基于高层次综合的卷积神经网络设计与优化方法研究

彭鑫磊 [1]，余 乐 [1,2]

（1 北京工商大学 计算机与信息工程学院，北京 100048；

2 食品安全大数据技术北京市重点实验室，北京 100048）

摘　要：本文基于 FPGA 高层次综合的设计方法学，在 ZYNQ-7020 上实现了一个卷积神经网络加速器.采用循环展开和并行流水的设计方法对卷积核运算进行优化，均衡了所占用逻辑资源及运算效率，从而实现加速器的最优性能.通过 MINST 数据集在 100MHz 的工作频率下对加速器进行性能测试，结果表明：对单张图片，该加速器相对于通用平台 ARM A9 可实现 3.77 倍加速，而对 1000 张图片的流式处理可实现高达 6.14 倍加速.

关键词：卷积神经网络；嵌入式；ZYNQ-7020；加速器

HLS-based design and optimization methodology for

convolutional neural network

PENG Xin-Lei [1], YU Le [1，2]

（1 School of Computer and Information Engineering, Beijing Technology and Business University,

Beijing 100048, China;2 Beijing Key Laboratory of Big Data Technology for Food Safety, Beijing 100048, China）

Abstract：Based on High Level Synthesis (HLS) design methodology of FPGA, this paper implements a convolutional neural network accelerator on ZYNQ-7020. The design method of cyclic unroll and pipelinling is used to optimize the convolution kernel operation, and the occupied logic resources and operation efficiency are balanced to achieve the optimal performance of the accelerator. The performance of the accelerator is tested by the MINST dataset at 100MHz working frequency. The results show that: the accelerator can achieve 3.77 times acceleration compared to the general platform ARM A9for a single picture, and the streaming processing of thousands of pictures can achieve up to 6.14 times acceleration.

Key words：convolutional neural networks；embedded；ZYNQ-7020；accelerator

作者简介：

彭鑫磊 男，（1993-），硕士研究生.研究方向为数字集成电路设计、嵌入式开发.

余 乐（通讯作者） 男，（1983-），博士，讲师.研究方向为计算视觉与类脑芯片.E-mail:ladd_u@163.com