Cambricon-F: Machine Learning Computers with Fractal von Neumann Architecture

Corresponding author

Yunji Chen, SKL of Computer Architecture

Keywords

Machine learning hardware, programming productivity, model support

ML通用架构，流片验证

Summary

*Challenge*

With the fast development in silicon technology, **programming productivity**, including programming itself and software stack development, becomes the vital reason instead of performance and power efficiency that hinders the application of machine learning computer.

如何支持各种机器学习算法

*Contribution*

1. Propose a series of **homogeneous**, sequential, multi-layer, layer-similar, machine learning computers with the same **ISA.** (架构设计)
2. Implement two instances at different scales. (流片验证)

*Result*

Compared to GPU based machines (DGX-1 and 1080Ti), Cambricon-F instances achieve 2.82x, 5.14x better performance, 8.37x, 11.39x better efficiency on average, with 74.5%, 93.8% smaller area costs, respectively