



林钰翔 13510595062 Email: yuxiang.lin@gatech.edu

求职意向：算法工程师 | 个人主页: <https://lum1104.github.io/> | Github: <https://github.com/Lum1104/>

教育背景

2024-08 ~ 2025-12	佐治亚理工学院 – 电子与计算机工程	硕士
2020-09 ~ 2024-06	深圳技术大学 – 物联网工程	本科

实习经验

MER-Factory – 大模型数据生成 Agent 算法开发工程师 2025-06 ~ 至今

背景介绍：因多模态情感识别及推理（MERR）研究中数据集构建复杂、标注成本高的问题，开发自动化的情感数据处理与分析工具链，目前开源至 Github 平台上，获约 70 Star 仍持续增长中<https://github.com/Lum1104/MER-Factory>。

- 开发了多模态情感识别与推理数据集构建工厂：利用多 Agent 协作使用 OpenFace、FFmpeg 和多种 MLLM 模型（API、Ollama、HuggingFace）自动化提取和分析视频、音频、图像的情感特征，针对情感任务设计情感尖峰帧检测及时序情感描述。
- 设计基于 LangGraph 的状态驱动处理架构：实现模块化计算图工作流，支持 AU 提取、音视频分析、图像识别和完整 MER 处理等多种管道 Pipeline，通过异步并发处理提升大规模数据集处理效率。
- 构建标注中-标注后自动化质量评估体系：设计 Gate Agent 使用 Plan-Solve 的 ReAct 方式在数据生成过程中把控数据质量，构建 Reference-free 的评估指标，利用 CLIP-score、CLAP-score 等评估数据集整体生成质量。

腾讯科技（深圳）有限公司 – 大模型算法工程师 2025-05 ~ 2025-09

背景介绍：腾讯云团队每天面临海量大客户售后工单，发掘这些工单背后的共性问题不仅能够提效降低工单，还能提前发现内部隐患。共性问题分析分为——已知共性问题分类及未知共性问题聚类。目前架构只能分类出固定的共性工单，例如产品体验、产品稳定性等。

- 设计开放文本的共性工单分类架构：利用混元大模型 Hunyuan Embedding 获得工单、分类的向量特征，通过构建模版以及计算相似度的形式，完成无需预定义分类类别的工单分类任务。
- 基于云专业知识的数据清洗与预处理：结合腾讯云专家&官方文档提取关键词清洗工单原始信息中包含的沟通、转单无用内容，并编写 Prompt 转写工单内容为更精准描述具体事件的内容，人类满意度从45%提升至70%。
- 主导共性工单发现系统的线上部署与应用：实现了对每日新增数万工单的准实时分析，自动挖掘并预警了超过 20 类未知潜在问题，将新问题发现准确率提升了 15%。

百度在线网络技术（北京）有限公司 – 图图搜索策略算法工程师 2024-01 ~ 2024-04

背景介绍：图搜内部测试发现检索模型在全局语义信息上有所匮乏，例如会误将白猫白狗/红椅子红垃圾桶，不同类别物体召回。

- 全局+局部表征能力：使用百度点攒日志数据集，改进图像检索 ViT 模型使其具备图片语义信息+局部色彩信息捕捉能力
- 构建层次损失函数：设计训练损失函数多层次约束检索模型，通过累加分类 logit 的方式构建多层标签信息，结合 CE loss 完成训练
- 显著提升性能：在内部测试集上显著改进特征的语义表征能力 bad case 如黑猫黑狗原无法区分现可以，检索效果 Recall@K 提升 2%

Item.com – 时间序列预测机器学习算法工程师 2025-09 ~ 2025-12

背景介绍：负责供应链核心场景下的需求预测算法研发与工程化落地，针对极度稀疏的订单数据特性，构建了从特征工程、模型建模到生产级服务部署的全栈解决方案。

- 稀疏数据算法架构创新：针对供应链数据中普遍存在的“零值膨胀”难题，设计并落地二阶段预测架构。通过级联 LightGBM 分类模型（判别下单概率）与回归模型（预测具体销量），有效解决了传统回归模型在长尾稀疏数据上失效的问题；
- 多维特征工程与评估体系：构建了包含Target Encoding、Rolling Feature等特征，大幅增强模型对时间序列模式的捕捉能力。引入 Log 变换处理数据偏态，并建立包含中位数 MAE、Log-R2 及离群点剔除 R2 的多维评估指标，量化模型在业务场景下的泛化性能。
- 高可用 MLOps 模型服务平台：基于 FastAPI 和 MinIO 构建生产级模型推理服务，设计了原子化热更新与异常自动回退机制。实现了模型文件的哈希校验、版本控制及在线无感切换，确保模型迭代发布期间服务零宕机

Everlyn – AIGC扩散模型算法工程师 2024-08 ~ 2025-02

背景介绍：本项目主要负责地理定位任务，经由地面视角全景图匹配检索至卫星图（含有经纬度信息）完成。

- 将检索任务与生成任务结合：负责跨视角图片生成项目，利用 Stable Diffusion 模型将地面视角图片（全景图）生成卫星视角的图片
- 改进输入condition特征：将地面视角图片转换为鸟瞰图，减轻生成过程中的显著视角差异，完成图片生成任务
- 多任务学习策略：提出检索+生成多任务学习，完成跨视角图片生成 (diffusion loss) 和地理定位 (InfoNCE loss)，均达到 SOTA 效果

- ACM MM 多模态情感识别挑战赛亚军；全国大学生数学建模大赛国家二等奖；腾讯云从业者资格证
- 浙江大华杰出奖学金（4000元）、工匠之星（3000元）、研究创新一等奖（3000元）、深圳技术大学优秀新生（6000元）

科研经历及论文列表

华盛顿大学 – 音视频多模态大模型情感理解算法工程师

2023-02 ~ 2024-07

背景介绍：主要负责多模态情感推理任务，将视频拆分为音频、视频、文本台词后利用大模型逐步推理出内在情感状态

- **论文发表**：发表 Emotion-LLaMA (NeurIPS, CCF-A) , ExpMAE (ACM MM, CCF-A) , RT-CAN (CVIU, CCF-B) 等多篇文章
- **多模态数据集构建**：通过心理学知识人脸微动作组合，从70000数据集构建伪标签输入 Emotion-LLaMA 大模型完成情感推理任务
- **大模型评测任务**：设计 EIBench 评测 14 种大模型开源、闭源在多面情感解释上的能力，提出由粗到细的 CFSA 自我提问流程辅助大模型标注，显著提升标注质量构建高质量评测集1665条，在 CVPR NeXD 上作为 1/3 篇 Oral 文章报告。

Emotion-llama: Multimodal emotion recognition and reasoning with instruction tuning, NeurIPS, CCF-A

第二个学生作者，2024/12/16

https://proceedings.neurips.cc/paper_files/paper/2024/file/c7f43ada17acc234f568dc66da527418-Paper-Conference.pdf

本文提出了首个专注于多模态情感推理的开源大模型 Emotion-LLaMA。通过设计特定的指令微调 (Instruction Tuning) 策略，将音频、视频与文本模态对齐，使模型不仅能精准识别情感类别，更能结合多模态线索推理出情感背后的成因，显著弥补了通用大模型在垂直情感理解与解释能力上的不足。

Semi-supervised multimodal emotion recognition with expression mae, ACM MM, CCF-A

二作，2023/10/26 <https://dl.acm.org/doi/pdf/10.1145/3581783.3612840>

针对多模态情感识别中高质量标注数据稀缺的痛点，提出了一种基于表情感知的掩码自编码器 (Expression MAE) 半监督学习框架。通过设计针对面部特征的掩码策略，有效利用海量无标签数据增强了模态间的交互与表征学习能力，在少样本场景下实现了优于全监督方法的性能。

Mips at semeval-2024 task 3: Multimodal emotion-cause pair extraction in conversations with multimodal language models, SemEval

三作，2024/3/31 <https://aclanthology.org/2024.semeval-1.97/>

在 SemEval 国际评测中，针对对话场景下的情感-原因对抽取 (ECPE) 任务，设计了基于多模态大语言模型的解决方案。通过联合分析对话上下文、声学特征与视觉线索，实现了对引发特定情感的具体话语或事件的精准定位，验证了 MLLM 在处理复杂对话归因与上下文理解任务中的有效性。

Invisible gas detection: An RGB-thermal cross attention network and a new benchmark, CVIU, CCF-B, JCR Q1

共一，2024/11/1

<https://www.sciencedirect.com/science/article/pii/S1077314224001802>

针对工业场景中不可见气体泄漏检测的难题，构建了首个RGB-热成像多光谱气体检测基准数据集，并提出了一种跨模态交叉注意力网络 (RT-CAN)。该方法通过深度融合可见光的环境语义与热成像的温度特征，有效解决了复杂背景下的干扰问题，在气体检测精度上达到了 SOTA 水平。

FlexEdit: Marrying Free-Shape Masks to VLLM for Flexible Image Editing, PR在投, CCF-B, JCR Q1

共一，2024/8/22 <https://arxiv.org/pdf/2408.12429>

探索了视觉大语言模型 (VLLM) 在精细化图像编辑中的应用，提出了结合自由形状掩码 (Free-Shape Masks) 的 FlexEdit 框架。该方法突破了传统矩形掩码的限制，允许用户通过自然语言指令配合任意形状选区，实现对图像特定区域的精准修改与内容生成，提升了编辑的灵活性与语义一致性。

Why We Feel: Breaking Boundaries in Emotional Reasoning with Multimodal Large Language Models, CVPR W, Oral

一作，2025/7/1

https://openaccess.thecvf.com/content/CVPR2025W/NeXD/papers/Lin_Why_We_Feel_Breaking_Boundaries_in_Emotional_Reasoning_with_Multimodal_CVPRW_2025_paper.pdf

聚焦于多模态大模型的情感因果推理能力，构建了综合评测基准 EIBench，并提出了“由粗到细自我提问” (CFSA) 的推理链方法。该工作通过引导模型逐步分解视频中的多模态线索来解释情感成因，显著提升了模型生成解释的逻辑性与准确度，该成果入选 CVPR Workshop Oral 报告。