域具有丰富的经验，能为本中心后续工作的开展提供强大的技术支持。中心在学校的指导下，以智能信息处理与嵌入式人工智能为主，面向智慧教育领域、面向智慧医疗领域、面向工业互联网领域等相关技术领域整体突破。在智慧教育领域，通过开展智慧教育信息化资源建设，及相应的教学培训、产品研发和成果转化，为广东省产业升级提供复合型高素质人才支撑的同时，争取成为省内智慧教育模范单位。在智慧医疗领域，为社会培养输送具有基于智能科技的医工结合应用技术水平和相关管理及服务理念的创业和工程创新型人才，同时加强新一代信息技术在医疗体系中的

应用，推动社会医疗服务进一步走向智能化，为未来更多的患者带来福音。在工业互联网领域，从智能设备、智能系统和智能决策三个层次上为广东省乃至国家的工业 4.0 及中国制造 2025 战略培养具有新兴工业互联网理念、技术及管理技能的应用型复合人才。中心采用聘用与培养相结合的方式，受聘专家的学术待遇与同级的工程中心人员相同，以广泛吸引国内外本领域的优秀人才；在管理上实行严格的末位淘汰制，建立灵活完善的人员激励机制，形成一支年龄梯次、专业结构合理的优秀工程技术队伍；中心鼓励科技人员自我创业，将科技成果带入市场，建立创业企业，推广科技成果；中心将利用自身资源从资金、管理、工程化等方面予以优惠支持，以促进优秀人才走向市场的进程，为社会提供服务。

（二）研究内容

（1）提出一种全新的数据集。包括各类经过标注的高质量描述文本和在一段时空内连续，严谨，合理的符号图像。用以弥补各类模型训练过程中缺少的阳性样本，从而提升此类模型的符号生成质量。

（2）将设计一种新的基准测试，皆在评价各类模型对符号系统的生成质量。这将有助于学术界寻找到最佳的符号生成算法。

（3）将完成以下配套工具：数据集可视化工具，数据集格式转换工具，与 pytorch/tensorflow 集成的数据集加载器，tiny 型号的数据集用于测试。

（4）通过电子与信息学院的数字孪生平台可以实时收集物理实体的运行数据，然后在虚拟机中构建相应的数字模型。这些数字模型不仅可以反映物理实体的实时状态，还可以通过算法进行预测和优化。这种交互融合的方式，使得数字孪生技术能够实现对物理世界的精准把握和有效管理，使我们数据集得到的数据更准确。

（5）家电实训中心软件平台拥有大量先进技术的代码资源可支撑数据集的制作，同时平台支持 JAVA、C++和 python 等多种语言，可对前端、数据处理、AI 算法等多个方面进行二次开发。

（三）国内外研究现状和发展动态

从提示词中生成质量良好的符号是目前的研究重点，此类 AI 具有广泛的现实应用，例如在教育学科中辅助制图和帮助学生理解复杂的数学问题，在文化设计领域中提供直接可用的设计原稿。最近提出的 sora[1]文生视频模型更是将这一领域推向了连续生成的视频阶段。

然而，文生视频模型并不是一日而成的大厦。传统的图像生成方法基于 GANs，生成的图像只适用于简单且特定的几类。近年来，像 GLIDE[2],DALLE[3],Imagen[4]等基于 transformer 架构的模型使在多模态图像生成上表现出了令人印象深刻的效果，展示了此类架构的潜力。

此类模型的训练依赖图像数据集，其具有输入的提示词和输出的图像。目前，已有许多数据集被设计来评估 AI 系统的数学能力。如 ChartQA[5]是一种具有视觉和逻辑推理的图表问答基准数据集。然而，这些数据集关注的问题过于片面，往往局限在特定的任务或者数据格式中。最近的 ICLR2024 中新发布的 mathvista[6]是一个综合的数学推理基准数据集。在数学问题的覆盖率达到了前所未有的广度和深度。

但是，此类数据集仍然缺少时空上的关联性。在文生视频模型中，我们希望输入一段提示文字，并得到一段视频输出。这要求训练的数据集也拥有时空上的连续性，即视频或动画

数据。

现有的文本-视频数据集在规模或质量上受限，如 WebVid-10M[7]虽然拥有 10M 数量的切片，但是分辨率相对较低，而且数据带有水印。HD-VR-130M[8]实现了同时包含大分辨率和大数据量，但是其数据主要来自视频网站，对于符号类视频数据量则是不足的。

数据集	切片数量	分辨率	范围	标注	水印
MSR-VTT	10k	240p	Open	Caption	No
UCF101	13k	240p	Human	Class label	No
HowTo100M	136M	240p	Instructional	Subtitle	No
HDVILA	103M	720p	Open	Subtitle	No
WebVid-10M	10M	360p	Open	Caption	Yes
HDVG130M	130M	720p	Open	Caption	No

不同视频数据集的比较

- **目前文生视频发展：**

（1）国内：

百度其实在很久之前就推出了文生视频的能力，在百度的百家号中，当用户上传文章之后，会有一部分文章被百度精选出来，自动生成视频，而在最近也发布了一款名为“UniVG”的视频生成模型，相关效果也位于除 Sora 之外的前列。

同时，百度作为国内深耕 AI 行业最深的企业，无论是算力的充足、数据的丰富还是工程能力的先进，都处于国内第一梯队，只要其以正常的速度进行推进，那么百度版的能力更强的文生视频模型，也将于未来不久上线。

除百度外，科大讯飞作为专精 AI 赛道的公司，也是大语言模型竞争中的佼佼者，1 月底，星火认知大模型刚完成了 V3.5 的升级，并在华为的帮助之下，相关算力与工程能力得到了较快的提升。也有接近科大讯飞人士透露，科大讯飞目前内部已经开始文生视频进一步攻关研发。

在“传统”领先的大模型企业外，字节跳动或将借助存储数据的优势弯道超车。

字节跳动在短视频和社交媒体方面的海量数据资源，使会其在文生视频模型的研发上占据独特优势。MagicVideo-V2 的发布及其效果上的显著提升，已经证明了字节跳动在该领域的技术实力与创新能力。

随着火山引擎大模型服务平台“火山方舟”的推出，以及与多家合作伙伴共建的生态体系不断完善，字节跳动不仅能够利用自身的庞大用户基础产生的实时、多样的数据流进行训练优化，还有望通过高效的模型迭代和协同创新，在未来开发出能与 Sora 匹敌甚至超越的新一代文生视频模型。

但这样的优势也未曾不是一种包袱，作为数据层面最占优势的字节，又能否快速补上工程能力上的短板，摘下国内首个正式开放文生视频的桂冠，仍需要时间来证明。

2 月 19 日虹软科技官微宣，其核心大模型技术引擎——虹软 ArcMuse 再次升级。而此次升级将支持面向商拍的商业视频自动生成。

据介绍，与 Open AI Sora 类似，虹软 ArcMuse 大模型视频生成基于 diffusion-transformer 技术架构，具备丰富多样的创意力和想象力。通过图像，ArcMuse 大模型能够捕捉到商品的细节特征、质感、色彩等方面的精确信息，生成更能展示商品真实面貌的动态商拍视频。

而因赛集团则在与记者的交流中表示，其 AIGC 项目团队按照计划，将在三月进行文生视频功能的开发，等待时机成熟后投入公测。

在大模型的主流玩家行列里，字节跳动早在年初就发布了超高清文生视频模型MagicVideo-V2。据悉，该模型输出的视频在高清度、润滑度、连贯性、文本语义还原等方面，比目前主流的文生视频模型Gen-2、Stable Video Diffusion、Pika1.0等更出色。

就在前几日，阿里云旗下魔搭社区（Model-Scope）上线文本生成视频大模型。目前由文本特征提取、文本特征到视频隐空间扩散模型、视频隐空间到视频视觉空间这3个子网络组成，整体模型参数约17亿。

整体看下来，除去字节跳动的MagicVideo-V2有一定的水平之外，其他大多都处于一言难尽，甚至还无法看到效果的阶段，同Sora的距离还有很远很远。

（2）国外：

自生成式人工智能ChatGPT发布以来，全球范围内掀起了一股强劲的人工智能热潮。在国内市场，人工智能犹如春笋般快速涌现。然而，我国的人工智能大模型相较于ChatGPT仍存在显著差距。百度作为国内较早涉足AI领域的领军企业，其推出的文心一言在文本和代码生成方面尚未能媲美ChatGPT的用户体验，同时，在图片生成效果上也被Midjourney拉开差距。

如今Sora模型推出，国内暂未有与之跟进的大模型出来。一些人也因此认为，和OpenAI相比，我们的大模型能力差距没有缩小，反而在扩大。

赛道火热，产品欠佳。中国的人工智能发展进入至暗时刻。缺的不仅是有实力的大模型，还有与之直接相关的人才和AI“三算”，即算力、算法、算据。

在算据上，国产大模型也与GPT-4存在差距。在自然语言大模型中，参数是衡量一个深度学习模型复杂度和能力的重要指标。参数多，意味着模型能够处理更多的数据，学习更多的知识。国外公司在训练数据的质量和多样性方面往往占有优势，能够获取到更丰富、跨语言、跨领域的全球数据资源。而国内企业受限于数据隐私保护政策、地域性等因素，可能在数据质量和规模上面临挑战，尽管国内数据总量庞大，但在数据处理规范、标注质量等方面需要进一步提升。

国外有研究人员将GPT参数规模与大脑神经元做类比，GPT-3的规模与刺猬大脑类似，GPT-4拥有100万亿个参数，基本达到人类大脑的规模。

再看国产大模型，目前百度的文心一言，华为的盘古大模型参数量在千亿规模的级别，与GPT-3相近，而即使是排名靠前的阿里巴巴的M6大模型，其参数规模也仍与GPT-4相差一个数量级，更多的大模型仍在“原始阶段”。

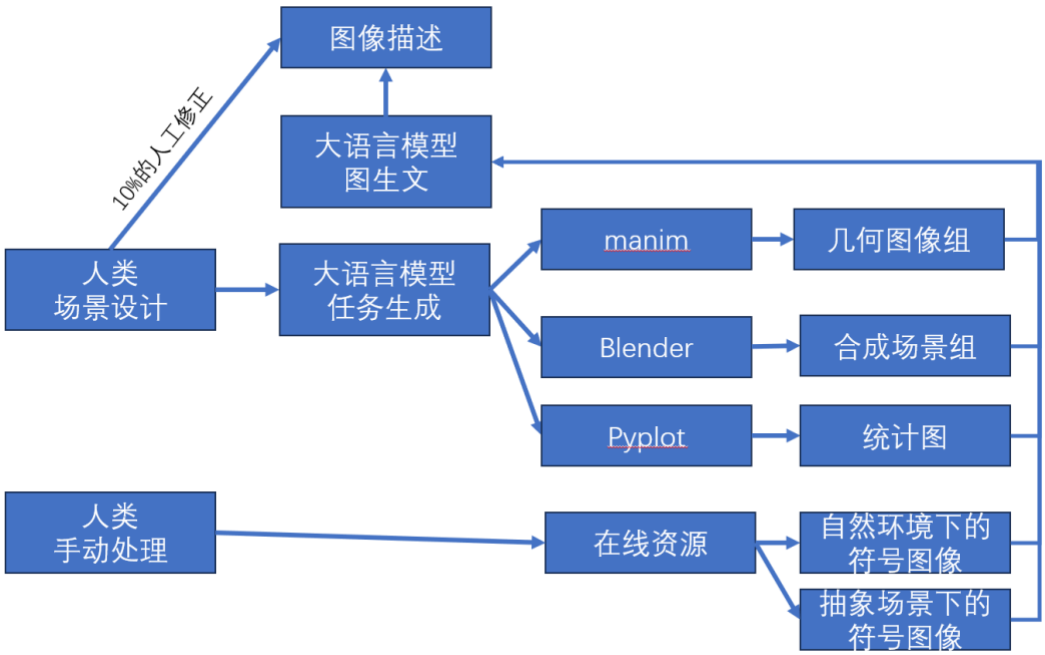
公司	大模型	参数
OpenAI	GPT-4	百万亿
	GPT-3	千亿
	GPT-2	百亿
阿里巴巴	M6	十万亿
百度	文心大模型	千亿
腾讯	混元大模型	千亿
华为	盘古大模型	千亿

在算法创新方面，虽然中国企业在算法领域有显著进步，尤其在模型架构优化、知识融合、多模态学习等方面取得了一系列重要成果，但在某些核心技术突破上，如自监督学习机制、模型并行和数据并行优化技术等方面，国外研究团队仍保持一定的领先优势

（四）技术路线

（1）制作数据集
（2）设计基准测试
（3）各类模型基准测试报告
（4）消融实验

- **制作数据集计划**



图表 1—数据集制作流程

我们的数据集将由自动系统生成，主要流程如下：

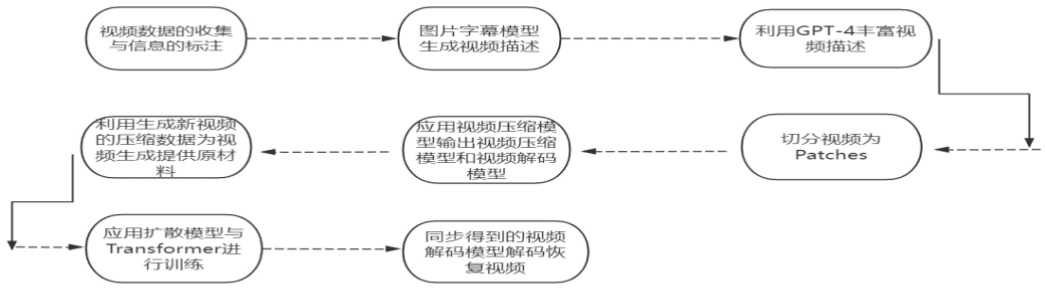
①由人类完成场景设计，LLM 负责将抽象的文字描述转换成可执行代码。

②调用下游生成引擎，根据不同的任务类型生成不同风格的符号图像

③将所有生成的图像组进行反向图生文，获得标注

④数据微调。我们将对其中 10%的数据重新进行人工标注，以提升数据集的准确性。

我们制作数据集的流程，灵感来自 sora 官方发布的技术报告[9]中提到的训练方法，见下表：



Sora的训练流程图

- **基准测试设计计划**

评价一个模型的生成质量通常由三部分组成：帧质量，时间相关性和文本匹配性

我们的初步设想是一种匹配函数

$$Match = b1 * Mf + b2 * Mc + b3 * Mt$$

其中，Mf 代表了每帧的生成质量，Mc 代表了生成帧之间的在时间上的相关性，Mt 代表了生成帧序列与输入提示词的匹配性。$b1, b2, b3$是相关系数，表示各个子项对于整体质量的占比，属于超参数的一种，需要我们依据后期的研究进行调整。

对于 Mf 的设想，在符号生成的目标图像中，例如文字，其边界一般是锐利且分明的，所以我们计划使用拉普拉斯函数 Laplacian，计算图像的模糊程度。

对于 Mc 的设想，基于 FVD 算法，在连续的视频帧之间，通常其像素的变化是有规律可寻的，上下文的变化不会太大。所以我们通过计算帧间差异获取这种变化。同时，我们会尝试跟踪一些特征明显的像素，计算其在一段时间内的向量变化，以探索这种变化的逻辑性。

$$|u_r - u_g|^2 + Tr(\sum R + \sum G - 2(\sum R \sum G)^{\frac{1}{2}}$$

对于 Mt 的设想，基于 CLIPSIM，计算每个 Clip 输出文本和视频的相似度。

- **各类模型基准测试计划**

目前，市面上提供了训练和测试代码的文生视频模型有以下几类：

模型	预训练权重	分辨率	CLIPSIM	FVD
Open-Sora	Yes	16*512*512	N/A	N/A
VedioGPT	Yes	16*128*128		2880.6
MoCoGAN	No			2886.8
DIGAN	Yes			1630.2
StyleGAN-V	No	16*256*256		1431.0
PVDM	Yes			343.6
CogVedio	Yes	32*?*?	0.2631	
LVDM	Yes		0.2381	
ModelScope			0.2795	

注：FVD/CLIPSIM 结果来自 https://arxiv.org/abs/2305.10874

我们计划对其发布的预训练模型进行额外训练，使用我们的数据集进行微调。然后通过评价指标估计其符号生成能力。

- **消融实验设想**

我们将使用各类模型提供的预训练模型进行测试，然后与我们微调之后的模型进行对比，最终得到差异值。

- **符号系统的理论研究**

众多模型难以生成优质的符号的其中一个原因是映射的不明确。符号是发散的数据信息，而自然语言本身就有表达的局限性和不准确性。在反向图生文时，我们往往只能选取可能性最高的文本，而忽视了其他存在的可能性。

（五）创新点与项目特色

（1）提出了一种新的数据集

（2）研究了 Ai 系统中符号生成问题的原理

（3）提出一种通过改进数据集来改进模型生成能力的办法

（六）拟解决的关键问题

整个项目的难点在于：

（1）构建自动生成数据的工作流

（2）验证基准测试方法的可行性

（3）验证制作的数据集的有效性

（4）探索符号生成系统的底层原理

（七）预期成果（专利、学术论文等）

我们将发布数据集技术报告，可同步发布至 arxiv 等预印本收录网站。

（八）项目进度安排

2024 年 3-4 月：1.确定项目研究方向和目标，进行文献调研和资料收集。2.制定详细的项目计划书，包括项目背景、研究内容、方法、预期成果等。

2024 年 5-6 月：1.根据项目计划书，开展具体的研究工作，包括实验设计、数据采集、分析处理等。2.定期召开项目进展会议，讨论研究过程中遇到的问题和解决方案，确保项目按计划进行，撰写初期报告。

2024 年 7-9 月：撰写中期报告，总结前半段研究工作的成果和不足，提出改进措施。

2024 年 10-12 月：进行成果定基，举行成果座谈，并进行阶段性成果交流。

2025 年 1 月-结题：完成论文的研究报告，并撰写定稿论。

（九）已有基础

1.与本项目有关的研究积累和已取得的成绩

（1）团队基础：

在学业方面：本团队的成员学业成绩都不错，并且对编程语言有一定了解，如 Python、Java、C 语言等，并且团队成员都对编程语言甚至人工智能十分感兴趣。

在竞赛方面：有成员在高中时期便参与过项目，也有成员正在参加广东省大学生计算机设计大赛。

（2）与本项目有关的研究积累与成绩：

①OpenAI 公司推出了新一代的大型语言模型——ChatGLM，并对其进行了更新。这次更新的亮点是推出了一个评测长文本理解能力的数据集——LongBench，以及一个支持 32k 上下文的 ChatGLM2-6B-32K 模型。LongBench 是一个专门为长文本理解能力而设计的数据集，它包含了各种类型的长文档，如新闻文章、小说、百科全书等。这些文档都被精心挑选并按照一定的标准进行了标注和处理，以确保模型在进行长文本理解时能够获得最佳的训练效果。与传统的长文本数据集相比，LongBench 具有更高的质量和更广泛的覆盖范围。它不仅包含了不同领域、不同语言的长文本数据，还包括了人类对长文本理解的评估标注。这些评估标注可以帮助模型更好地理解人类的意图和情感，从而在处理长文本时更加准确和自然。

②DeepMind 近日发布了一个新型数据集，包含大量不同类型的数学问题（练习题级别），旨在考察模型的数学学习和代数推理能力。

数据集地址：https://github.com/deepmind/mathematics_dataset

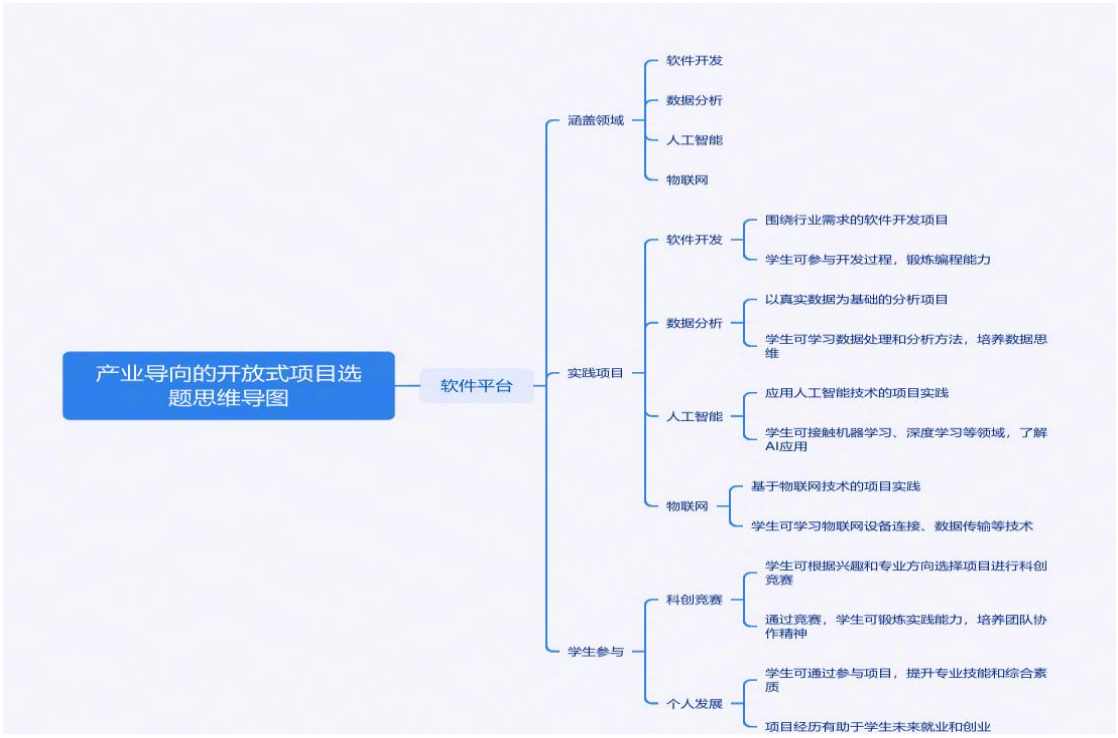
目前该数据集发布了 1.0 版，其每个模块包含 200 万（问题答案）对和 10000 个预生成测试样本，问题的长度限制为 160 字符，答案的长度限制为 30 字符。每个问题类型中的训练数据被分为「容易训练」、「中等训练难度」和「较难训练」三个级别。这允许通过课程来训练模型。

③微软研究团队引领着教育技术领域的不断创新，近日推出了一款名为 Orca-Math 的前沿工具，它是一款小语言模型（SLM），拥有 7 亿参数，并基于 Mistral-7B 架构微调而来。这一创新方法重新定义了传统数学单词问题教学的策略，彻底改变了学生参与和掌握这一学科的方式。Orca-Math 的方法论的核心是一个由 20 万道数学问题组成的精心制作的合成数据集。然而，Orca-Math 的真正巧妙之处在于其迭代学习过程。在模型遍历这个数据集时，它尝试解决问题并获得对其努力的详细反馈。这个反馈循环丰富了偏好对比，将模型的解决方案与专家反馈进行对比，促进了一个学习环境，使模型不断完善其解决问题的能力。

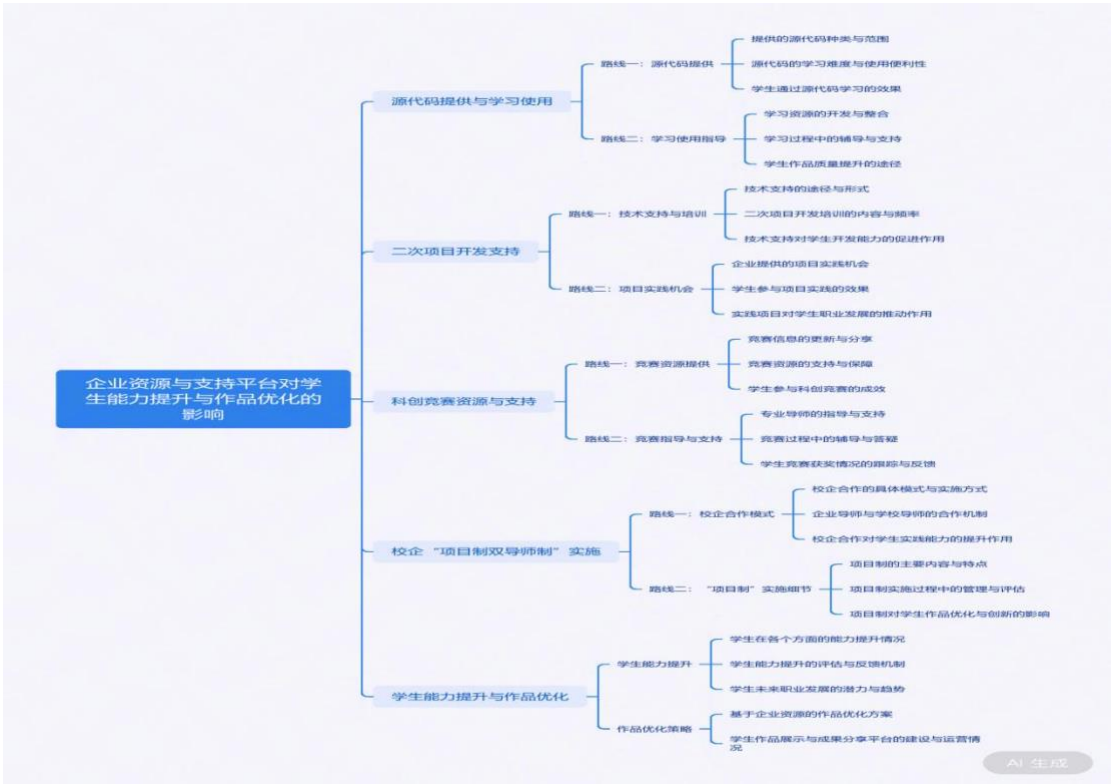
（3）电子与信息学院家电实训中心软件平台基础：

电子与信息学院家电实训中心软件平台为学生提供了丰富的学习和学科竞赛资源。学生依托平台发布的一系列开源代码和硬件接口，不仅能够深入了解软件开发技术的原理和应用，更能够降低参加学科竞赛的难度，增强学生信心，从而在实践中提升专业能力和创新能力。

①产业导向的开放式项目选题：软件平台提供了一系列与行业紧密结合的实践项目，涵盖了软件开发、数据分析、人工智能以及物联网等多种领域和技术方向，学生可以根据自己的兴趣和专业方向选择适合的项目进行科创竞赛。



②企业资源与支持：平台整合并提供了实训中心大部分模块的源代码供学生学习使用和二次项目开发，为参与科创竞赛的学生提供竞赛资源和支持，包括竞赛信息、培训资料、导师指导等。此外，面向学生实施校企“项目制双导师制”，学生可以获得来自企业工程师的实践指导和技术支持。企业导师根据学生需求，不定期进行线上指导，帮助学生解决在项目开发过程中遇到的技术难题，提高参赛作品质量。



③实践经验积累：平台拥有大量先进技术的代码资源可支撑包括软件开发在内的多种创新项目和技术竞赛。以 AI 科创平台为例，该平台支持 JAVA、C++和 python 等多种语言，可

对前端、数据处理、AI 算法等多个方面进行二次开发。通过在平台上进行项目开发和实践操作，学生可以积累丰富的实践经验，提升自己的技术能力和创新能力。这些经验对于参加科创竞赛非常有帮助，可以让学生更加游刃有余地应对各种挑战。

2. 已具备的条件，尚缺少的条件及解决方法

- **研究支持条件**

（1）硬件条件：一切需要的硬件条件，如 4090 等 AI 训练常用芯片

（2）软件条件：Python、人工智能等

（3）研究环境：本小组由电子与信息学院欧阳明俊老师指导，电子与信息学院拥有多个人工智能相关的实验室，以及大量人工智能相关设备，在这方面积累的海量研究资料与数据。

- **本项目尚缺少的条件：**

（1）数据量过大，人工训练费时费力

（2）低质量的分类

（3）不平衡的分类

（4）数据不平衡

- **对应的解决办法：**

（1）利用脚本自动训练

（2）选择正确的粒度级别进行分类

（3）收集代表性不足的分类的更多样本

（4）对数据进行过度/不足的采样

（5）裁剪或拉伸数据，使其具有与其他样本相同的方面或格式

（6）规范化数据，使每个样本的数据都在相同的值范围内

（十）参考文献

[1]Sora, https://github.com/hpcaitech/Open-Sora

[2]GLIDE, Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models, arXiv:2112.10741v1[cs.CV]

[3]DALLE, ImprovingImage Generation with Better Caption, https://openai.com/dall-e-3

[4]Imagen

[5] ChartQA Lu et al., 2021a;Dahlgren Lindstr¨ om & Abraham, 2022;Masry et al., 2022

[6] MathVista: Evaluating Mathematical Reasoning of Foundation Models in Visual Contexts arXiv:2310.02255 [cs.CV]

[7] WebVid-10M

[8] HD-VR-130M

[9]《OpenAI 最新文生视频模型 Sora 技术能力解密：基于 Patch 的数据规范性、多模态 Prompt 支持、物体持久性和远程相干性能力》，Garvin Li

[10]《Sora 模型发布，哪些行业要变天？》，IT 魔术师

[11]《国内复现 Sora 能力几何？李维：不存在跨不过的技术门槛》，中证金牛座

[12]《Sora 技术文档》，OpenAI

[13]人工智能文生视频大模型 Sora 的核心技术、运行机理及未来场景。 夏德元，文化艺术研究，2024

[14]Sora 出世 人工智能将引领新一轮行业变革。 罗茂林，2024

三、经费预算

开支科目	预算经费（元）	主要用途	阶段下达经费计划（元）	
			前半阶段	后半阶段
预算经费总额	8000	科研	4350	3650
1.业务费	4900	项目相关业务处理	2400	2500
（1）计算、分析、测试费	1000	外网租用，sora等模型租用，实际场景的测试等	400	600
（2）能源动力费	1000	能源	500	500
（3）会议、差旅费	900	人员外派深入专业技术学习	500	400
（4）文献检索费	0	无	0	0
（5）论文出版费	2000	预计发表1-2篇学术论文	1000	1000
2.仪器设备购置费	1000	购置设备	700	300
3.实验装置试制费	700	实验装备试制	400	300
4.材料费	700	购置材料	500	200
5 其他.	700	供于学习与研究	350	350
学校批准经费				

四、指导教师意见

该项目具备较好的前沿性，对于解决文生图模型符号质量问题有较大的帮助。同时，学生在参与该项目过程中，可以进一步的提高个人能力，因此，我同意推荐该项目申报。

指导教师（签名）： 

2024 年 3 月 28 日

五、院系大学生创新创业训练计划专家组意见

经过学院专家组审阅和讨论，该项目选题合理，内容丰富，有助于增强学生创新和实践能力。请指导老师对该项目的实施认真指导和把关，同意申报。

专家组组长（签章）：

2024 年 4 月 3 日

六、学校大学生创新创业训练计划专家组意见

负责人（签章）：

年　月　日

七、学校大学生创新创业训练计划工作小组审批意见

负责人（签章）：

年　月　日