

个人信息

- 章川/男/1992
- 硕士：2014年~2017年 华中科技大学 图像识别与人工智能研究所
- 本科：2010年~2014年 华中科技大学 自动化专业
- 工作年限：7.5年
- Github: <https://github.com/ZhangChuann>
- 技术博客: https://blog.csdn.net/neo_qiye
- 期望城市：深圳



联系方式

- 手机：13342966865
- Email: zhangchuanpro@outlook.com
- QQ/微信号：1003851855
- 通讯地址：广东省深圳市宝安区西乡街道

工作经历

腾讯公司（2020年10月 ~ 至今）

沧海V2 AI子系统算子开发和网络优化

(2022年6月~至今)

- 沧海V2是腾讯第二代自研VPU芯片，其中的AI子系统为全自研NPU IP，参与架构设计，关键指令集设计，微架构设计，编程模型设计
- 构建了一套算法到硬件自动性能分析的软件框架，实现架构和微架构的快速寻优探索；同时负责算子开发和网络优化工作，包括底层算子优化，定制融合优化和结合图层的深度融合优化

AI芯片选型测试和性能分析

(2023年6月~至今)

在CSIG-星星海实验室兼职，参与多家AI芯片的对接，测试，评估和数据分析，包括：Intel Gaudi系列，昇腾系列，昆仑芯系列，AMD MI300X系列，NVIDIA H20等；参与构建理论性能评估平台，指导服务器整体架构设计。

AI处理器高性能算子库开发

(2022年1月~至今)

作为紫霄V2自定义高性能AOT算子库和沧海V2 AI子系统高性能算子库的开发负责人，设计整体代码开发思路，开发框架和任务管理，并负责关键算子的方案设计和开发：主要有矩阵类算子Conv, MatMul方案设计，模版拆分和Auto-Tiling设计；Softmax, Cast, Pooling, Resize等关键算子的方案设计和开发。

紫霄V1 AI芯片算子开发

(2020年10月~2021年04月)

熟悉紫霄芯片架构和SIP Core架构，并基于SIP架构编写和优化OCR网络中的算子，包括：图像预处理：bilinear resize, normalization; 图像类型转换：Rgb2Gray; NN网络中Feature Map的上采样：Upsample Layer; NN网络中的global average pooling。

华为公司（2017年07月~2020年10月）

Davinci Mobile AICore 架构设计与验证

(2019年09月~2020年03月)

结合手机应用场景和Neural Network自身特点，负责设计专业的NN网络卷积后处理模块，以支撑下一代NPU处理器的多任务并行处理架构的目标。在该项目中负责手机端的专用网络需求分析，算法性能分析，主持NN后处理模块的架构设计和指令定义，同时负责相应的网络算子编写和性能验证，支撑未来手机NPU应用需求。

Davinci Cloud 网络性能优化支撑

(2019年06月~2020年02月)

作为芯片架构团队成员，支撑Cloud解决方案团队基于昇腾910芯片的网络性能优化专项，促进商业目标达成，其中ResNet50网络训练一分钟内完成收敛，Bert网络性能达到GPU V100性能的2倍以上。在ResNet50优化专项中，主要负责AICore侧功能和性能问题定位，多核方案拆分优化，卷积Tiling方案优化等工作；在Bert网络优化专项中，主要负责功能和性能问题定位，矢量算子方案设计和性能优化，总体性能提升6倍+。

DaVinci Cloud AICore 网络训练性能验证

(2018年10月~2019年05月)

我在此项目中负责GNMT (Google's Neural Machine Translation System) 网络在DaVinci Cloud芯片上的网络训练性能验证。负责网络中的核心算子的前向，反向性能理论分析，算子CCE编写，性能回归，探索架构的缺点，为架构性能达标贡献力量。

Davinci 芯片神经网络应用精度问题攻关

(2018年03月~2018年05月)

在第一代产品问世时，神经网络应用有部分网络无法与精度标杆完全匹配，组织专项定位攻关，最终实现精度问题收敛，网络性能达标。在该项目中，主要负责分析基本单指令的精度特征及其精度提升方案，后续持续支撑网络算子开发团队精度问题分析和解决方案制定。

DaVinci Edge AICore 架构性能验证

(2017年11月~2018年02月)

我在这个项目中主要负责基于DaVinci指令集完成目标检测中roipooling, roialign算子，以及轻量级网络算子：Shuffle Net算子的编写；由于这是第一版AICore，指令集和编译器等基础设施都处于快速迭代阶段，算子性能验证只能采用汇编编写的形式来完成，初期依照cuda和cpu的实现方式，算子性能严重低于预期，后期探索适用于架构特点的程序策略，并对架构提出有效建议，最终使得roi类算子有近10倍的性能提升，有效的支撑了第一代AICore在目标检测领域的性能目标。

DaVinci AICore Model VALU模块

(2017年07月~2017年12月)

我在这个项目中，主要负责float16算子的基本实现，边界条件判断等，作为RTL实现的reference model，用于功能bit级校准。由于float16算子的表示范围较小，很容易越界，为了适用于通用的AI计算，DaVinci的float16并没有完全按照IEEE的标准，存在很多不同的边界条件的检测；在这个工作的过程中，通过快速迭代，通用化代码，使得该模块快速稳定；同时，在这一过程中，也对计算机体系结构有了一些新的认识，了解了代码版本管理，是真正进入实际项目的第一课。

工作荣誉

- 腾讯公司2022 H2 Outstanding Contributor
- 腾讯公司2021 H2 Outstanding Contributor
- 腾讯部门三次月度之星奖励
- 在Davinci微架构验证，算子性能优化，算子精度验证，TIK编程方法探索, 网络训练性能分析获得五次部门芯星奖奖励
- 华为公司图灵核2019年度明日之星
- 华为公司图灵核2018年度明日之星

校园经历

2015.10-2016.12 红外图像气动光学效应校正、目标识别方法研究

该项目主要是研究在高速飞行条件下的图像增强算法，我在这个项目主要负责图像热辐射校正算法，去模糊算法的研究和软件系统的实现。

2014.10-2015.10 红外成像制导技术（DSP+FPGA+ASICs嵌入式目标识别系统）

该项目主要是研究自动目标识别实时处理技术，我在这个项目中主要负责两个部分：上位机模拟发图软件编写和嵌入式系统DSP图像处理及识别算法的实现，目标识别处理速度50fps+。

技能清单

- C/C++: 熟悉基本的C/C++开发，在华为和腾讯工作期间和研究生期间都有C/C++开发经历；
- linux: 熟悉常用的linux命令；在工作中有使用远程ubuntu/centos计算云的工作经历；
- 深度学习: 了解经典的网络；工作中对常用网络有一些了解；
- 图像处理: 了解基本的Low level的图像处理技术，研究生阶段有图像增强学习工作经历；
- Python: 了解Python基本语法，可以编写基本的Python应用和脚本；
- 底层算子性能优化: 在工作期间，有基于多种自研架构指令集的汇编和intrinsic C代码编写经历；
- 英语: 通过英语四、六级，会基本的听，读，写；
- 版本管理: github 基本命令

自评 & 致谢

一个普通的有志青年，热爱计算机&AI技术，希望通过自己的努力可以为这个世界带来一些便利，一些快乐，一些惊喜,一些美好；相信人工智能技术，云计算技术,自动驾驶技术可以让我们的生活变得更美好，希望在将来的工作中能够从事相关工作。

感谢您花时间阅读我的简历，期待能有机会和您共事。