

BP 神经网络图像压缩算法乘累加单元的 FPGA 设计

杨 隽¹, 周 詮², 张敏瑞¹

(1. 西安科技大学 通信与信息学院 陕西 西安 710054;

2. 中国空间技术研究院 西安分院 国家级重点实验室 陕西 西安 710000)

摘 要:提出一种基于三层前馈 BP 神经网络实现图像压缩算法的方案,该方案采用可重载 IP 核和 VHDL 代码相结合的设计方式。对方案中重要单元-乘累加单元进行了 FPGA 设计,该模块设计采用流水线处理方式,增大了数据吞吐量,减小了系统延时,提高了时钟频率,并完成了该单元的行为级功能仿真。仿真结果验证了 FPGA 设计的可行性。

关键词:FPGA;神经网络;图像压缩;乘累加单元

中图分类号:TP183

文献标识码:A

文章编号:1004-373X(2009)19-038-04

FPGA Design of Multiply - accumulate Module in Image Compression Algorithm Based on BP Neural Network

YANG Jun¹, ZHOU Quan², ZHANG Minrui¹

(1. School of Electronic and Information Engineering, Xi'an University of Science and Technology, Xi'an, 710054, China;

2. National Key Laboratory, China Academy of Space Technology, Xi'an, 710000, China)

Abstract: A realization scheme for image compression algorithm based on three layers Back Propagation (BP) neural network is presented, which adopts the method combining the universal IP core with the VHDL language. The FPGA design of multiply - accumulate module which is one of the important components is carried out, it is designed with pipeline, which could increase the data of processing, decrease the latency of system, raise the frequency of clock. The behavior of functional simulation for MAC is completed. The simulation results show that the design is practicable.

Keywords: FPGA; neural network; image compression; MAC

0 引 言

神经网络(Neural Networks)是人工神经网络(Artificial Neural Networks)的简称,是当前的研究热点之一。人脑在接受视觉感官传来的大量图像信息后,能迅速做出反应,并能在脑海中重现这些图像信息,这不仅与人脑的海量信息存储能力有关,还与人脑的信息处理能力,包括数据压缩能力有关。在各种神经网络中,多层前馈神经网络具有很强的信息处理能力,由于其采用 BP 算法,因此也称为 BP 神经网络。采用 BP 神经网络模型能完成图像数据的压缩处理。在图像压缩中,神经网络的处理优势在于^[1]:巨量并行性;信息处理和存储单元结合在一起;自组织自学习功能。

与传统的数字信号处理器 DSP (Digital Signal Processor)相比,现场可编程门阵列(Field Programmable Gate Array, FPGA)在神经网络的实现上更具优势。DSP 处理器在处理时采用指令顺序执行的方式,

而且其数据位宽是固定的,因而资源的利用率不高,限制了处理器的数据吞吐量,还需要较大的存储空间^[2-4]。FPGA 处理数据的方式是基于硬件的并行处理方式,即一个时钟周期内可并行完成多次运算,特别适合于神经网络的并行特点,而且它还可以根据设计要求配置硬件结构,例如根据实际需要,可灵活设计数据的位宽等。随着数字集成电路技术的飞速发展,FPGA 芯片的处理能力得到了极大的提升,已经完全可以承担神经网络数据压缩处理的运算量和数据吞吐量。图像压缩是信息传输和存储系统的关键技术,然而如何进行 FPGA 设计,以实现给定的功能已经成为神经网络应用的关键^[5,6]。

基于以上原因,选择 FPGA 作为三层 BP 神经网络图像压缩算法的实现方式,提出了具体的一种实现方案,并对其中的重点单元进行了 FPGA 设计与仿真验证。

1 BP 神经网络图像压缩算法

一般习惯将单隐层前馈网称为三层前馈网,它包括输入层、隐含层和输出层。三层 BP 神经网络结构如

收稿日期:2009-03-17

基金项目:国家重点实验室基金项目(9140C5305020706)

图 1 所示,原始数据节点和重建数据节点构成节点数较大的外层(输入层和输出层),而中间的具有较小节点数的细腰层即构成压缩结果。其基本思想是强迫原始数据通过细腰型网络瓶颈,并期望在网络的瓶颈处能获得较为紧凑的数据表示,以达到压缩的目的。在网络的学习过程中,通过 BP 训练算法,调整网络的权重,使重建图像在均方误差意义上尽可能近似于训练图像。经过训练的网络即可用来执行数据压缩任务,网络输入层与隐含层之间的加权值相当于一个编码器,隐含层与输出层之间的加权相当于一个解码器。从输入端输入的原始图像数据经过神经网络的处理,在隐含层得到的输出数据就是原始图像的压缩编码,而输出层矢量即为解压后重建的图像数据^[7]。

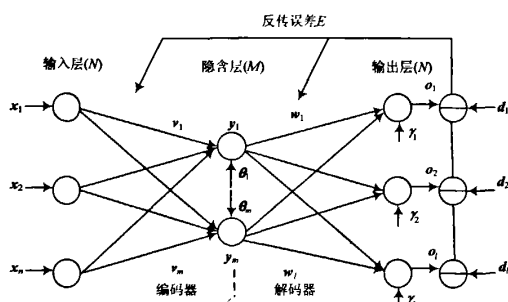


图 1 三层前馈 BP 神经网络压缩结构

BP 神经网络用于图像编码的压缩比与输入层和隐含层的节点数有关:

$$\text{压缩比} = \text{输入层节点数}(n) / \text{隐含层节点数}(m)$$

因此一般来说采用不同数目的隐含层神经元就可实现同一图像的不同压缩比。

三层 BP 前馈网中输入向量 $X = x(x_1, x_2, \dots, x_i, \dots, x_n)^T$, 隐含层输出向量 $Y = y(y_1, y_2, \dots, y_j, \dots, y_m)^T$, 输出层输出向量 $O = o(o_1, o_2, \dots, o_k, \dots, o_l)^T$, 期望输出向量 $d = d(d_1, d_2, \dots, d_k, \dots, d_l)^T$, 输入层到隐含层的权值向量 $V = v(v_1, v_2, \dots, v_j, \dots, v_m)^T$, 其中 v_j 为隐含层第 j 个神经元对应的权值向量; 隐含层到输出层的权值向量 $W = w(w_1, w_2, \dots, w_k, \dots, w_l)^T$, 其中 w_k 为输出层第 k 个神经元对应的权值向量; 隐含层的阈值向量 $\theta = (\theta_1, \theta_2, \dots, \theta_j, \dots, \theta_m)^T$; 输出层的阈值向量 $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_k, \dots, \gamma_l)^T$ 。

(1) 用小的随机数对每一层的权值和偏差初始化, 以保证网络不被大的加权输入饱和, 并进行以下参数的设定或初始化: 期望误差最小值; 最大循环次数; 修正权值的学习速率;

(2) 将原始图像分为 4×4 或 8×8 大小的块, 选取其中一块的像素值作为训练样本接入到输入层, 计算各层输出:

$$y_j = f(\text{net}_j) \quad (1)$$

$$\text{net}_j = \sum_{i=1}^n v_{ij} x_i + \theta_j, \quad j = 1, 2, \dots, m; \quad (2)$$

$$o_k = f(\text{net}_k) \quad (3)$$

$$\text{net}_k = \sum_{j=1}^m w_{kj} y_j + \gamma_k, \quad k = 1, 2, \dots, l \quad (4)$$

其中: $f(\cdot)$ 为 BP 网络中各层的传输函数。

(3) 计算网络输出与期望输出之间的误差, 判断是否小于期望误差, 是则训练结束, 否则至下一步, 其中反传误差的计算式为:

$$E = \frac{1}{2} (d - O)^2 = \frac{1}{2} \sum_{k=1}^l (d_k - o_k)^2 \quad (5)$$

(4) 计算各层误差反传信号;

(5) 调整各层权值和阈值;

(6) 检查是否对所有样本完成一次训练, 是则返回步骤(2), 否则至步骤(7);

(7) 检查网络是否达到最大循环次数, 是则训练结束, 否则返回步骤(2)。

经过多次训练, 最后找出最好的一组权值和阈值, 组成三层前馈神经网络, 用于该算法的 FPGA 设计。

2 基于 BP 神经网络的 FPGA 设计

2.1 硬件设计方案

本文提出了一种基于三层前馈 BP 神经网络的图像压缩算法的实现方案, 包括三层 BP 神经网络图像压缩算法编码器以及译码器的实现方案。

由三层 BP 前馈神经网络图像压缩的结构图可以看出, 其解码器是编码器的逆过程, 在实现编码器之后, 实现解码器非常方便, 故本文重点在于解码器的 FPGA 设计。

基于三层 BP 前馈神经网络图像压缩算法的编码器的 FPGA 设计方案如图 2 所示。

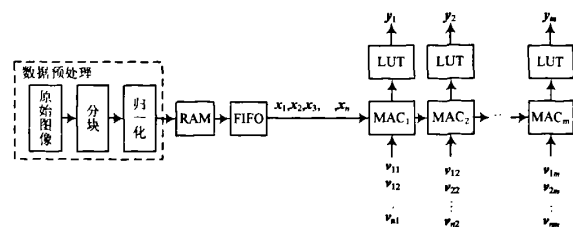


图 2 三层 BP 神经网络图像压缩算法编码器结构图

其中, 在数据预处理部分, 首先将原始图像分成 $n \times n$ 的小块, 以每一小块为单位进行归一化^[8]。归一化的目的, 主要有以下两点:

(1) BP 网络的神经元均采用 Sigmoid 转移函数, 变换后可防止因净输入的绝对值过大而使神经元输出

饱和,继而使权值调整进入误差曲面的平坦区;

(2) Sigmoid 转移函数的输出在 $-1 \sim +1$ 之间, 作为信号的输出数据如不进行变换处理, 势必使数值大的输出分量绝对误差大, 数值小的输出分量绝对误差小。网络训练时只针对输出的总误差调整权值, 其结果是在总误差中占份额小的输出分量相对误差较大, 对输出量进行尺度变化后这个问题可迎刃而解。

归一化后得到以每小块的灰度值为列向量组成的待压缩矩阵,将该矩阵存储在 RAM 里,然后以每一列为单位发送给先入先出寄存器 FIFO(First Input First Output);由 FIFO 将向量 x_1, x_2, \dots, x_n 以流水(pipe-line)方式依次传入各乘累加器 MAC(Multiply - Accumulate),相乘累加求和后,送入 LUT(Lookup Table)得到隐层相应的节点值,这里 LUT 是实现 Sigmoid 函数及其导函数的映射^[9]。

在整个电路的设计中,采用 IP(Intellectual Property)核及 VHDL 代码相结合的设计方法,可重载 IP 软核,具有通用性好,便于移植等优点,但很多是收费的,比如说一个高性能流水线设计的 MAC 软核,所以基于成本考虑,使用 VHDL 语言完成 MAC 模块的设计,而 RAM 和 FIFO 模块则采用免费的可重载 IP 软核,使整个系统的设计达到最佳性价比。

在压缩算法的实现中,乘累加单元是共同部分,也是编码和译码器 FPGA 实现的关键。

2.2 乘累加器 MAC 的流水线设计及其仿真

流水线设计是指将组合逻辑延时路径系统地分割,并在各个部分(分级)之间插入寄存器暂存中间数据的方法。流水线缩短了在一个时钟周期内信号通过的组合逻辑电路延时路径长度,从而提高时钟频率。对于同步电路,其速度指同步电路时钟的频率。同步时钟愈快,电路处理数据的时间间隔越短,电路在单位时间内处理的数据量就愈大,即电路的吞吐量就越大。理论而言,采用流水线技术能够提高同步电路的运行速度^[10]。

MAC 电路是实现 BP 神经网络的重要组成部分,在许多数字信号处理领域也有着广泛应用,比如数字解调器、数字滤波器和均衡器,所以如何提高 MAC 的效率和运算速度具有极高的使用价值。

本方案采用的 MAC 设计以四输入为例。

四输入的 MAC 电路必须执行四次乘法操作和两次加法操作,以及最后的两次累加操作。如果按照非流水线设计,完成一次对输入的处理,需要这三步延迟时间的总和,这会降低一个高性能系统的效率。而采用流水线设计,则可以避免这种延迟,将 MAC 的操作安排的像一条装配线一样,也就是说,通过这种设计它可以使系统执行的时钟周期减小到流水线中最慢步骤所需

的操作时间,而不是各步骤延迟时间之和,如图 3 所示。

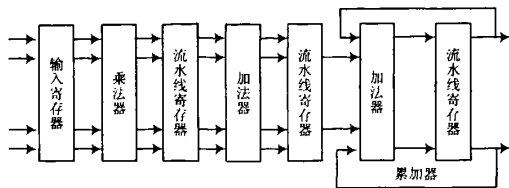


图3 基于流水线设计的四输入 MAC 结构图

在第一个时钟边沿,第一对数据被存储在输入寄存器中。在第一个时钟周期,乘法器对第一对数据进行乘法运算,同时系统为下一对数据的输入作准备。在第二个时钟边沿,第一对数据的积存储在第一个流水线寄存器,且第二对数据已经进入输入寄存器。在第二个时钟周期,完成对第一对数据积的两次加法操作,而乘法器完成第二对数据的积运算,同时准备接收第三队数据。在第三个时钟边沿,这些数据分别存放在第二个流水线寄存器,第一个流水线寄存器,以及输入寄存器中。在第三个时钟周期,完成对第一对数据和之前数据的累加求和,对第二对数据的两次加法操作,对第一对数据的乘法运算,并准备接收第四对数据。在第四个始终边沿,累加器中的和将被更新。

在本设计方案中,测试仿真平台选用的 FPGA 芯片为 ALTERA 公司 Cyclone II 系列的 EP2C8 芯片,它采用 90 nm 的制造工艺,拥有 8 256 个逻辑单元,36 个 M4K 随机只读存储器,2 个数字锁相环,以及 18 个硬乘法器等丰富资源。仿真工具使用业界流行的 Mentor Graphics 公司的仿真软件 Modelsim 6.1f。

对设计进行验证时,常见的方法是在模拟时施加输入激励信号,然后“读”该设计的输出信号,它的主要缺点是随着模拟器的不同而不同。为了克服此缺点,采用的测试方法是用 VHDL 编写一个测试模型发生器,称为 Testbench,它的优点是通用性好,灵活性强,可以随时更改输入激励,已得到不同的仿真结果。

在对该 MAC 模块进行测试的过程中,涉及输入数据的转化问题,如前所述,在本神经网络中,输入数据归一化后,集中在 $-1 \sim +1$ 之间,所以处理时必须进行转化,最后采用 16 位补码形式的定点二进制表示法,由于在求和中可能会产生溢出,还必须包含一个溢出状态信号。

输入数据转换 16 位补码的仿真波形如图 4 所示。

图 4 输入数据转换 16 位补码仿真波形图

16 位补码转换原输入实数的仿真波形如图 5 所示。



图5 16位补码转换为实数的仿真波形图

在完成了对输入、输出数据的转换之后,编写 Testbench(测试台)程序,对基于流水线设计的四输入 MAC 进行行为级仿真,仿真波形如图 6 所示。

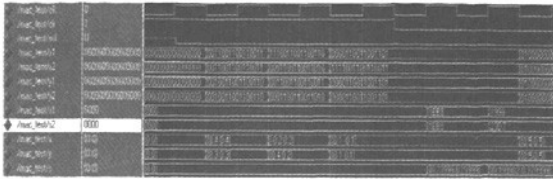


图6 MAC 行为级功能仿真图

综上所述,在基于流水线的乘法设计中,虽然每一步操作后都加入了寄存器,消耗了更多的资源,但却可以将系统延时降低到最慢步骤所需要的时间,极大地提高了同步电路的运算速度。

3 结 语

介绍了基于三层前馈 BP 神经网络的图像压缩算法,提出了基于 FPGA 的实现验证方案,详细讨论了实现该压缩网络组成的重要模块 MAC 电路的流水线设计。在对 BP 神经网络的电路设计中,对传输函数及其导函数的线性逼近也是近来研究的热点之一,本文使用的压缩查找表虽然能够满足设计要求,但仍然消耗了大

量资源。该研究结果对整个压缩解压缩算法的实现以及多层神经网络的相关研究工作提供了参考。

参 考 文 献

(上接第 29 页)

3 结 语

根据 WMN 的节点准静态特性,基于 DSR 设计了一种跨层多径路由协议 CMRP,在其中添加主动路由的思想,增加了链路质量判决、拥塞避免及多径路由、路由断链判定等机制。仿真结果表明,比起原来的 DSR,CMRP 在平均吞吐率、平均时延、平均抖动等网络性能上都有较大的提高,达到了预期的目标,可以更好地适用于 WMN。

参 考 文 献

- [1] 方旭明. 下一代无线因特网技术: 无线 Mesh 网络[M]. 北京: 人民邮电出版社, 2005.
- [2] 罗明胜, 黄联芬, 姚彦. 无线网络跨层设计的研究现状及展望[J]. 移动通信, 2005, 29(7): 95-98.
- [3] 方旭明, 马忠建. 无线 Mesh 网络的跨层设计理论与关键技术[J]. 西南交通大学学报, 2005(12): 711-719.
- [4] Bruno R, Conti M, Gregori E. Mesh Networks: Commodity Multihop Ad Hoc Networks [J]. IEEE Communications

- [1] 董长虹. Matlab 神经网络与应用[M]. 2 版. 北京: 国防工业出版社, 2007.
- [2] 赵宏怡. DSP 技术与应用实例[M]. 北京: 电子工业出版社, 2003.
- [3] 刘艳萍. DSP 技术原理及应用教程[M]. 北京: 北京航空航天大学出版社, 2005.
- [4] 苏涛, 蔺丽华. DSP 实用技术[M]. 西安: 西安电子科技大学出版社, 2005.
- [5] 田华, 冯勤群, 胡喜飞. 基于 DSP 和 FPGA 的高速图像压缩系统设计[J]. 电子工程师, 2005, 31(8): 51-52.
- [6] 张秀艳. 基于 FPGA 的神经网络硬件实现中的关键问题研究[J]. 科技情报开发与经济, 2005, 15(5): 273-275.
- [7] 冯春丽. 基于改进型 BP 网络图像压缩方法的研究[D]. 锦州: 辽宁工学院, 2007.
- [8] 张锐菊, 周诠. 神经网络用于遥感图像压缩的一些研究结果[J]. 中国体视学与图像分析, 2003, 8(3): 183-186.
- [9] Ferreira P, Ribeiro P, Antunes A, et al. Artificial Neural Networks Processor: A Hardware Implementation Using a FPGA[A]. Proceedings of the 4th International Conference on Field-Programmable Logic and Applications[C]. Antwerp, 2004: 1 084-1 086.
- [10] Ashenden P J. The Designer's Guide to VHDL[M]. Second Edition. USA: Elsevier Science, 2002.

Magazine, 2005, 43(3): 123-131.

- [5] Koksal C E, Balakrishnan H. Quality Aware Routing Metrics for Time-varying Wireless Mesh Networks[J]. IEEE Journal on Selected Areas in Communications, 2006, 1 984-1 994.
- [6] Ekram Hossain, Kin K Leung. Wireless Mesh Networks: Architectures and Protocols[M]. Springer, 2007.
- [7] 沈强, 方旭明. 无线 Mesh 网中一种基于综合准则的 DSR 扩展路由方法[J]. 电子学报, 2007, 35(4): 614-620.
- [8] 郭嘉丰, 张信明, 谢飞, 等. 基于节点空闲度的自适应移动 Ad Hoc 网络路由协议[J]. 软件学报, 2005, 16(5): 960-969.
- [9] 于明刚. 论无线网状网路由分组接收率的提高与网络性能的改进[D]. 长春: 吉林大学, 2007.
- [10] Scalable Network Technologies [EB/OL]. <http://www.QualNet.com>, 2005.
- [11] Hans-Peter Schwefel, Lester Lipsky. Impact of Aggregated, Self-Similar ON/OFF traffic on Delay in Stationary Queueing Models (Ext. Version)[A]. Performance Evaluation[OL]. 2000.

BP神经网络图像压缩算法乘累加单元的FPGA设计

作者: [杨隼](#), [周詮](#), [张敏瑞](#), [YANG Jun](#), [ZHOU Quan](#), [ZHANG Minrui](#)
作者单位: [杨隼, 张敏瑞, YANG Jun, ZHANG Minrui \(西安科技大学, 通信与信息学院, 陕西, 西安, 710054\)](#), [周詮, ZHOU Quan \(中国空间技术研究院, 西安分院, 国家级重点实验室, 陕西, 西安, 710000\)](#)
刊名: [现代电子技术](#) **ISTIC**
英文刊名: [MODERN ELECTRONICS TECHNIQUE](#)
年, 卷(期): 2009, 32(19)
被引用次数: 0次

参考文献(10条)

1. [董长虹](#) [Matlab神经网络与应用](#) 2007
2. [赵宏怡](#) [DSP技术与应用实例](#) 2003
3. [刘艳萍](#) [DSP技术原理及应用教程](#) 2005
4. [苏涛. 蔺丽华](#) [DSP实用技术](#) 2005
5. [田华. 冯勤群. 胡喜飞](#) [基于DSP和FPGA的高速图像压缩系统设计](#)[期刊论文]-[电子工程师](#) 2005(08)
6. [张秀艳](#) [基于FPGA的神经网络硬件实现中的关键问题研究](#)[期刊论文]-[科技情报开发与经济](#) 2005(05)
7. [冯春丽](#) [基于改进型BP网络图像压缩方法的研究](#) 2007
8. [张锐菊. 周詮](#) [神经网络用于遥感图像压缩的一些研究结果](#)[期刊论文]-[中国电视学与图像分析](#) 2003(03)
9. [Ferreira P. Ribeiro P. Antunes A](#) [Artificial Neural Networks Processor, A Hardware Implementation Using a FPGA](#) 2004
10. [Ashenden P J](#) [The Designer's Guide to VHDL](#) 2002

相似文献(10条)

1. 学位论文 [闫明](#) [基于FPGA的神经网络硬件实现](#) 2008
神经网络是由多个非常简单的处理单元彼此按某种方式相互连接而形成的计算系统, 该系统是靠其状态对外部输入信息的动态响应来处理信息的。神经网络是一个非线性的动态系统, 可在任意精度内实现变量间的非线性关系的映像, 具备解决非线性问题能力、网络学习能力和系统拟合能力, 因此, 人工神经网络已在我国科研、生产和生活等方面产生了普遍而巨大的影响。
本文主要的工作是研究神经网络的硬件实现问题, 神经网络的硬件实现是神经网络研究的基本问题之一, 在构造神经网络的实际应用系统时, 必然要研究和解决其硬件实现的问题。神经网络专用硬件可提供高速度, 并具有比通用串、并行机高得多的性能价格比, 所以, 特定应用下的高性能专用神经网络硬件是神经网络研究的热点。本文在比较了几种神经网络的可行性基础上, 选用了BP神经网络作为硬件实现的神经网络模型。BP神经网络对输入输出非线性关系的高精度映射能力、较强的包容性、良好的推广能力和泛化能力, 使得它们在实际应用中表现出了强大的生命力, 成为当今的研究热点之一。作为BP神经网络中的激励函数之一的双曲正切S型(tan-s)函数适用于变化剧烈的场合, 能够加快网络学习收敛速度。可编程技术的迅猛发展, 在EDA技术中占有举足轻重的地位。FPGA可以充分发挥神经网络并行执行的优点, 二者的结合也成为当今的研究热点之一。文中详细叙述了用STAM算法在FPGA上对tan-s函数的高精度实现方式, 并对仿真结果进行了讨论。基于FPGA的可重构技术, 就是利用FPGA的可以多次重复配置的特点, 通过时分复用的方式, 充分利用器件的逻辑资源, 在同一FPGA中自主地交替实现多种逻辑结构及其功能的技术。基于这种技术的可重构系统既可像通用微处理器系统那样设计灵活、易升级, 又可像专用集成电路系统那样速度快、效率高。
本文采用基于FPGA的可重构技术, 以BP(BackPropagation)型误差反传多层前馈神经网络作为典型的数学模型来展开。设计中依据成熟的BP算法公式, 采用原理图输入的设计方法, 按照前向-反传-更新三个阶段分别设计与仿真, 通过验证, 逻辑功能正确。该方法可以推广至更多类型的神经网络, 为神经网络的硬件实现提供了可靠的基础。
2. 期刊论文 [张智明. 张仁杰. ZHANG Zhiming. ZHANG Renjie](#) [神经网络激活函数及其导数的FPGA实现](#) -[现代电子技术](#) 2008, 31(18)
神经元的激活函数是神经网络的重要组成部分。采用折线斜率为2的次幂的分段线性逼近方法实现激活函数(sigmoid函数)及其导数的映射。该方法在FPGA实现时不需要使用硬件乘法器, 而且可以节约大量的RAM单元。由于神经网络的并行计算需要消耗大量的硬件乘法器和RAM, 因此, 与其他方法相比, 该方法为整个神经网络的FPGA实现有效地节省了大量宝贵的FPGA资源, 可以较好地应用在BP神经网络的在线训练中。
3. 期刊论文 [刘晋明. 刘年生. LIU Jin-ming. LIU Nian-sheng](#) [基于神经网络混沌吸引子公钥加密算法的FPGA实现](#) -[厦门大学学报\(自然科学版\)](#) 2010, 49(2)
利用具有顺序和并行执行的特点的VHDL语言, 设计并实现了基于神经网络混沌吸引子的公钥加密算法, 在编解码器设计中采用专用的控制模块来控制加密和解密操作;同时, 在RAM模块中自主设计了具有并行读写功能的子模块, 以进一步提高算法的数据加密速度;整个系统在DE2实验平台中经过反复实验测试和试运行, 结果表明该算法是可以硬件实现的, 并且具有较高的数据加密速度, 时钟频率可达50 MHz以上。
4. 期刊论文 [李利歌. 阎保定. 侯忠. LI Li-Ge. YAN Bao-Ding. HOU Zhong](#) [基于FPGA的神经网络硬件可重构实现](#) -[河南科技大学学报\(自然科学版\)](#) 2009, 30(1)
针对软件实现神经网络速度慢的缺点, 介绍了一种神经网络在FPGA上可重构实现的设计方法。设计中依据成熟的BP算法公式, 以一个三层的BP网络为例, 利用Verilog HDL硬件语言自顶向下设计各个模块, 使网络训练时将前向模块、误差反传模块和相应的控制模块同时配置到FPGA中进行网络训练;而当训练好的网络正常工作时, 只在FPGA中配置前向模块和相应的控制模块就可以高速运行该神经网络。实验结果表明, 该系统结构能极大地提高BP网络的学习速

度.

5. 学位论文 冯俊 基于FPGA的神经网络的研究与实现-K均值聚类算法的硬件加速 2009

近年来,神经网络的应用越来越广泛,同时人们对其研究也随之更加深入,RBF神经网络是其中一种比较典型的神经网络,其结构简单、训练简洁而且学习收敛速度快,在时间序列分析,模式识别,非线性控制和图像处理等方面有着广泛应用。聚类算法中具有代表性的学习算法-K均值聚类算法是研究RBF神经网络的经典学习算法,通过软件对该算法的改进已经达到一定深度,所以本文另辟蹊径在软件研究的基础上通过硬件对该算法进行加速。

FPGA(Field Programmable Gate Array)是一种可编程的数字集成电路IC(IntegratedCircuit),凭借其自身的优势目前在国内外研究相当广泛,其结构非常灵活,这个特性使得它适合于流水线结构应用和大寄存器的应用,FPGA可以用来实现任何硬件设计,FPGA还能用来实现“处理器加软件”的应用,特别是那些需要快速处理输入数据流的应用中,所以对该算法的研究首先考虑利用FPGA对该算法进行硬件加速,相比于硬件加速来说软件受到处理过程复杂以及软件耗时的局限,使得对算法本身的改进还有很大的空间。

本文从分析算法本身特点着手,从两个方面论证了利用FPGA对该算法进行硬件加速的可行性,首先算法必须具备需要快速处理输入的大量数据流,其次算法本身具有一定的并行性,K均值聚类算法本身的特性非常符合以上特点,其中对于并行性方面,软件对算法的数据只能逐个依次循环处理,直至处理结束,而在硬件处理过程中可以同时处理一个以上的数据,即硬件可以实现算法的并行处理,当然这需要消耗一定的硬件资源,在对资源与速度的综合分析后可得到最优值,在这种前提下可以使得K均值聚类算法在很大程度上提高速度,并提升处理效率,在算法本身的速度得到相当的提高同时,其必然对辅助研究RBF神经网络、图像处理等聚类算法起着至关重要的领域有相当的帮助。本文第四章详细介绍了硬件加速K均值聚类算法的细节,设计采用自上而下的设计方法,在模块的设计中应用了经典的有限状态机的原理,文章对总的大模块以及重要的小模块分别进行了仿真测试与逻辑综合,整个设计是在Xilinx公司的ISE环境下编写的,文章末尾对系统的仿真测试结果与算法在软件上的实现进行了比较,验证了对算法进行硬件加速的可行性和必要性,文章的最后对未来的工作进行了展望。

6. 学位论文 杨志刚 基于FPGA的协处理器设计及RBF神经网络硬件实现的研究 2007

FPGA协处理器是一种新的协处理器结构,它的研究开始于90年代。总的来说FPGA的计算核心是复杂的可编程逻辑集成电路矩阵,区别于传统计算机的顺序指令执行,改变了传统计算及计算模式,为一些高速运算实现提供了新的方法。

本文主要研究的内容是基于FPGA的可重复配置并行协处理器的设计和RBF径向基神经网络硬件实现。通过在测试系统上的实验数据和绘制的辨识曲线,说明了FPGA协处理器硬件实现算法的计算效果。

全文主要内容简要介绍如下:

首先,介绍FPGA的结构、设计方法和神经网络实现技术的现状。

其次,描述了基于PMC板规范的FPGA并行协处理器的设计,给出了系统整体结构设计构想,详细叙述了各个硬件单元的设计,实现了PCB电路板设计。

。

再次,分析了RBF径向基神经网络运算的并行性,并且改进了RBF网络算法的计算流程,提高了算法的并行处理程度。按照自顶向下的设计思想,设计了RBF网络算法硬件实现的系统结构,并对分解后的基本功能单元模块进行了详细的阐述和设计。实现了RBF网络的FPGA硬件设计。

另外,集成了硬件实现的RBF网络算法和系统局部总线接口功能块,构成了协处理器算法程序,程序下载到FPGA协处理器板。并开发了上位操作系统的驱动程序。

最后,以MEN公司的A15PowerPC板为母板,集成FPGA协处理器构成了一个并行处理系统。并以此为核心连接两台开发计算机构成了测试运行系统。通过模型辨识实验,对并行处理系统性能进行了分析。

7. 期刊论文 张海燕,李欣,田书峰, Zhang Hai-yan, Li Xin, Tian Shu-feng 基于BP神经网络的仿真线设计及其

FPGA实现 -电子与信息学报2007, 29(5)

该文提出了一种采用BP神经网络实现仿真线的方法.首先采用遗传算法优化神经网络结构,用离线训练后的BP神经网络逼近传输线的传递函数,然后用STAM算法以较少的存储空间实现BP神经网络的激励函数近似,进而用FPGA和D/A转换器进行硬件实现.文中基于FPGA对长度为10000 m,特性阻抗为55 Ω 的同轴电缆进行了仿真线的硬件实现,实验结果验证了该方法的有效性.该方法可以推广到传递函数未知的传输网络的仿真应用中.

8. 学位论文 袁娟娟 求解二次规划问题的基于LVI的原一对偶神经网络FPGA设计和实现 2007

在实际应用中,很多问题可以归结为二次规划问题的求解.反馈神经网络是实时求解二次规划问题的一条非常有效的途径.反馈神经网络的实现包括软件实现和全硬件实现.由于FPGA器件工作速度快,一般可以达到几百兆赫兹,基于FPGA的反馈神经网络可以快速求解二次规划问题,因此本文研究基于FPGA的反馈神经网络的全硬件实现具有非常重要的意义.本文的主要工作如下:

1. 结合二次规划问题,对描述基于LVI的原、对偶神经网络模型的动态方程进行推导,并在此基础上给出详细的模块框图。

2. 讨论了电子设计自动化技术(Electronic. Design Automation, EDA)、设计语言Verilog HDL和VHDL、FPGA器件厂商和型号的选择、系统开发环境Quartus II和仿真工具ModelSim。

3. 根据第二章中详细的模块框图,给出网络模型的顶层结构图,采用硬件描述语言Verilog: HDL和VHDL对各个子模块和系统顶层模块进行设计,并采用ModelSim仿真工具对各个子模块进行仿真,最后完成顶层模块的仿真。

4. 给出二次规划问题实例进行实验,验证设计的正确性。

9. 期刊论文 许先斌,胡婧,吴玉芹, XU Xian-bin, HU Jing, WU Yu-qin 基于神经网络的视频压缩的FPGA实现 -计算机

工程与设计2009, 30(13)

基于PC的传统的视频压缩实时性较差,给视频的实时压缩和传输带来了困难.针对实时视频压缩传输的难点,提出了在FPGA平台上实现基于神经网络的视频压缩的策略.在建立了基于神经网络的视频压缩系统结构的基础上,给出了神经网络算法的设计,同时完成了非线性的神经元激励函数的线性逼近,并给出了典型FPGA模块的设计,最后通过DSPBuilder和Matlab工具对编写的verilog模块加以验证,给出了实验结果。

10. 学位论文 石凯贺 改善纳米铁酸镧与高分子复合湿敏元件输出特性的神经网络方法及其FPGA实现 2005

纳米铁酸镧与高分子制成复合材料湿敏元件.测量元件的灵敏度、湿滞特性、电容特性、阻抗特性、响应恢复时间特性、温度特性.其中测试电压为1V、测试频率为10Hz~100KHz、测试温度为5℃~30℃.结果表明,元件灵敏度较高,在相对湿度11%~97%范围内,元件的电阻变化了四个数量级,最大湿滞约为4%RH,测试频率在100Hz~10KHz范围内,具有较好的线性度,响应时间为90s,恢复时间为100s.温度漂移问题是湿度传感器应用的主要障碍之一.纳米铁酸镧与高分子复合湿敏元件的温度漂移系数在2.5~3%RH/℃.本文用BP神经网络方法改善湿敏元件的温度漂移.首先基于Matlab做定性的分析,通过比较采用不同隐层神经元个数、不同传递函数以及不同学习算法的BP神经网络,得出一个可以使精度达到 $\pm 2\%$ RH的神经网络,网络结构是单隐层、隐层含5个神经元、传递函数是双曲正切函数、学习算法采用加动量项的梯度下降法.接着讨论了硬件实现中有限精度的问题,确定了具体实现时,网络输入、权值、阈值、传递函数的精度.得出网络输入为8位有符号数、权值为16位有符号数、阈值为18位有符号数、传递函数采用CORDIC算法实现、输出是8位有符号数的神经网络.位数的统一,可以进行乘法器、加法器和传递函数的复用,以节约资源.接着应用Matlab做算法级的仿真,得出具体硬件系统实现是正确可行的.通过FPGA硬件实现该BP神经网络.首先采用ModelSim做功能验证和时序验证、Leonard 2002a做综合、Quartus II做布局布线,然后在Altera公司的ACEX1k EP1K50208-3 FPGA验证板上进行验证,其工作频率设置为12MHz.得出一个离线的人工神经网络硬件系统,其精度可以达到 $\pm 3\%$ RH.此系统可以有有效的解决温度漂移对测量带来的影响。

