

HPC AI TECH
潞晨科技

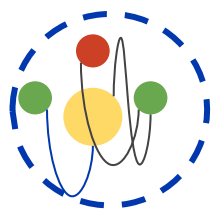
打造下一代AI基础设施
解放AI生产力

<http://www.luchentech.com/>



Colossal-AI

面向大模型时代的通用深度学习系统



核心技术

高效多维并行

异构内存管理

- 高效快速部署AI大模型训练和推理
- 显著降低训练和推理成本
- 减少学习和部署的人力成本

自适应任务
调度

大规模优化库



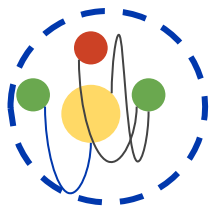
<https://github.com/hpcaitech/ColossalAI>





Colossal-AI

面向大模型时代的通用深度学习系统



核心技术

◆ 多维自适应切分，最大化并行能力，提升计算效率

Colossal-AI在兼容现有并行方案的基础上，独创序列并行，并首先把张量并行从1维扩展到2维、2.5维及3维。易用性极佳，仅需少量修改即可快速部署各种复杂并行策略。

◆ 高效异构内存管理，降低硬件成本

AI大模型训练需要大量计算资源，Colossal-AI通过异构内存管理，利用更低廉的CPU内存在硬盘空间，可节省GPU等高端硬件需求50%以上。

◆ ONES调度系统，实现动态自适应优化，智能把控企业需求

集群调度系统是并行计算的必要基础设施，Colossal-AI提供更灵活的优先级调度，缩短任务等候时间，减少每个GPU和线程的闲置时间，使算力效率最大化，提高训练速度。



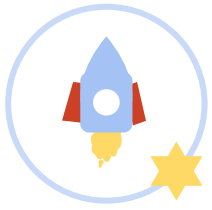
<https://github.com/hpcaitech/ColossalAI>



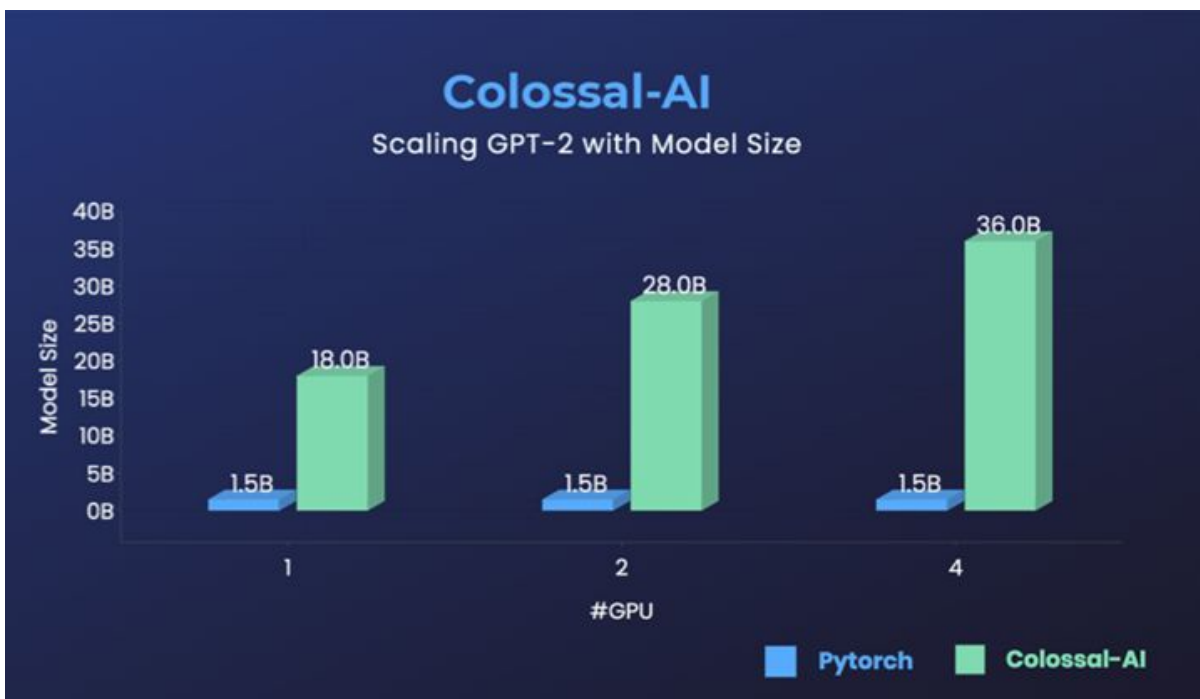


Colossal-AI

面向大模型时代的通用深度学习系统

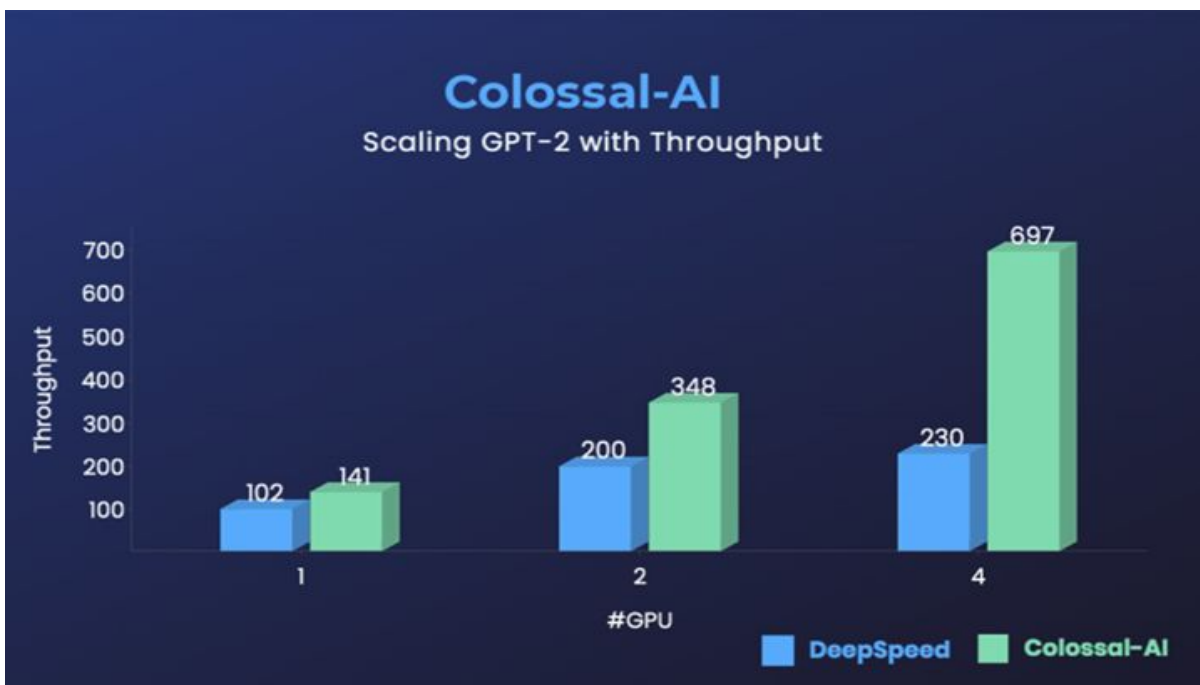


技术优势



GPT-2

- 用相同硬件训练24倍大的模型



- 超3倍的吞吐量



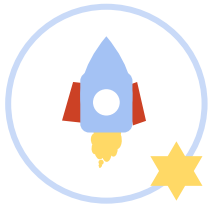
<https://github.com/hpcaitech/ColossalAI>



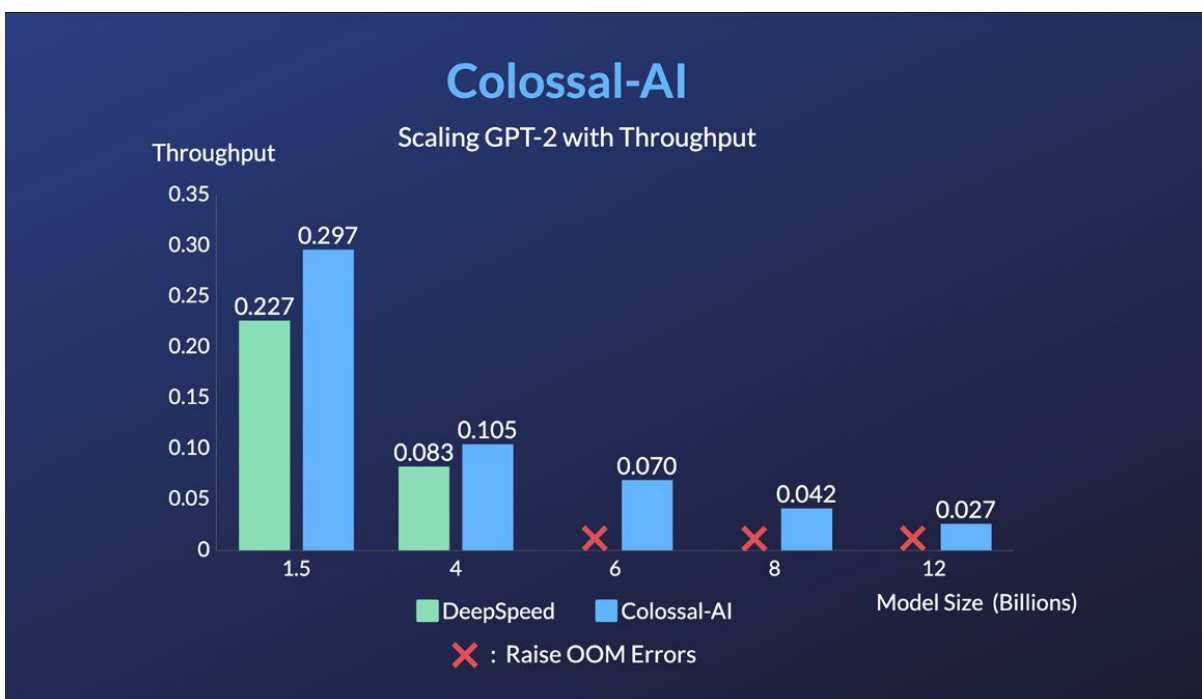
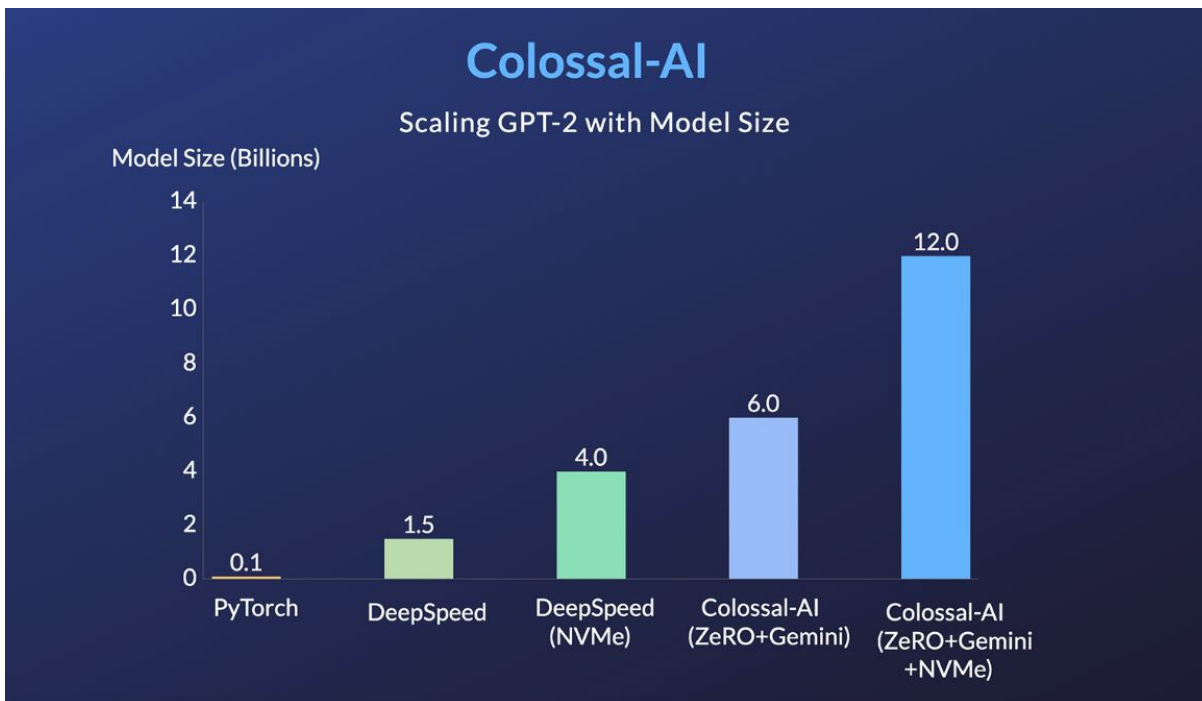


Colossal-AI

面向大模型时代的通用深度学习系统



技术优势



GPT-2

- 用相同的硬件训练120倍大的模型 (RTX 3080)



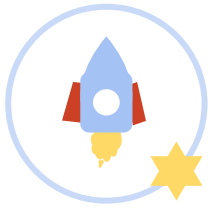
<https://github.com/hpcaitech/ColossalAI>



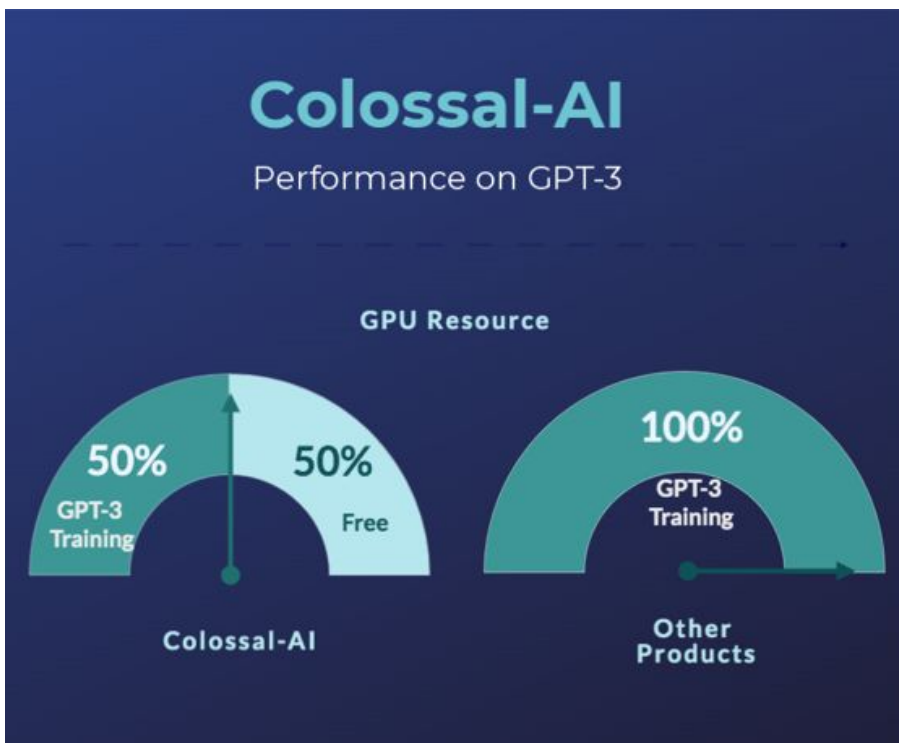


Colossal-AI

面向大模型时代的通用深度学习系统

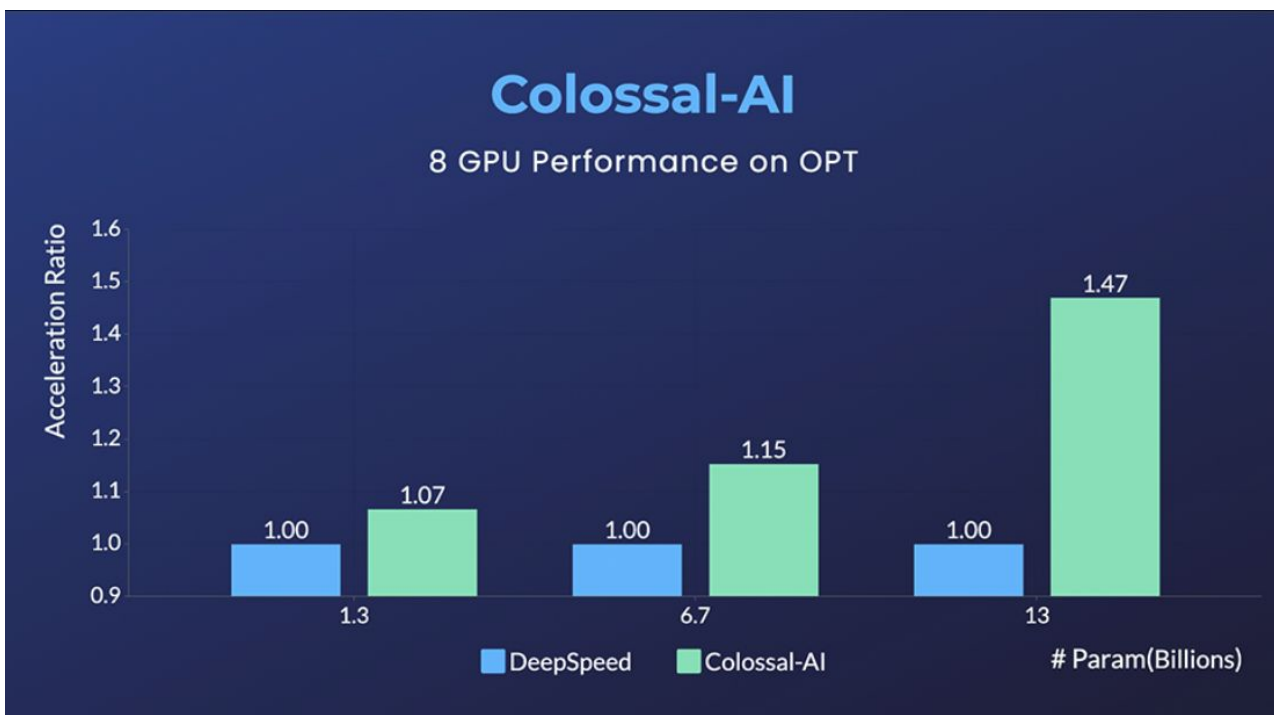


技术优势



- 半数 GPU 训练 GPT-3

GPT-3



- 加速47%，仅需几行代码低成本微调 OPT

OPT



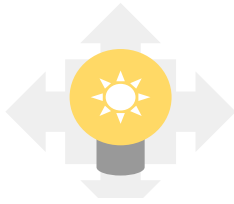
<https://github.com/hpcaitech/ColossalAI>





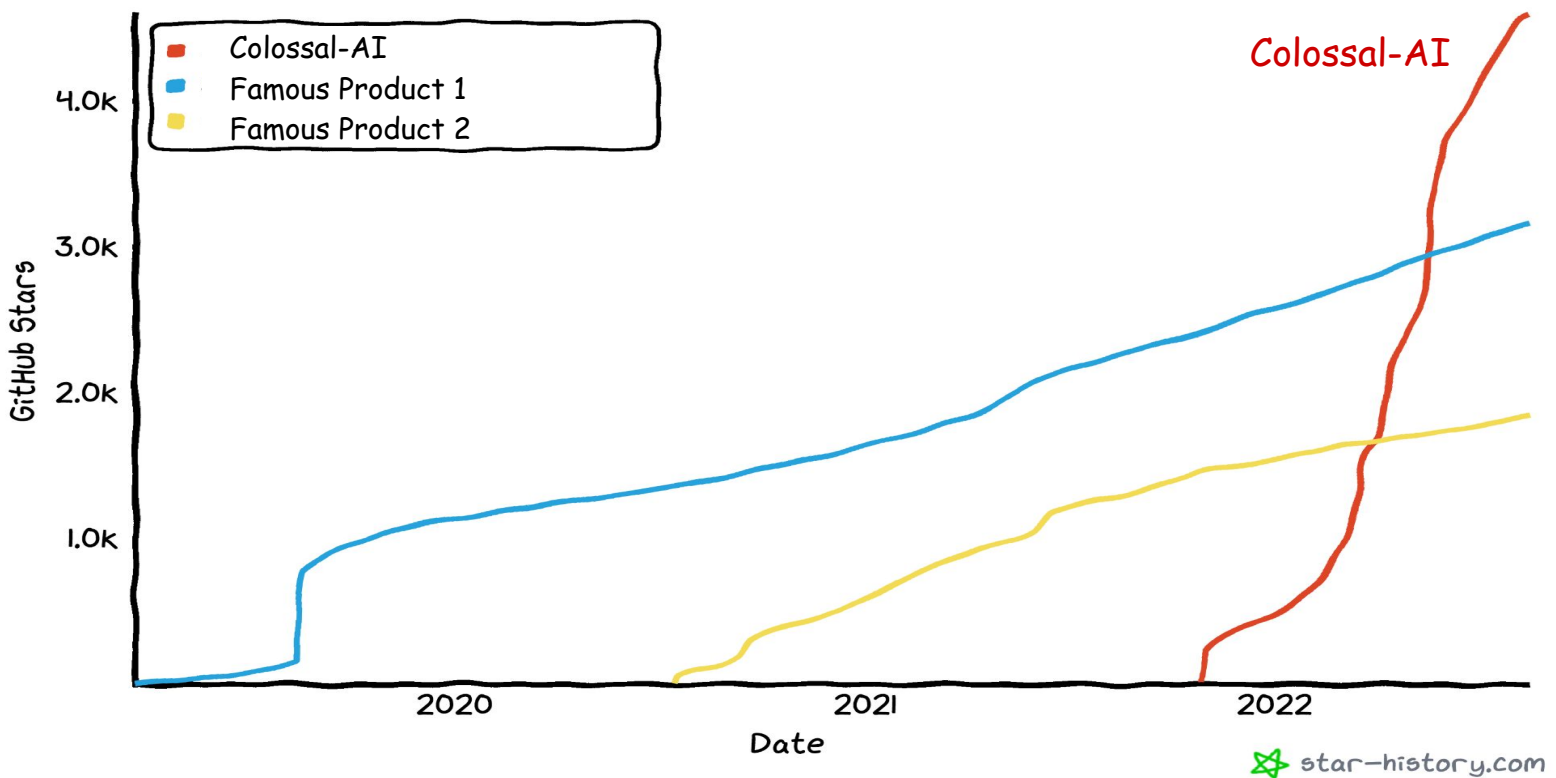
Colossal-AI

面向大模型时代的通用深度学习系统



开源社区

Star History



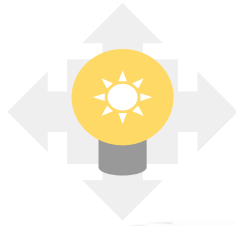
<https://github.com/hpcaitech/ColossalAI>





Colossal-AI

面向大模型时代的通用深度学习系统

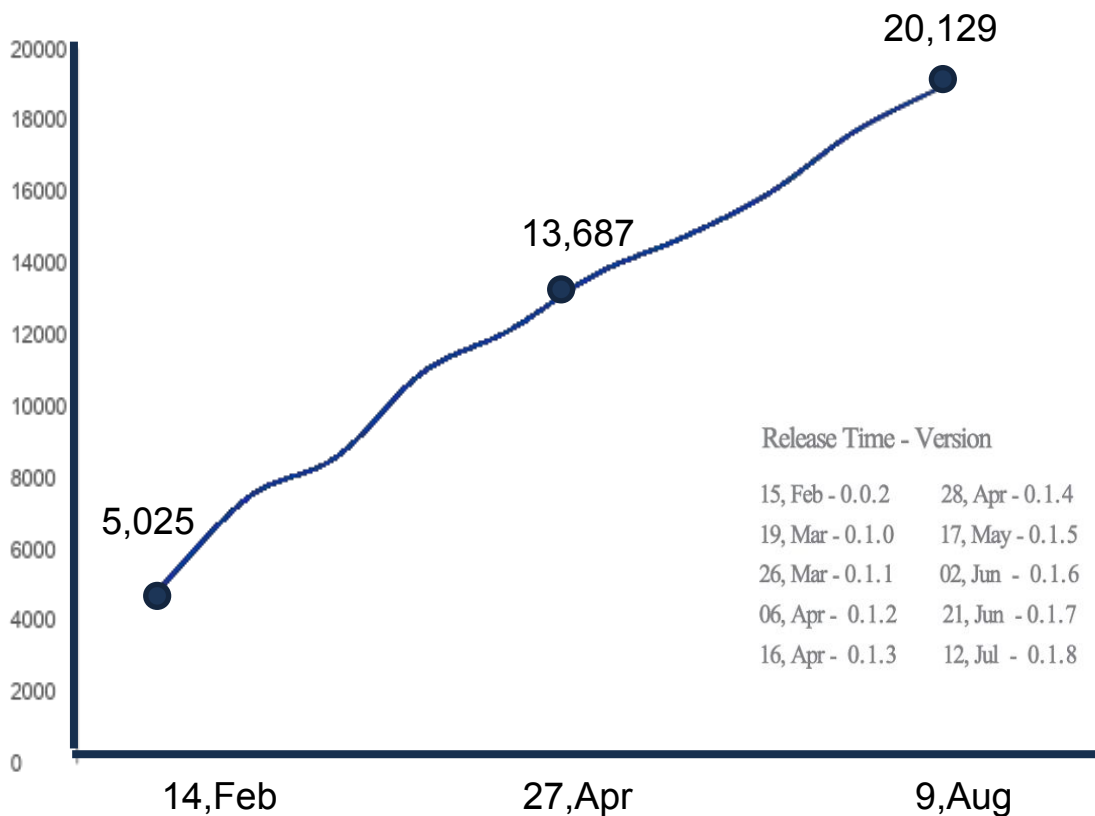


开源社区

开放的开源软件生态



开源至今下载量攀升明显



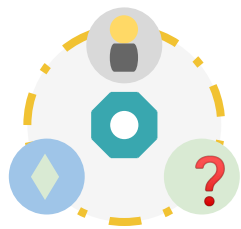
<https://github.com/hpcaitech/ColossalAI>





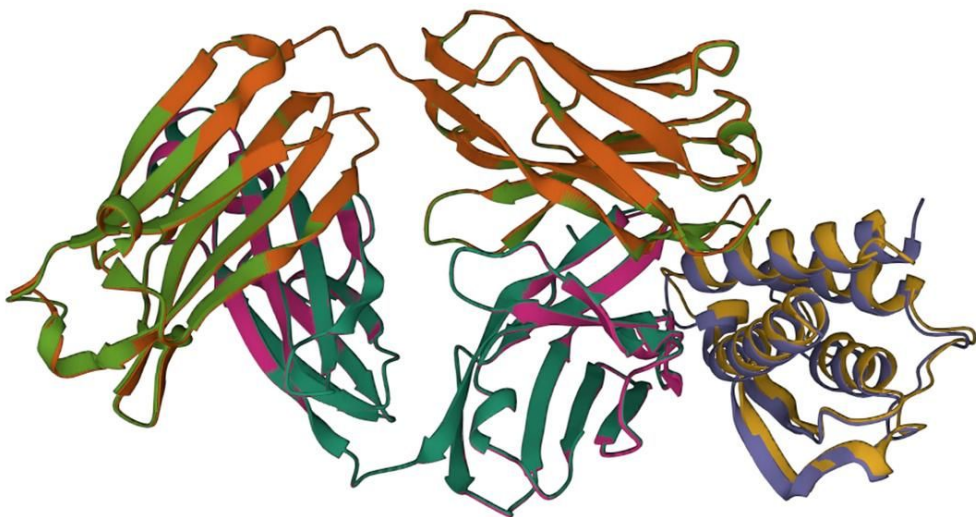
ColossalAI

面向大模型时代的通用深度学习系统



解决方案

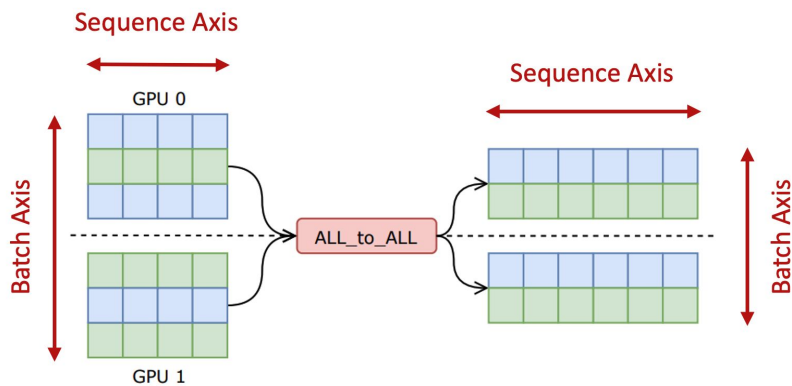
xTrimoMultimer: 11倍加速蛋白质单体与复合物结构预测



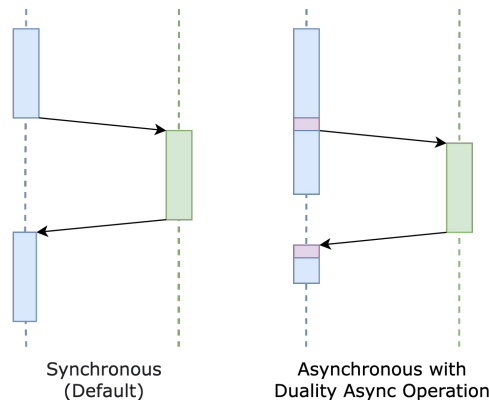
Prediction Result of AlphaFold2-Multimer and xTrimo-Multimer (Protein 3G6D)

FastFold: 整体训练时间从 11 天减少到 67 小时, 成本显著降低

并行 Evoformer



动态轴并行
一种新颖的高效模型并行方法

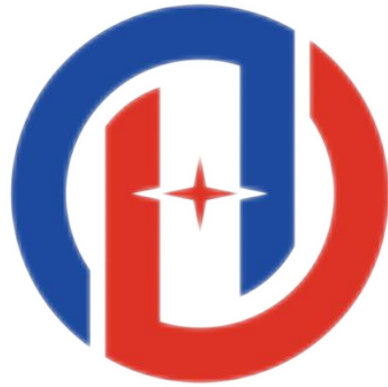


对偶异步算子
一个用于通信优化的PyTorch扩展



<https://github.com/hpcaitech/ColossalAI>





HPC AI TECH

瀚晨科技

Email: contact@luchentech.com

Website: www.luchentech.com

项目地址



加入用户群

