

谢竞成

(+86) 139-7338-7126 | jingchengxie0131@gmail.com¹

教育经历

西安电子科技大学 智能科学与技术, 人工智能学院 本科	2020.09—2024.7
• Rank: 7/163	

中国科学技术大学 || 人工智能, 信息科学学院 | 在读研究生
• 类脑国家工程实验室
• 方向: 大模型高效化与轻量化

实习经历

vivo 维沃移动通信有限公司 助理算法工程师, 影像效果项目部 影像效果自动化调试项目	2023.09—2024.02
• 参与高通平台视频自动化调试模块开发, 负责高通平台自动化调试代码重构, 以及并行图像匹配算子开发。 • 预研并轻量化部署修复算法, 产生定向像素级图像增强模块, 取代自动化调试中的竞品机。	

科研经历

探索无需训练的稀疏后大模型恢复	2024.11—2025.06
• 提出了 EWIR 方法, 针对 Llama 等大模型非结构化剪枝后性能损失的问题, 本文提出一种无需再训练的微调策略, 基于诺伊曼级数快速求逆, 10min 内高效微调恢复剪枝后的 Llama2-7B 剪枝后损失的性能。	

• Jingcheng Xie, Xiaoyu Liu, Yinda Chen, Zhiwei Xiong: Training-Free Fine-tuning of Sparse Large Models via Efficient Iterative Weight Recovery. Under-review

探索保留泛化能力的大模型微调方式	2024.6—2024.11
• 提出了 SHERPA 方法, 针对大模型在特定任务中易受过拟合影响、导致泛化能力下降的问题, 通过利用小模型的任务相关注意力来指导大模型微调, 使泛化能力保留提升最多达 11.1%, 任务特定性能提升最多达 2.2%。 • Jingcheng Xie, Yinda Chen, Xiaoyu Liu, Zhiwei Xiong: SHERPA: Fine-tuning Large Segmentation Models with Task-relevant Guidance. Under-review	

于西电 IPIU 实验室科研实习	2021.10—2023.03
------------------	-----------------

项目经历

联合稀疏量化加速算子开发	2025.4—至今
基于三级内存层次优化与 Tensor-Core-Aware 位图编码, 共享内存位图解码完成稀疏 W4A16 乘法算子。	

Kaggle: G2Net Detecting Continuous Gravitational Waves	2022.10—2023.1
银牌 private 榜单 45/936 • 引入大卷积核和自耦去偏置操作, 以捕捉引力波功率谱形状信息, 同时通过 CUDA-FFT 加速计算, 并采用随机噪声和信号注入实现在线生成数据, 防止模型拟合噪声。	

Kaggle: Novozymes Enzyme Stability Prediction	2022.10—2023.1
private 榜单 45/2482 • 基于蛋白质一维序列预测三维结构, 提取体素特征编码、蛋白质 pose 信息、溶剂可及表面积与能量分数, 集成这些特征以 XGBoost 预测单点突变蛋白质稳定性, 得到优于蛋白质稳定性预测网络 ThermoNet 的结果 (45/2482)。	

Kaggle: RSNA Screening Mammography Breast Cancer Detection	2023.1—2023.3
--	---------------

荣誉奖项

• Kaggle: G2Net Detecting Continuous Gravitational Waves 银牌 (Top 5%)	2023.01
• Kaggle Expert	2023.01
• 中国科学技术大学一等奖学金	2024.9
• 西安电子科技大学一等奖学金三次	2021.09 2022.09

掌握技能

- 编程语言: Python, C, CUDA
- 框架: Pytorch, Deepspeed, Keras
- 软件: PyCharm, Linux and shell, Git, Docker

¹ 下划线内容包含超链接。