



(12) 发明专利

(10) 授权公告号 CN 111788552 B

(45) 授权公告日 2023. 10. 24

(21) 申请号 201980016230.8

(22) 申请日 2019.02.11

(65) 同一申请的已公布的文献号
申请公布号 CN 111788552 A

(43) 申请公布日 2020.10.16

(30) 优先权数据
62/630,096 2018.02.13 US
16/000,248 2018.06.05 US

(85) PCT国际申请进入国家阶段日
2020.08.31

(86) PCT国际申请的申请数据
PCT/US2019/017472 2019.02.11

(87) PCT国际申请的公布数据
W02019/160804 EN 2019.08.22

(73) 专利权人 特斯拉公司
地址 美国得克萨斯州

(72) 发明人 P·J·班农 K·A·胡德
E·塔尔佩斯

(74) 专利代理机构 北京市金杜律师事务所
11256

专利代理师 马明月

(51) Int.Cl.
G06F 9/06 (2006.01)
G06F 13/00 (2006.01)
G06F 15/80 (2006.01)
G06F 16/53 (2019.01)
G06F 21/71 (2013.01)

(56) 对比文件
US 2006047865 A1, 2006.03.02
US 2006132822 A1, 2006.06.22
US 2006187031 A1, 2006.08.24
US 2008222380 A1, 2008.09.11
US 2015081967 A1, 2015.03.19
US 2016162505 A1, 2016.06.09
US 6725321 B1, 2004.04.20
黄长宾. 二维动态可划分内存多核硬件支持. 《万方数据》. 2008, 全文.

审查员 宋泽宇

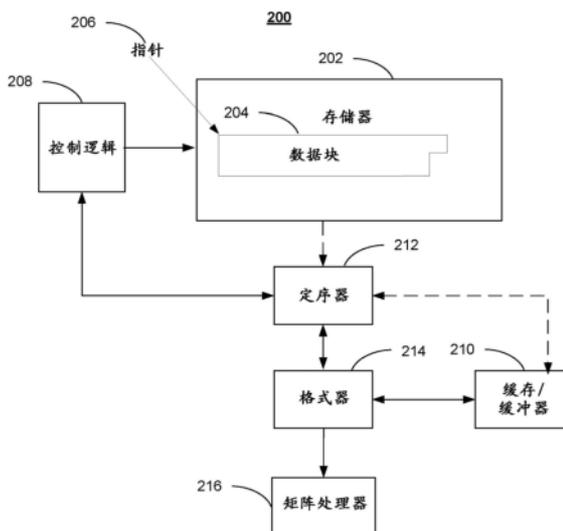
权利要求书2页 说明书7页 附图4页

(54) 发明名称

用于低等待时间硬件存储器的系统和方法

(57) 摘要

提出了允许有效的数据处理的系统和方法, 该系统和方法减少了数据等待时间, 并且因此减少功耗和数据管理成本。在各种实施例中, 这通过使用定序器来实现, 该定序器标识存储器内的第一数据块的地址指针、以及数据的长度, 该数据包括该数据块并且与矩阵处理器的输入有关。定序器随后基于块长度、输入长度和存储器映射计算块计数, 该块计数是要从存储器检索的数据块的数目的代表。使用地址指针, 定序器可以在取决于数据块是否连续的若干周期中从存储器中检索多个数据块。



1. 一种用于在硬件加速处理系统内检索数据的方法,所述方法包括:

根据与所述硬件加速处理系统的至少一个参数有关的所定义的程序,将数据文件存储在存储器内,所述数据文件被划分为具有对应的多个地址指针的多个固定长度块;

生成存储器映射,所述存储器映射标识在多个固定长度块内的被连续地存储的所述多个固定长度块;

响应于读取请求,标识:

数据的长度,所述长度与进入硬件加速处理器中的输入的长度有关;以及

地址指针,所述地址指针在所述存储器内标识用于所述读取请求的第一数据块;

基于所述多个固定长度块的长度、所述输入的所述长度和所述存储器映射中的至少一项,计算块计数和掩码,所述块计数是要从所述存储器检索的数据块的数目的代表,所述掩码是用于所述读取请求的要检索的若干数据块的代表;以及使用所述地址指针从所述存储器检索一个或多个数据块。

2. 根据权利要求1所述的方法,其中所述硬件加速处理器是矩阵处理器。

3. 根据权利要求1所述的方法,进一步包括步骤:将所述一个或多个数据块映射到所述硬件加速处理器的所述输入,使得所述一个或多个数据块是呈多维格式的,并且所述输入是一维的。

4. 根据权利要求1所述的方法,其中所述固定长度块与所述硬件加速处理器的所述输入的所述长度有关。

5. 根据权利要求1所述的方法,其中所述读取请求与固定长度数据块的第一组相对应,所述第一组被连续地存储并且使用单个初始地址指针被读取。

6. 根据权利要求5所述的方法,其中在单个时钟周期中执行所述读取请求。

7. 根据权利要求1所述的方法,其中所述硬件加速处理器的所述长度是固定长度数据块的整数倍。

8. 根据权利要求1所述的方法,进一步包括步骤:计算结束地址,所述结束地址与对应于所述读取请求的所述多个固定长度数据块相关联。

9. 根据权利要求8所述的方法,其中使用所述第一数据块的所述地址指针和所述固定长度数据块的所述长度计算所述结束地址。

10. 一种用于将数据从存储器映射到硬件加速处理系统的系统,所述系统包括:

存储器,所述存储器根据与所述硬件加速处理系统的至少一个参数有关的所定义的程序来存储数据文件,所述数据文件被划分为具有对应的多个地址指针的多个固定长度数据块;

耦合到所述存储器的控制器,所述控制器被配置为根据所述所定义的程序来存储数据并且生成所存储的数据的虚拟映射;

耦合到所述存储器的定序器,所述定序器响应于恢复到读取请求,执行以下步骤:

标识数据的长度和地址指针,所述长度与进入硬件加速处理器中的输入的长度有关,所述地址指针在所述存储器内标识用于所述读取请求的第一数据块;

基于所述多个固定长度数据块的所述长度、所述地址指针和存储器映射中的至少一项,计算块计数和掩码,所述块计数是要从所述存储器检索的数据块的数目的代表,所述掩码是用于所述读取请求的要检索的若干数据块的代表;以及

检索所述多个固定长度数据块的与所述读取请求相对应的子组。

11. 根据权利要求10所述的系统,进一步包括:耦合到所述定序器的格式器,所述格式器基于所述数据的长度,将从存储器检索到的所述多个固定长度数据块映射到所述硬件加速处理器的一维输入。

12. 根据权利要求11所述的系统,其中所述硬件加速处理器是矩阵处理器。

13. 根据权利要求11所述的系统,进一步包括耦合到所述格式器的缓存,所述缓存被配置为存储跨所述硬件加速处理器的多个操作的冗余的数据。

14. 根据权利要求10所述的系统,其中所述存储器的虚拟映射被存储在所述定序器或所述控制器中。

15. 根据权利要求10所述的系统,其中所述多个固定长度数据块与所述硬件加速处理器的所述输入的所述长度有关。

16. 根据权利要求10所述的系统,其中所述读取请求与所述多个固定长度数据块的第一组相对应,所述第一组被连续地存储并且使用单个初始地址指针被读取。

17. 根据权利要求16所述的系统,其中在单个时钟周期中执行所述读取请求。

18. 根据权利要求10所述的系统,进一步包括步骤:计算结束地址,所述结束地址与对应于所述读取请求的所述多个固定长度数据块相关联。

19. 一种用于加速数据处理的系统,所述系统包括:

存储器,所述存储器根据与硬件加速处理系统的至少一个参数有关的所定义的程序来存储多个数据文件;

耦合到所述存储器的控制器,所述控制器被配置为根据所述所定义的程序来存储数据并且生成所存储的数据的虚拟映射;

耦合到所述存储器的定序器,所述定序器被配置为生成与要输入到多维矩阵处理器中的数据块有关的读取请求,所述输入是一维的并且具有第一长度;

耦合到所述定序器的格式器,所述格式器将从所述存储器中检索的多个数据块转换为具有所述第一长度的一维向量;以及

多维矩阵处理器,所述多维矩阵处理器对以第一方向输入的第一组数据和以第二方向输入的第二组数据执行多个数学运算。

20. 根据权利要求19所述的系统,其中所述存储器将多个数据块存储在固定长度数据块中,所述固定长度数据块与所述多维矩阵处理器的至少一个输入的输入长度具有关系。

用于低等待时间硬件存储器的系统和方法

[0001] 相关申请的交叉引用

[0002] 本申请根据35USC§119(e)要求于2018年2月13日提交的、发明名称为“低等待时间硬件存储器管理的系统和方法”、美国专利申请号为62/630,096【代理人案号为20150-2153P(P0827-1PUS)】、将Peter Joseph Bannon、Kevin Altair Hurd和Emil Talpes列为发明人的美国专利申请的优先权,该申请通过引用整体并入本文。

[0003] 本申请还要求于2018年6月5日提交的、发明名称为“低等待时间硬件存储器管理的系统和方法”、美国专利申请号为16/000,248【20150-2153(P0827-2NUS)】、并且将Peter Joseph Bannon、Kevin Altair Hurd和Emil Talpes列为发明人的美国专利申请的优先权,该申请通过引用整体并入本文。

技术领域

[0004] 本公开涉及改进的存储器读取操作,并且更特别地,涉及有效的存储器管理方法和系统,该方法和系统将来自存储器的数据检索与硬件加速处理设备的特性对齐。

背景技术

[0005] 本领域技术人员将认识到在使用硬件加速处理器时来自存储器的高速数据检索的重要性。硬件加速处理系统使得时间敏感应用(诸如自动驾驶、安全性和用户身份验证过程以及本领域技术人员已知的其他技术)内的数据的分析成为可能。在设计这些系统时,重要的是要以足够的速度处理从存储器检索到的数据,以使硬件加速处理器不需要针对该新数据而等待延长的时间段。在许多情况下,归因于处理和实现读取请求所需要的时间,传统的存储器体系结构和方法可以在加速系统中造成瓶颈。

[0006] 图1示出了本领域中已知的普通存储器管理系统。系统100包括中央处理单元(CPU)104、控制器106和存储器110,该中央处理单元(CPU)104通常在通用计算机系统的单片芯片上实现。CPU104是接收和处理图像数据102的普通处理器。图像数据102通常是文件,该文件包含相对大量的像素,该像素被表示为二进制数字并且以数据114的组或块的形式被处理。

[0007] 存储器110是物理存储器设备,例如随机存取存储器(RAM),该物理存储器设备被设计为存储数字数据。存储器110包含独立数据存储位置114,该独立数据存储位置114使用例如特定于硬件的协议被访问,该特定于硬件的协议在存储器地址空间存储器(memory address spacy memory)110中定义和索引物理地址,使得不同的体系结构经常在存储器内低效地存储图像数据102并且在检索该数据中需要多个时钟周期。

[0008] 在操作中,响应于通过CPU 104接收写入命令,存储器控制器106将图像数据102分割为可以在各个独立存储位置处被存储的独立部分114,这些独立部分114可以被分组为可以在存储器110内分散的块。存储位置通常基于存储器110的体系结构、先前已经存储在存储器110中的任何预先存在的数据、所存储的数据的量以及其他情况。各个位置114各自与在存储器110内的存储器地址相关联。

[0009] 将图像数据102存储在存储器110中通常需要使用相对大量的指针124,这些指针124例如以某个字节大小的整数倍来区分物理存储器地址。每个指针114可以指示特定的存储器块124、或在存储器块124内的用于将图像数据102的一部分存储在存储器110内的位置。存储器控制器106将对应于图像102的不同部分的指针124分配给存储器位置114,从而,将单个图像分割为较小的数据的分块,这些分块然后被分布在存储器110内的各个位置处。例如,基于第一指针124,读取第一块114,基于第二指针126,读取第二块116,等等,直到完整的图像数据102被写入并且存储到存储器110中为止。

[0010] 这种现有技术的存储器结构和数据检索方法可以导致不必要地减慢了处理系统的速度的等候时间。类似地,随后如果期望需要从存储器110访问和从存储器110检索或恢复图像,则每个指针114将必须被处理,以便可以读取出独立块114并且可以重编独立块114以从图像数据102重建完整的图像。

[0011] 因此,需要允许减少数据等待时间并且因此减少数据管理成本和功耗的有效的存储器管理的系统和方法。

附图说明

[0012] 将参考本发明的实施例,这些实施例的示例可以在附图中示出。这些图仅是说明性的,而非限制性的。尽管一般在这些实施例的上下文中描述了本发明,但是应当理解,其并不旨在将本发明的范围限制为这些特定实施例。

[0013] 图1示出了本领域中已知的普通存储器管理系统。

[0014] 图2示出了根据本公开的各种实施例的用于将数据从存储器映射到矩阵处理器的存储器管理系统。

[0015] 图3示出了根据本公开的各种实施例的定序器。

[0016] 图4是根据本公开的各种实施例的用于管理存储器的说明性过程的流程图。

具体实施方式

[0017] 在下面的描述中,出于解释的目的,阐述了具体细节以便提供对本发明的理解。然而,对于本领域的技术人员将显而易见的是,可以在没有这些细节的情况下实践本发明。此外,本领域的技术人员将认识到,以下描述的本发明的实施例可以以多种方式来实现,诸如过程、装置、系统、设备或有形计算机可读介质上的方法。

[0018] 在图中示出的部件或模块是本发明的示例性实施例的说明,并且意在避免使本发明晦涩难懂。还应当理解,在整个讨论中,部件可以被描述为单独的功能单元,其可以包括子单元,但是本领域技术人员将认识到,各种组件或它们的一部分可以被分割为分开的部件或可以集成在一起,包括集成在单个系统或部件中。应当注意,本文讨论的功能或操作可以被实现为部件。部件可以以软件、硬件或它们的组合来实现。

[0019] 此外,附图内的部件或系统之间的连接不旨在限于直接连接。而是,这些部件之间的数据可以被中间部件修改、重新格式化或以其他方式改变。同样,可以使用附加的或更少的连接。还应注意,术语“耦合”、“连接”或“通信耦合”应理解为包括直接连接、通过一个或多个中间设备的间接连接以及无线连接。

[0020] 在说明书中对“一实施例”、“优选实施例”、“一个实施例”或“实施例”的引用意味

着结合该实施例描述的特定特征、结构、特性或功能被包括在本发明的至少一个实施例中，并且可以在多于一个实施例中。同样，上述短语在说明书中各个地方的出现不一定全部指代一个或多个相同的实施例。

[0021] 在说明书中的各个地方使用某些术语是为了说明，并且不应解释为限制性的。服务、功能或资源不限于单个服务、功能或资源；这些术语的使用可以是指相关服务、功能或资源的分组，它们可以是分布式的或聚合的。此外，在本文中可以使用存储器、数据库、信息库、数据存储、表和硬件等的使用来指代可以输入信息或以其他方式记录信息的一个或多个系统部件。

[0022] 此外，应注意，本文描述的实施例是在安全微控制器的背景下给出的，但是本领域技术人员将认识到，本公开的教导不限于任何类型的微控制器，并且对于任何想要在其他背景和应用程序中使用IP保护服务的软件提供商来说，本公开的教导同样可以是感兴趣的。

[0023] 在本文中，术语“非易失性存储器”是指在断电之后保持其状态的电数据存储设备（例如，闪存存储器）。术语“物理存储器”是指电气数据存储设备、与由软件或处理器定义的虚拟或逻辑存储器相反，它能够实现对存储器进行写和读。术语“处理器”和“微控制器”可互换使用，并且是指执行一个或多个数据处理和/或控制操作的电路。类似地，术语“缓冲器”和“存储器”是指物理存储器的区域，其可以包括通过物理地址可寻址的一个或多个存储位置。如本文所使用的，术语“指针”是指存储器设备中的物理地址的开始。

[0024] 根据本发明的各种实施例，硬件加速处理系统以特定模式在存储器内存储数据，以减少读取操作的等待时间。在某些实施例中，图像文件被划分成固定块长度，这些固定块长度与加速处理器（例如，矩阵处理器）的一个或多个输入长度和/或系统的其他特性有关。如果系统正在执行卷积运算，则该固定块长度可以与卷积块大小、步幅长度、通道的数量和本领域技术人员已知的其他卷积参数有关。这种存储数据的方法允许系统在减少指针的数量和相应的时钟周期的情况下从存储器检索数据的块。理想地，将要被输入到矩阵处理器中的数据块使用单个指针被检索。然而，在某些情况中，取决于数据在存储器内是否是连续的、矩阵处理器输入内所需的特定数据等，可以使用多个时钟周期来检索数据。通过使存储器内的数据块的大小和位置更好地与用于对数据进行操作的加速处理器的特定特性对齐，系统能够减少存储器读取的等待时间。

[0025] 图2示出了根据本公开的各种实施例的用于将数据从存储器映射到硬件加速处理器（诸如矩阵处理器）的存储器管理系统。系统200表示专用计算机系统的部件，该专用计算机系统被设计为根据硬件加速处理器的硬件要求执行特定操作。在实施例中，系统200包括存储器202、控制逻辑208、定序器212、格式器214、缓存或缓冲器210和矩阵处理器216。本领域技术人员将理解，矩阵处理器216是适用于本公开的各种实施例的各种不同的硬件加速处理器的代表。

[0026] 存储器202可以是本领域中已知的存储数据块204（一个或多个）的任何类型的存储器，例如SRAM。存储器202可以根据控制器来存储数据，在该控制器中，使用虚拟存储器映射，跨存储器而映射数据。这种虚拟化引起更有效的方式来映射在存储器202内内部地存储的数据。如将参考图3更详细地讨论的，定序器212可以包括用于存储虚拟化的存储器映射和/或与虚拟化的存储器映射相连接的部件、块大小到向量对齐器(block size-to-vector

aligner) 以及块计数到数据结构计算器(block count-to-date structure calculator)。定序器212可以与存储器202、缓存/缓冲器210或它们的组合相连接。

[0027] 应当理解,由图2中的在系统200中的部件中的一个部件执行的一个或多个功能可以由其他部件分享或执行。例如,定序器212和格式器214的一个或多个功能可以集成到单个部件中,或者虚拟化的存储器映射可以存储在控制器208中和/或定序器212中。

[0028] 在操作中,数据存储器202存储每个可以具有固定大小的数据块204,并且在实施例中,数据块204表示与图像相关联的二维(2D)或三维(3D)数据结构。在实施例中,从与地址指针206相关联的初始存储器地址开始,数据块204可以以若干行(虚拟的或实际的)的方式作为连续数据被存储在存储器202中。数据块204的大小可以变化,可以是可编程的,并且在实施例中,可以由与硬件加速系统有关的转换参数(诸如步幅、形状(例如,行和列)、通道的数量等)来确定。

[0029] 在实施例中,数据块204存储在存储器202中的存储器地址可以取决于存储器202的体系结构和每个图像的大小。例如,存储器202可以使用虚拟存储器技术来将数据块204(并且通过扩展、图像数据)存储在定义的(例如分布式的)存储器位置中的连续虚拟地址中。

[0030] 在实施例中,表示例如在图像中的点的集合的数据可以在一个或多个循环中从存储器202检索,例如以便帮助对图像数据进行组装。在实施例中,在存储器202中连续的数据块204可以例如在单个周期中立刻被检索,单个周期在由指针206确定的地址处开始,因为系统能够使用块的固定长度将连续数据块映射到单个指针,以标识结束地址。相反地,如果数据块204或其中的数据是不连续的,通过针对每个后续的周期生成新地址(带有相应的指针),数据块204或其中的数据可以在两个或更多周期中被检索,在该新地址处,来自另一个数据块204的固定量的数据可以被检索。

[0031] 因为数据块204的起点和大小是已知的,所以可以例如通过使用定序器212来计算位置,后续的数据块204在这些位置处开始和结束。定序器212可以基于以下项来查明针对特定读取请求的端点:图像的行和列的数量、以及数据在定序器212中的存储方式。在实施例中,可以例如从诸如2D结构的行和列的数量的常数或通过使用任何数量的参数来推断或计算端点。这种预定的数据存储体系结构允许基于系统的某些特征将数据存储在存储器中,从而可以实现更有效、及时的检索过程。

[0032] 在实施例中,控制逻辑208可以向定序器212提供一组这样的参数,这些参数可以定义一组固定的数据块204,这些数据块204各自都可以由指示存储器202中的某个地址的专用指针(未示出)来标识。在实施例中,定序器212从控制逻辑208接收卷积指令,该卷积指令包括可以存储在例如基本寄存器中的若干卷积参数。卷积指令可以包括滤波器参数,例如滤波器大小、步长、输出通道的数量等。在实施例中,定序器212使用卷积参数来标识卷积运算的输入或卷积运算的输入的地址,并且定序器212从存储器202中的相应地址位置提取数据。

[0033] 在实施例中,格式器214可以通过如下方式将从存储器202检索的一些或全部数据线性化:将2D或3D数据转换成可以由行或列表示的单个向量或字符串,从而将数据线性化或向量化。在实施例中,格式器214通过如下方式准备存储器202中的图像数据以供矩阵处理器216处理:根据卷积参数,根据矩阵处理器216的硬件要求,将图像数据映射为合适的格

式,使得矩阵处理器216可以执行矩阵乘法作为卷积计算的部分,例如以生成输出像素。

[0034] 在实施例中,数据格式器216例如经由状态机来标识重叠的(即相同的或冗余的)输入,并且可能存在于一个或多个位置,该位置对于给定的卷积运算必须被访问两次或更多次。状态机可以被配置为:使用滤波器参数(诸如滤波器大小和步幅),以将重叠数据标识为可重用数据,使得矩阵处理器216在不必重新访问和传输来自存储器202的数据的情况下可以重用操作数(operands)。相反,在实施例中,可从例如存储在缓存210中的本地副本加载可重用数据,从而减少计算工作量、时间和功耗。如果从存储器202检索到数据,则系统尝试通过减少指针的数量以尽可能快地检索该数据,指针用于读取在矩阵处理器216内的输入所需的数据。

[0035] 在实施例中,数据定序器212可以例如针对每个卷积运算,根据卷积运算的每个周期中由矩阵处理器216所预计的位置来布置经检索的输入,例如以在矩阵处理器216实行点乘法时匹配矩阵处理器216的给定的输入格式。定序器212可以在执行卷积运算时生成用于读取数据的地址、写入结果以及记录系统200的状态。在实施例中,定序器212使用该信息中的一些或全部来确定从存储器202的哪个地址获得数据、以及如何以如下的方式处理数据:该方式使得数据可以例如在后续的卷积步骤中由矩阵处理器216正确使用。在实施例中,定序器212耦合到数据格式器214,该数据格式器214根据给定的输入格式,以预定的顺序使经检索和经同步的图像数据与矩阵处理器216对齐。

[0036] 在实施例中,格式器214确定经检索的数据是否包含足够的数据以在给定的周期中馈送矩阵处理器216,并且如果没有,则格式器214可以指令矩阵处理器216等待例如一个周期以接收附加数据。在实施例中,格式器214可以指令定序器212标识新的地址指针执行附加读取操作,以从存储器202提取一批新的数据,来获得可以提供给矩阵处理器216的附加数据。以这种方式,尽管存储器设备202可以在任何给定的周期中处理单个读取操作,但是定序器212可以在将数据传递到矩阵处理器216之前在任何数量的周期上处理读取操作。在实施例中,在使数据对矩阵处理器216可用之前,格式器214例如通过根据矩阵处理器216的一组硬件要求对数据进行过滤来处理数据。例如,格式器214可以输出具有块长度的数据,该块长度与矩阵处理器216的输入对齐。

[0037] 在卷积应用中,定序器212使数据对齐以用于并行处理,以便可以在单个步骤中执行整个卷积,而不是GPU等待主机(例如CPU)一次发送一条指令以执行卷积的单个步骤。基于数据路径的大小以及卷积运算的特性,读取在此并行和流水线处理器中使用的数据。在实施例中,系统使用单个指针和预定大小的数据块,以在减少数量的时钟周期中,检索要和矩阵处理器对齐的适当数据。经这样做,系统通过使指针206和固定长度块的数量与读取请求对齐来标识要检索的数据的结束地址。换句话说,可以在不必读取每个数据块、查明下一个指针以读取下一个块的情况下读取图像,以此类推。

[0038] 在某些实施例中,从存储器202或缓存/缓冲器210中检索数据块206,在存储器202或缓存/缓冲器210中仅期望某些子块输入到格式器214中。在这种情况下,定序器212生成经检索的块的掩码,在该经检索的块的掩码中,相关的子块在其中被标识。结果,这些子块可以从数据块被提取并且被提供到格式器214。

[0039] 图3示出了根据本公开的各种实施例的定序器。系统300包括虚拟存储器映射302、块大小到输入长度的对齐器304、块计数到数据结构计算器306和指针存储308。在实施例

中,定序器300可以是可编程的并且包括不同或附加的部件,以实现本发明的目的。例如,定序器300可以包括状态机,该状态机例如通过使用单个总线寄存器从存储器提取数据并且标记数据的可用性。而且,可以在存储器控制器中实现存储器映射302和/或指针存储308。

[0040] 在实施例中,定序器300可以例如对于卷积运算,使用块大小到输入长度的对齐器304来标识并且对齐要从存储器中被读取到格式器的数据,该格式器将这些数据向量化为加速处理器的输入。该对齐可以考虑输入的向量长度、来自先前操作(在某些情况下是本地缓存)的冗余的特定数据以及确保读取请求标识适当数据的其他参数。块大小到输入长度的对齐器304还可以标识与该读取请求的初始比特相对应的指针。经这样做,块大小到输入长度对齐器304可以与指针存储308通信,该指针存储308存储用于数据的多个指针,数据以与上述描述一致的方式被存储。

[0041] 块计数到数据结构计算器306标识与读取请求相对应的结束地址。经这样做,使用如下项来计算结束地址:连续存储在存储器内的块的特定长度、和读取请求的长度。如前所述,如果一组连续存储的数据块可以满足读取请求,则可以使用单个时钟周期来检索该请求。然而,如果相应的数据以非连续的方式存储,则使用多个时钟周期和指针来检索数据。块计数到数据结构计算器可以与虚拟化的存储器映射302(位于定序器或存储器控制器内)通信,以确定适当的块是否是连续的。

[0042] 本领域技术人员将认识到,定序器300可以是分立设备或与其他部件集成在一起。附加地,本领域技术人员将认识到,定序器内的部件可以跨其他部件(诸如存储控制器208或格式器214)呈分布式。

[0043] 图4是根据本公开的各种实施例的用于管理存储器的说明性过程的流程图。当在步骤402处,基于一组参数将具有固定和可编程的块长度的数据块根据存储器映射存储在存储器中时,进程400开始。在实施例中,参数可以包括卷积参数,诸如卷积块大小、步幅和通道的数量,从而标识出优选的块长度以用于将数据存储存储在存储器中。

[0044] 在步骤404处,响应于读取请求,标识与请求有关的数据的长度和地址指针。在实施例中,地址指针标识与读取请求相对应的第一数据块在存储器中的初始位置。

[0045] 在步骤406处,基于块长度、输入长度和/或存储器映射,块计数(该块计数表示要从存储器检索的数据块的数目)被计算为满足读取请求。如前所述,如果在计数中被标识的数据块被连续地存储,则可以在单个时钟周期中执行读取请求。如果数据块是分布式的,则将需要多个时钟周期。

[0046] 最后,在步骤408处,在一个或多个周期中从存储器检索数据块,该周期的数量可以取决于数据块或数据块中的数据是否是连续的。

[0047] 本发明的实施例可以利用用于一个或多个处理器或处理单元的指令来在一个或多个非暂时性计算机可读介质上编码,以使得步骤被执行。应当注意,一个或多个非暂时性计算机可读介质应包括易失性和非易失性存储器。应当注意,备选的实施方式是可能的,包括硬件实施方式或软件/硬件实施方式。可以使用ASIC(一个或多个)、可编程阵列、数字信号处理电路等来实现硬件实施方式的功能。因此,任何权利要求中的“装置”术语旨在涵盖软件和硬件实施方式。类似地,本文所使用的术语“一种或多种计算机可读介质”包括其上包含指令程序或其组合的软件和/或硬件。考虑到这些实现方式的备选,应当理解,附图和随附的说明书提供了本领域技术人员将需要编写程序代码(即软件)和/或制造电路(即

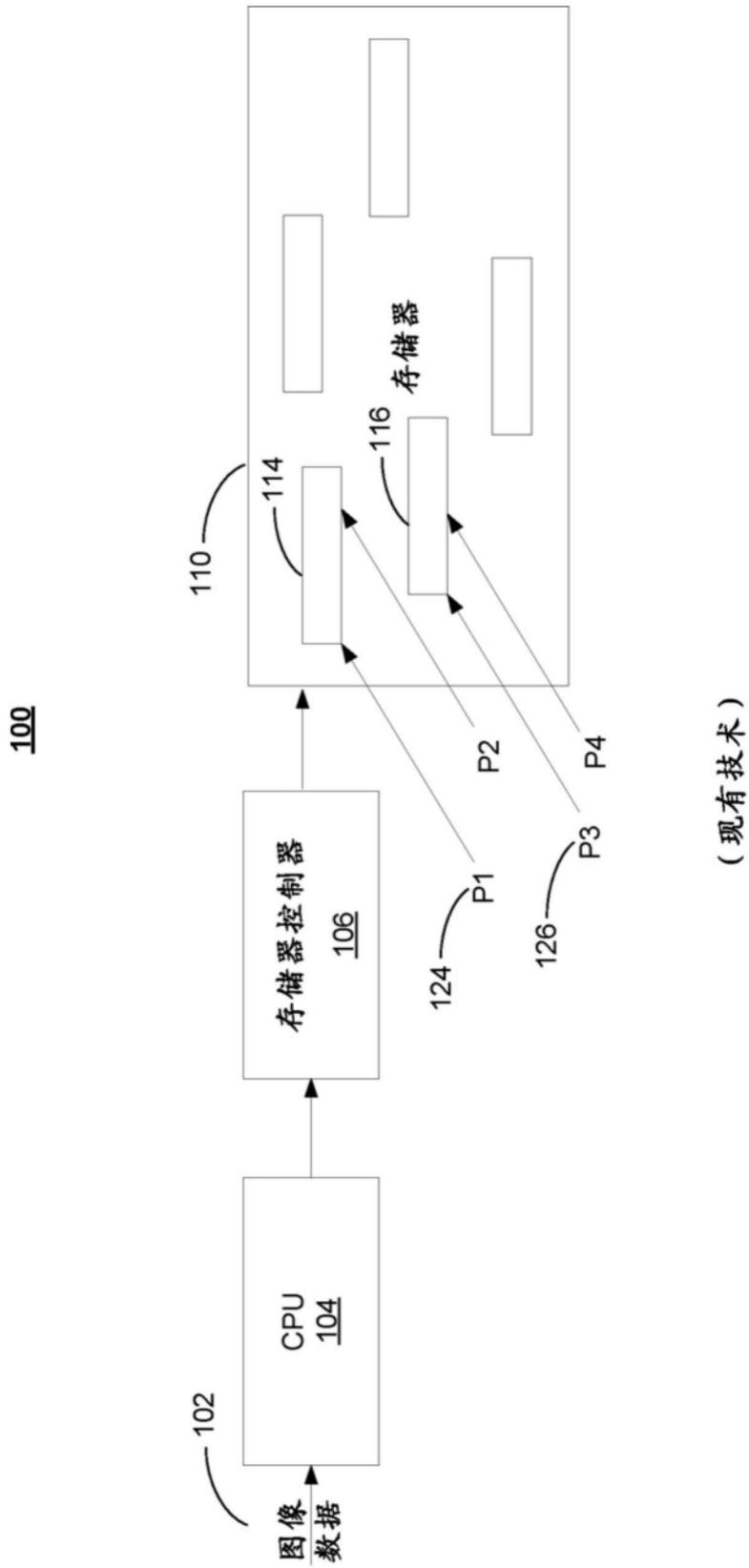
硬件)以执行所需的处理的功能信息。

[0048] 应当注意,本发明的实施例可以进一步涉及具有非暂时性有形计算机可读介质的计算机产品,其上具有用于执行各种计算机实现的操作的计算机代码。介质和计算机代码可以是出于本发明的目的而专门设计和构造的那些,或者它们可以是相关领域技术人员已知或可获得的那些。有形的计算机可读介质的示例包括但不限于:磁性介质(诸如硬盘、软盘和磁带);光学介质(诸如CD-ROM和全息设备);磁光介质;以及专门配置用于存储或存储并且执行程序代码的硬件设备(诸如专用集成电路(ASIC)、可编程逻辑设备(PLD)、闪存设备以及ROM和RAM设备)。计算机代码的示例包括机器代码(诸如由编译器生成的机器代码)以及包含由计算机使用解释器执行的更高级别代码的文件。本发明的实施例可以全部或部分地实现为机器可执行指令,该可由机器执行的指令可以在由处理设备执行的程序模块中。程序模块的示例包括库、程序、例程、对象、组件和数据结构。在分布式的计算环境中,程序模块可以在物理上位于本地、远程或二者兼有的设置中。

[0049] 本领域的技术人员将认识到,没有任何计算系统或编程语言对于本发明的实施是至关重要的。本领域的技术人员还将认识到,上述若干元件可以在物理上和/或功能上分离为子模块或组合在一起。

[0050] 应当注意,下面的权利要求的元件可以被不同地布置,包括具有多个引用关系、配置和组合。例如,在实施例中,各个权利要求的主题可以与其他权利要求结合。

[0051] 本领域技术人员将理解,前述示例和实施例是示例性的,并且不限制本发明的范围。其意图在于,对本领域技术人员而言,在阅读说明书和研究附图之后显而易见的所有排列、改进、等同、组合和改进都包括在本发明的真正精神和范围内。



(现有技术)

图1

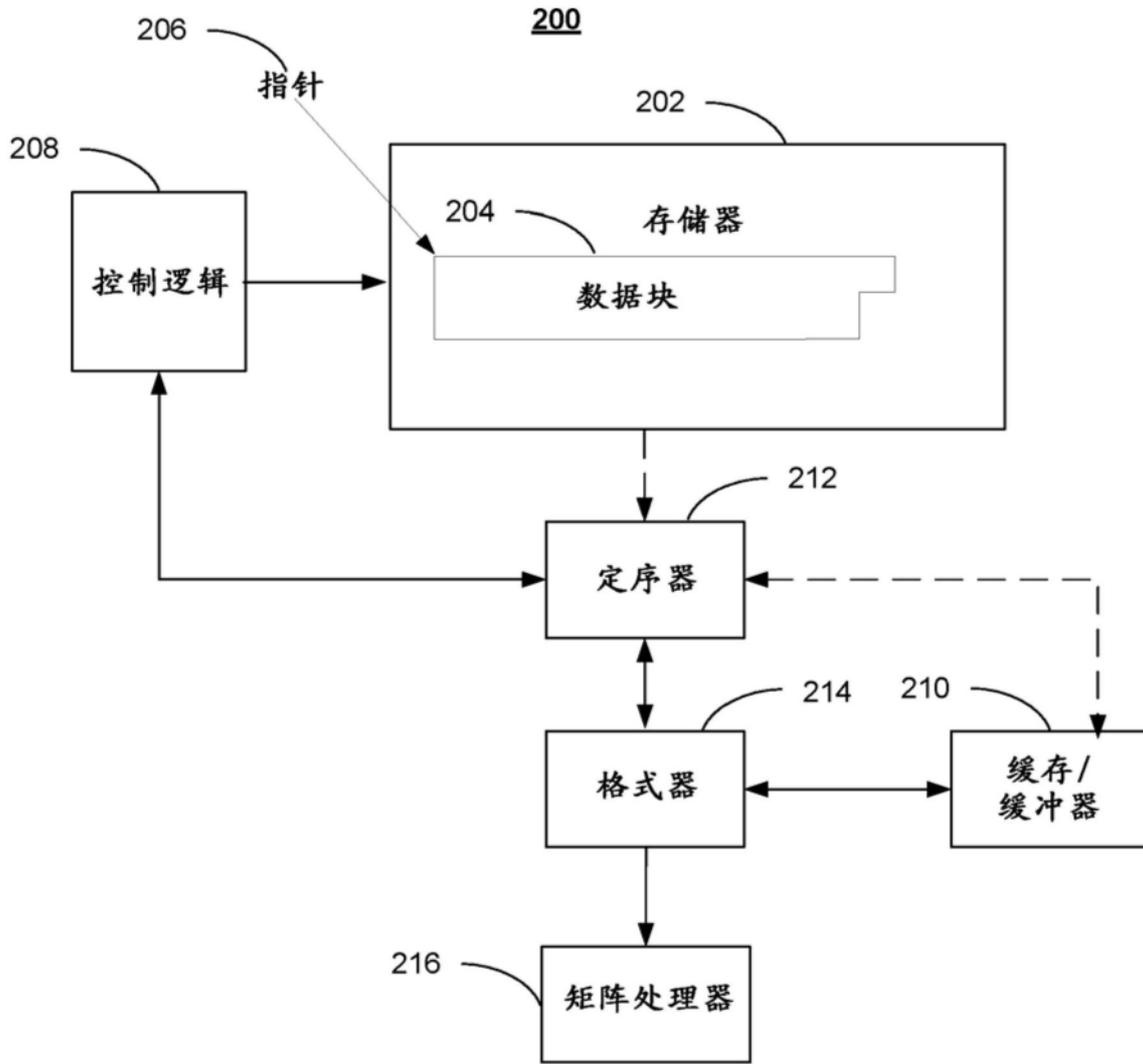


图2

300

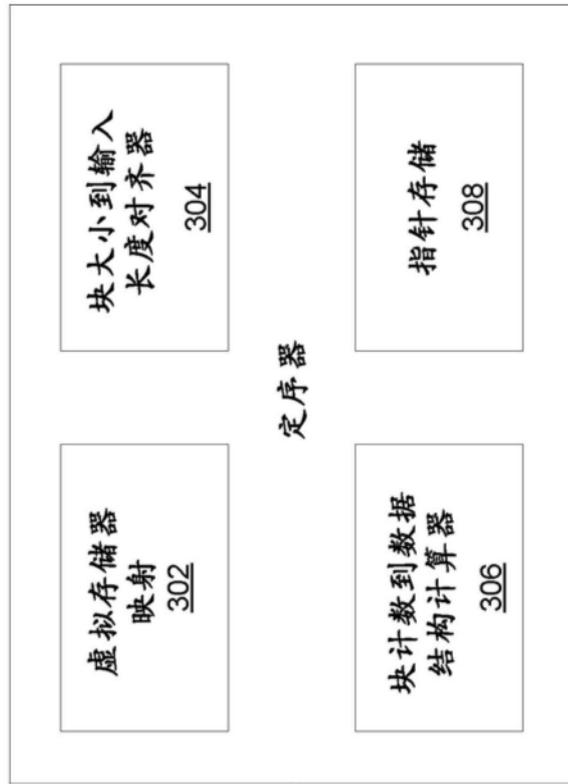


图3

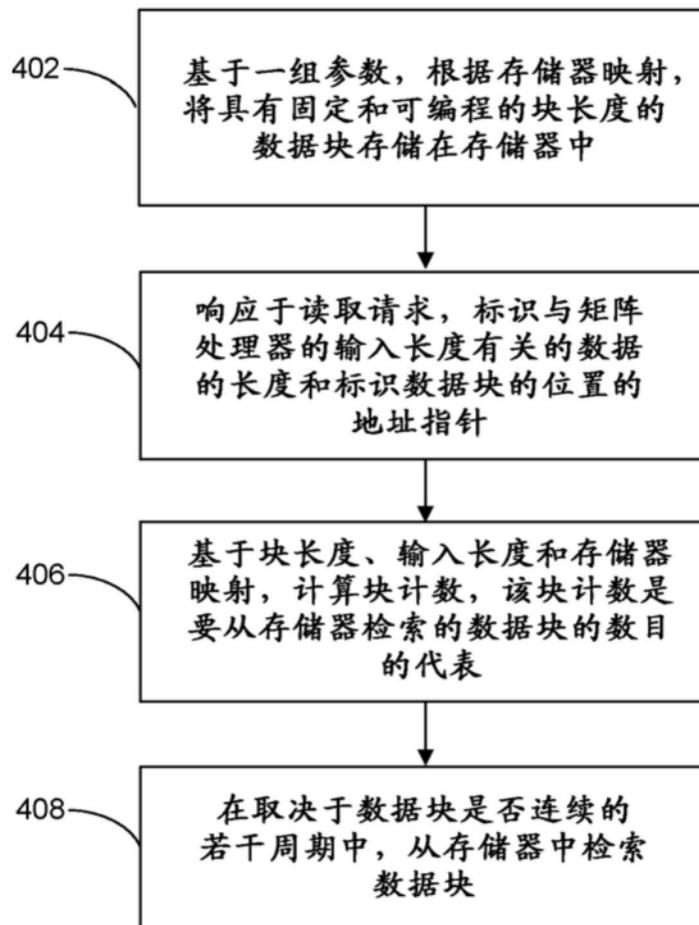
400

图4