一种基于国产嵌入式 CPU 核的 BP 神经网络 SoC 设计

徐文亮
(杭州电子科技大学 电子信息学院,浙江 杭州 310018)

摘 要:基于国产嵌入式 CPU 核 CK803S 及其 SoC 设计平台,设计一款 BP 神经网络 SoC。给出了 SoC 的设计结构及 BP 神经网络硬件加速器的设计方案,针对 BP 神经网络硬件加速器中非线性的 Sigmod 和 Guass 激活函数,选择了一种既不影响速度又节约资源的方法来实现,并对其性能、功耗进行优化。验证结果表明,设计满足要求。

关键词:BP 神经网络;国产嵌入式处理器 CK803S;SoC 设计平台;FPGA 实现
中图分类号:TN47;TN492 **文献标识码**:A DOI:10.16157/j.issn.0258-7998.200949

中文引用格式:徐文亮. 一种基于国产嵌入式 CPU 核的 BP 神经网络 SoC 设计[J].电子技术应用,2021,47(4):63-66.
英文引用格式:Xu Wenliang. Design of a BP neural network SoC based on domestic embedded CPU[J]. Application of Electronic Technique,2021,47(4):63-66.

Design of a BP neural network SoC based on domestic embedded CPU

Xu Wenliang
(School of Electronic Information, Hangzhou Dianzi University, Hangzhou 310018, China)

Abstract: The paper designs a Back Propagation(BP)neural network system on chip(SoC) based on the domestic embedded Central Processing Unit(CPU) CK803S and its SoC design platform. The design structure of SoC and the design scheme of BP neural network hardware accelerator are given, and for the non-linear BP activation functions Sigmod and Guass, a method that can save hardware resources while not affect the speed is selected to implement them,and optimize accelerator's performance and power consumption. The verification results show that the design can meet the requirements.

Key words: BP neural network; domestic embedded processor CK803S; SoC design platform; FPGA implement

0 引言

人工神经网络的实现方法主要分为硬件实现[1]和软件实现[2]两种。神经网络软件实现的方法具有并行度低和实现速度慢的特点,并且不能满足神经网络对实时运算的要求。除此之外,最大的缺点是用软件模拟实现的方法需要庞大体积的计算机作支持,这样就很不适合应用于嵌入式场景。基于硬件实现的神经网络具有运算速度快、并行性高等优点[3],并且在实时运算方面也能满足要求。综合考虑,本文采用硬件实现的方法来设计人工神经网络。

本文设计的目的是找到一种方法——硬件实现的神经网络能够进行动态调节,既可以实现神经网拓扑结构的动态调节,即每层网络和每层神经元的个数动态可调,也可以实现输入权值和阈值的自动更新。本文以 BP 神经网络为例,使用国产嵌入式 CPU CK803S 及其 SoC 设计平台 SmartL-Prime,实现一款 BP 神经网络 SoC 的设计。

1 SoC 结构设计

本文设计的 BP 神经网络 SoC 采用平头哥(杭州中天微)提供的基于 CK803S 嵌入式 CPU 的 SmartL-Prime 平台。CK803S 是面向控制领域的 32 位高能效嵌入式 CPU 核[4],采用了精简的 3 级流水线结构,具有低成本、低功耗等特点。BP 神经网络 SoC 的系统结构图如图 1 所示。

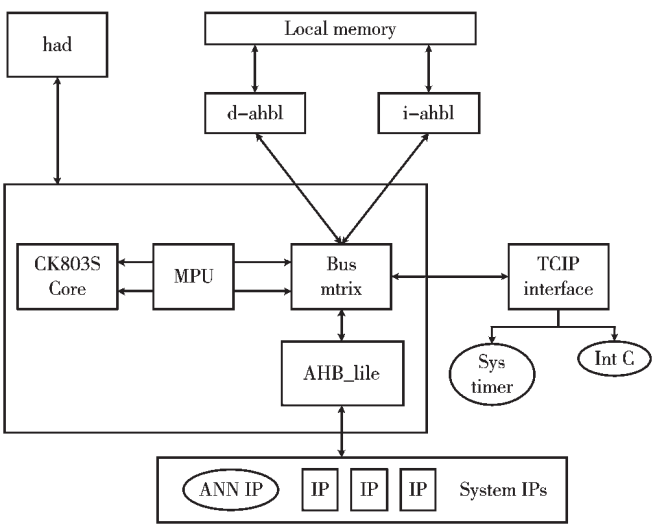

图 1 BP 神经网络 SoC 的系统框图

图 1 中 ANN IP 即为 BP 神经网络加速器，通过系统总线挂载到 SmartL-Prime 平台上，系统总线使用了 AHB-Lite 总线为单主机结构[5]。将 CK803S 处理器作为主机，BP 神经网络加速器作为从机，在处理器的控制下，通过 AHB-Lite 总线实现对 BP 神经网络加速器 IP 核的访问和数据的交互。设计的 IP 经过封装打包完成后完成兼容 AHB-Lite 的协议，将其挂载到总线上后，即可通过 CK803S 作为主机，实现对神经网络 IP 核的访问。

BP 神经网络加速器的 AHB-Lite 总线连接方式如图 2 所示。BP 神经网络加速器为从机，主机 CK803S 输出 BP 神经网络加速器的地址，译码器产生了选通信号使能 BP 神经网络加速器，其余的地址输出到多路复用器中，用于选通从机的输出，从而通过系统总线读取到 BP 神经网络加速器的数据。对 BP 神经网络加速器的写操作与读操作类似，地址稳定之后有效，给 BP 神经网络加速器发送写使能信号，接收到就绪响应后，往写数据总线上输出数据，即可完成对 BP 神经网络加速器的写操作。

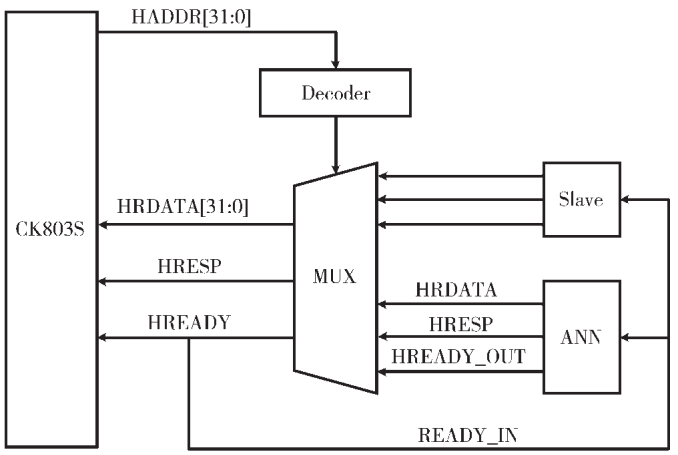

图 2 主机和 BP 神经网络加速器的 AHB-Lite 总线连接

图 3 所示为从机的选通信号通过译码器生成框图。AMBA 的交互总线结构和统一编址设计方便了系统的寻址，AHB-Lite 的单主结构也简化了总线的复杂度，减少了复杂的仲裁逻辑。

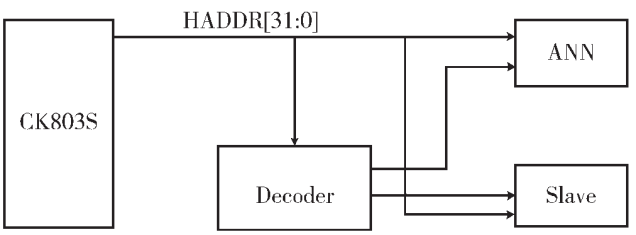

图 3 从机选通信号生成图

2 BP 神经网络加速器的设计

2.1 神经元设计

为实现 BP 神经网络的硬件设计，首先应该完成神经网络的基本单元，即人工神经元的硬件设计[6-8]。图 4 是

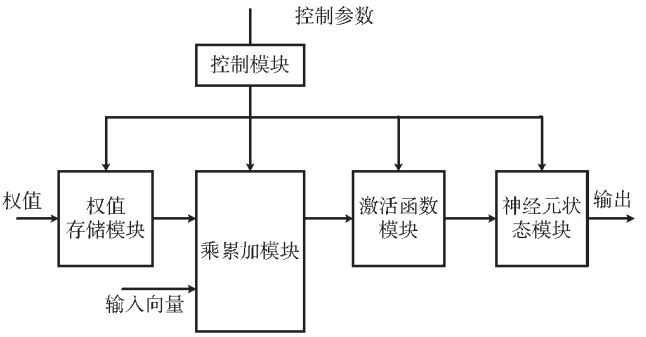

图 4 神经元运算过程示意图

单个神经元的运算过程示意图。

控制模块的控制参数通过软件可配置，来控制数据流在模块间的传输，并最后将数据通过神经元状态模块传输出去。首先控制模块控制数据流从权值存储模块开始，权值和输入向量一起被加载到乘累加模块进行乘累加运算，接着将计算结果输入到激活函数模块，经过该模块计算后输出到神经元状态模块。

CPS(Connections-Per-Second)即每秒的连接率，是对神经网络硬件性能评估的一种重要的方法。为了测试训练的速度，往往通过每秒钟联接的更新率来评估。此外学习的速率也与选用的学习算法有关。

针对不同神经网络的设计，激活函数的选择可以是不同的。本设计中，使用的激活函数是 Sigmoid 函数和 Guass 函数[9-10]。此外学习的速率也与选用的学习算法有关。神经元具体的硬件设计如图 5 所示。

2.2 BP 神经网络设计

一个三层的 BP 神经网络由输入层、隐含层和输出层组成[11]。如果具有足够的隐含层神经元数，它就能以任意精度逼近任何连续的非线性函数，所以 BP 神经网络通常用来进行函数逼近和分类问题[12-13]。神经网络每进行一次完整的训练，BP 神经网络硬件都会进行一次误差反向运算和前向运算，并会修改相应的权值矩阵。隐含层和输出层使用的函数分别是 Sigmoid 函数和线性函数。本设计中将 BP 神经网络的整个计算过程总结为以下几个功能：激活函数运算 Sigmod 模块，误差运算 Error 模块，权值修正与更新 Updata 模块，输入输出层 RAM 存储模块以及功能可复用的神经元乘累加模块。

BP 神经网络的硬件实现整体结构如图 6 所示。

从硬件实现框图可以看出，BP 神经网络的层内的执行过程是并行的。整个神经网络的硬件实现过程如下：

(1)开始时把初始权值分别存放于隐含层权值 RAM 和输出层权值 RAM 中以及训练用到的样本集存放在输入层 RAM 中，作为训练使用。

(2)根据对输入层 RAM 中的输入和对于隐含层权值 RAM 中的数值，对隐含层神经元进行计算，送到 MAC 模块进行累加计算，并将运算完成的结果输出到激活函数模块。

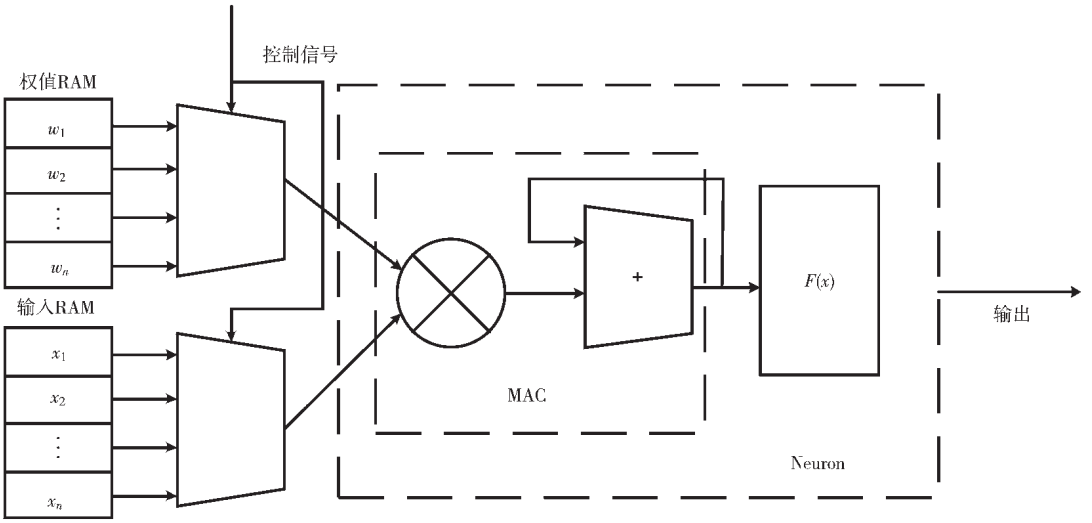

图 5 神经元的硬件实现结构

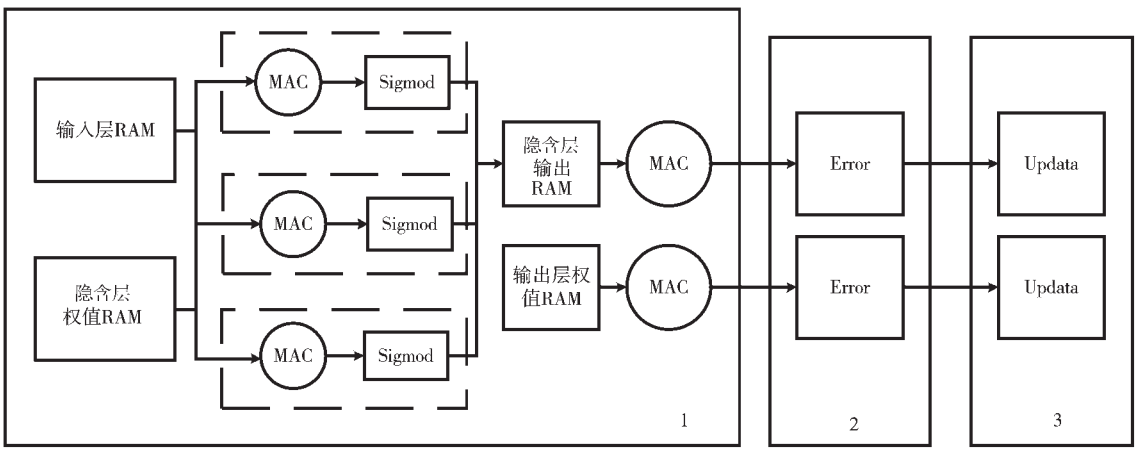

图 6 BP 神经网络的硬件实现整体结构图

(3)把激活函数的计算结果存放到隐含层输出 RAM 中。

(4)分别读取隐含层输出 RAM 中的数据和输出层权值 RAM 中的权值数据,输出到 MAC 模块进行第二次乘累加运算,最终得到输出层的输出数据。至此,前向运算阶段结束。

(5)误差运算单元 Error 用来计算隐含层的输出和输出层的输出数据。至此,误差反向传播阶段完成。

(6)将得到的误差、输入以及隐含层输出送到权值修正和更新单元,得到新的权值,重新存入权值 RAM 中[13]。

至此,整个神经网络完成一次完整的训练过程,接着重复训练其他样本,直到满足指定的训练步数或者是误差满足要求为止。这样完成了对整个 BP 神经网络硬件的实现。

3 验证和分析

神经网络加速器使用 VHDL 实现,通过 Synopsys VCS 的验证环境,验证神经网络 IP 核的功能和逻辑[14-15],并使用了 Assertion 方式进行了误差分析。对于 SoC 的测试软硬件的协同设计,本文使用基于 FPGA 的 FMX7AR3B 平台,在基于 CK803S 的 SmartL-Prime 平台进行设计之后,使用 FPGA 平台进行配置实现。使用 CDK 工具链进行软硬件协同设计和验证[16]。

SoC 设计和验证完成之后,本文使用 Vivado 分析工具进行 PPA(Perfromance Power Area,即性能、功耗和面积)分析。性能分析结果如图 7 所示,是 Dhrystone 的基准跑分测试;当波特率设置为 115 200Bd,数据位为 8 位,无校验位,停止位为 1 位的情况下,结果为 0.34 DMIPS/MHz,满足设计要求。

图 8 所示为在 Vivado 下查看的资源使用情况。从图中可以看出 Look Up Table 的资源使用率约为 18%,BRAM 的资源使用率约为 78%,IO 的资源使用率约为 87%。

参考文献

[1] 周齐国.基于 ARM 和 FPGA 的神经网络处理系统的设计与实现[D].福州:福建师范大学,2014.

[2] 陈霁威,乐慧丰,黄道.基于神经网络和遗传算法的在线优化软件设计与实现[J].华东理工大学学报,2002(4):

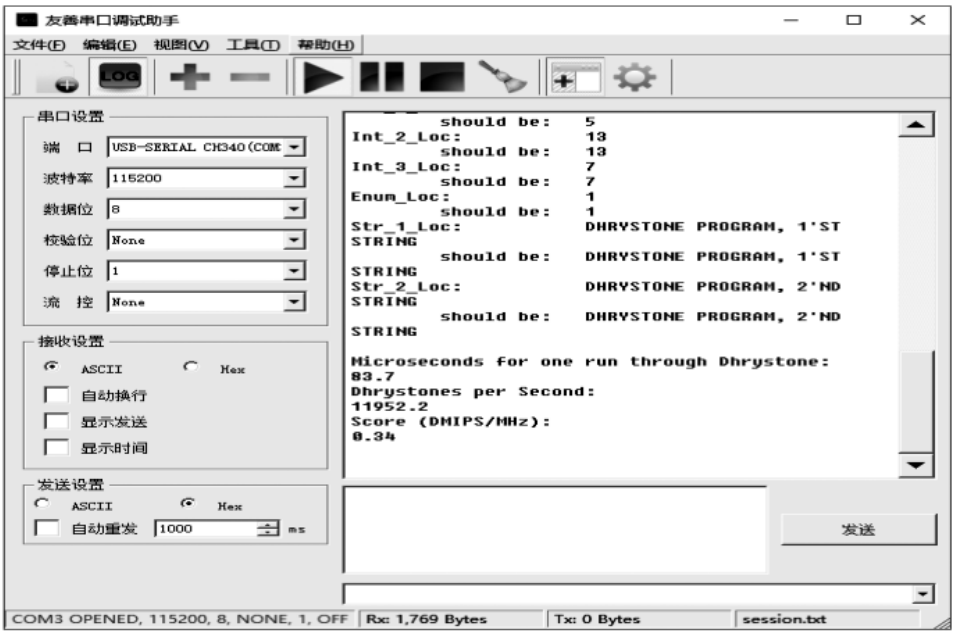

图 7 benchmark 跑分结果

Utilization Post-Synthesis | Post-Implementation

Graph | Table

Resource	Utilization	Available	Utilization %
LUT	24059	133800	17.98
LUTRAM	22	46200	0.05
FF	9320	267600	3.48
BRAM	284.50	365	77.95
DSP	10	740	1.35
IO	248	285	87.02
BUFG	2	32	6.25

图 8 资源使用情况

419-422.

[3] 李安新.BP 神经网络研究与硬件实现[D].青岛:山东科技大学,2010.

[4] 陈亮.基于嵌入式 CPU 的数据加解密子系统的设计研究[D].杭州:浙江大学,2013.

[5] 刘政林,张振华,陈飞,等.基于 AHB-Lite 总线的祖冲之密码算法 IP 核研究[J].微电子学与计算机,2015,32(8):88-92.

[6] 刘培龙.基于 FPGA 的神经网络硬件实现的研究与设计[D].成都:电子科技大学,2012.

[7] 杨程.基于 FPGA 的人工神经网络的研究与实现[D].西安:西安电子科技大学,2016.

[8] 李昂,王沁,李占才,等.基于 FPGA 的神经网络硬件实现方法[J].北京科技大学学报,2007(1):90-95.

[9] 钱玉多.基于 FPGA 的神经网络硬件实现研究[D].武汉:华中科技大学,2012.

[10] 杨银涛,汪海波,张志,等.基于 FPGA 的人工神经网络实现方法的研究[J].现代电子技术,2009,32(18):170-174.

[11] 洪启飞.面向深度学习的 FPGA 硬件加速平台的研究[D].成都:电子科技大学,2018.

[12] 王蒙,常胜,王豪.一种自适应训练的 BP 神经网络 FPGA 设计[J].现代电子技术,2016(15):115-118.

[13] 余子健,马德,严晓浪,等.基于 FPGA 的卷积神经网络加速器[D].杭州:浙江大学,2017.

[14] 李永红.基于 OVM 的 SoC 验证平台的设计与实现[D].西安:西安电子科技大学,2013.

[15] 张强.UVM 实战[M].北京:机械工业出版社,2014.

[16] 程翼胜.SoC 芯片 FPGA 原型的软硬件协同验证[J].单片机与嵌入式系统应用,2017,17(11):7-10,13.

(收稿日期:2020-09-26)

作者简介:

徐文亮(1995-),男,硕士研究生,主要研究方向:数字 IC 设计与验证。

(上接第 62 页)

情感分类方法[J].济南大学学报(自然科学版),2019,33(6):483-488.

[18] 刘铭,咎红英,原慧斌.基于 SVM 与 RNN 的文本情感关键句判定与抽取[J].山东大学学报(理学版),2014,49(11):68-73.

(收稿日期:2020-07-14)

作者简介:

杨戈(1974-),通信作者,男,博士,副教授,主要研究方向:计算机视觉、视觉跟踪技术,E-mail:yangge@pkusz.edu.cn。

杨麓涛(1998-),男,本科,主要研究方向:计算机视觉。

版权声明