

引用格式:薛天志,刘百成,陈松.一种脉冲卷积神经网络 VLSI 硬件架构设计[J].微电子学与计算机,2019,36(10):37-41.  
[XUE T Z, LIU B C, CHEN S. A hardware implementation of a spiking convolutional neural network[J]. Microelectronics & Computer, 2019, 36(10): 37-41.]

# 一种脉冲卷积神经网络 VLSI 硬件架构设计

薛天志,刘百成,陈 松

(中国科学技术大学,安徽 合肥 230026)

**摘 要:** 本文设计了一种识别手写体数字的脉冲卷积神经网络数字电路,使用脉冲神经元代替卷积核,并分别对卷积层和池化层设计相应的电路结构,实现全流水线并行.相比于传统的卷积神经网络,在识别 MNIST 数据集时,卷积神经网络的精确度为 98.61%时,脉冲卷积神经网络的精度能达到 98.04%.与相同流水线结构的卷积神经网络相比,脉冲神经网络平均能耗减少约 50%.

**关键词:** 脉冲卷积神经网络电路;手写体识别;数字集成电路

**中图分类号:** TN492

**文献标识码:** A

**文章编号:** 1000-7180(2019)10-0037-05

## A hardware implementation of a spiking convolutional neural network

XUE Tian-zhi, LIU Bai-cheng, CHEN Song

(University of Science and Technology of China, Hefei, 230026, China)

**Abstract:** This paper designed a digital circuit of spiking convolutional neural network for recognizing handwriting number. Spiking neurons are used to replace the convolution kernels in CNN. Corresponding circuit structures are designed for convolution layer and pooling layer respectively to achieve full pipeline parallelism. Compared with the traditional convolutional neural network, when recognizing the MNIST dataset, the accuracy of the two is 98.61% and 98.04% respectively. Compared with CNN that has similar pipeline architecture, spiking convolutional neural network has a 50% reduction in average energy consumption.

**Key words:** spiking neural network; handwriting recognition; Digital IC

### 1 引言

近年来,随着计算能力的提升,神经网络在众多研究领域引起关注,其中发展较为迅速的是卷积神经网络(CNN),但是卷积神经网络的工作方式和人体内的神经传递方式完全不同,所以另一种更接近生物神经元信息传导方式的脉冲神经网络(SC-NN)<sup>[1]</sup>逐渐开始发展起来.脉冲神经网络相比于卷积神经网络,具有更少的计算量,由此带来了高能效,但是由于缺乏有效的训练算法,所以目前并没有

很好的应用场景.目前脉冲神经网络的应用之一是处理卷积神经网络的计算任务,以此来降低功耗,提升能效.

相比于卷积神经网络,脉冲神经网络的理论计算量能减少两个数量级左右<sup>[1]</sup>,所以使用脉冲神经网络处理卷积神经网络的计算任务是一种很有前景的高能效移动端解决方案.一种实现脉冲神经网络的思路是借用卷积神经网的训练方法,训练好网络后将网络权重转换为脉冲神经元之间的连接强度,由此得到一个脉冲卷积神经网络.转换后会伴随着

网络识别率的下降,如何降低网络性能的下降是转换算法的关键问题,一种常用的算法是对网络的权重进行归一化<sup>[2]</sup>,这样在网络中的脉冲分布会更均衡一些,从而使性能的下降不会过大.

在卷积神经网络中,卷积层和池化层是主要的组成部分,设计更适合进行卷积和池化的硬件加速电路对卷积神经网络的硬件性能有重要的作用.同理,在设计脉冲卷积神经网络的时候,如何实现相对应的卷积层和池化层也是重要的研究问题.

2 脉冲卷积神经网络背景知识

2.1 脉冲神经元和卷积核工作原理

为了能够实现脉冲神经元和卷积核的一一对应,对卷积核和神经元都要进行一定的约束,卷积核的计算可以表示成下面的公式:

$$Z = \sum_m \sum_n x(i+m, j+n)w(m, n) + b \quad (1)$$

脉冲神经元功能可以表示成下面的公式:

$$V(t) = V(t-1) + \sum_f w_i s_i(t-1) \quad (2)$$

$$\begin{cases} s(t) = 1, V(t) > V_{th} \\ s(t) = 0, V(t) \leq V_{th} \end{cases} \quad (3)$$

卷积核的功能是将输入  $x$  和卷积核权重  $w$  对应相乘,并加上偏置值后输出到下一级.而脉冲神经元的功能则是在接收到突触前神经元发射过来的脉冲后,将自身的膜电位  $V(t-1)$  依据突触权重改变相应的值,若膜电位超过了阈值电压  $V_{th}$ ,则会向下一级脉冲神经元发射脉冲,否则不发射脉冲.

对比两个结构的公式,脉冲神经元中不含偏置项;脉冲神经元的工作方式决定了输出只能是正值,所以对输出含有负值的卷积核需要双倍的神经元来实现功能.为了避免过高的硬件代价和精度损失,对卷积神经网络模型需要进行约束:(1)取消偏置项;(2)使用 ReLU 激活函数.

在池化层中,由于脉冲神经元很难单独实现最大化池化,所以池化层只能使用平均值池化的方式,这样用单个脉冲神经元即可完成池化功能.

2.2 脉冲卷积神经网络转换算法原理

由卷积神经网络转换成脉冲卷积神经网络的算法有两种,一种是直接进行转换的方法,这种方法对小规模的网络效果很好,而且速度较快.但是在网络规模变大时,精度会迅速下降.另一种算法是基于归一化的思想,对每一层的神经元权重进行归一化,这样在网络中主要的信息会更容易通过脉冲传递到下一层.

进行直接转换时,脉冲神经元会产生过激活和欠激活.过激活是指脉冲神经元很容易发射脉冲,在同一时刻,网络中有数量较多的脉冲,关键的信息淹没在大量不重要的信息中,这既导致网络的精度收到影响,也使网络有很多无意义的能耗产生.而欠激活则是指神经元要经过很长的时间才能发射脉冲,当网络的层数很大时,这种时延影响是按指数增长的,导致网络的识别时间过长甚至无法识别,也导致在硬件实现的时候,静态功耗占比过大,整个网络的能效反而很低.这二者均是严重影响脉冲卷积神经网络性能的原因.归一化算法则能很好的降低这二者的影响,在延迟不是很大的情况下,能达到和卷积神经网络相差不大的精确度.

由于直接转换的方法太不稳定,以至于在手写体规模的网络上精度即开始大幅度下降,所以我们采用归一化的方法来进行卷积核权重到脉冲神经元连接强度的转换,归一化算法如图 1 所示,可以简单的描述如下:

- (1)求出卷积神经网络某一层中每一个卷积核的正权重和,并选出其中最大的和.
- (2)将这一层的权重均除以第一步得到的最大权重和.
- (3)将转换后的权重作为对应的脉冲神经元的突触权重.

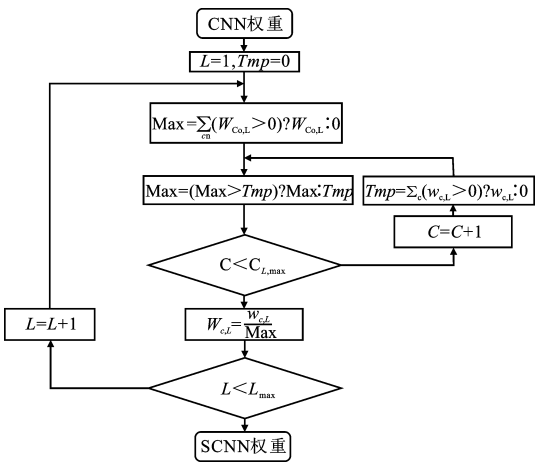


图 1 归一化算法流程

进行归一化算法后,把脉冲神经元的阈值电压设为 1,即能保证一个时钟周期最多只能发射一个脉冲,与卷积神经网络相比传递的数据的损失不大,对分类有重要作用的脉冲更易早一些发射出来,而不重要的脉冲则晚一些才会发射,对精度有很大的提升.

这种归一化算法只和训练好的网络模型有关,

还有一种思路是按照数据集来进行归一化,即网络模型训练好后,对数据集进行一次识别,统计在识别时每一层能达到的最大值,并以这个值代替原归一化算法中的权重和最大值,进行归一化.这种改进的归一化算法虽然在测试集上能有较好的性能,但是在识别非数据集的数据时,稳定性并不如原始的归一化算法,所以在本文中,采用的算法是更有稳定性的原始的归一化算法.

3 脉冲卷积神经网络电路设计

本文所用到的网络模型为手写体数字识别的经典模型 LeNet-5<sup>[3]</sup>的简化模型,以其为基础进行脉冲卷积神经网络转换.网络的基本结构:第一层由 6 个卷积核组成;第二层为  $2 \times 2$  池化;第三层是部分连接层,由 60 个卷积核进行部分连接,最终产生 16 个输出;第四层为  $2 \times 2$  池化;第五层是全连接卷积层,有 120 个输出,最后一层为全连接层,直接连接到 10 个输出.下面将分别从基本的神经元电路和总体电路来介绍本文的脉冲卷积神经网络电路架构.

3.1 卷积层和全连接层脉冲神经元电路结构

脉冲神经元的电路结构如图 2 所示,整个神经元的权重采用分布式存储的方式,将每个通道的权重分别存在一个小的片上存储器中.在电路工作的时候,对应通道的神经元突触权值暂存在寄存器  $w_1-w_{25}$  中,有脉冲输入时,二选一选择器会选通,将权值经过加法树累加,和上一时刻的膜电位  $V(t-1)$  相加,得到新的膜电位.若新的膜电位大于阈值  $V_{th}$ ,则发射脉冲(outspike),同时将膜电位重置为  $V_r$ ,否则不发射脉冲,保留新的膜电位.更新后的膜电位将存储到片上存储器中,在下一次该神经元进行动作的时候作为旧的膜电位缓存到  $V(t-1)$ .

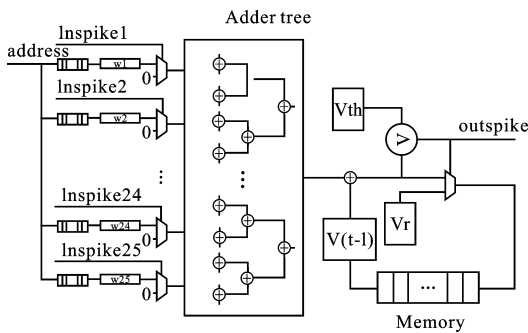


图 2 脉冲神经元电路结构

采用分布式存储可以很容易的对硬件进行分时

复用,从而在有限的硬件资源上实现大量的脉冲神经元.处理完数据后,将  $w_1-w_{25}$  的值清零,并重新在存储器中读取新的权重,  $V(t-1)$  也同时读取对应的上一时刻神经元膜电位,即可实现多个神经元共享相同的硬件资源.

在卷积神经网络中,全连接层也是一个重要的组成部分,全连接层的特点通常是输入数目很大,但是对数据没有复用,我们仍可以使用这种结构来处理,通过对  $w_1-w_{25}$  寄存器的值进行配置,每个周期处理全连接层的一部分数据,然后重新读取权重到  $w_1-w_{25}$  中,用多个时钟周期来实现全连接层的功能.

在脉冲卷积神经网络中,并没有乘法操作,只有膜电位的加减,为了简化电路,降低电路的功耗,一种方法是将浮点的权重进行定点化,根据前人的研究<sup>[4]</sup>,8 bit 数定点化对网络精确度的影响可以忽略不计.另一种办法是采用补码运算,通过补码的方式可以将加法和减法的操作统一起来,对加法器的要求降低,从而降低加法器电路的复杂度,进而降低电路的功耗.

3.2 池化层电路结构

卷积神经网络中有两种常见的池化方式:最大值池化和平均值池化.最大值池化的特点是选择区域内最大的值传到下一层,即将对分类关键的信息传送到下一层.在脉冲神经网络中,实现类似的功能需要将最早发出脉冲的脉冲序列发送到下一级.但是在实际测试中,当采用最大值池化的方式时,将在一定程度上影响整个网络的精确度.

所以采用平均值池化的方式在电路上更容易设计,实现电路和卷积层的电路相似,但是可以进一步简化,以  $2 \times 2$  池化为例,一个池化电路只需要有四个输入,每个输入的权重为 1,当膜电位超过 4 后即向下一层发射一个脉冲.在池化的时候权重并不需要变动,但是相应的膜电位还是要进行存储以便于分时处理.

3.3 总体电路结构

为了在网络中实现数据的复用,需要一些中间电路来对数据进行缓存,在用 LeNet-5 网络处理 MNIST 数据集时,输入端首先要缓存好五行数据,然后处理前五列数据,处理后丢掉第一列数据,并读入一列新数据,中间四列在下次处理还会用到.当处理完一行数据后,同样将第一行的数据丢掉,读入新的一行,中间的四行仍需要复用.在总体图中的数据输入部分的电路就是实现这个功能的.通过五行首



冲神经元只有接收到脉冲才会动作,网络中的脉冲数很少,所以动态功耗很低,相比之下静态功耗的占比显著提高. 对识别一幅图的平均能耗进行分析,采用脉冲率编码对图像编码后,脉冲网络处理一幅图的时间大约是卷积网络的 80 倍,但由于脉冲神经网络中的运算数明显少于卷积神经网络,所以在 MNIST 数据集上平均识别一幅图的能耗大概是卷积网络的 50%,有很明显的能效优势.

5 结束语

介绍了脉冲神经网络的基本原理,重点介绍了脉冲神经网络的硬件电路实现. 设计一种可以时分复用的脉冲神经元电路来实现卷积层和全连接层,并针对不同的池化方式设计了池化层电路. 在 SMIC40nm 工艺库下,本文介绍的脉冲卷积神经网络硬件电路相比于卷积神经网络电路能耗降低 50%左右.

参考文献:

[1] CAO Y, CHEN Y, KHOSLA D. Spiking deep convolutional neural networks for energy-efficient object recognition[J]. International Journal of Computer Vision, 2015, 113(1): 54-66.

[2] DIEHL P U, NEIL D, BINAS J, et al. Fast-classif-

ying, high-accuracy spiking deep networks through weight and threshold balancing[C]//2015 International Joint Conference on Neural Networks (IJCNN). IEEE, 2015: 1-8.

[3] LECUN Y, JACKEL L D, BOTTOU L, et al. Learning algorithms for classification: A comparison on handwritten digit recognition[J]. Neural networks: the statistical mechanics perspective, 1995, 261: 276.

[4] ANWAR S, HWANG K, SUNG W. Fixed point optimization of deep convolutional neural networks for object recognition[C]//2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2015: 1131-1135.

[5] JIA Y, SHELHAMER E, DONAHUE J, et al. Caffe: Convolutional architecture for fast feature embedding[C]//Proceedings of the 22nd ACM international conference on Multimedia. ACM, 2014: 675-678.

作者简介:

薛天志 男,(1993),硕士研究生. 研究方向为数字集成电路设计. E-mail:qtqtyyy@mail.ustc.edu.cn

刘百成 男,(1994),硕士研究生. 研究方向为数字集成电路设计.

陈松 男,(1979),博士,副教授. 研究方向为高效能计算, VLSI 计算机辅助设计.