

赖彦涛

电话: 13033081901(微信同号) | 邮箱: yantaolai@nuaa.edu.cn | 中共党员

个人主页: yantaolai.github.io



教育经历

南京航空航天大学 (推免)

计算机技术 硕士 计算机科学与技术学院 模式分析与机器智能 (工信部) 重点实验室

2023 年 9 月-2026 年 4 月

安徽大学

智能科学与技术 本科 互联网学院

2019 年 9 月-2023 年 6 月

绩点: 3.65/5, 排名 14/118

实习经历

北京小米移动有限公司南京分公司

AIGC 与大模型算法工程师

2024 年 12 月-2025 年 3 月

南京

- 实习内容: 参与到了小米文生图大模型梵高的训练工作, 包含对开源文生图大模型 FLUX.1 的 LORA 微调与全参微调, 以及对基于 FLUX.1 的图像编辑功能 (如 Inpainting, Outpainting, 超分) 进行了探索与训练。目前该模型的生图效果超过了可灵 AI V1.5 (快手) 等模型, 并在内部投入了使用。
- 涉及技术: FLUX.1, Stable Diffusion, ControlNet, IP-Adapter, VAE, SFT, LORA, DiT, UNet, Flow Match, DDPM, DDIM。

合肥科大讯飞教育发展有限公司

数据资源开发工程师

2023 年 3 月-2023 年 6 月

合肥

- 实习内容: 参与讯飞星火认知大模型的 AI 数据标注、数据清理、数据生成等工作。累计处理数据量达 10 万条。

科研经历

Pathformer3D: A 3D Scanpath Transformer for 360° Images (与导师共一)

ECCV 2024

- 该论文提出了一种基于 SphereNet、Transformer 以及三维混合密度网络架构的全景图像扫视路径预测模型。该模型使用 SphereNet 模型来提取全景图像信息, 并通过 Transformer 编码器融合位置信息, 还采用了 Transformer 解码器来处理长序列历史注视点信息, 并设计了一种三维混合密度网络来生成新的注视点。实验结果表明, 该模型在 4 个数据集以及 6 个评价指标上均达到了 SOTA 的水平。

CLIPGaze: Zero-Shot Goal-Directed Scanpath Prediction Using CLIP (第一作者)

ICASSP 2025

- 该论文提出了一种基于 CLIP 的目标导向的扫视路径预测模型。该模型首先利用 CLIP 模型来提取文本和图像特征, 并设计了一种层级式视觉-语义融合模块进行特征融合, 然后采用类似 DETR 的 Transformer 解码器结构来并行预测人类的视线转移情况。该模型还尝试采用图像分割作为预训练任务来提升模型性能。实验结果表明, 该模型在零样本、传统目标存在、传统目标缺失等三种情况下都取得了 SOTA 的水平。

Speech-directed Human Attention Prediction with Large Multi-Modal Model ACM MM 2025(计划)

- 该研究旨在探索语言的文本描述中每一个单词对于人类注意力的引导情况。近几年, 多模态大语言模型得到了飞速的发展, 并被运用到多种视觉-语言任务中, 如视觉问答、视觉定位、图像分割, 但在扫视路径预测任务中仍然缺乏探索。我们对部分位置信息敏感的多模态大语言模型 (如: Sa2VA、Qwen2.5-VL) 进行修改, 并采用了 LORA 微调来提升在该任务下的效果, 初步实验证明我们的模型效果已经并超过了 ECCV 2024 的基准模型。

一种面向全景图像的人眼扫视轨迹预测方法 (导师一作, 本人二作) 专利号: ZL 2024 1 0512599.0 专利已授权

一种目标导向的扫视路径预测方法 (导师一作, 本人二作) 专利申请号: 202411720546.4

专利已受理

项目经历:

基于三维点云的牙齿修复研究

- 项目描述: 针对针对医疗患者牙齿缺失修复过程中依赖医生手动设计的问题, 提出了一种基于三维点云的牙齿修复方法。通过输入医疗患者牙齿的点云数据, 自动生成缺失牙齿的三维模型, 显著提升了修复效率和精度。我们通过点云补全技术生成完整的完整点云数据, 然后采用点云分割分割出需要部分的牙齿点云, 最后通过点云上采样来生成密度均匀与光滑的点云数据。

- 相关技术: 点云补全, 点云上采样, 点云分割

C++ Web 高性能服务器 (Tiny-WebServer)

- 项目描述: 一种基于 C++ 的轻量级 Web 服务器

- 相关技术: 线性池, 非阻塞 socket, epoll

技术细节 & 荣誉

- 研究领域: 计算机视觉, 深度学习, 多模态学习, 扫视路径预测, 点云, 文生图
- 主要课程: 计算机视觉, 机器学习, 模式识别, 图像处理与分析, 智能控制理论, 智能软件开发与测试技术, 数据库, 自然语言处理
- 编程语言 & 框架: Python, C++, C, PyTorch, Linux Shell, Hugging face, Git, LATEX, OpenCV, Java(了解), Tensorflow(了解), Matlab(了解)
- 本硕期间荣誉: 同竣奖学金, 校一/二等奖学金, 校三好学生、校优秀共青团员、研究生一等奖学金、三好研究生、科研创新先进个人
- 其他: CET-4/6