



2024 年 CAAI-联想蓝天科研基金申报主题

方向一、大模型及智能体

1. 技术应用类：多语言和跨语言的文本向量模型优化
2. 技术应用类：AI Agent 中复杂任务的可控性分解
3. 技术应用类：基于端侧 LLM 的 Agent 技术研究及优化
4. 技术应用类：多模态 RAG 系统构建与实现
5. 技术应用类：文生图模型对应 Control 模型的高效训练
6. 技术应用类：高效 AGI 推理服务平台设计方案与验证

方向二、空间智能及具身智能

1. 科学探索类：足式机器人本体能力认知和安全强化学习
2. 技术应用类：基于 3D 分割与补全技术的三维重建中目标单体化技术
3. 技术应用类：基于固定 3D 场景与视觉融合的高质量远程视频技术
4. 技术应用类：鲁棒强化学习在足式机器人运动控制中应用
5. 技术应用类：足式机器人关节容错控制的强化学习算法
6. 技术应用类：机器人云台引导中红外图像匹配与实时配准技术
7. 技术应用类：面向知识增强任务的人机回环交互系统算法研究
8. 技术应用类：基于多模态的 3D 数字人生成和驱动
9. 技术应用类：跨应用渲染引擎技术中单视点到双视点转换方案研究与空间感知增强
10. 技术应用类：裸眼 3D 场景下 2D 视频转 3D 技术研究

方向一、大模型及智能体

课题 1：多语言和跨语言的文本向量模型优化

1、研究背景：

大语言模型在端侧设备的应用越来越广泛，联想已经全面推出了带有端侧大模型推理功能的 AIPC 产品。RAG 系统是增强大语言模型输出的重要技术，本课题主要目标是在端侧的有限资源条件下，实现端侧文本向量模型的在多语言业务场景中的高召回率以及高准确率。

2、研究内容：

在多语言文本向量模型的基础上，持续提升其在业务场景下的多语言和跨语言的检索能力。业务场景为用户提问- 检索-返回数据库相关文本切片，用户提问类型包括事实、推理、总结等，文档没有领域限制。

3、研究目标：

提升文本向量模型在检索任务上的性能，语言集合包括（中、日、英、德、葡、意、法、西）

在检索能力上比原始 SOTA 模型在各语言的召回指标提升 10%以上

在跨语言检索能力上召回率（前三）达到 80%

4、产出及交付物：

- ① 代码：算法代码与文档；
- ② 数据集：双方合作共建多语言数据集；
- ③ 研究报告：输出研究报告，完成报告分享；
- ④ 论文专利：发表 CCF-A 类会议/期刊论文 1 篇，申请专利 2 篇；

课题 2：AI Agent 中复杂任务的可控性分解

1、研究背景：

在 3C 领域中，用户任务专业性高、复杂度强，使得 AI Agent 在执行这类任务时，面临着精度低、可控性弱等不足，导致最终的任务执行输出结果不能很好地满足用户的实际需求，而如何提升 AI Agent 面向复杂任务进行分解的可控性是学术界和工业界的共同难题。基于此，提出 AI Agent 中复杂任务可控性分解这一即具有实际价值又具有研究挑战的技术课题。

2、研究内容：

在人工智能领域，AI Agent 对处理复杂任务的需求日益增长。本研究致力于构建一个端到端的框架，实现将复杂任务以合理且有序的方式分解为高质量、可独立解决的子任务，从而在 AI Agent 中完成对复杂任务的可控性分解。

3、研究目标：

本研究将涵盖以下几个关键方面

研究含约束的伪分解样本生成算法及预训练重组模型；

研究子任务分解顺序优化机制；

设计智能的样本过滤机制；

研究上下文增强的训练与推理方法。

4、产出及交付物：

- ① 代码：算法代码；
- ② 开发完整的可控任务分解框架，构建对复杂任务的可控、高质量分解能力；算法能应用到实际场景中；
- ③ 提升可控任务分解的质量，主观上任务分解效果相比现存的最佳方案提升至少 5%；
- ④ 发表人工智能领域顶级期刊会议 1-2 篇，申请发明专利 2 项。

课题 3：基于端侧 LLM 的 Agent 技术研究及优化

1、研究背景：

随着人工智能技术的发展，大语言模型在自然语言处理领域取得了显著的进展，为各种语言任务提供了强大的基础。联想已经全面推出了带有端侧大模型推理功能的 AIPC 产品，使得端侧设备能够利用大语言模型提供更智能的服务。对话系统是一种典型的端侧大语言模型的应用场景，它可以让用户通过自然语言与设备进行交互，实现各种功能和需求。对话系统的核心组件之一是对话管理，它负责根据用户的输入和上下文信息，生成合适的系统回应，维持和引导对话的流畅和有效。对话管理的性能直接影响了对话系统的用户体验和满意度。

2、研究内容：

本研究的主要内容是在端侧业务场景下，通过优化 Agent 技术，实现对话管理的 Agent 化并将该技术在端侧落地。具体来说，我们将基于端侧的对话 APP 软件系统，探索如何利用端侧的对话记录、对话历史信息、端侧用户文档等信息，将用户的对话信息进行有效的语义理解与管理，实现更优的对话性能。

3、研究目标：

本研究的主要目标是基于端侧的对话 APP 软件系统，实现高效的对话管理。具体的目标如下：

- ① 探索对话管理 Agent 在端侧系统的高效实现，包括端侧大语言模型的调用、端侧数据的利用、端侧计算资源的分配等方面。
- ② 通过技术优化，实现多轮对话准确率提升 10%，即在给定的对话场景和目标下，系统能够正确地识别用户的意图、提供相关的信息、完成任务、处理异常等。

4、产出及交付物：

本研究的主要产出和交付物如下：

- ① 代码：算法代码与文档，包括对话管理 Agent 的设计、实现、测试和评估等。
- ② 实现端侧对话系统用户满意度提升的评价基准，包括对话系统的性能指标、用户满意度问卷、用户反馈等。
- ③ 实现端侧对话系统对用户对话管理的技术方案，包括对话管理 Agent 的架构、流程、算法、优化等。
- ④ 论文专利：发表 CCF-A 类会议/期刊论文 1 篇，申请专利 2 篇，

课题 4：多模态 RAG 系统构建与实现

1、研究背景：

多模态大模型发展迅速，特别是随着 gpt-4o 等模型的推出，如何在端侧设备上实现更好的多模态用户体验也成为了产品的热点需求。该课目标是通过构建和优化多模态 RAG 系统流程，从而全面提升用户的多模态交互体验。

2、研究内容：

针对多模态大模型的特性，面向 AIPC 中用户个人知识库中的多模态信息问答场

景，构建基于多模态大模型的 RAG 系统以及相关的用户交互，提升基于多模态大模型的用户体验以及系统文档的召回率和准确率。

3、研究目标：

提升多模态大模型在端侧部署使用时的用户体验：

系统融合的模态包括但是不限于：语音、文本、图像、视频

对比现有 SOTA 的 RAG 系统，多模态 RAG 系统召回率指标提升 10%

4、产出及交付物：

① 代码：算法代码与软件工具；

② 研究报告：输出研究报告，完成报告分享；

③ 论文专利：发表 CCF-A 类会议/期刊论文 1 篇，申请专利 2 篇；

课题 5：文生图模型对应 Control 模型的高效训练

1. 研究背景：

文生图模型迭代更新迅速，从 SD1.5 到 SDXL 再到 SD3，模型架构不停在优化演变。Control 模型在文生图产品中可以有效提升用户交互体验，优化用户出图效果。但是不同的 Control 模型要适配不同的文生图基模，这也就导致切换基模后该基模的其他辅助生图模型都要进行迭代更新，需要花费很大的精力进行适配。

2、研究内容：

探讨高效的 Control 模型训练方案，在数据利用和训练效率上都能得到有效提升，从而快速实现适配新架构基模的 Control 模型的训练及部署

3、研究目标：

加速端侧文生图 Control 模型的训练，Control 模型包括但不限于：

Inpainting, Tile, scribble, canny, pose, depth, ID 等

设计高效的模型训练方案，比现有 SOTA 流程在训练数据利用率和性能上提升 10%

4. 产出及交付物：

-
- ① 代码：算法代码与文档；
 - ② 数据集：双方合作共建相关数据集；
 - ③ 论文专利：发表 CCF-A 类会议/期刊论文 1 篇，申请专利 2 篇；

课题 6：高效 AGI 推理服务平台设计方案与验证

1、研究背景：

AGI 技术正被快速应用到各个领域，有效提高了社会生产力。AGI 应用依赖大模型的高效推理，而推理输出的精度与参数大小成正比。仅在端侧运行的模型受到算力和续航能力的限制，不能采用大参数量的模型。因此通过云端模型与端侧一起为用户提供推理服务成为必然趋势。云端模型服务可以通过数据中心支持大模型，在通过技术手段确保用户数据安全的前提下，具有模型精度高，用户免升级，可靠性高和跨终端体验一致等优势。

但现有推理服务平台为了保证模型运行效率，无法充分利用云原生服务框架提供的便利，需要人工管理所有模型副本，占用的显卡及其它资源。这些限制严重影响了平台算力利用率，用户体验和平台的规模扩展。因此，现阶段，迫切需要能够支持云原生管理能力的高效 AGI 推理服务平台设计和实现方案。该课题研究成果预计未来为联想 AI Now 提供云端推理服务。

2、研究内容：

通过模型结构、优化器、算子融合、并行和系统的全栈协同优化实现大模型推理服务平台设计与实现；大模型推理服务接口设计；实现通过监控组件对集群模块及事件进行实时监测，及时发现问题并快速恢复，确保稳定性的方案；大模型推理平台网络拓扑设计。

3、研究目标：

- ① 通过研究工作提出并验证大模型推理平台设计方案，创建大模型服务系统评估指标，该系列指标能综合评估推理平台的 QPS，成本，最大并行数，算力利用率，服务可靠性等
- ② 验证平台搭建：在业界开源框架基础上实现模型管理自动化、并行计算和系统的全栈协同优化；设计并开发大模型推理服务接口；实现通过监控组件进行微服务及事件的实时监测和快速恢复

- ③ 大模型推理平台网络拓扑设计

4、产出及交付物：

- ① 大模型推理服务验证平台 1 套；基于该平台展示 2 项；
- ② 平台性能系统性评估指标定义，理论值与实际值分析差异报告
- ③ 申请 2 项相关专利，发表人工智能领域 CCF-A 类会议、期刊论文 1 篇

方向二、空间智能及具身智能

课题 1：足式机器人本体能力认知和安全强化学习

1、研究背景：

现状:近年来，强化学习算法因其适应性和泛化能力，使足式机器人的运动能力取得了长足的进步，可以应对各种复杂的结构化和非结构化地形，极大拓展了足式机器人的应用范围。

问题:然而，当前的强化学习算法并未考虑到足式机器人本身的能力边界和安全性，因此在执行复杂和危险任务时，机器人无法对可能的危险情况进行预判，从而发生严重事故，造成自身的伤害和损坏。因此，开展足式机器人本体能力认知和安全强化学习算法研究和探索，使机器人能够根据传感器信号实时评估环境和自身状态，了解自己的能力边界和可能的安全问题，从而做出安全和有效的决策，是非常有必要的。

可能应用:未来可应用到联想晨星六足机器人产品上，提高机器人产品的可靠性和寿命。

2、研究内容：

- 1) 研究安全约束建模与表示，包括：本体约束：能够反映机器人的动力学特性，满足物理规律的强约束。安全集约束：能够反映机器人在未来时间窗口内的安全性需求，采用概率形式描述。行为引导约束：用于引导机器人的行为模式，以形成安全且有效的运动模式。通过建立这些约束关系，在强化学习训练过程中帮助机器人学习和遵循安全的行为规范。
- 2) 构建安全强化学习框架，通过考虑安全约束，利用奖励和惩罚机制来引导机器人的行为。研究约束集成与优化手段：将各类安全约束集成到强化学习框架中，通过优化算法（如约束优化或者安全性增强方法）来确保机器人在学习过

程中始终遵守安全规定。

3、研究目标:

- (1) 构建安全强化学习框架：研发利用奖励和惩罚机制来引导机器人行为的安全强化学习算法，实现机器人在未知环境中对危险的预判和紧急处理机制；
- (2) 机器人具有安全保护，能够自动恢复安全状态，包括任意姿态恢复，动态运动中恢复，恢复时间小于 2s；

4、产出及交付物:

- ① 代码：算法示例及说明；
- ② 研究报告：输出研究报告，完成报告分享；

课题 2：基于 3D 分割与补全技术的三维重建中目标单体化技术

1、研究背景

三维展厅或 XR 展览展示应用中，三维数字内容包括三维场景和三维目标单体至关重要，随着 AI 技术的不断进步，特别是新的基于 NeRF, 3DGS (3D Gaussian Splatting) 等方法的三维重建技术的发展，通过重建构建高拟真的三维模型变成可能，这里面，三维单体目标的快速构建对数字资产的积累具有重大意义。针对三维重建中单体目标的构建，一般的方法是通过特定干净背景下扫描建模来实现，并且还存在未扫描到的底部无法建模的问题，这限制了应用条件，因此希望通过自然场景中目标物进行扫描重建，利用视觉大模型等分割技术和 3D 补全 (inpainting) 来实现单体目标的分割与完整建模。

2、研究内容:

基于 3D 分割与补全技术的三维重建中目标单体化技术，包括：

- ① 研究基于视觉大模型如 SAM 等的目标三维信息提取技术，实现场景中单体目标高精度重建技术。
- ② 研究三维模型的补全方法，特别是 3DGS 表达的三维补全方法，实现目标未扫描区域的三维信息自动补全。

3、研究目标:

研究一套基于 3D 分割与补全技术的三维重建中目标单体化技术，实现高精度的三维单体分割与补全效果，典型场景下目标分割结果 mIoU 优于 93%，目标补全

结果 FID 优于 105。实现一套基于三维场景中目标单体模型的自动提取系统原型 demo，可以通过点击等交互方法实现完整单体的提取。

4、产出及交付物：

- ① 代码：算法代码与文档；
- ② 研究报告：输出研究报告，完成报告分享；
- ③ 论文专利：发表相关领域高水平期刊或会议论文 1 篇，专利 2 个

课题 3：基于固定 3D 场景与实时视频融合的高质量远程视频技术

1、研究背景：

通过 XR 远程技术可以实现高效的沟通，具有很高的应用价值。由于 XR 设备本身具有的高速运动的特点，同时还受到应用场景光照、采集设备的传感器质量、传输带宽的影响，XR 远程视频面临一些列挑战。XR 远程视频中很多情况是在某个特定场景/区域进行视频通话，例如在线会议、AR 远程设备维修、远程医疗，针对上述场景和现有的远程视频技术问题，研究如何有效利用这些场景/区域的 3D 时空先验信息，例如基于大模型训练获得固定场景下的三维信息或图像数据分布信息，再基于少量实时动态信息结合已训练的场景模型来恢复高质量的图像画面，来提升实时视频中的图像质量或减少带宽要求是一个非常具有研究价值和应用价值的课题，有望极大提升固定场景下的 XR 远程视频通话的图像质量。

2、研究内容：

研究固定场景中结合三维场景信息与实时视频流信息的高质量远程视频技术，包括：

- ① 研究针对比较固定的远程视频通话场景/区域，研究大模型训练 3D 时空信息提升远程视频技术的现有相关技术和技术路线。
- ② 研究和优化在低带宽、低光照等不同条件下结合已有固定三维场景先验信息提升远程视频质量的方法。

3、研究目标：

研究结合固定场景先验信息的远程视频相关方法，实现固定场景下远程视频画质的提升，实现相同画质条件下带宽要求减少 3 倍；实现一套结合固定场景先

验信息的远程视频 demo 原型，能够反映在低带宽或低光照等条件下的性能优势。

4、产出及交付物：

- ① 代码：算法代码与文档；
- ② 研究报告：输出研究报告，完成报告分享；
- ③ 论文专利：发表相关领域高水平期刊或会议论文 1 篇，专利 2 个

课题 4：鲁棒强化学习在足式机器人运动控制中应用

研究背景

现状：近年来，强化学习算法因其适应性和泛化能力，使足式机器人的运动能力取得了长足的进步，可以应对各种复杂的结构化和非结构化地形，极大拓展了足式机器人的应用范围。

问题：足式机器人在现实世界中面临着多样化、不确定性和动态变化的环境挑战，传感器受到光照、温度、故障等多方面的影响，其感知的数据具有不可靠性。然而，传统的强化学习对环境变化和噪声极其敏感，导致学习的不稳定性和低效性，且难以迁移。因此，鲁棒强化学习的研究显得尤为重要，其旨在提高机器人对环境变化的适应能力和学习效率，以实现更安全、可靠的运动控制和智能决策。

可能应用：未来可应用到联想晨星六足机器人产品上，提高机器人产品的可靠性和寿命。

研究内容：

- 1) 研究环境模型和不确定性建模，开发适应于足式机器人的环境模型，设计并优化适应于不确定环境的鲁棒强化学习算法，以提高机器人在实时决策和行动执行中的鲁棒性和稳定性。结合深度强化学习和模型预测控制等技术，实现机器人在动态和不确定环境中的优化决策与执行。最后在实际机器人上部署和验证。
- 2) 开发多模态传感器数据融合技术，利用融合后的数据为鲁棒强化学习算法提供更多样化的输入，增强机器人对环境变化的响应能力和决策准确性。最后在实际机器人上部署和验证。

研究目标:

- (1) 设计适应于不确定传感器输入的鲁棒强化学习算法，结合深度强化学习和模型预测控制等技术，实现机器人在动态和不确定环境中的优化决策与执行；
- (2) 对测量噪声的鲁棒性，对比普通强化学习算法，运动性能提高 20%，能够抵抗自身重量 20%以上的外力扰动；

产出及交付物:

- ① 代码：算法示例及说明；
- ② 研究报告：输出研究报告，完成报告分享；
- ③ 论文专利：CCF-A 类会议或 IEEE/ACM Transactions 发表 1-2 篇论文，专利 2 个；

课题 5：足式机器人关节容错控制的强化学习算法**1、研究背景**

现状：足式机器人由于其腿足式结构，可以适应各种复杂地形环境，特别适合代替人类到各种恶劣和危险的环境中执行任务，如火灾、核电站、行星表面等环境。然而，相比于轮式机器人，足式机器人的机械和电气复杂度也大大增加。

问题：在恶劣环境中，足式机器人的机械和电气元件难以始终保持最佳状态。特别是足式机器人的驱动关节会出现各种磨损和故障，如关节失效、关节锁死、机械零位不准、间隙增大等。而传统的机器人强化学习算法在这种情况下，将直接失去行走能力。为了确保在这些关键场景中的任务成功率，需要研究一种能够在关节故障或机械零位不准情况下使机器人自主适应关节故障情形，并恢复行走能力的强化学习算法。

可能应用：

未来可应用到联想晨星六足机器人产品上，提高机器人产品的可靠性和寿命。

2、研究内容：

- (1) 针对关节失效故障情况，建立关节失效模型和机器人变胞动力学模型，提出关节失效判别算法，使机器人可以自主感知正常情况和不同的关节失效情况，从而自主调整行走策略；建立包含关节失效模型和机器人变胞动力学模型

的仿真模拟环境，在该模型环境上设计针对关节失效故障的强化学习控制算法。最后在实际机器人上部署和验证。

(2) 针对关节间隙增大和零位不准情况，建立关节间隙和零位偏移模型，在此基础上设计基于在线关节间隙和零位辨识的强化学习控制算法，使机器人能够在关节间隙增大情况和零位不准下自动调整步态。最后在实际机器人上部署和验证。

3、研究目标：

(1) 机器人能够在关节故障（锁死、失效）情况下自动调整步态，恢复行走能力。确保机器人在 80% 的单关节失效和 50% 的双关节失效情况下依然能保持稳定行走；

(2) 机器人能够在关节间隙增大情况和零位不准下自动调整步态。确保机器人在零位偏移 15° 和间隙 1.5° 的情况下，仍具有平地、坡面、台阶等地形正常行走能力，行走速度不低于正常情况下行走速度的 90%。

4、产出及交付物：

- ① 代码：算法示例及说明；
- ② 研究报告：输出研究报告，完成报告分享；
- ③ 论文专利：CCF-A 类会议或 IEEE/ACM Transactions 发表 1-2 篇论文，专利 2 个；

课题 6：机器人云台引导中红外图像匹配与实时配准技术研究

1、研究背景：

红外图像在机器人巡检过程中可以用于感知温度信息，暗光下信号等重要信息，在机器人巡检过程中，由于虚拟点拍摄预置位的不准确导致拍摄会存在偏差，因此需要在拍摄点位进行云台引导来对准拍摄和检测的目标，但由于红外图像缺乏可见光图像丰富的纹理，因此用于云台引导的图像匹配算法存在失败的情况。另外还可以借助可见光图像进行云台引导，但由于可见光相机对焦变化等原因，可见光图像与红外图像之间实际应用中需要实时配准，以便通过可见光图像进行目标检测与匹配后在红外图像中能找到对应像素或区域。

2、研究内容：

研究红外图像和可见光图像的配准技术，充分利用两者图像信息则需要进行精确配准，包括：

- ① 研究红外图像与红外图像的图像配准技术，包括全局图像的配准和图像中目标的匹配方法。
- ② 研究红外图像与可见光图像的跨模态配准技术，调研与分析跨模态图像配准主要方法和技术路径。
- ③ 研究匹配模型轻量化或加速方法，优化实时性能以应用到实际场景。

3、研究目标：

研究红外图像和可见光图像的配准技术，包括红外与红外图区域匹配、红外与可见光配准，实现联想机器人红外图像巡检中精确云台引导能力，与现有 baseline 相比，提升 15% 的指标。探索跨模态图像配准技术，为多模态目标匹配与配准研究提供技术参考。

4、产出及交付物：

- ① 代码：算法代码与文档；
- ② 研究报告：输出研究报告，完成报告分享；
- ③ 论文专利：CCF-A 类会议或 IEEE/ACM Transactions 发表 1 篇论文，专利 2 个；

课题 7：面向知识增强任务的人机回环交互系统算法研究

1、研究背景：

知识增强指在当前知识加工、处理的过程中，给用户提供额外知识，或将现有知识与用户已有内容建立连接。知识增强在办公、学习等场景具有广泛的应用，通过大模型结合检索增强生成（RAG）等技术，可以实现一定程度的知识增强。但是现有大模型知识增强方法中，鲜有将人类反馈纳入其中，因此很难做到系统随着用户不断使用，不断学习用户习惯，进而不断优化提升知识增强效果和使用体验。结合人机回环，可以解决知识增强中的诸多问题，例如（1）用户意图理解不准确导致后续知识增强方向不对；（2）检索增强生成各环节中不符合用户预期导致结果偏离实际需求；（3）知识增强的展示时机和形式不符合使用场景，导致知识未能被有效接收和使用，甚至影响主线任务等。通过引入

人机回环方法，可在上述三个重要阶段提升用户意图理解准确性、知识增强有效性、知识使用的可用性，进而实现系统的“越用越好用”。

2、研究内容：

本课题研究方向包括但不限于大模型知识增强意图理解、大模型知识增强交互算法，基于人机回环的知识增强算法，以人为中心的大模型知识增强交互等方向。研究基于知识增强与人机回环的结合，设计可自我学习用户偏好与场景、意图、任务感知的算法机制，构造大模型时代“越用越好用的”知识增强系统。解决知识增强中的用户意图理解、增强知识检索/生成、人机交互及人机回环反馈等问题。提供有效可用、越用越好用的大模型知识增强交互系统。

3、研究目标：

- 1) 构建多样化的用户行为数据集，包括用户在不同交互场景中的点击、输入、选择和反馈等数据。并设计一套创新的发用户行为特征提取算法，用于从多样化的用户行为数据集中提取特征。通过上述算法，构建高效的用户行为驱动引擎，以提升用户行为预测和理解的准确性。在用户行为预测和理解方面，在 MIntREC 等主流数据集上，比现有 SOTA (85.51 Accuracy) 提高至少 20%。
- 2) 构建一种主动式自适应算法，该算法在运行时能够根据用户的实时行为特征，动态调整系统响应策略。并结合人机回环机制设计一种基于用户反馈的自我强化学习算法，通过算法能够在用户使用过程中，逐步学习和适应用户的行为习惯，实现对用户需求的精准理解，实现自适应机制的动态可调性，从而在不同使用场景下保持最佳性能。实现在知识增强相关任务下，提升知识检索和增强的效率和效果。在知识检索和增强领域，在 Quality 等主流数据集上，比现有 SOTA (82.6 Accuracy) 提升 20%

4、产出及交付物：

- ① 算法原型：完成大模型知识增强交互中的人机回环算法设计，构造全新的知识增强交互范式，交付一套大模型知识增强交互中的人机回环系统原型及其源代码
- ② 论文、专利：发表联想认可的 CCF-A 类会议、期刊论文 1-2 篇，申请 2 项相关专利

课题 8：基于多模态的 3D 数字人生成和驱动

1、研究背景：

3D 人物生成和驱动在许多直播、虚拟教学、游戏等诸多场景中都有广泛的应用。现有解决方案大多受限于 生成质量 → 可编辑性 → 可驱动范围。然而现有的人物生成算法在部分场景下表现出精度不高、Mesh 不规整、包含伪影等问题；与此同时，业界也暂无成熟、快速的 AI 生成方式来驱动虚拟人的动作与表情（较多的都是预置动作）。为了解决上述问题，提高人体生成精度以及 AI 驱动的生动性和动作丰富度，是当前研究的热点与难点进一步地，为了满足实际应用中的实时性要求，探索在端侧 GPU 等算力条件下完成高质量生成，并结合云端算力进行优化，具有重要的理论和实际意义。

2、研究内容：

- 1、文/图/单目视频 生成 3D 数字人：研究基于文/图/单目视频生成（或重建）全身可驱动 Avatar
- 2、多模态驱动动作生成：研究基于图像、文本、语音等输入的数字人肢体动作生成，并可选择多种动作风格
- 3、多模态驱动面部表情：研究基于图像、文本、语音等输入的数字人面部表情生成，面部表情生动形象

3、研究目标：

本课题旨在解决高质量三维人体生成，和全身 AI 驱动技术的关键问题，具体目标包括：

- 1、提高人物生成的精度及质量，并可转为四边形 mesh 导入主流制作软件进行编辑，具备合理几何结构以及较高质量纹理
- 2、提升多模态驱动动作及面部表情的准确度，动作丰富度、自然度，
 $BC > 7.724$, $Diversity > 13.47$, $FGD < 5.51$ (参考 EMAGE2024)
- 3、快速生成能力：实现基于端侧 GPU 的准实时驱动框架，算力需求小于 12GB 的显存 (<RTX4070)，其中人体单目重建耗时不超过 1 小时 3

4、产出及交付物:

- ① 技术原型及实验报告:3D 数字人生成以及多模态驱动的完整技术原型以及相关实验报告
- ② 集成原型与演示:提供数字人生成+多模态准实时驱动 3D 数字人动作以及表情的演示 Demo。
- ③ 研究论文与专利:在 CCF-A 国际期刊或会议上发表 1 篇研究论文, 专利 2 项

课题 9：跨应用渲染引擎技术中单视点到双视点转换方案研究与空间感知增强

1、研究背景:

随着裸眼 3D 显示器的兴起, 现在主流 3D 游戏和 3D 模型设计软件仍然以单视点显示为主, 难以充分利用裸眼 3D 显示器的 3D 效果。这种局限导致 3D 内容 (左右视图格式) 短缺, 影响了用户在裸眼 3D 显示器上的沉浸式体验。因此, 研究如何实现从单视点到双目视点的转换, 提升 3D 游戏和模型设计软件在裸眼 3D 显示器上的表现, 具有重要的理论和实际意义。

2、研究内容:

本课题旨在探索和开发跨应用渲染引擎中从单视点到双视点转换的技术方案。

主要研究内容包括:

视点转换理论模型: 建立从单视点到双视点转换的理论模型, 分析其在不同应用场景中的适用性。

视点转换算法设计: 设计高效的视点转换算法, 确保在不同应用中的实时性和准确性。

渲染引擎集成: 将视点转换技术集成到现有渲染引擎 (Directx、OpenGL、Vulkan 等) 中, 优化其在跨应用场景中的表现。

用户体验评估: 通过实验和用户测试, 评估视点转换技术对用户体验的提升效果。

3、研究目标:

视点转换技术的验证: 通过实验验证视点转换技术的有效性和实用性。

视点转换算法优化: 优化视点转换算法, 提高其在不同应用中的性能表现。

跨应用集成方案: 实现视点转换技术在不同渲染引擎中的无缝集成, 提升渲染

引擎的适用范围和用户体验。

4、产出及交付物：

- ① 技术报告：包括视点转换理论模型、算法设计与优化、实验设计与结果分析，视点转换后图像与原图像 $PSNR > 40$ ，转换后运行帧率 $\geq 60FPS$ 。
- ② 算法实现与测试结果：提供视点转换算法的实现代码及其在主流 3D 游戏及 3D 模型设计软件中的测试结果。
- ③ 集成原型与演示：提供集成视点转换技术的游戏和 3D 模型设计演示 Demo。
- ④ 用户体验评估报告：详细记录用户测试的设计、过程与结果，分析视点转换技术对用户体验的提升效果。
- ⑤ 专利：申请 2 项技术专利

课题 10：裸眼 3D 场景下 2D 视频转 3D 技术研究

1、研究背景：

裸眼 3D 技术作为一种不需要穿戴特殊设备即可观看立体图像的技术，近年来受到广泛关注。为了丰富裸眼 3D 的内容，将 2D 视频转化为 3D 视频成为关键挑战。传统的 2D 视频转换由于缺乏准确的深度信息和前景分割技术，导致立体感不足、时序稳定性差。在此背景下，通过构建相应的数据集，提升通用视频深度预测和前景分割模型的时序稳定性和精度，成为当前研究的重点。

2、研究内容：

本研究旨在通过构建高质量的数据集，提高通用视频深度预测和前景分割模型的时序稳定性和精度，从而实现 2D 视频向裸眼 3D 视频的高效转换。具体目标包括：

1. 提高视频深度预测模型的时序稳定性和精度，使物体帧间的深度变化规律与其 Z 轴运动轨迹一致。
2. 提高视频前景分割模型的时序稳定性和精度，确保分割结果的连续性和一致性。

3、研究目标：

1. 数据集构建：收集多种场景下的高质量 3D 视频数据；注视频中的深度信息和前景分割信息，形成用于训练和测试的数据集。

2. 深度预测模型研究：开发和优化基于深度学习的深度预测模型，提升模型的精度和时序稳定性；设计算法确保物体帧间的深度变化规律与物体的 Z 轴运动轨迹一致，避免深度跳跃和闪烁现象。深度预测时序稳定性：运动意图与真实运动的平均端点误差（EPE） ≤ 5.0 ；深度预测精度： $AbsRel \leq 0.07$

3. 前景分割模型研究：开发和优化基于深度学习的前景分割模型，提升模型的精度和时序稳定性；确保分割结果在连续帧间的稳定性，避免分割边界的不稳定和不连续现象。分割精度：交并比（Intersection over Union, IoU） ≥ 0.88 。

4、产出及交付物：

- ① 数据集：构建经过标注的高质量视频数据集，用于训练和测试深度预测和前景分割模型。
- ② 算法模型：提供经过优化的深度预测和前景分割算法模型。
- ③ 学术论文与专利申请：1 篇 CCF-A 类文章；2 篇相关技术专利，保护研究成果。