杜根 **Du Gen**

联系方式 (+86)186-4285-6913

电子邮件 ddd1695963186@gmail.com

求职意向 **AI编译器研发工程师**

工作经验

理想汽车有限公司 **算子编译工程师** **2023.04 - 至今**

- 参与自研NPU芯片从设计到成功流片及bring up全流程，**负责关键算子开发与性能优化**
- **构建基于MLIR的算子编译框架**，实现算子融合、拆分与并行优化，完成高层IR到硬件特定IR的转换流程
- 设计卷积类和采样类算子的编译优化方案，降低硬件感知门槛，提升团队研发效率

项目经历

算子编译框架开发 | 2024.05 - 至今

采样类算子编译优化 **项目负责人** **2024.12 - 2025.02**

- **实现GridSample与Mul、ReduceSum算子的融合优化方案**
- 抽象**GridSample、Resize、Rotate**等算子的计算流程，**将采样类算子拆分为Sample2D+DepthwiseConv两阶段计算模式**，通过配置组合实现多种图像处理功能

Conv2d算子编译优化 **项目负责人** **2024.05 - 2024.11**

- 利用MLIR生态**构建多层嵌套scf.for与scf.forall结构，精确映射NPU硬件行为**，完整表达卷积计算模式
- **通过Transform dialect实现硬件无关的算子表达**，为分块优化、计算重排和bank conflict消除提供统一框架

高性能算子开发与优化 | 2023.04 - 2024.04

采样类算子开发与优化 **项目负责人** **2023.11 - 2024.04**

- 实现支持**BEV、UniAD**等网络的**GridSample、Resize、Rotate**算子，支持bilinear和nearest插值模式
- **设计Spatial切分策略**，通过cross-tile数据访问**消除冗余的global_transpose操作**，显著提升整网性能
- 协助优化IPU模块精度(fp16→fp24→fp32)，**解决normalize过程中的精度损失**，实现与PyTorch bitwise对齐

卷积类算子开发与优化 **项目负责人** **2023.06 - 2023.10**

- 基于自研NPU架构设计并实现**Conv2d**算子，在**BEV、FPN、UniAD**等模型中实现卷积算子**功能100%覆盖**
- **优化ConvTranspose2d实现方案**，将传统input swelling方法改进为weight pattern拆分策略，在2倍上采样场景中**降低75%内存占用并提升50%计算效率**
- 实现Conv算子**WrapBack优化策略**，重排计算顺序构建细粒度pipeline，解决RegNet网络边缘数据依赖瓶颈

专业技能

- **编程语言:** 熟悉C++，具备复杂算法设计与实现能力
- **AI编译:** 熟悉MLIR框架及**核心Dialect(Linalg、Affine、SCF、Transform)**，具备算子融合、拆分、并行优化及Pass开发经验
- **算子开发:** 深入理解pytorch各类算子原理，**具备NPU平台算子设计、性能调优及精度优化经验**
- **开发工具:** 熟练使用**Git、Vim、GDB**等工具，具备Linux环境高效协作开发调试能力

教育经历

南京理工大学	机械工程	硕士	Rank:5%	2020.09 - 2023.04
大连交通大学	车辆工程	学士	Rank:10%	2016.09 - 2020.06

个人评价

- **技术驱动:** 热衷技术挑战，具备快速学习能力和技术创新思维
- **问题解决:** 善于分析复杂问题，设计高效解决方案
- **团队协作:** 具备良好的沟通能力和团队合作精神
- **职业规划:** 致力于AI编译领域的技术深耕与创新