

# 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)

2023 年 9 月-2026 年 4 月

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

### 安徽大学

2019 年 9 月-2023 年 6 月

智能科学与技术 本科 互联网学院

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

2024 年 12 月-2025 年 2 月

AIGC 与大模型算法工程师

南京

• 实习内容: 参与小米文生图大模型梵高的训练工作, 主要涉及文生图大模型 FLUX.1 的 LORA 微调与全参微调。涉及 FLUX.1, Stable Diffusion, Control Net, Flow Match, VAE, DiT, LORA 等相关知识

### 合肥科大讯飞教育发展有限公司

2023 年 3 月-2023 年 6 月

数据资源开发工程师

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

## 科研经历

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

ECCV 2024

• 该论文旨在研究全景图像中人类视线/注意力的转移情况。针对全景图像的扭曲问题, 我们采用了一种为全景图像定制的三维卷积网络 SphereNet 来提取全景图像信息; 然后将其通过 Transoformer 编码器来上下文化图像信息; 再将其和历史注视点信息一起输入 Transformer 解码器中, 并采用自回归的方式来生成新注视点的隐藏状态; 最后该隐藏状态通过我们所设计的一种新颖的三维混合密度网络, 这样操作可以将注视点的隐藏状态转化为相应的注视点分布信息; 最后我们通过采样的方式来生成新的注视点。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

• 该论文旨在研究目标导向的人类视线/注意力的转移情况。针对语义对齐问题, 我们采用 CLIP 来提取目标的语义信息和多层级场景图像信息; 然后我们采用一种新颖的层级式视觉-语义融合模块来对器进行融合; 再将融合后的信息和注视点的查询信息一起输入并行的 Transformer 的解码器 (类似 DETR) 中来预测注视点的隐藏状态; 最后采用了多个 MLP 来将其转化为注视点的分布信息。我们还采用了图像分割作为代理任务来进行预训练, 这显著提升了扫视路径预测效果。实验结果证明, 我们的模型在零样本、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

进行中

• 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, 多模态视觉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。由于该任务需要预测精确的位置坐标, 我们将对部分像素级位置信息敏感的视觉语言模型 (如: Sa2VA、Qwen2.5-VL) 进行部分修改, 然后对其进行 LORA 微调, 探索其在当前任务下的效果表现。

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) 专利公布号: CN 118397686 A 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:

### 牙齿点云补全

• 项目描述: 针对医疗患者缺少部分牙齿的点云数据, 设计一种自动对牙齿点云进行补全的方法, 针对性的生成缺失部分牙齿的点云数据

• 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