简 历

基本信息

姓名：徐浩　　求职岗位：算法工程师、HPC
性别：男　　联系方式：18913-028-323
年龄：24岁　　邮箱：haoxu.live@outlook.com

教育经历

2024.09 – 2025.08　爱丁堡大学 (QS27)　**高性能计算与数据科学专业**　**硕士**　**英国爱丁堡**

主要课程：线程编程 (OpenMP)、高级消息传递编程 (MPI)、加速系统-原理与实践 (CUDA)、HPC架构、性能编程、大规模机器学习、数据管理基础、实用软件开发

2022.09 – 2024.06　萨塞克斯大学　计算机科学与人工智能专业　本科 2+2　**英国布莱顿**

2019.09 – 2021.06　苏州大学 (211)　**人工智能专业　本科** 2+2　**江苏苏州**

主要课程：数据库、机器学习基础、神经网络、计算机视觉、自然语言工程、程序分析

个人简介

- 211 本科学历，QS27 硕士学历，人工智能和高性能计算专业科班背景
- 本科阶段主要为机器学习，计算机视觉，自然语言工程等人工智能领域研究
- 硕士阶段主要为计算机结构，并行计算，GPU加速等高性能计算领域研究
- 校内助教与实习经历，锻炼了较强的解决问题和沟通能力
- 编程语言: Python, C/C++
- 技术：Linux, LaTeX, Git, MySQL, MongoDB, OpenMP, MPI, CUDA, SYCL, Hadoop, Spark, PyTorch
- 英国留学多年，英文可以作为工作语言

工作经历

2023.10 – 2023.12　英国萨塞克斯大学　自然语言工程专业助教　助教　英国布莱顿

- 向本科二年级学生讲授自然语言工程课程的基本原理
- 教授自然语言工程课程实验，带领学生完成代码问题集
- 帮助有困难的学生解答问题

2023.07 – 2023.09　英国萨塞克斯大学　初级助理研究员　全职　英国布莱顿

- 使用 Cellpose 对细胞显微照片进行分割
- 尝试通过 VAE 基于缺失通道的细胞显微照片生成完整的四通道图片
- 对细胞周期阶段进行分类

2021.09 – 2022.06　焦点科技有限公司　人工智能研发部　实习　江苏南京

- 使用 CNN 识别违规商品图片
- 通过 Yolo 对敏感图标进行检测
- 复现 Image Matting 来细化分割出的物品边缘，使分割出的商品图像更加清晰
- 训练 CLIP 模型参加阿里天池的多模态商品检索竞赛

2021.06 – 2022.09　　南京迈煜创信息技术有限公司　　数据分析实习生　　实习　　江苏南京

- 参与舆情分析系统的开发工作
- 使用 Scrapy 抓取新闻并预处理新闻数据
- 训练 BERT 实现情感分析和文本分类

项目经历

2025.03 – 2025.04　　Vison Transformer 训练加速　　大规模机器学习作业　　爱丁堡大学

项目描述: 在 Cirrus HPC 平台上对 ViT 训练做端到端性能剖析与优化

- 性能剖析：使用 DLProf、Nsight Systems 打点 CUDA/cuDNN 与 PyTorch 流程，定位 CPU/GPU 与内存瓶颈
- 代码优化：启用 AMP、cudnn.benchmark，并将损失累积从 CPU 搬到 GPU，显著提升内核执行效率
- 并行策略：对比 DP、DDP 和手动 Pipeline Parallel，在单节点多 GPU 与多进程场景下测试伸缩性
- 超参调优：测试 Batch Size (16/32/64) 与 num_workers (0/2/4/8)，最终选定 BS=64、num_workers=4
- 训练吞吐从单 GPU FP32 的 ~6.2 samples/s 提升到 ~36 samples/s（4×加速），且精度无损

2025.03　　脑模拟系统的并行优化　　并行设计模式作业　　爱丁堡大学

项目描述: 并行化串行脑模拟代码，基于事件驱动协调模式，使用 C/MPI 在多节点集群上实现分布式仿真

- 设计并实现可复用的通用事件驱动协调框架，机制（事件队列、调度）与策略（神经模型行为）分离
- 利用 MPI_Allgatherv 实现跨进程信号交换，本地队列处理 Update / Signal 事件
- 支持 single/multi handler 模式，动态伸缩 FIFO 队列以优化内存与顺序
- 在 Cirrus 集群上对 small/medium/large/massive 四种规模做弱 / 强扩展测试

2025.02　　渗流模型 CUDA 并行优化　　加速系统-原理与实践作业　　爱丁堡大学

项目描述: 在GPU上并行化二维渗流模型，将串行CPU版本迁移至CUDA，实现高效计算并分析性能

- 基于16×16线程块映射二维网格，优化全局内存访问，实现模型并行更新
- 使用 atomic 操作结合 host-device 同步，准确检测迭代终止，并引入 pinned memory 加速数据传输
- 对 512^2–16384^2 不同规模网格进行基准测试，NVIDIA V100 上最高达 52.8× 加速

2024.11 – 2024.12　　细胞自动机问题的二维分解　　消息传递编程作业　　爱丁堡大学

项目描述: 使用MPI对细胞自动机 C 语言串行程序进行二维分解以实现并行加速

- 为细胞自动机问题设计二维分解策略并使用MPI拓扑实现
- 手动管理数据的分发与收集
- 使用非阻塞通信实现进程间网格边界的数据交换
- 根据问题规模，进程数量对程序的可扩展性进行实验分析

2024.10 – 2025.04　　文献检索分析系统　　实用软件开发小组作业　　爱丁堡大学

项目描述: 基于 Python/Flask + SQLite，构建的一站式学术文献检索与分析平台，支持实时 arXiv 检索、本地缓存、论文情感打分、主题聚类及元数据统计可视化，满足研究人员对论文搜索与深度分析的需求

- 设计并实现 /search 与 /analysis 后端接口，支持按关键词/作者/标题检索及元数据统计
- 调用 arXiv API 并将结果持久化至 SQLite 数据库，兼顾实时性与离线可用性，减少 API 调用延迟
- 使用 NLTK VADER 对论文摘要进行整体情感打分，并在搜索结果中展示情感标签，提高信息筛选效率
- 采用 TF-IDF 提取摘要特征，基于 KMeans 聚类并通过 PCA 降维，在前端以散点图形式呈现论文主题分布
- 使用 HTML + Tailwind CSS + jQuery 配合 ECharts，开发交互式界面，实现检索列表、聚类图、情感雷达及元数据统计图的动态渲染
- 采用 Docker 容器化与 GitLab CI/CD 自动化部署，部署效率提升 80%，自动化测试覆盖率达 85%

2023.10 – 2024.06　　基于LLM的复习助理　　毕业设计　　萨塞克斯大学

项目描述: 通过LLM开发一个自动测试生成的网页应用，基于用户上传的书籍、笔记、上课音视频生成测验来帮助用户巩固知识

- 处理各种格式文件输入 LLM，包括使用 Whisper 模型转录音视频文件
- 研究 Map Reduce 与 Clustering 等处理LLM长文本输入的方式以生成帮助学生巩固知识的测验
- 通过 RAG 检索文本中相关信息对用户的错误答案基于反馈
- 处理 LLM 的输出，并通过 StreamLit 构建网页应用

2023.02 – 2023.05　　面部对齐与关键点检测系统　　计算机视觉作业　　萨塞克斯大学

项目描述: 设计、构建、测试和评估一个用于执行面部对齐的系统，即在图像中定位面部44个关键点。
利用完整的数据集和一个仅包含 5 个关键点坐标的子集，设计和实施一个可以利用这些额外训练数据的程序，并评估整体任务的性能提升

- 构建级联模型，在第一阶段预测 5 个关键点坐标后修正人脸角度，在第二阶段预测所有 44 个关键点
- 对数据集进行大量的数据增强
- 实验了大量的试验参数：模型选择，数据增强，MSE Loss 和 Wing Loss，学习率，Batch Size 等
- 对实验结果和失败案例进行充分的定量分析和定性分析，最终 NME 降至 5 %