

施晓强

多模态大模型 / Agent / 3D 视觉方向博士研究生

13081232083

shixiaoqiang21@mails.ucas.ac.cn

GitHub

Google Scholar

个人概况

中国科学院大学计算机应用技术博士在读。具备 LLM/MLLM 预训练、SFT/LoRA 微调、强化学习训练、视觉语言对齐与 Agent 工具调用等实践经验；同时长期从事计算机视觉与基于 3D Gaussian Splatting 的 3D 人体重建研究，在可驱动人体/头部 Avatar 建模和高保真渲染方面有深入积累。

核心技能

- **LLM 训练与微调**: 熟悉 Decoder-only Transformer 训练闭环，具备 tokenizer、数据构建、预训练、SFT/LoRA 微调、强化学习训练、KV Cache 与采样解码等实践经验。
- **多模态大模型与 Agent**: 实践过视觉编码器-投影层-LLM 的多模态对齐、图文指令数据组织与 VQA/Caption 原型；了解 Agent 任务规划、工具调用、记忆管理和多轮交互流程，实现过 Coding Agent 原型。
- **计算机视觉与 3D 重建**: 熟悉语义分割、目标检测等 2D 视觉任务，长期研究基于 3D Gaussian Splatting 的可驱动人体/头部重建、几何约束、外观建模与高保真渲染。

教育背景

中国科学院大学	2021.09–至今
计算机应用技术，硕博连续（博士在读）	
辽宁科技大学	2017.09–2021.06
机械设计制造及其自动化，工学学士	

项目与工程实践

- **轻量级大模型全链路复现与多模态扩展实践**: 系统复现小参数 Decoder-only LLM 从预训练、SFT/LoRA 微调到强化学习训练与推理部署的完整流程，并进一步扩展到视觉语言模型和语音-视觉-语言多模态模型。
- **Coding Agent 开发**: 基于大语言模型构建面向代码任务的 Agent 原型，设计任务解析、上下文管理、工具调用、命令执行与结果反馈流程，支持根据用户需求完成代码理解、文件修改、运行验证和错误修复等闭环操作。
- **高保真 3D Gaussian 人体/头部重建**: 围绕单目/少样本输入构建 3D Gaussian Avatar 表示，设计可变形高斯、拓扑约束、分层增密与先验引导外观学习等方法，形成 3DGA、CANVAS、HiFiAvatar 等系列工作。
- **实时道路场景语义分割与轻量化网络**: 围绕道路理解场景设计 BSSNet 实时语义分割网络，关注精度、速度和模型复杂度之间的平衡；相关成果发表于 TCSVT 期刊，具备从网络结构设计到实验对比和论文产出的完整经验。

代表性成果

- **3DGA: 3D Avatar Animation from Monocular Video via Deformable Gaussian Splatting**, *Pattern Recognition*, 2026.01, IF: 7.6, 本人共一，排名第二（导师排名第一），SCI 1 区 Top, CCF B。
- **BSSNet: A Real-Time Semantic Segmentation Network for Road Scenes Inspired from AutoEncoder**, *TCSVT*, 2023.10, IF: 11.1, 第一作者, SCI 1 区 Top, CCF B。
- **CANVAS: Any-Shot Animatable 3D Gaussian Head Avatars from 1-K Images**, *ACMMM 2026* 投稿, 第一作者, CCF A。
- **HiFiAvatar: Prior-Guided Gaussian Appearance Learning for High-Fidelity Animatable Head Avatars from Monocular Video**, *ACMMM 2026* 投稿, 第一作者, CCF A。
- **3DGA: Topology-Aware 3D Gaussian Head Avatars with Barycentric Parameterization and Layered Den-sification**, *TIP* 投稿, 第一作者, SCI 1 区 Top, IF: 13.7, CCF A。

荣誉与服务

- 担任 TCSVT、KBS、ESWA、AAAI、ACMMM 等国际期刊和会议审稿人。
- 2024–2025 学年中国科学院大学三好学生。
- “中国光谷·华为杯”第十九届中国研究生数学建模竞赛二等奖。

自我评价

具备较强的问题拆解、实验推进和跨方向学习能力，能够围绕 LLM/MLLM、Agent 应用原型和多模态视觉理解快速建立可运行方案并持续迭代；沟通协作意识较强，能够主动承担任务并推进结果交付。