

孙爱华

男 | 年龄: 22 岁 | 经验: 1 年 | 学历: 本科
求职意向: 深度学习 | 期望薪资: 30 - 50k | 期望城市: 杭州

🏠 <https://aeeeeep> 🔗 aeeeeep ✉ aeeeeep@proton.me ☎ +86 153 9689 7213

个人优势

- 熟悉大规模网络分布式训练框架源码及其二次开发, 如 DDP, deepspeed, Megatron 等
- 掌握 GPU 异构编程, 显存优化, 性能调优
- 维护个人开源项目 objwatch, 目前下载使用量超 5k
- 在个人服务器上管理运维博客: <https://aeeeeep.top>

工作经历

上海壁仞智能科技有限公司

分布式训练工程师 | 软件生态部

大模型训练算法 · PyTorch · Python · C++

2023.07 -- 至今

- 内容:** 基于壁仞生态, 针对 NLP/CV/音视频/多模态/推广搜等场景, 构建大规模的分布式机器学习系统; 负责解决业务交付流程中遇到的单机多卡, 多机多卡的精度、性能问题, 研究行业领先的超大规模分布式策略。
- 业绩:**
 - 分布式策略优化:** 如通过在 deepspeed 分区 tensor h2d 后异步 reorder, 并修改计算流程减少计算量, 节省单 step 通信同步等待耗时, 加速 6%; 消除 dataloader 时 broadcast, 单 step 加速 3%。在 megatron lora 微调特性研发中, 实现 tp 并行场景下的分布式融合算子, 单 step 加速 10%。
 - 双 die 芯片场景下的分布式优化器设计:** 在双 die 设计的芯片上, 研发具有特殊存储结构的优化器组件, 打通 12 项关键框架功能与算子节点, 并提出针对优化器状态底层存储结构的优化点, 整体降低显存利用率 30%, 利用自己开发的 objwatch 追踪库, 实现在一周跑通, 两周功能完善, 性能超预期的效益。
 - pytorch profiler 解析工具开发:** 利用 pandas, multiprocessing 等库研发解析工具, 实现在半小时内解析千卡规模下的 pytorch profile 百 GB 级别的性能数据, 输出自动分析的 bubble 耗时, 各算子带宽与真实耗时, 各模块耗时比率, 峰值显存等性能数据, 大大推动部门算子优化和显存优化效率。

项目经历

objwatch 开源项目

Owner

链接: <https://github.com/aeeeeep/objwatch>

2024.12 -- 至今

- 内容:** 为了解决在阅读和 debug 复杂的项目时, 加速理解遇到的多达十几层的嵌套调用, 与多进程场景下, 在单个进程上的调试往往会导致其他进程等待超时, 需要不断重复启动调试程序的问题, 开发了一个用于简化复杂项目调试和监控的 Python 工具库。利用 trace 模块, 通过实时追踪对象属性和方法调用, 加速深入了解代码库, 帮助识别问题、优化性能并提升代码质量。
- 业绩:**
 - 方便观察数千行级别的代码对 tensor 的处理逻辑, 加速了工作时框架相关工作 50% 的效率。
 - 利用自定义 wrapper 能够快速解释在工作中需要大量工程量排查的疑难杂症, 比如在不同网络 topo 的机器上, 框架的显存利用不一致的问题, 是由于 topo 的改变导致 cpu 与 gpu 的处理时序发生改变, 会有显存碎片堆积的现象, 在 cpu 端添加同步点可解决; 比如通过显存 wrapper 可以高效观察每个事件造成的显存变化, 辅助优化显存。

基于多模态特征的大黄蜂入侵物种防治算法

校科研项目

链接: <https://github.com/aeeeeep/BumblebeeSightings>

2023.03 -- 2023.05

- **内容:** 本项目提出了一种基于多模态特征的防治算法。通过结合图像识别、用户文本描述和上报地点等多模态信息, 实现对大黄蜂的防治与繁殖预测, 为后续的防治工作提供可靠的依据。
- **业绩:** 对于图像模态, 使用 Convnext 模型, 对收集到的大黄蜂图像进行了训练和分类。识别大黄蜂的存在与否。对于文本特征, 使用 Fasttext 对用户的文本描述进行情感分析和关键词提取, 再编码识别。最后使用熵权法研究不同特征对大黄蜂繁殖的重要程度, 进行加权判定, 在测试集上达到了 97% 的准确率。

电力生产环境员工安检系统

校科研项目

链接: <https://github.com/aeeeeep/ESDSalgo>

2022.12 -- 2023.03

- **内容:** 基于现场监控摄像头视频进行智能分析, 作业前,进行人员着装与绝缘防护判断, 对作业的配电柜与设备初始状态, 做时态信息和票面信息对比研判, 对错误的配电柜和错误初始状态及时告警。
- **业绩:** 使用 Yolo 目标检测网络进行实时着装检测, 使用 OpenCV 传统视觉方法对设备状态进行检测。并对算法使用 Tensorrt 量化, k8s 集群上的部署, 实现 50 fps (96.8 acc) 的检测速度。设备检测算法实现 500 fps (99.5 acc) 的检测速度。

竞赛经历

- 2022 微信大数据挑战赛 多模态短视频分类 全国三等奖
 - 参赛代码: <https://github.com/aeeeeep/2022WBDC-semi>
- 2022 第十八届挑战杯揭榜挂帅专项赛 RCS 数据目标识别技术研究 Rank 25
- kaggle - LLM Science Exam TOP 13%
- kaggle: RSNA 2022 Cervical Spine Fracture Detection TOP 24%
- kaggle: DFL - Bundesliga Data Shootout TOP 37%
- 2023 第四届全国人工智能大赛 AI+视觉特征编码赛道 Rank 30
- 2023 全球人工智能技术创新挑战赛 医学影像诊断报告生成 Rank 123
 - 参赛代码: <https://github.com/aeeeeep/2023GAIIC>
- 2022 MathorCup 高校数学建模挑战赛 全国二等奖
- 2022 五一数学建模竞赛 全国三等奖
- 2022 全国大学生新媒体大赛(摄影组)全国二等奖
- 2021 APMCM 亚太地区大学生数学建模竞赛 全国二等奖
 - 参赛代码: <https://aeeeeep.top/p/ae4fbfca>

教育经历

盐城师范学院·信息工程学院

本科

数字媒体技术

2020.09 -- 2024.06

- **毕业论文:**
 - **基于自动并行策略的轻量化深度学习框架研究:** 基于 jax 框架, 设计并实现 3d 并行的分布式训练框架, 基于 gpt2 验证, 总体性能达到不使用融合算子的 megatron 框架。
- **在校经历:**
 - 机器视觉实验室负责人 2021.06 -- 2023.12
 - 摄影协会会长 2020.10 -- 2021.10