基于 RISC-V 的卷积神经网络处理器设计与实现

傅思扬，陈　华，郁发新
（浙江大学 航空航天学院，浙江 杭州 310058）

摘　要：针对卷积神经网络对于运算资源需求的不断增长，和传统的硬件卷积加速方案在功耗、面积敏感的边缘计算领域难以应用的问题，设计并实现了一个低功耗嵌入式卷积神经网络加速处理器.目标处理器基于 RISC-V 指令集架构，内核扩展 4 条自定义神经网络指令，并在硬件层面实现加速处理.该卷积神经网络处理器最大程度的复用了原 RISC-V 的数据通路和功能模块，减小了额外的功耗和芯片面积等资源开销.目标处理器通过 RISC-V 官方标准测试集验证，并对 MNIST 手写数据集进行识别测试，正确率达到 97.23%.在 TSMC 40nm 标准数字工艺下，目标处理器面积仅为 0.34 mm2,，动态功耗仅为 11.1 μw/MHz，与同期处理器相比，面积和功耗方面均具有一定优势.

关键词：处理器；卷积神经网络；定制指令集；RISC-V

Design and implementation of convolutional neural network processor based on RISC-V instruction set

FU Si-yang, CHEN Hua, YU Fa-xin
(college of Aeronautics and Astronautics, Zhejiang University, Hangzhou, 310058, China)

Abstract：A low power embedded convolutional neural network acceleration processor is designed and implemented in response to the growing demand for computational resources in convolutional neural networks and the difficulty of applying traditional hardware convolution acceleration schemes in power- and area-sensitive edge computation applications. The target processor based on the RISC-V instruction set architecture, extend four custom neural network instructions, and accelerates the processing on the hardware architecture. The convolutional neural network processor maximizes the reuse of the original RISC-V data path and functional modules, reducing additional resourceoverhead. The target processor is verified by the RISC-V official standard test set, and the MNIST handwritten data set is identified and tested, with a correct rate of 97.23%. The target processor occupies 0.34mm2 area and consume 11.1μw/MHz dynamic power using TSMC 40nm technology. Compared with the correlation processor, it has certain advantages in area and power consumption.

Key words：processor；convolutional neural network；custom instruction set；RISC-V

作者简介：
傅思扬　男，(1995-)，硕士研究生.研究方向为数字集成电路设计.
陈　华(通讯作者)　男，(1986-)，副研究员.研究方向为耐辐照集成电路设计．E-mail：chenhua@zju.edu.cn