郑晋图

Email：zhengjintu22@mails.ucas.ac.cn

性别：男 | **年龄：**26 | **求职意愿：**不限行业的算法工程师、研究员；Base 优先在：广州、深圳、上海

个人简介

有端对端模型的训练开发经验，从事过数据闭环系统的开发，熟悉 BEV 感知任务（Map、Det3d）的精度优化。有模型训练加速优化经历。从事过推理加速的研究。有多模态大模型在 3 维空间实现有效感知的经历。

算法研究经历丰富，在顶刊顶会有多篇文章发表（ECCV / NeurIPS / Sensor Journal / ICPR / MMAsia / MIA），有 **CV 三大会 Oral 第一作者**（接收几率 2%）的工作，擅长探索新技术和新方向。

算法工程开发技术广泛（Python，C++），有多种模型调参经验，了解基础的 HPC 技术、CUDA 算子开发，熟悉 mm 系列，Xtuner，DeepSpeed、Lightning 等相关开发，使用过 TensorRT。此外，对 RL、图像生成方向有一定接触。

工作经历

2024.10 – 2025.4(至今) **零一汽车 | 智能驾驶部门**，（感知算法工程师，实习）

- **方向：**端对端大模型的复现；训练并行加速优化；数据闭环系统开发；离线感知优化；
- **期间：**复现 Waymo 的 EMMA 模型；对已有端对端框架的感知编码训练以无损精度的前提下加速了 40%，节省了模型开发迭代的时间；负责数据闭环中离线感知标注的开发和优化；增强 VLM 对 3 维空间的感知能力；

2024.02 – 2024.10 **AMD (超威半导体) | AIG-Models**，（算法研究员，实习）

- **方向：**车道拓扑预测、在线高精地图重建；LLM 剪枝、扩散模型剪枝
- **期间：**参加 CVPR 2024 AGC Mapless Driving 赛道排名第 5，共 120 个参赛队伍；会议文章 2 篇（一作 1 篇）

教育经历

2022.09 - 2025.06 **中国科学院大学 | 深圳先进技术研究院，**电子信息-计算机技术（硕士）

- **方向：**AI4Science，显微图像分割、跟踪，亚细胞结构预测，显微三维重建；
- **期间：**GPA：3.75/4；领域知名期刊、会议文章 6 篇，其中第一作者 4 篇；

2018.09 - 2022.06 **华南农业大学（双一流） | 数学与信息学院，**计算机科学与技术（本科）

- **期间：**GPA：3.51/4；2022 年优秀毕业论文；软件学报 1 篇、专利 1 个；大创国家级项目负责人，省级学科一等奖 2 个

主要从事项目描述

1. **EMMA 端对端大模型复现：（零一汽车）**

   项目描述：根据 Google Waymo 的公开文档复现 EMMA ，使用 nuplan 评估。成功复现了单独任务的回答（Drivable area、Det3d、Reasoning）和合并训练，设计优化了 EMMA 对感知任务中坐标的纯文本输出，降低 tokens 输出为实时推理做铺垫。

2. **端对端驾驶模型的训练、开发、加速：（零一汽车）**

   项目描述：需求是现有的模型版本训练迭代时间过长。分析出目前感知特征部分的更新存在瓶颈，用 profiler 分析各个模块瓶颈，优化了地图解码器的栅格缓存；对并行多卡训练的实现做优化；将在线地图的 geloss 用 CUDA C 重写（需要推导反向传播并用 c 重写），在 fp32 和 fp16 中验证了和原有训练表现对齐，实现了 140%的加速比。此外，设计了一种在 3D 空间感知位置 token，引入到 VLM，同时引入在 EMMA 复现项目的探索出来优化技术，提高现有 planning 的任务表现。

3. **数据闭环开发：（零一汽车）**

   项目描述：为公司的数据闭环系统开发高精度的离线感知标注模型。为克服数据不同源，设计了一些策略实现更高泛化性，开发自动化知识蒸馏系统。

4. **车道拓扑预测：（AMD）**

   项目描述：参加 CVPR 2024 AGC Mapless Driving 赛道，实现在线地图（Camera only）车道拓扑预测；最后排名第 5 名（共 120 支队伍），打败滴滴等公司的队伍；在 mmdet3d 框架开发比赛的训练、推理；引入 SD Map 先验用于优化车道预测；增加车道中心线分类作为辅助损失；引入已有的 BEV 预测结果优化拓扑关系输出；引入 BEV 上的时序信息；交通要素数据分析、清洗、增强；测试不同目标检测在交通要素检测任务上的实际表现；ViT-Large 作为 Multi-View Image Encoder 训练的调参，解决全量训练 OOM（逐层 warmup 解冻训练策略，搜索对下游关键的层参数）；迁移 MapTR、MapTRv2、StreamMapNet、MapTracker 关键提点的组件；利用车道中心线输出的 z 轴坐标拟合 BEV 曲面，优化道路边界

5. **推理加速：（AMD）**

   项目描述：为 LLM 实现无微调推理加速；提出全新的无微调推理加速算法，在稀疏度 22%的 Llama2、Llama3/3.1、Qwen2 上保留 99%性能，远优于竞争方法。目前 ACL 在审，一作。测试不同 token 特征对推理运行时的影响；开发动态 Router 完成层间的实时推理路径判断；开发并行超参搜索框架；测试在 KV Cache 上的实际加速；实现 Router 的模拟直通，解决离散结果无法参与梯度反传的问题；SFT 微调、采用蒸馏实现特征对齐恢复。

6. **Diffusion 推理加速：（AMD）**

   项目描述：在无需微调的剪枝方法在 SD 1.5 上实现 4.4 倍的推理加速；中 NeurIPS 2024 Poster；

7. **亚细胞预测：（硕士在校研究方向）**

   项目描述：中 ECCV 2024 Oral（一作）；提出稀疏视图的亚细胞结构预测，最大可降低 87.5%的生物成像成本，对显微动态活细胞成像有重大意义；分析现有亚细胞结构预测方法的成像成本；基于稀疏成像数据开发训练和推理框架；比较不同拓扑架构策略的差异；

8. **视频对象跟踪、分割：（本科毕设研究方向）**

   项目描述：提出一种联合解码阶段共同优化的策略，并对 memory bank 的 token 做不同尺度的分配，解决 Memory based 方法普遍存在的记忆错误匹配问题。（ICPR 2024），第一作者

获奖 & 竞赛经历

硕士期间：

CVPR 2024 Autonomous Grand Challenge (AGC), Mapless Driving Track，第五名，共 120 个队伍参赛

本科期间：

- **2020 年大学生创新训练 项目负责人（国家级）**《基于姿态评估和时空特征分析的母猪分娩时间精准预测研究》
- **2020 年“丁颖杯”暨“挑战杯”广东省课外学术科技竞赛一等奖（省级）**《群养环境下生猪的自动识别和体况计算研究》
- **2021 年（第十六届）泛珠三角+ 大学生计算机作品赛总决赛一等奖（省级）**《基于自适应模板更新孪生网络的目标跟踪》
- **2022 届本科优秀毕业论文（校级）**《基于图注意力特征记忆和定位聚焦网络的视频对象分割方法》
- **DAVIS 2017 test-dev 提交，排行第 4. (公开竞赛)**

更多

[1] 研究工作主页： Jintu Zheng | OpenReview，ORCID，dblp: Jintu Zheng；

[2] AGC 获奖结果主页： CVPR 2024 AGC 排行榜，（AMD）