自由探索计划“天目启航”专项项目选题征集表

|  |  |  |  |
| --- | --- | --- | --- |
| **教师姓名** | **韩丽霞** | **学 院** | **集成电路学院** |
| **职 称** | **副研究员** | **联系方式** | **18810913961** |
| **邮 箱** | **lixiahan@nuaa.edu.cn** | **研究方向** | **存算一体AI加速芯片** |
| **项目名称** | **面向存算一体AI加速芯片的高效部署算法研究** | | |
| **项 目 简 介（200字左右）** | **随着人工智能技术的快速发展，AI模型规模不断扩大，传统冯诺依曼架构下的计算模式难以满足日益增长的计算需求。面向存算一体AI加速芯片的高效部署算法研究旨在突破传统架构限制，探索基于存内计算范式的创新解决方案。本项目以新型存算一体架构为基础，重点研究深度学习模型在片上存储与计算资源的协同优化方法。**  **本研究将围绕三个核心方向展开：首先，设计面向存算一体架构的模型压缩与量化算法，包括基于权重共享的模型压缩方法和混合精度量化策略，在保证模型精度的前提下实现高效存储和计算；其次，提出适应存算一体特性的算子映射与调度策略，通过分析存算一体架构下典型算子的计算特征与访存模式，设计基于数据复用的算子分解与融合方法，优化计算流水与数据移动；最后，开发支持异构计算单元协同的部署框架，构建支持异构计算单元的中间表示，实现算子到硬件的灵活映射，同时提出存算一体架构专用的编译优化方法，实现端到端的模型优化与加速。**  **本项目的研究成果将为AI模型在存算一体架构上的高效部署提供理论与技术支撑，有望大幅提升AI芯片的算力与能效比，对推动存算一体架构在边缘计算、自动驾驶等场景下的快速应用具有重要意义。** | | |
| **人 员 技 术 需 求** | **主要职责、任务** | **需求人数** | **专业及技能要求** |
| **AI算法训练** | **2** | **高等数学，Python**  **感兴趣，学有余力** |
| **算子分解融合** | **2** | **高等数学，线性代数**  **感兴趣，学有余力** |
| **算法芯片映射** | **1** | **高等数学，线性代数**  **感兴趣，学有余力** |
|  |  |  |
|  |  |  |
| **备 注** | **可根据个人兴趣和项目进展，调整项目细节和任务划分** | | |