方悦

电话：18938813098 | 邮箱：1206137707@qq.com

个人评价：工作认真，自我驱动能力与团队沟通能力强，能基于经验及自学快速提升个人能力。本科专业排名前 8%，具有 1 段自驾相关实习（蔚来），1 段大模型幻觉相关实习（中国电子），1 篇多模态大模型方向 CCF 论文在投，一项自驾相关专利。

教育背景

2025.09-2028.06（已保研，未入学）　　东南大学（985 双一流）-网络空间安全学院 学术硕士

主要荣誉：学习优秀奖学金

2021.09-2025.06　　北京交通大学（211 双一流）-电子信息工程学院　　本科

本科成绩：89.0/100　　专业排名：6/73（8%）

主修课程：自然语言处理（96）、ACM（98）、模式识别与机器学习（96）、信号与系统（94）、数字信号处理（97）。

主要荣誉：二等学习优秀奖学金（10%），校三好学生，二等社会工作奖学金（10%），校优秀团干，校优秀学生干部。

实习经历

蔚来 | 自动驾驶研发数据算法部门 - 车端挖掘团队 - 大模型数据算法实习生　　2025.02-2025.06

- 选用多个基准及最新 RL 技术（VLM-R1、R1V、Viusal-RFT、Video-Chat-R1），实验验证强化微调相较于监督微调在增量学习、模型泛化、知识遗忘、资源消耗等领域的优劣势。基于 open-r1 框架设计强化学习算法，进一步提升 base 模型目标检测能力。
- 面向红绿灯检测任务（TLD）及施工场景交通设施检测任务，基于大模型部门设计的视觉语言模型进行“两阶段监督微调+强化微调”训练。数据集部分采用 YOLO-V12 选框、特殊标记转换、动态目标裁剪、坐标归一化等处理；监督微调上依托 Llama Factory 平台进行参数调整对比；强化微调上设计适用于攻击内部模型 openmlm 的 GRPO 算法。项目最终实现 FN、FW 视角下的高精度目标检测，在部分检测任务中达到 93%的准确率。
- 为提升 VLM 模型路径规划能力（速度与方向），调用 NIO 数据库及 NuScence 平台上的数据，通过丰富提问范式、构造思维链、设计驾驶状态真值生成算法等方法制作丰富的驾驶数据，设计多个基于 GRPO 的强化学习奖励（如规划准确性奖励、动作加权奖励、规划多样性奖励等），算法最终在部分路径规划任务中达到 97%的准确率。
- 为充分释放自车与仿真环境交互生成的海量数据价值，参考 Self-Play 等前沿技术设计 PPO 算法，验证仿真环境能力潜力，并以此推动路径规划仿真数据库的建设。
- 针对长尾数据缺失，利用 DriveGEN 生成极端天气多视角数据集；针对标注成本及多维数据整合问题，应用 MIM4D 实现时空四维数据到二维的高效映射，应用 MetaVQA 生成 30 类包含 CoT 的问答对。
- 探索 VLA+RL 在自动驾驶领域的工程应用，调研并复现 DriveMoE 和 SimLingo 等最新技术，并在 Carla 仿真环境中调用 Bench2Drive 进行训练与能力评估。

中国电子云 | 大模型算法实习生　　2025.01-2025.02

- 针对基于大模型的人物政党记及大事记生成任务中附带的幻觉问题，主持设计基于 Qwen 大模型及提示词工程的结构化自动幻觉纠正工具。工具涵盖原文时间核对模块、主体一致性判断模块、主客观分析模块、独立声明提取模块及幻觉自纠正模块等，针对不同幻觉分类提供细粒度评分及纠正建议，**成功降低幻觉率至 2%。**

项目经历

多模态大语言模型幻觉的检测与量化评估 | 独立完成　　2024.09-2025.06

- 针对语义对大视觉语言模型模型输出内容幻觉评估的影响，从横纵两个方向增强语义协作并减少语义干扰，实现基于不确定性量化的**零参考黑盒幻觉量化评估**；
- 语义定位器（横向）：从 UD 语料库中随机抽取若干符合先决条件的单词对独立子句中的每个单词进行替换，基于 BertScore 评估替换前后句子语义波动幅度，进而衡量单词语义重要性，使用重要性得分对幻觉分数进行重加权；
- 语义提纯器（纵向）：设计并对比三种语义聚合方法对每个令牌输出的概率分布进行语义聚类，同时在此阶段排除语句风格及表达顺序对幻觉量化的影响；
- 设计人工数据标注程序及自动标注工具并验证自动标注效果可靠性，降低人力成本；调研并选取若干最新的性能优越的幻觉评估数据集、幻觉评估技术及多模态大语言模型进行实验验证；
- **项目获校级优秀毕业论文，已列入华为技术有限公司“东南大学鲲鹏昇腾科教创新孵化中心”合作协议项目，目前 IMAVIS（中科院二区，CCF-C）在投。**

清华大学猛狮无人驾驶实验室：自动驾驶视觉感知课题 | 联合培养实习生 **2024.05-2024.11**

- 面向**计算机视觉领域的增量学习**，旨在提升开放域增量学习的实时性、提出高准确度陌生目标识别方法；
- 实现基于 R-CNN 的目标检测，基于 SAM 的目标分割以及基于元学习技术的特征原型生成与增量学习；
- 实现基于强化学习的机器人路径规划；
- 负责调研整理领域新兴技术、攥写技术报告等文书，环境配置与代码调试，**项目已转化为专利《一种基于增量式学习的感知模型的目标检测方法及装置》，申请号 CN202411715425。**

教材撰写：大模型安全 | 独立完成 **2024.09-2024.12**

- 负责撰写教材第 10 章《大模型安全》共计 26 页内容；
- 内容涵盖大模型幻觉、提示注入攻击（直接与间接攻击）、大模型投毒（数据投毒与模型投毒）三大大模型核心安全问题的典型案例、分类、评估与检测方法、防御策略、代表性技术和新兴技术等。

安全生产智能巡检机器人关键技术研发 | 核心成员 **2023.01-2024.04**

- 为解决大型厂矿巡检任务中画面受雾色及噪声干扰大、小目标识别困难问题，设计高准确度图像处理及目标检测技术；
- 调研并选用暗通道先验算法、高斯模糊算法等克服画面干扰，基于 YOLOV5 进行目标检测，利用阈值分割、霍夫变换、DBNet、CRNN 等实现 21 种目标的检测与信息读取，设计 UI 界面并与部署多种运动算法的 Turtobot3 实现实时通信；
- mAP@0.5 达 87.6%（超基线 11.4%），获中国国际大学生竞赛校级奖项、“小挑”校级奖项，大创校级立项。

关于交通标志识别与检测系统的优化 | 核心成员 **2024.03-2024.06**

- 面向交通场景中交通标志的漏检、误检问题设计两阶段交通标志检测与识别系统设计；
- 选用 6 个评估指标，设计消融实验对比并选用性能最优的图像处理技术，基于 YOLOV5 对处理后的数据进行目标检测；
- 选取 LoRa 微调大模型，实现模型对交通标志具体类别的识别。

轨道交通控制与安全国家重点实验室 | 参研 **2023.09-2024.03**

- 负责太赫兹成像技术调研工作，并通过 Meshmixer 对不同材质的粗糙体进行构建，基于 FEKO 实现太赫兹散射模型的全波仿真，基于 Matlab 生成参数文件等。

肺部 MRI 图像肺实质分割 | 独立完成 **2024.03-2024.05**

- 采用 U-Net 完成基本语义分割任务，将其收缩路径用 VGG16 替代以提升性能；
- 通过数据增强、形态学处理、空间一致性处理等克服医学图像匮乏问题。

校园经历

北京交通大学合唱团 | 声部长 **2022.09-2023.08**

- 带领部员共计 46 人，多次组织策划演出十余场，包括长征组歌专场演出（4000+人流量）、北京市大学生音乐节等，获评“校级优秀团干”、“一类一等文艺活动优秀奖学金，“北京大学生音乐节声乐类展演”优秀表演奖等。

BAS 线上跨国支教 | 优秀个人，获国际志愿者证书 **2023.01-2023.02**

其他技能

语言：英语（CET6（596），IELTS（7），全国大学生英语竞赛（省级二等奖），中文（普通话二甲）；

编程能力：较好把握 python、C++、Matlab、Latex 等，可熟练操作 Linux 平台，熟悉常见机器学习及深度学习基础理论、常见机器学习及深度学习框架（Pytorch、Tensorslow），**并具有相关项目开发经验**；

办公能力：计算机二级（WPS）。