



(12) 发明专利

(10) 授权公告号 CN 108647045 B

(45) 授权公告日 2021.10.01

(21) 申请号 201810229334.4

G06N 20/00 (2019.01)

(22) 申请日 2018.03.20

(56) 对比文件

(65) 同一申请的已公布的文献号
申请公布号 CN 108647045 A

CN 104915386 A, 2015.09.16

CN 107122798 A, 2017.09.01

CN 106875013 A, 2017.06.20

(43) 申请公布日 2018.10.12

CN 103050116 A, 2013.04.17

(73) 专利权人 科大讯飞股份有限公司

CN 107480771 A, 2017.12.15

地址 230000 安徽省合肥市高新开发区望江西路666号

CN 107729984 A, 2018.02.23

CN 106227851 A, 2016.12.14

(72) 发明人 江宏武 于振华 张致江 王智国 胡国平

CN 107437245 A, 2017.12.05

US 2018005676 A1, 2018.01.04

(74) 专利代理机构 北京维澳专利代理有限公司 11252

王羽. 基于FPGA的卷积神经网络应用研究. 《中国优秀硕士学位论文全文数据库 信息科技辑》. 2017,

代理人 王立民

审查员 郑攀

(51) Int. Cl.

G06F 9/38 (2006.01)

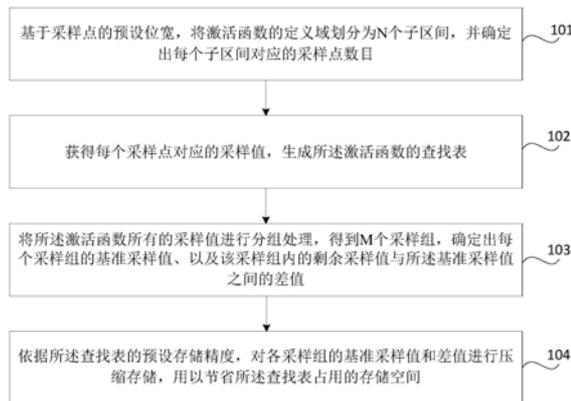
权利要求书2页 说明书8页 附图4页

(54) 发明名称

激活函数的实现方法及装置、存储介质、电子设备

(57) 摘要

本公开提供一种激活函数的实现方法及装置、存储介质、电子设备。该方法包括：基于采样点的预设位宽，将激活函数的定义域划分为N个子区间，并确定出每个子区间对应的采样点数目；获得每个采样点对应的采样值，生成所述激活函数的查找表；将所述激活函数所有的采样值进行分组处理，得到M个采样组，确定出每个采样组的基准采样值、以及该采样组内的剩余采样值与所述基准采样值之间的差值，所述差值消耗的存储位宽小于该差值对应的剩余采样值消耗的存储位宽；依据所述查找表的预设存储精度，对各采样组的基准采样值和差值进行压缩存储，用以节省所述查找表占用的存储空间。如此方案，可以在确保输出精度的同时，降低资源消耗，节省硅片面积。



1. 一种激活函数的实现方法,其特征在于,所述方法包括:

基于采样点的预设位宽,将激活函数的定义域划分为N个子区间,并确定出每个子区间对应的采样点数目;

获得每个采样点对应的采样值,生成所述激活函数的查找表;

将所述激活函数所有的采样值进行分组处理,得到M个采样组,确定出每个采样组的基准采样值、以及该采样组内的剩余采样值与所述基准采样值之间的差值,所述差值消耗的存储位宽小于该差值对应的剩余采样值消耗的存储位宽;

依据所述查找表的预设存储精度,对各采样组的基准采样值和差值进行压缩存储,用以节省所述查找表占用的存储空间。

2. 根据权利要求1所述的方法,其特征在于,所述将所述激活函数所有的采样值进行分组处理,得到M个采样组,包括:

根据保存所述查找表的物理存储器的位宽,确定所述采样组包括的采样值的数目,所述物理存储器的位宽不小于所述采样组的整体位宽,所述采样组的整体位宽为所述基准采样值的存储位宽以及该采样组中所有差值的存储位宽之和。

3. 根据权利要求1或2所述的方法,其特征在于,针对与所述激活函数具有映射关系的关联激活函数,所述方法还包括:

基于所述激活函数与所述关联激活函数之间的映射关系、以及所述激活函数的查找表,实现所述关联激活函数。

4. 一种激活函数的实现装置,其特征在于,所述装置包括:

子区间划分模块,用于基于采样点的预设位宽,将激活函数的定义域划分为N个子区间,并确定出每个子区间对应的采样点数目;

查找表生成模块,用于获得每个采样点对应的采样值,生成所述激活函数的查找表;

采样组分组模块,用于将所述激活函数所有的采样值进行分组处理,得到M个采样组,确定出每个采样组的基准采样值、以及该采样组内的剩余采样值与所述基准采样值之间的差值,所述差值消耗的存储位宽小于该差值对应的剩余采样值消耗的存储位宽;

查找表压缩模块,用于依据所述查找表的预设存储精度,对各采样组的基准采样值和差值进行压缩存储,用以节省所述查找表占用的存储空间。

5. 根据权利要求4所述的装置,其特征在于,

所述采样组分组模块,用于根据保存所述查找表的物理存储器的位宽,确定所述采样组包括的采样值的数目,所述物理存储器的位宽不小于所述采样组的整体位宽,所述采样组的整体位宽为所述基准采样值的存储位宽以及该采样组中所有差值的存储位宽之和。

6. 根据权利要求4或5所述的装置,其特征在于,针对与所述激活函数具有映射关系的关联激活函数,所述装置还包括:

关联激活函数实现模块,用于基于所述激活函数与所述关联激活函数之间的映射关系、以及所述激活函数的查找表,实现所述关联激活函数。

7. 一种存储介质,其中存储有多条指令,其特征在于,所述指令由处理器加载,执行权利要求1至3任一项所述方法的步骤。

8. 一种电子设备,其特征在于,所述电子设备包括:

权利要求7所述的存储介质;以及

处理器,用于执行所述存储介质中的指令。

激活函数的实现方法及装置、存储介质、电子设备

技术领域

[0001] 本公开涉及深度学习技术领域,具体地,涉及一种激活函数的实现方法及装置、存储介质、电子设备。

背景技术

[0002] 目前,云计算、大数据和人工智能技术“铁三角”正广泛应用于人们的日常生活中,为了提高用户的使用体验,企业也在不断地提升相关业务的并发响应能力。基于专用集成电路(英文:Application Specific Integrated Circuit,简称:ASIC)、现场可编程门阵列(英文:Field Programmable Gate Array,简称:FPGA)的异构系统,因具有较高的能效比和吞吐率成为业界研究焦点之一,这包括宏观的产业化和技术生态研究,也包括微观的业务加速逻辑的研究。

[0003] 在微观的业务加速逻辑方面,激活函数的非线性运算电路的实现是绕不开的话题,以Sigmoid函数、Tanh函数为例,用数字电路实现激活函数的方法可以有:泰勒级数展开法、查表法、坐标旋转数字计算机法、分段线性逼近。在物理实现时,上述实现方法均存在以下问题:为了使输出达到可用精度,会消耗较多的资源,占用大量的硅片面积,因此在深度学习领域缺乏实用性。

[0004] 以查表法为例,可以预先将函数值,即采样点对应的采样值保存于RAM或ROM中,这势必会消耗大量的存储器资源。例如,针对Sigmoid函数建立查找表时,若采样区间设为 $[0, 12]$ 、采样间隔设为 1×10^{-3} 、采样精度设为16bit,对应的查找表需要消耗192Kbit存储空间。

[0005] 此外,从输出精度的角度看,查表法的表容量、取样间隔和取样精度之间需要做好权衡,取样间隔越小,取样精度越高,表容量越大,较大容量的表容易获得较高的输出精度,但却需要消耗大量的存储资源;反之,若表容量过低则会降低输出精度。

发明内容

[0006] 本公开的主要目的是提供一种激活函数的实现方法及装置、存储介质、电子设备,有助于在确保输出精度的同时,降低资源消耗,节省硅片面积。

[0007] 为了实现上述目的,本公开提供一种激活函数的实现方法,所述方法包括:

[0008] 基于采样点的预设位宽,将激活函数的定义域划分为N个子区间,并确定出每个子区间对应的采样点数目;

[0009] 获得每个采样点对应的采样值,生成所述激活函数的查找表;

[0010] 将所述激活函数所有的采样值进行分组处理,得到M个采样组,确定出每个采样组的基准采样值、以及该采样组内的剩余采样值与所述基准采样值之间的差值,所述差值消耗的存储位宽小于该差值对应的剩余采样值消耗的存储位宽;

[0011] 依据所述查找表的预设存储精度,对各采样组的基准采样值和差值进行压缩存储,用以节省所述查找表占用的存储空间。

[0012] 可选地,如果所述激活函数以Y轴为对称轴,针对 $X \geq 0$ 的区间划分出的子区间数目

$$N = (W_0 - W_1 - W_2) + 1;$$

[0013] 第*i*个子区间对应的取值范围为 $[X_{i\min}, X_{i\max})$:

$$[0014] \quad X_{i\min} = \begin{cases} 0, & i = 0 \\ i - 2^{i-(W_0-W_1-W_2)}, & \text{others} \end{cases}$$

$$[0015] \quad X_{i\max} = \begin{cases} \infty, & i = (W_0 - W_1 - W_2) \\ i - 2^{i-(W_0-W_1-W_2)+1}, & \text{others} \end{cases}$$

[0016] 第*i*个子区间对应的采样点数目为 K_i :

$$[0017] \quad K_i = \begin{cases} 1, & i = (W_0 - W_1 - W_2) \\ 2^{(W_0-W_1-W_2)-i}, & \text{others} \end{cases}$$

[0018] 其中, W_0 为预设位宽, W_1 为符号位的位宽, W_2 为整数位的位宽。

[0019] 可选地,所述将所述激活函数所有的采样值进行分组处理,得到M个采样组,包括:

[0020] 根据保存所述查找表的物理存储器的位宽,确定所述采样组包括的采样值的数目,所述物理存储器的位宽不小于所述采样组的整体位宽,所述采样组的整体位宽为所述基准采样值的存储位宽以及该采样组中所有差值的存储位宽之和。

[0021] 可选地,针对与所述激活函数具有映射关系的关联激活函数,所述方法还包括:

[0022] 基于所述激活函数与所述关联激活函数之间的映射关系、以及所述激活函数的查找表,实现所述关联激活函数。

[0023] 本公开提供一种激活函数的实现装置,所述装置包括:

[0024] 子区间划分模块,用于基于采样点的预设位宽,将激活函数的定义域划分为N个子区间,并确定出每个子区间对应的采样点数目;

[0025] 查找表生成模块,用于获得每个采样点对应的采样值,生成所述激活函数的查找表;

[0026] 采样组分组模块,用于将所述激活函数所有的采样值进行分组处理,得到M个采样组,确定出每个采样组的基准采样值、以及该采样组内的剩余采样值与所述基准采样值之间的差值,所述差值消耗的存储位宽小于该差值对应的剩余采样值消耗的存储位宽;

[0027] 查找表压缩模块,用于依据所述查找表的预设存储精度,对各采样组的基准采样值和差值进行压缩存储,用以节省所述查找表占用的存储空间。

[0028] 可选地,如果所述激活函数以Y轴为对称轴,针对 $X \geq 0$ 的区间划分出的子区间数目

$$N = (W_0 - W_1 - W_2) + 1;$$

[0029] 第*i*个子区间对应的取值范围为 $[X_{i\min}, X_{i\max})$:

$$[0030] \quad X_{i\min} = \begin{cases} 0, & i = 0 \\ i - 2^{i-(W_0-W_1-W_2)}, & \text{others} \end{cases}$$

$$[0031] \quad X_{i\max} = \begin{cases} \infty, & i = (W_0 - W_1 - W_2) \\ i - 2^{i-(W_0-W_1-W_2)+1}, & \text{others} \end{cases}$$

[0032] 第*i*个子区间对应的采样点数目为 K_i :

$$[0033] \quad K_i = \begin{cases} 1, & i = (W_0 - W_1 - W_2) \\ 2^{(W_0-W_1-W_2)-i}, & \text{others} \end{cases}$$

[0034] 其中, w_0 为预设位宽, w_1 为符号位的位宽, w_2 为整数位的位宽。

[0035] 可选地, 所述采样组分组模块, 用于根据保存所述查找表的物理存储器的位宽, 确定所述采样组包括的采样值的数目, 所述物理存储器的位宽不小于所述采样组的整体位宽, 所述采样组的整体位宽为所述基准采样值的存储位宽以及该采样组中所有差值的存储位宽之和。

[0036] 可选地, 针对与所述激活函数具有映射关系的关联激活函数, 所述装置还包括:

[0037] 关联激活函数实现模块, 用于基于所述激活函数与所述关联激活函数之间的映射关系、以及所述激活函数的查找表, 实现所述关联激活函数。

[0038] 本公开提供一种存储介质, 其中存储有多条指令, 所述指令由处理器加载, 执行上述激活函数的实现方法的步骤。

[0039] 本公开提供一种电子设备, 所述电子设备包括:

[0040] 上述的存储介质; 以及

[0041] 处理器, 用于执行所述存储介质中的指令。

[0042] 本公开方案中, 可以结合预设位宽, 对激活函数进行子区间划分, 进而确定出每个子区间对应的采样点数目, 以此来权衡查找表的表容量、取样间隔和取样精度; 此外, 还可以通过分组保存基准采样值、剩余采样值与基准采样值之间差值的方式, 实现压缩查找表。如此方案, 相对现有技术, 可以在确保输出精度的同时, 降低资源消耗, 节省硅片面积。

[0043] 本公开的其他特征和优点将在随后的具体实施方式部分予以详细说明。

附图说明

[0044] 附图是用来提供对本公开的进一步理解, 并且构成说明书的一部分, 与下面的具体实施方式一起用于解释本公开, 但并不构成对本公开的限制。在附图中:

[0045] 图1为本公开方案激活函数的实现方法的流程示意图;

[0046] 图2为现有技术实现Sigmoid函数时的误差分布示意图;

[0047] 图3为现有技术实现Sigmoid函数时的面积占用情况示意图;

[0048] 图4为本公开方案实现Sigmoid函数时的误差分布示意图;

[0049] 图5为本公开方案实现Sigmoid函数和Tanh函数时的面积占用情况示意图;

[0050] 图6为本公开方案激活函数的实现装置的构成示意图;

[0051] 图7为本公开方案用于实现激活函数的电子设备的结构示意图。

具体实施方式

[0052] 以下结合附图对本公开的具体实施方式进行详细说明。应当理解的是, 此处所描述的具体实施方式仅用于说明和解释本公开, 并不用于限制本公开。

[0053] 参见图1, 示出了本公开激活函数的实现方法的流程示意图。可以包括以下步骤:

[0054] S101, 基于采样点的预设位宽, 将激活函数的定义域划分为N个子区间, 并确定出每个子区间对应的采样点数目。

[0055] 作为一种示例, 本公开方案中的激活函数可以是非线性曲线拟合的函数, 例如, Sigmoid函数、Tanh函数等, 本公开方案对此可不做具体限定。

[0056] 本公开方案中, 采样点的预设位宽指的是用多少个二进制表示采样点在X轴的值,

也即激活函数中参数 x 的二进制位数。举例来说,预设位宽可以是16bit、8bit等,具体可结合实际应用需要而定,本公开方案对此可不做限定。

[0057] 通常,参数 x 可以包括以下三部分:符号位、整数位、小数位。其中,符号位用于表示参数 x 的取值位于 X 轴的正半轴,还是负半轴,例如,可以用符号“0”表示 X 轴的正半轴,用符号“1”表示 X 轴的负半轴,本公开方案对此可不做具体限定。

[0058] 作为一种示例,可以根据激活函数的饱和状态,确定整数位的位宽。以Sigmoid函数为例,在 $|x| > 12$ 以后处于饱和状态,无需再采样,故可以用4bit来表示参数 x 的整数部分。

[0059] 对应地,可以根据预设位宽 W_0 、符号位的位宽 W_1 、整数位的位宽 W_2 ,确定出小数位的位宽为 $(W_0 - W_1 - W_2)$ 。仍以Sigmoid函数为例,若预设位宽为16bit,符号位的位宽为1bit,整数位的位宽为4bit,则小数位的位宽可以为11bit。

[0060] 本公开方案中,对于参数 x 的小数部分来说,每个比特位相对于 X 轴上的一个标记,即 X 轴上共有 $(W_0 - W_1 - W_2)$ 个标记,为了得到 $(W_0 - W_1 - W_2)$ 个标记,就需要划分出 $(W_0 - W_1 - W_2) + 1$ 个小区间。以上文所举Sigmoid函数为例,考虑到函数以 Y 轴为对称轴,故可针对 $X \geq 0$ 的区间进行采样,具体可以将 $X \geq 0$ 的区间划分成 $(W_0 - 5) + 1$ 个子区间。

[0061] 作为一种示例,子区间 i 对应的参数 x 的取值范围可以为 $[X_{i\min}, X_{i\max})$,子区间 i 对应的采样点数目为 K_i 。其中,

$$[0062] \quad X_{i\min} = \begin{cases} 0, & i = 0 \\ i - 2^{i-(W_0-W_1-W_2)}, & \text{others} \end{cases}$$

$$[0063] \quad X_{i\max} = \begin{cases} \infty, & i = (W_0 - W_1 - W_2) \\ i - 2^{i-(W_0-W_1-W_2)+1}, & \text{others} \end{cases}$$

$$[0064] \quad K_i = \begin{cases} 1, & i = (W_0 - W_1 - W_2) \\ 2^{(W_0-W_1-W_2)-i}, & \text{others} \end{cases} \circ$$

[0065] 也就是说,对于每个子区间来说,可能具有不同的长度,但子区间内各采样点之间可以是均匀分布的。

[0066] 上文所举Sigmoid函数的示例中,针对 $X \geq 0$ 的区间可以划分出12个子区间,各子区间对应的取值范围和采样点数可参见下表1所示。

[0067] 表1

子区间 i	参数 x 的取值范围		采样点数 K_i
	X_{imin}	X_{imax}	
0	0	0.999511719	2048
1	0.999511719	1.999023438	1024
2	1.999023438	2.998046875	512
3	2.998046875	3.99609375	256
4	3.99609375	4.9921875	128
5	4.9921875	5.984375	64
6	5.984375	6.96875	32
7	6.96875	7.9375	16
8	7.9375	8.875	8
9	8.875	9.75	4
10	9.75	10.5	2
11	10.5	11	1
12	11	∞	1

[0069] S102, 获得每个采样点对应的采样值, 生成所述激活函数的查找表。

[0070] 确定出采样点后, 便可通过激活函数的表达式计算出采样点对应的采样值, 进而基于采样值生成激活函数对应的查找表。激活函数的表达式、计算采样值的过程可参照相关技术实现, 此处不做详述。

[0071] 针对表1所举示例, 对应第12个子区间时, 激活函数已达饱和状态, 该子区间内采样点对应的采样值可以不存入查找表中, 本公开方案对此可不做具体限定, 可以结合实际应用需求确定是否将该子区间对应的采样值写入查找表中。

[0072] S103, 将所述激活函数所有的采样值进行分组处理, 得到M个采样组, 确定出每个采样组的基准采样值、以及该采样组内的剩余采样值与所述基准采样值之间的差值, 所述差值消耗的存储位宽小于该差值对应的剩余采样值消耗的存储位宽。

[0073] 得到激活函数的查找表后, 为了尽量节省保存查找表的物理存储器消耗的硅片面积, 可以对查找表进行压缩处理。

[0074] 作为一种示例, 可以将激活函数所有的采样点分成M个采样组, 通过保存基准采样值、剩余采样值与基准采样值之间的差值的方式, 实现压缩查找表的目的。如此处理, 可以减小查找表的大小, 进而节省物理存储器占用硅片的面积。

[0075] 在实际应用过程中, 可以采用等分方式得到M个采样组, 即每组包括的采样值的数目相同; 或者, 也可以不进行等分, 本公开方案对此可不做具体限定。

[0076] 需要说明的是,本公开方案中同一采样组内的采样值压缩,都是基于同一个基准采样值实现的,为了以最小的代价进行采样值还原,通常需要一次性从查找表中读出全组的采样值。对应于此,可以根据保存查找表的物理存储器位宽,来确定采样组包括的采样值的数目,通常,物理存储器的位宽不小于采样组的整体位宽,其中,采样组的整体位宽=基准采样值的存储位宽+本采样组中所有差值的存储位宽。

[0077] 以上表1为例,共计4096个采样点,即查找表中需要保存4096个采样值,若以等分方式进行分组,且每个采样组包括8个采样值,则可划分得到 $4096/8=512$ 个采样组。对于第 j 个采样组来说,作为一种示例,可以将组内第1个采样点对应的采样值确定为基准采样值,记为 B_j ;第2~8个采样点对应的采样值作为剩余采样值,与 B_j 之间的差值分别记为 Δ_{j1} 、 Δ_{j2} 、 \dots 、 Δ_{j7} 。

[0078] S104,依据所述查找表的预设存储精度,对各采样组的基准采样值和差值进行压缩存储,用以节省所述查找表占用的存储空间。

[0079] 本公开方案中,查找表的预设存储精度指的是采样值精确保存到什么位,例如,可以保存到小数点后6位,具体可结合实际应用需要而定,本公开方案对此可不做限定。通常,存储精度越高,需要消耗的存储位宽就越大。

[0080] 此外,需要说明的是,本公开方案中组内的差值,均具有共同的基准采样值,故可以节省差值消耗的存储位宽。举例来说,针对采样值序列{0.000113,0.000125,0.000139,0.000146},存储精度达到小数点后第6位,序列中每个元素需要20bit才能直接存储,即每个元素需要消耗20bit,若将基准采样值定为0.0001,则上述序列可以变为{13,25,39,46},此时序列中每个元素只需6bit存储,实现了查找表的大幅压缩。

[0081] 由上文介绍可知,查找表的压缩效率主要受以下因素的影响:采样组包括的采样值的数目、基准采样值的存储位宽、差值的存储位宽。

[0082] 上文所举512个采样组的示例中,如果预设存储精度对应的比特位为16bit,采样组内用16bit存储 B_j ,用8bit存储各差值 Δ_{j1} 、 Δ_{j2} 、 \dots 、 Δ_{j7} ,则第 j 个采样组占用72bit存储空间,整个查找表压缩后占用的存储空间为 $72\text{bit} \times 512 = 36\text{Kbit}$ 。

[0083] 作为一种示例,本公开方案还可以实现至少两个激活函数的兼容,进一步节省激活函数消耗的硅片面积。具体地,按照图1所示方法得到激活函数的压缩后查找表,针对与该激活函数具有映射关系的关联激活函数,可以基于激活函数与关联激活函数之间的映射关系、以及激活函数的查找表,实现关联激活函数。

[0084] 举例来说,Sigmoid函数、Tanh函数之间具有如下映射关系:

[0085] $\text{Sigmoid}(x) - 0.5 = 0.5 * \text{Tanh}(0.5x)$

[0086] 本公开方案可以基于上述映射关系实现Sigmoid函数与Tanh函数的兼容,即输入一个参数 x ,可以先通过查找表得到 $\text{Sigmoid}(x)$ 的值或者 $\text{Tanh}(x)$ 的值,然后通过移位和加法运算求得另一激活函数的值,也就是说Sigmoid函数、Tanh函数可以共用一个查找表,相对于现有技术需要单独实现各激活函数,本公开方案有助于节省多个激活函数消耗的硅片面积。此外,利用一个查找表实现Sigmoid函数、Tanh函数的兼容,也使得本公开方案在同时具有Sigmoid函数、Tanh函数的深度学习网络中更具优势。

[0087] 下面以Sigmoid函数为例,进一步对本公开方案的有益效果进行解释说明。

[0088] 现有技术对Sigmoid函数建立查找表时,若采样区间设为 $[0,12]$ 、采样间隔设为1

$\times 10^{-3}$ 、采样精度设为16bit,则查找表需要消耗192Kbit存储空间,对应的误差分布可参见图2所示示意图。

[0089] 现有技术为了使输出达到可用精度,在ASIC或FPGA中实现Sigmoid函数时一般都会占用大量的硅片面积。以在FPGA(型号:KU040,Xilinx)中实现Sigmoid函数为例,面积占用情况可参见图3所示,包括:Sigmoid函数本体占用的面积、测试Sigmoid函数功能占用的面积、剩余的空闲未被占用的面积。

[0090] 对应地,采用本公开方案实现Sigmoid函数时,若参数x的预设位宽为16bit,由表1可知,查找表中需要存储4096个采样值,若查找表的预设存储精度为16bit为例,则需要占用的存储空间为64Kbit,对应的误差分布可参见图4所示示意图。若继续进行查找表压缩,由上文介绍可知,整个查找表压缩后占用的存储空间为36Kbit。

[0091] 对比图2、图4可知,在保证同等误差情况下,现有技术需要192Kbit存储空间,本公开方案只需36Kbit存储空间,有效的降低了对查找表容量的需求,降低了对存储资源的消耗。

[0092] 另外,以本公开方案在FPGA(型号:KU040,Xilinx)中实现Sigmoid函数和Tanh函数为例,面积占用情况可参见图5所示,包括:Sigmoid函数和Tanh函数本体占用的面积、测试Sigmoid函数和Tanh函数功能占用的面积、剩余的空闲未被占用的面积。

[0093] 对比图3、图5可知,本公开方案在物理实现上只需占用少量存储资源和逻辑资源,有效节省了硅片面积,非常适合在ASIC或FPGA中实现。

[0094] 参见图6,示出了本公开激活函数的实现装置的构成示意图。所述装置可以包括:

[0095] 子区间划分模块201,用于基于采样点的预设位宽,将激活函数的定义域划分为N个子区间,并确定出每个子区间对应的采样点数目;

[0096] 查找表生成模块202,用于获得每个采样点对应的采样值,生成所述激活函数的查找表;

[0097] 采样组分组模块203,用于将所述激活函数所有的采样值进行分组处理,得到M个采样组,确定出每个采样组的基准采样值、以及该采样组内的剩余采样值与所述基准采样值之间的差值,所述差值消耗的存储位宽小于该差值对应的剩余采样值消耗的存储位宽;

[0098] 查找表压缩模块204,用于依据所述查找表的预设存储精度,对各采样组的基准采样值和差值进行压缩存储,用以节省所述查找表占用的存储空间。

[0099] 可选地,如果所述激活函数以Y轴为对称轴,针对 $X \geq 0$ 的区间划分出的子区间数目 $N = (W_0 - W_1 - W_2) + 1$;

[0100] 第i个子区间对应的取值范围为 $[X_{i\min}, X_{i\max})$:

$$[0101] \quad X_{i\min} = \begin{cases} 0, & i = 0 \\ i - 2^{i-(W_0-W_1-W_2)}, & \text{others} \end{cases}$$

$$[0102] \quad X_{i\max} = \begin{cases} \infty, & i = (W_0 - W_1 - W_2) \\ i - 2^{i-(W_0-W_1-W_2)+1}, & \text{others} \end{cases}$$

[0103] 第i个子区间对应的采样点数目为 K_i :

$$[0104] \quad K_i = \begin{cases} 1, & i = (W_0 - W_1 - W_2) \\ 2^{(W_0-W_1-W_2)-i}, & \text{others} \end{cases}$$

[0105] 其中, w_0 为预设位宽, w_1 为符号位的位宽, w_2 为整数位的位宽。

[0106] 可选地, 所述采样组分组模块, 用于根据保存所述查找表的物理存储器的位宽, 确定所述采样组包括的采样值的数目, 所述物理存储器的位宽不小于所述采样组的整体位宽, 所述采样组的整体位宽为所述基准采样值的存储位宽以及该采样组中所有差值的存储位宽之和。

[0107] 可选地, 针对与所述激活函数具有映射关系的关联激活函数, 所述装置还包括:

[0108] 关联激活函数实现模块, 用于基于所述激活函数与所述关联激活函数之间的映射关系、以及所述激活函数的查找表, 实现所述关联激活函数。

[0109] 关于上述实施例中的装置, 其中各个模块执行操作的具体方式已经在有关该方法的实施例中进行了详细描述, 此处将不做详细阐述说明。

[0110] 参见图7, 示出了本公开用于实现激活函数的电子设备300的结构示意图。电子设备300至少可以包括处理器301和存储介质302, 作为一种示例, 处理器301和存储介质302可以通过总线或其它方式连接, 图7所示以通过总线连接为例。处理器301的数量可以为一个或者多个, 图7所示以一个处理器为例。存储介质302所代表的存储设备资源, 用于存储可由处理器301执行的指令, 例如应用程序。此外, 处理器301可以被配置为加载存储介质中的指令, 以执行上述激活函数的实现方法。

[0111] 以上结合附图详细描述了本公开的优选实施方式, 但是, 本公开并不限于上述实施方式中的具体细节, 在本公开的技术构思范围内, 可以对本公开的技术方案进行多种简单变型, 这些简单变型均属于本公开的保护范围。

[0112] 另外需要说明的是, 在上述具体实施方式中所描述的各个具体技术特征, 在不矛盾的情况下, 可以通过任何合适的方式进行组合, 为了避免不必要的重复, 本公开对各种可能的组合方式不再另行说明。

[0113] 此外, 本公开的各种不同的实施方式之间也可以进行任意组合, 只要其不违背本公开的思想, 其同样应当视为本公开所公开的内容。

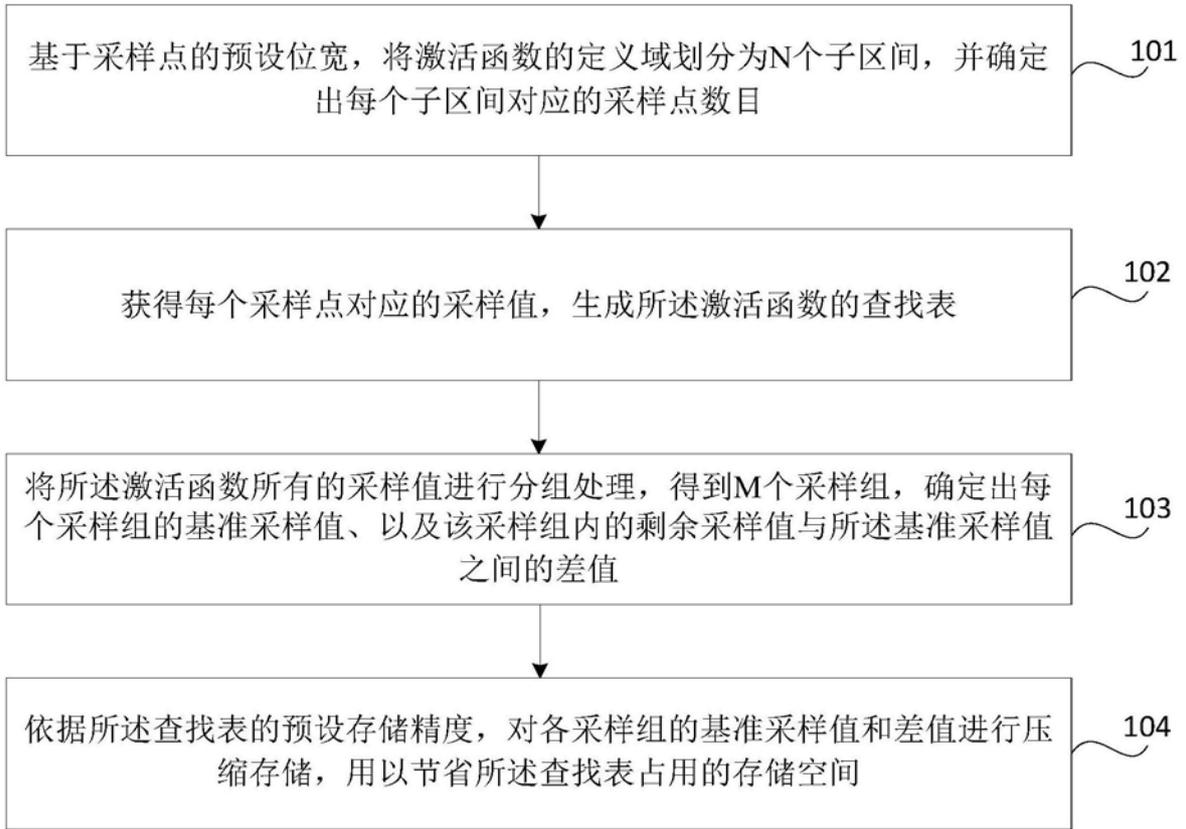


图1

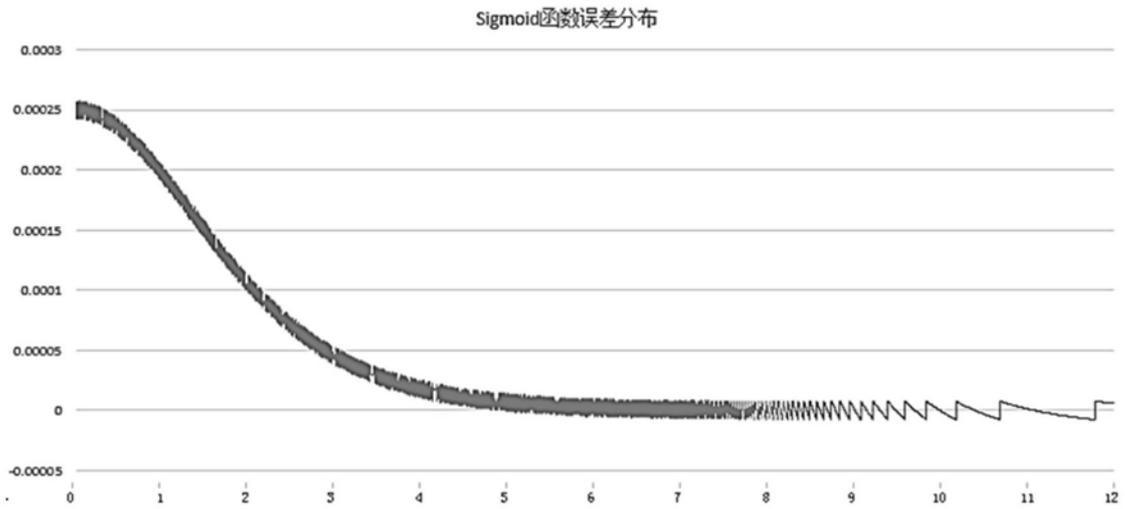


图2

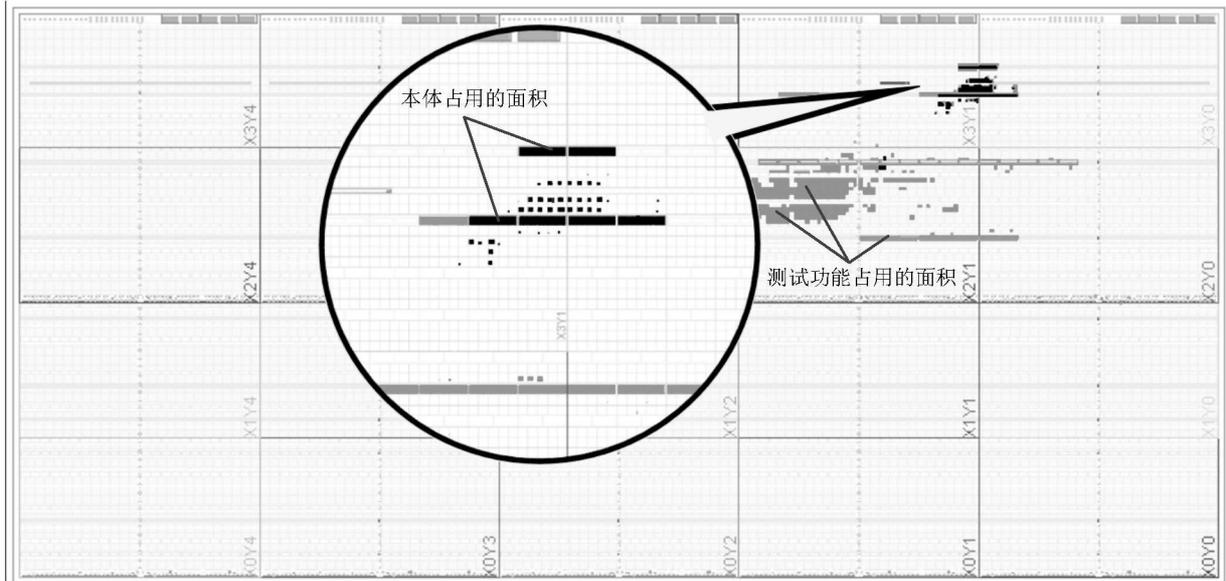


图3

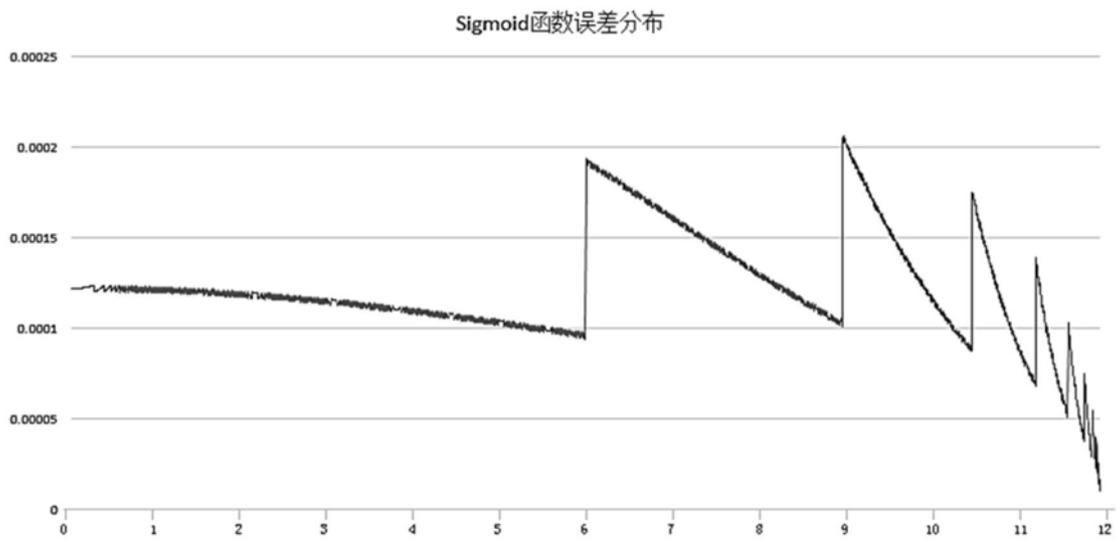


图4

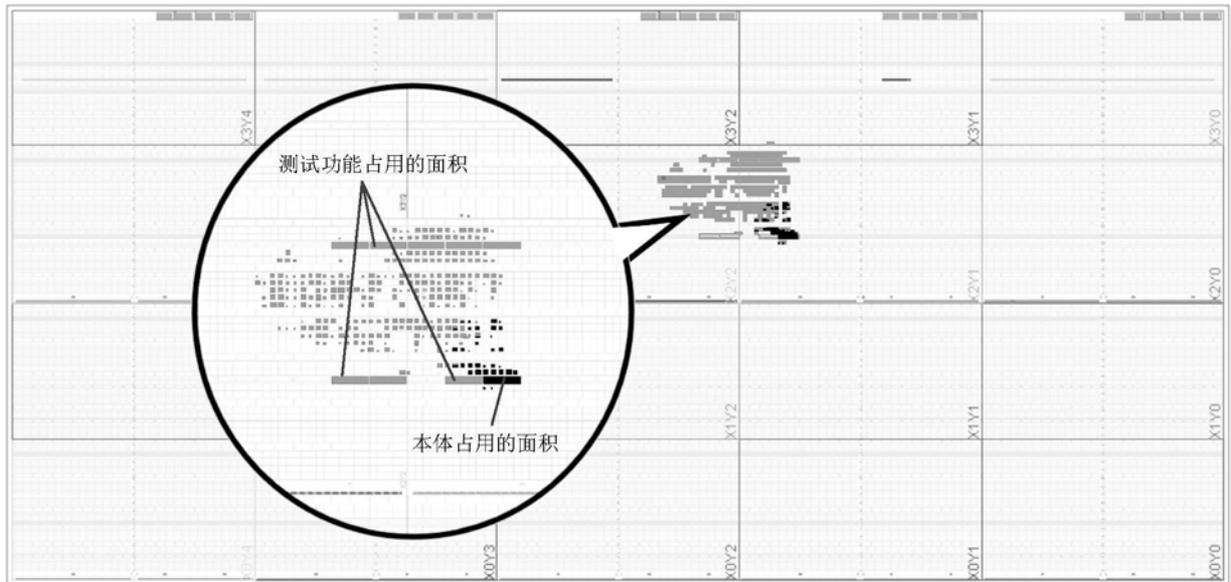


图5



图6

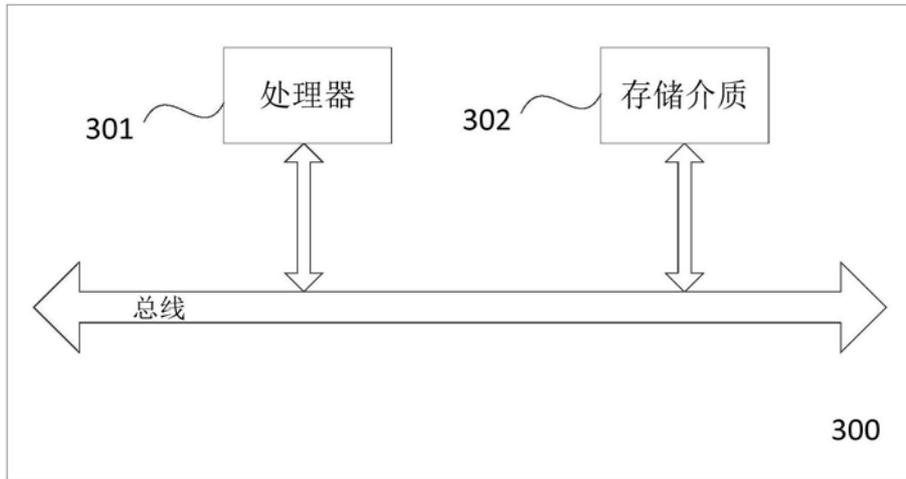


图7