

吴梦林

上海市浦东新区达尔文路88号
云从科技 200000

电话: +86-177xxxx4729
Email: menglinwoo@163.com

github: github.com/MIWoo
出生年月: 1990.XX

教育背景

哈尔滨工业大学(HIT)	机械工程硕士	09/2013 至07/2015
哈尔滨工业大学(HIT)	机械设计制造及其自动化学士	09/2009 至08/2013

获奖荣誉

云从科技优秀团队奖	2018
Intel STO 部门奖	2016
Intel Design Win (Pikazo)	2016
中兴通讯公司2015 级优秀新员工	2015
“飞航杯”全国飞行器设计大赛黑龙江省赛区“巽翼杯”飞行器设计大赛第四名	2014
第八届全国周培源力学竞赛获全国三等奖(黑龙江省第七名)	2011
国家励志奖学金 (前3%)	2010-2011

工作和研究

项目: TTS (语音合成) 开发、优化与部署	云从科技	技术组长	06/2018 至今
项目职责: 熟知行业状况, 分析发展趋势, 拆解项目任务, 量化工作指标, 快速验证想法, 带教培养新人, 填坑推动进度。			

- 复现并改良了高质量和高性能声码器Autoregressive/Parallel WaveNet, 并对parallel WaveNet模型结构进行精简优化, PyTorch版本性能提升约4倍, 达到4MHz的合成速度 (200-300X实时速度于16KHz采样率的音频)。
- 复现和改进高质量的声学模型Tacotron2, 与Parallel WaveNet声码器一起使用可达到业内一流水准。同时增添了说话者embedding vector, 模型实验结果证实单模型可以控制输出不同的说话者音色的合成效果。采取Forward Attention解决了长语句合成时跳帧问题。并提出了解决合成时结尾重音, 某些字词效果欠佳的等trick。
- 掌握TTS的从研发到部署的技术路线。

项目: PyTorch深度学习框架的开发与优化	Intel	开发人员	09/2017至06/2018
-------------------------	-------	------	-----------------

- PyTorch **Contributor**, element-wise优化(PR#2764), Conv3D优化(PR#4824), OpenMP开销阈值调优(PR#5584), Bernoulli随机数生成优化(PR#7171), 贡献代码增量累计约**3.0K**行, 删除累计约**1.5K**行。其中element-wise优化策略仍是CPU上优化上的重要策略。定位并指出因错误使用BLAS接口导致性能下降的bug(Issue#5047 和Issue#3525), 且修改意见均被官方采纳。
- 以较低额外计算开销实现element-wise操作的OpenMP并行化, 并对OpenMP开销阈值调优做了详细的测试和分析, 在某些典型CPU上得到6X-30X的加速。
- Conv3D优化, 修改计算策略增大了每个核上的chunk size减少循环, 在某case的CNN部分在某典型的CPU上得到高达**8.43X**的加速。该思路可推广在Conv2D、Conv3D和及其变种Conv2D、Conv3D和及其变种。
- Intel-PyTorch icc分支主要开发人员。修改工程CMake文件, 使其支持Intel编译器编译优化和持AVX512的primitive等优化。

项目: Intel-Torch深度学习框架的开发、优化和应用	Intel	参与者	05/2016至06/2018
--------------------------------	-------	-----	-----------------

- 性能优化: 集成MKLML深度学习加速库; 从多线程OpenMP, 向量化SIMD, 访存等多方面优化element-wise操作; 支持Intel编译器编译优化; 支持AVX512源语优化; ImageNet典型topo性能在Intel-Torch得到了**6.4X到14X**的加速。
- 框架功能: 支持MKLML加速库特定layout模型的读取和保存, 实现MKLML加速库内部内存的管理。实现官方Torch模型到Intel-Torch模型的自动转换。编写不同层次的测试脚本, 完成了自下而上的测试架构。
- 模型训练: 实现ImageNet训练时数据读取和模型训练的生产者消费者模型, 读取数据开销降低了**75%**, 并提出了一些在Intel Xeon & Xeon Phi上提高训练速度的技巧。
- 成功案例: Intel-Torch的优化使得Pikazo项目性能提高**7X**, 云计算成本降低**2/3**, 争取客户从原先的NVidia GPUs切换到Intel Xeon & Xeon Phi, 该项目作为**Intel Design win**在SC16, US AI Day和PRC AI Day上多次展示。
- 训练案例: 多节点OpenNMT(<http://opennmt.net>)训练, 修改默认的optimizer(SGD)为adam, 采用大batch size缩短训练时间。首个epoch进行warmup, 之后小幅度提升learning rate, 使得原先近40个epoch的训练精度可在8个epoch内达到。

项目: 交换机二层协议的开发与维护	中兴通讯公司	参与者	07/2015 至05/2016
-------------------	--------	-----	------------------

- ZTE版EAPS协议开发与维护, 新增了环路状态变化时50ms以内(EAPS要求50ms到100ms)的网络链路的自动快速切换。
- IPTV协议的开发与维护, 支持全局, VLAN, 用户密码等多种IPTV用户登录与管理模式。

技能和兴趣

- 英语: CET-4: 555 CET-6: 477 CET-口语考试: C+
- 编程语言: C, C++, Lua, Python, CMake, Shell, Matlab
- 其他软件: Solidworks, proE, AutoCAD, CAXA
- 自我评价: 对新技术保持有好奇心, 自学能力强, 耐心坚韧