

教育背景

- 2017.09 北京邮电大学, 信息工程, 学士学位
- 2021.06 研究方向: 人工智能, 信号处理
- 2021.09 北京邮电大学, 人工智能, 硕士学位
- 2024.06 研究方向: 计算机视觉, 语音识别与合成

工作经历

- 2024.07 奇虎 360, 高级语音工程师, 负责语音识别、语音理解大模型
- 至今 负责语音识别方向能力建设, 为输入法、搜索、录音笔等业务提供服务; 探索语音理解大模型。
- 2022.10 联想研究院, 算法实习生, 负责图像分割、新视点合成
- 2024.03 负责图像分割、3D 重建算法, 实现业界首个支持 4K 超高清实时 2D 转 3D 技术。

专业技能

- 专业能力 语音识别、多模态语音大模型
- 编程能力 Python、Funasr、Git、Transformers

项目经历

语音理解大模型方向 (2024-2025)

- 2024.10 语音理解大模型与数据, 项目负责人, 奇虎 360
- 2025.06
 - 解决数据不足问题, 通过利用设计数据挖掘 Pipeline, 提供 15 万小时高质量数据。
 - 搭建语音理解大模型, 支持多语种 ASR、多语种 AST、语音问答等通用语音任务。
 - 多语种 ASR/翻译作为重点能力, 提升实用性与覆盖面, 翻译和识别能力超过 Qwen2-Audio。
 - 360 极客大赛: “大模型语音识别引擎” 课题获得 “创始人奖”。

语音识别方向 (2024-2025)

- 2024.07 输入法流式语音识别, 项目负责人, 奇虎 360
- 2025.06
 - 构建输入法语音识别能力, 提升模型在室外场景的抗噪能力以及口音识别能力。
 - 设计业务数据自动标注和筛选流程, 通过筛选线上弱标签数据, 对模型快速优化。
 - 提升输入法场景下的模型识别准确度, 相对竞品准确率相对提升 13%, 支持中英文混杂输入。
 - 负责模型部署和服务维护, 输入法业务访问量提升 80%, 维护服务稳定运行。
- 2024.11 离线长音频语音识别, 项目负责人, 奇虎 360
- 2025.06
 - 为音视频解析业务提供语音理解能力, 支持长音频 ASR、时间戳预测、说话人预测能力。
 - 通过闭源大模型投票标注数十万小时数据, 以及 TTS 合成专有领域数据, 扩充知识覆盖面。
 - 优化训练速度和内存占用, 提升模型抗噪能力, 专有名词召回率超过竞品。
 - 优化各个模块并行推理和聚类 GPU 加速, 提升推理速度 100%, 支持 10 小时长音频文件。
- 2025.04 小语种语音识别, 项目负责人, 奇虎 360
- 2025.06
 - 打造东南亚语种语音识别能力, 为业务出海提供技术支持。
 - 通过采集互联网公开音视频, 进行弱标注获取大量小语种数据, 进行模型训练, 提升模型效果。
 - 通过 ASR+ 文本翻译大模型级联, 对外提供同声传译服务。

计算机视觉方向 (2022-2024)

- 2022.04 **图像分割后处理**, 项目负责人, 联想研究院
- 2023.03
 - 负责图像分割后处理算法代码实现、实验验证、论文撰写, 在 Cityscapes 数据集达到 sota。
 - 首次将类别信息和位置信息作为 Global Knowledge Prompt 辅助图像分割后处理任务。
- 2023.04 **2D to 3D 视频实时转换**, 项目负责人, 联想研究院
- 2024.03
 - 实现业界首个支持 4K 超高清实时 2D 转 3D 技术, 入选 2023 联想研究院十大科技。
 - 攻克了视频抖动、视差无法太大的技术难题, 并发表专利, 为裸眼 3D 显示器提供算法支持。

Publications

"Pyramid Attention CycleGAN for Non-Parallel Voice Conversion," (*ICCC*)

"Enhancing FPGA-Based YOLO Object Detection: Multi-Bank Storage Optimization and Model Refinement for Real-Time Applications," (*ICICM*)

"MRCI: Multi-range Context Interaction for Boundary Refinement in Image Segmentation." (*ICPR*)