****

工程硕士学位论文

**基于FPGA的卷积神经网络应用研究**

|  |  |
| --- | --- |
| 作者姓名 | 王羽 |
| 工程领域 | 电子与通信工程 |
| 校内指导教师 | 徐向民 |
| 校外指导教师 | 李辉宪 |
| 所在学院 | 电子与信息学院 |
| 论文提交日期 | 2015年12月28日 |

**FPGA based convolutional Neural Network Application Research**

A Dissertation Submitted for the Degree of Master

**Candidate：Yu Wang**

**Supervisor：Prof. XiangminXu**

SouthChinaUniversity of Technology

Guangzhou, China

**分类号：学校代号：10561**

**学号：**

华南理工大学硕士学位论文

**基于FPGA的卷积神经网络应用研究**

作者姓名：申请学位级别：工程硕士

工程领域名称：

校内指导教师姓名、职称：校外指导教师姓名、职称：

论文形式：ꇶ产品研发ꇶ工程设计ꇶ应用研究ꇶ工程/项目管理ꇶ调研报告

研究方向：

论文提交日期：年月日论文答辩日期：年月日

学位授予单位：华南理工大学学位授予日期：年月日

答辩委员会成员：

主席：

委员：

**华南理工大学**

**学位论文原创性声明**

本人郑重声明：所呈交的论文是本人在导师的指导下独立进行研究所取得的研究成果。除了文中特别加以标注引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写的成果作品。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律后果由本人承担。

作者签名：日期： 年 月 日

**学位论文版权使用授权书**

本学位论文作者完全了解学校有关保留、使用学位论文的规定，即：研究生在校攻读学位期间论文工作的知识产权单位属华南理工大学。学校有权保存并向国家有关部门或机构送交论文的复印件和电子版，允许学位论文被查阅（除在保密期内的保密论文外）；学校可以公布学位论文的全部或部分内容，可以允许采用影印、缩印或其它复制手段保存、汇编学位论文。本人电子文档的内容和纸质论文的内容相一致。

本学位论文属于：

□保密，在年解密后适用本授权书。

□不保密,同意在校园网上发布，供校内师生和与学校有共享协议的单位浏览；同意将本人学位论文提交中国学术期刊(光盘版)电子杂志社全文出版和编入CNKI《中国知识资源总库》，传播学位论文的全部或部分内容。

(请在以上相应方框内打“√”)

作者签名： 日期：

指导教师签名： 日期

作者联系电话： 电子邮箱：

联系地址(含邮编)：

# 摘 要

卷积神经网络算法是一种计算机视觉领域的新兴算法。在传统神经网络的基础上，通过权值共享等方式，降低参数数量，直接输入图片，进行特征提取和分类、识别等运算。在图像识别领域，卷积神经网络能够达到较高的识别率，而且能克服图像的位移和形变。由于卷积神经网络仍然是一种运算密集型算法，而且其运算由大量独立重复的乘法、加法运算组成，并行实现这种算法的需求日益突出。目前，大多数深层网络被运行在多块GPU组成的并行运算平台上，以实现对卷积神经网络并行加速的目的。然而，GPU运算平台体积大、功耗大，就使得卷积神经网络难以被更广泛地应用在小尺寸、低功耗的平台上。由于通用处理器和嵌入式处理器的串行架构无法充分发挥并行优势；专用DSP芯片虽有对图像处理的特殊优化仍然难以胜任卷积神经网络的密集计算规模，小尺寸、低功耗平台上卷积神经网络的实现仍处于探索阶段，少有成熟系统的解决方案。

现场可编程门阵列（Field Programmable Gate Array， FPGA）内部资源和结构能够应对卷积神经网络这种具有大量独立重复运算的要求，可以充分发挥卷积神经网络中的并行特性，并在小尺寸、低功耗的限制条件下，实现卷积神经网络的运算。在此理论基础上，本文做了如下工作：

1. 在考虑到FPGA芯片中基本逻辑单元的局限性之后，本文着重探讨围绕FPGA中大量存在的数字信号处理单元设计并行卷积神经网络基本运算单元。通过详细分析卷积神经网络，在卷积神经网络最基本的卷积、非线性激活函数和子采样等基本运算的基础上，本文抽象出“从特征映射图到特征映射图”的运算层次，设计与之相应的FPGA卷积神经网络模块并组成系统。
2. 在这个层次上，运算的重复性很强，使用FPGA实现这个运算层次的卷积器、非线性激活函数和子采样器等基本模块，且上述运算模块可以经过最少的修改而适应整个卷积神经网络中多个不同卷积层的运算。因此，以该层次运算为基本单元，运算模块通过复制可以实现并行加速，同时也可以通过重用，实现串行运算，节省硬件资源。

本文以该层次运算为基本运算模块在FPGA实现卷积神经网络具有并行程度灵活可调整，面积和速度之间能良好折衷的特性，同时，本文用FPGA实现的卷积神经网络具有功耗低的优势。综合得到寄存器传输级（Register Transfer Level，RTL）电路经分析表明，单个基本运算模块功耗1.05W，浅层网络系统功耗小于1.6W。因此，本文在FPGA上实现的卷积神经网络功能完整，功耗较低，能被应用于智能电视、机顶盒或者智能手机等小尺寸、低功耗平台上，作为卷积神经网络协处理器，实现多种识别、交互应用。

**关键词：**现场可编程门阵列；卷积神经网络；并行结构；芯片设计

**Abstract**

ANN

**Key Words:** CNN; FPGA