



(12) 发明专利

(10) 授权公告号 CN 111782577 B

(45) 授权公告日 2023. 03. 24

(21) 申请号 201910272625.6

(22) 申请日 2019.04.04

(65) 同一申请的已公布的文献号
申请公布号 CN 111782577 A

(43) 申请公布日 2020.10.16

(73) 专利权人 安徽寒武纪信息科技有限公司
地址 230000 安徽省合肥市高新区习友路
3333号中国(合肥)国际智能语音产业
园研发中心楼611-194室

(72) 发明人 请求不公布姓名

(74) 专利代理机构 北京林达刘知识产权代理事
务所(普通合伙) 11277
专利代理师 刘新宇

(51) Int. Cl.
G06F 15/78 (2006.01)

(56) 对比文件

CN 107077327 A, 2017.08.18

CN 109542515 A, 2019.03.29

US 2017131939 A1, 2017.05.11

CN 109543832 A, 2019.03.29

审查员 王楠

权利要求书4页 说明书16页 附图3页

(54) 发明名称

数据处理装置及方法以及相关产品

(57) 摘要

本公开涉及数据处理装置及方法以及相关产品,该产品包括控制模块,所述控制模块包括:指令缓存单元、指令处理单元和存储队列单元;所述指令缓存单元,用于存储所述人工神经网络运算关联的计算指令;所述指令处理单元,用于对所述计算指令解析得到多个运算指令;所述存储队列单元,用于存储指令队列,该指令队列包括:按该队列的前后顺序待执行的多个运算指令或计算指令。通过以上方法,本公开可以提高相关产品在进行神经网络模型的运算时的运算效率。



1. 一种数据处理装置,其特征在于,所述装置包括控制单元以及执行单元,所述控制单元被配置为:

在解码后的第一处理指令的操作数包括描述符的标识时,根据所述描述符的标识,从描述符存储空间获取所述描述符的内容,所述描述符用于指示张量的形状;

向所述执行单元发送所述描述符的内容及所述第一处理指令,以使所述执行单元根据所述描述符的内容执行所述第一处理指令;

所述执行单元被配置为:

根据接收到的描述符的内容,确定与所述第一处理指令的操作数对应的数据在数据存储空间中的数据地址;

其中,根据接收到的描述符的内容,确定与所述第一处理指令的操作数对应的数据在数据存储空间中的数据地址,包括:

在所述操作数还包括针对所述描述符的数据描述位置时,根据所述描述符的内容以及所述数据描述位置,确定所述操作数对应的数据在所述数据存储空间中的数据地址,数据描述位置是描述符所指示的张量数据中的点或区域的映射位置;

根据所述数据地址,执行与所述第一处理指令对应的数据处理;

其中,所述第一处理指令包括数据存取指令,所述操作数包括源数据和目的数据;

所述控制单元被配置为:在所述源数据包括第三描述符的标识时,从所述描述符存储空间获取所述第三描述符的内容;向所述执行单元发送所述第三描述符的内容及所述第一处理指令;

所述执行单元被配置为:根据接收到的所述第三描述符的内容,获取所述源数据的第三数据地址;从所述第三数据地址读取数据并写入到所述目的数据的数据地址;

或者,

所述控制单元被配置为:在所述目的数据包括第四描述符的标识时,从所述描述符存储空间获取所述第四描述符的内容;向所述执行单元发送所述第四描述符的内容及所述第一处理指令;

所述执行单元被配置为:根据接收到的所述第四描述符的内容,获取所述目的数据的第四数据地址;从所述源数据的数据地址读取数据并写入到所述第四数据地址。

2. 根据权利要求1所述的装置,其特征在于,所述描述符用于指示N维的张量数据的形状,N为正整数,

其中,所述描述符的内容包括表示张量数据的形状的至少一个形状参数。

3. 根据权利要求2所述的装置,其特征在于,所述描述符还用于指示N维的张量数据的地址,其中,所述描述符的内容还包括表示张量数据的地址的至少一个地址参数。

4. 根据权利要求3所述的装置,其特征在于,所述张量数据的地址参数包括所述描述符的数据基准点在所述张量数据的数据存储空间中的基准地址;

其中,所述张量数据的形状参数包括以下至少一种:

所述数据存储空间在N个维度方向的至少一个方向上的尺寸、所述张量数据的存储区域在N个维度方向的至少一个方向上的尺寸、所述存储区域在N个维度方向的至少一个方向上的偏移量、处于N个维度方向的对角位置的至少两个顶点相对于所述数据基准点的位置、所述描述符所指示的张量数据的数据描述位置与数据地址之间的映射关系。

5. 根据权利要求4所述的装置,其特征在于,所述基准地址包括所述数据存储空间的起始地址。

6. 根据权利要求1所述的装置,其特征在于,其中,所述控制单元被配置为:

在所述源数据包括第一描述符的标识,且所述目的数据包括第二描述符的标识时,从所述描述符存储空间分别获取所述第一描述符的内容和所述第二描述符的内容;

向所述执行单元发送所述第一描述符的内容、所述第二描述符的内容及所述第一处理指令;

其中,所述执行单元被配置为:

根据接收到的所述第一描述符的内容和所述第二描述符的内容,分别获取所述源数据的第一数据地址和所述目的数据的第二数据地址;

从所述第一数据地址读取数据并写入到所述第二数据地址。

7. 根据权利要求1所述的装置,其特征在于,所述第一处理指令包括运算指令,

其中,所述执行单元根据所述数据地址,执行与所述第一处理指令对应的数据处理,包括:

从所述数据地址读取数据并执行与所述运算指令对应的运算。

8. 根据权利要求1所述的装置,其特征在于,所述控制单元还被配置为:

对接收到的第一处理指令进行解码处理,得到解码后的第一处理指令,

其中,解码后的第一处理指令包括操作码以及一个或多个操作数,所述操作码用于指示与所述第一处理指令对应的处理类型。

9. 根据权利要求1-8中任意一项所述的装置,其特征在于,所述描述符存储空间为所述控制单元的内部存储器中的存储空间,所述数据存储空间为所述控制单元的内部存储器或与所述控制单元连接的外部存储器中的存储空间。

10. 一种数据处理方法,其特征在于,所述方法应用于处理器,所述方法包括:

在解码后的第一处理指令的操作数包括描述符的标识时,根据所述描述符的标识,从描述符存储空间获取所述描述符的内容,所述描述符用于指示张量的形状;

根据所述描述符的内容,确定与所述第一处理指令的操作数对应的数据在数据存储空间中的数据地址;

根据所述描述符的内容,确定与所述第一处理指令的操作数对应的数据在数据存储空间中的数据地址,包括:

在所述操作数还包括针对所述描述符的数据描述位置时,根据所述描述符的内容以及所述数据描述位置,确定所述操作数对应的数据在所述数据存储空间中的数据地址,数据描述位置是描述符所指示的张量数据中的点或区域的映射位置;

根据所述数据地址,执行与所述第一处理指令对应的数据处理;

其中,所述第一处理指令包括数据存取指令,所述操作数包括源数据和目的数据,

在解码后的第一处理指令的操作数包括描述符的标识时,根据所述描述符的标识,从描述符存储空间获取所述描述符的内容,包括:在所述源数据包括第三描述符的标识时,从所述描述符存储空间获取所述第三描述符的内容;

根据所述描述符的内容,确定与所述第一处理指令的操作数对应的数据在数据存储空间中的数据地址,包括:根据所述第三描述符的内容,获取所述源数据的第三数据地址;

根据所述数据地址,执行与所述第一处理指令对应的数据处理,包括:从所述第三数据地址读取数据并写入到所述目的数据的数据地址;

或者,

在解码后的第一处理指令的操作数包括描述符的标识时,根据所述描述符的标识,从描述符存储空间获取所述描述符的内容,包括:在所述目的数据包括第四描述符的标识时,从所述描述符存储空间获取所述第四描述符的内容;

根据所述描述符的内容,确定与所述第一处理指令的操作数对应的数据在数据存储空间中的数据地址,包括:

根据所述第四描述符的内容,获取所述目的数据的第四数据地址;

根据所述数据地址,执行与所述第一处理指令对应的数据处理,包括:从所述源数据的数据地址读取数据并写入到所述第四数据地址。

11. 根据权利要求10所述的方法,其特征在于,所述描述符用于指示N维的张量数据的形状,N为正整数,

其中,所述描述符的内容包括表示张量数据的形状的至少一个形状参数。

12. 根据权利要求11所述的方法,其特征在于,所述描述符还用于指示N维的张量数据的地址,其中,所述描述符的内容还包括表示张量数据的地址的至少一个地址参数。

13. 根据权利要求12所述的方法,其特征在于,所述张量数据的地址参数包括所述描述符的数据基准点在所述张量数据的数据存储空间中的基准地址;

其中,所述张量数据的形状参数包括以下至少一种:

所述数据存储空间在N个维度方向的至少一个方向上的尺寸、所述张量数据的存储区域在N个维度方向的至少一个方向上的尺寸、所述存储区域在N个维度方向的至少一个方向上的偏移量、处于N个维度方向的对角位置的至少两个顶点相对于所述数据基准点的位置、所述描述符所指示的张量数据的数据描述位置与数据地址之间的映射关系。

14. 根据权利要求13所述的方法,其特征在于,所述基准地址包括所述数据存储空间的起始地址。

15. 根据权利要求10所述的方法,其特征在于,在解码后的第一处理指令的操作数包括描述符的标识时,根据所述描述符的标识,从描述符存储空间获取所述描述符的内容,包括:在所述源数据包括第一描述符的标识,且所述目的数据包括第二描述符的标识时,从所述描述符存储空间分别获取所述第一描述符的内容和所述第二描述符的内容;

其中,根据所述描述符的内容,确定与所述第一处理指令的操作数对应的数据在数据存储空间中的数据地址,包括:根据所述第一描述符的内容和所述第二描述符的内容,分别获取所述源数据的第一数据地址和所述目的数据的第二数据地址;

其中,根据所述数据地址,执行与所述第一处理指令对应的数据处理,包括:从所述第一数据地址读取数据并写入到所述第二数据地址。

16. 根据权利要求10所述的方法,其特征在于,所述第一处理指令包括运算指令,

其中,根据所述数据地址,执行与所述第一处理指令对应的数据处理,包括:

从所述数据地址读取数据并执行与所述运算指令对应的运算。

17. 根据权利要求10所述的方法,其特征在于,所述方法还包括:

对接收到的第一处理指令进行解码处理,得到解码后的第一处理指令,

其中,解码后的第一处理指令包括操作码以及一个或多个操作数,所述操作码用于指示与所述第一处理指令对应的处理类型。

18.根据权利要求11-17中任意一项所述的方法,其特征在于,所述描述符存储空间为所述处理器的内部存储器中的存储空间,所述数据存储空间为所述处理器的内部存储器或外部存储器中的存储空间。

19.一种人工智能芯片,其特征在于,所述芯片包括如权利要求1-9中任意一项所述的数据处理装置。

20.一种电子设备,其特征在于,所述电子设备包括如权利要求19所述的人工智能芯片。

21.一种板卡,其特征在于,所述板卡包括:存储器件、接口装置和控制器件以及如权利要求19所述的人工智能芯片;

其中,所述人工智能芯片与所述存储器件、所述控制器件以及所述接口装置分别连接;

所述存储器件,用于存储数据;

所述接口装置,用于实现所述人工智能芯片与外部设备之间的数据传输;

所述控制器件,用于对所述人工智能芯片的状态进行监控。

22.根据权利要求21所述的板卡,其特征在于,

所述存储器件包括:多组存储单元,每一组所述存储单元与所述人工智能芯片通过总线连接,所述存储单元为:DDR SDRAM;

所述芯片包括:DDR控制器,用于对每个所述存储单元的数据传输与数据存储的控制;

所述接口装置为:标准PCIE接口。

数据处理装置及方法以及相关产品

技术领域

[0001] 本公开涉及计算机技术领域,尤其涉及一种数据处理装置及方法以及相关产品。

背景技术

[0002] 随着人工智能技术的不断发展,需要处理的数据量和数据维度都在不断增大。在相关技术中,处理器通常通过获取指令的参数来确定数据地址,然后根据数据地址完成数据的读取和使用。这就要求技术人员在设计参数时设定数据存取的相关参数(例如数据之间或数据维度之间的相互关系等),从而生成指令传递给处理器来完成数据的存取,这种方式降低了处理器的处理效率。

发明内容

[0003] 有鉴于此,本公开提出了一种数据处理装置及方法以及相关产品。

[0004] 根据本公开的一方面,提供了一种数据处理装置,所述装置包括控制单元以及执行单元,所述控制单元被配置为:

[0005] 在解码后的第一处理指令的操作数包括描述符的标识时,根据所述描述符的标识,从描述符存储空间获取所述描述符的内容,所述描述符用于指示张量的形状;

[0006] 向所述执行单元发送所述描述符的内容及所述第一处理指令,以使所述执行单元根据所述描述符的内容执行所述第一处理指令。

[0007] 根据本公开的另一方面,提供了一种数据处理方法,所述方法应用于处理器,所述方法包括:

[0008] 在解码后的第一处理指令的操作数包括描述符的标识时,根据所述描述符的标识,从描述符存储空间获取所述描述符的内容,所述描述符用于指示张量的形状;

[0009] 根据所述描述符的内容,确定与所述第一处理指令的操作数对应的数据在数据存储存储空间中的数据地址;

[0010] 根据所述数据地址,执行与所述第一处理指令对应的数据处理。

[0011] 根据本公开的另一方面,提供了一种人工智能芯片,所述人工智能芯片包括如上所述的数据处理装置。

[0012] 根据本公开的另一方面,提供了一种电子设备,所述电子设备包括如上所述的述人工智能芯片。

[0013] 根据本公开的另一方面,提供了一种板卡,所述板卡包括:存储器件、接口装置和控制器件以及上述人工智能芯片;

[0014] 其中,所述人工智能芯片与所述存储器件、所述控制器件以及所述接口装置分别连接;

[0015] 所述存储器件,用于存储数据;

[0016] 所述接口装置,用于实现所述人工智能芯片与外部设备之间的数据传输;

[0017] 所述控制器件,用于对所述人工智能芯片的状态进行监控。

[0018] 根据本公开的实施例,控制单元通过第一处理指令操作数中的描述符的标识获取描述符的内容,并向执行单元发送描述符内容以执行第一处理指令,使得执行处理指令时可以通过描述符来获取操作数,从而降低数据存取的复杂度,提高处理器的数据存取效率。

[0019] 根据下面参考附图对示例性实施例的详细说明,本公开的其它特征及方面将变得清楚。

附图说明

[0020] 包含在说明书中并且构成说明书的一部分的附图与说明书一起示出了本公开的示例性实施例、特征和方面,并且用于解释本公开的原理。

[0021] 图1示出根据本公开实施例的数据处理装置的框图。

[0022] 图2示出根据本公开实施例的数据处理装置的数据存储空间的示意图。

[0023] 图3示出根据本公开实施例的数据处理方法的流程图。

[0024] 图4示出根据本公开实施例的板卡的结构框图。

具体实施方式

[0025] 以下将参考附图详细说明本公开的各种示例性实施例、特征和方面。附图中相同的附图标记表示功能相同或相似的元件。尽管在附图中示出了实施例的各种方面,但是除非特别指出,不必按比例绘制附图。

[0026] 在这里专用的词“示例性”意为“用作例子、实施例或说明性”。这里作为“示例性”所说明的任何实施例不必解释为优于或好于其它实施例。

[0027] 另外,为了更好的说明本公开,在下文的具体实施方式中给出了众多的具体细节。本领域技术人员应当理解,没有某些具体细节,本公开同样可以实施。在一些实例中,对于本领域技术人员熟知的方法、手段、元件和电路未作详细描述,以便于凸显本公开的主旨。

[0028] 图1示出根据本公开实施例的数据处理装置的框图。如图1所示,该数据处理装置包括控制单元11以及执行单元12,所述控制单元11被配置为:

[0029] 在解码后的第一处理指令的操作数包括描述符的标识时,根据所述描述符的标识,从描述符存储空间获取所述描述符的内容,所述描述符用于指示张量的形状;

[0030] 向所述执行单元发送所述描述符的内容及所述第一处理指令,以使所述执行单元根据所述描述符的内容执行所述第一处理指令。

[0031] 根据本公开的实施例,控制单元通过第一处理指令操作数中的描述符的标识获取描述符的内容,并向执行单元发送描述符内容以执行第一处理指令,使得执行处理指令时可以通过描述符来获取操作数,从而降低数据存取的复杂度,提高处理器的数据存取效率。

[0032] 在一种可能的实现方式中,所述数据处理装置可应用于处理器,其中,处理器可包括通用处理器(例如中央处理器CPU、图形处理器GPU)和专用处理器(例如人工智能处理器、科学计算处理器或数字信号处理器)。本公开对处理器的类型不作限制。

[0033] 在一种可能的实现方式中,所述控制单元11可配置为对接收到的第一处理指令进行解码处理,得到解码后的第一处理指令,其中,解码后的第一处理指令包括操作码以及一个或多个操作数,所述操作码用于指示与所述第一处理指令对应的处理类型。其中,第一处理指令可包括数据存取指令、运算指令、描述符管理指令以及同步指令等。本公开对第一处

理指令的具体类型不作限制。

[0034] 在一种可能的实现方式中,控制单元11对接收到的第一处理指令进行解码处理后,在该指令的操作数包括描述符的标识时,根据描述符的标识从描述符存储空间获取描述符的内容,并将描述符的内容及第一处理指令发送给执行单元12;执行单元12在接收到描述符的内容及第一处理指令后,首先根据描述符的内容,确定与所述第一处理指令的操作数对应的数据在数据存储空间中的数据地址,然后根据数据地址,执行与第一处理指令对应的数据处理。其中,数据存储空间中的数据地址可以为实际的物理地址或虚拟地址,本公开对数据地址的类型不作限制。

[0035] 在一种可能的实现方式中,描述符存储空间可以为控制单元11的内部存储器中的存储空间,数据存储空间可以为控制单元11的内部存储器或与控制单元11连接的外部存储器中的存储空间。本公开对此不作限制。

[0036] 在一种可能的实现方式中,所述描述符用于指示N维的张量数据的形状,N为正整数,例如N=1、2或3。其中,张量可以包含多种形式的组成方式,张量可以是不同维度的,比如标量可以看作是0维张量,向量可以看作1维张量,而矩阵可以是2维或2维以上的张量。张量的形状包括张量的维度、张量各个维度的尺寸等信息。举例而言,对于张量:

$$[0037] \quad \begin{bmatrix} 1 & 2 & 3 & 4 \\ 11 & 22 & 33 & 44 \end{bmatrix}$$

[0038] 该张量的形状可以被描述符描述为(2,4),也即通过两个参数表示该张量为二维张量,且该张量的第一维度(列)的尺寸为2、第二维度(行)的尺寸为4。需要说明的是,本申请对于描述符指示张量形状的方式并不做限定。

[0039] 在一种可能的实现方式中,N的取值可根据张量数据的维数(阶数)来确定,也可以根据张量数据的使用需要进行设定。例如,在N的取值为3时,张量数据为三维的张量数据,描述符可用来指示该三维的张量数据在三个维度方向上的形状(例如偏移量、尺寸等)。应当理解,本领域技术人员可以根据实际需要设置N的取值,本公开对此不作限制。

[0040] 在一种可能的实现方式中,所述描述符可包括描述符的标识和描述符的内容。其中,描述符的标识用于对描述符进行区分,例如描述符的标识为其编号;描述符的内容包括表示张量数据的形状的至少一个形状参数。例如,张量数据为3维数据,在该张量数据的三个维度中,其中两个维度的形状参数固定不变,其描述符的内容可包括表示该张量数据的另一个维度的形状参数。

[0041] 在一种可能的实现方式中,描述符的标识和内容可存储在描述符存储空间(内部存储器),例如寄存器、片上的SRAM或其他介质缓存等。描述符所指示的张量数据可存储在数据存储空间(内部存储器或外部存储器),例如片上缓存或片下存储器等。本公开对描述符存储空间及数据存储空间的具体位置不作限制。

[0042] 在一种可能的实现方式中,描述符的标识、内容以及描述符所指示的张量数据可以存储在内部存储器的同一块区域,例如,可使用片上缓存的一块连续区域来存储描述符的相关内容,其地址为ADDR0-ADDR1023,其中,可将地址ADDR0-ADDR63作为描述符存储空间,存储描述符的标识和内容,地址ADDR64-ADDR1023作为数据存储空间,存储描述符所指示的张量数据。在描述符存储空间中,可用地址ADDR0-ADDR31存储描述符的标识,地址ADDR32-ADDR63存储描述符的内容。应当理解,地址ADDR并不限于1位或一个字节,此处用来

表示一个地址,是一个地址单位。本领域技术人员可以实际情况确定描述符存储空间、数据存储空间以及其具体地址,本公开对此不作限制。

[0043] 在一种可能的实现方式中,描述符的标识、内容以及描述符所指示的张量数据可以存储在内部存储器的不同区域。例如,可以将寄存器作为描述符存储空间,在寄存器中存储描述符的标识及内容,将片上缓存作为数据存储空间,存储描述符所指示的张量数据。

[0044] 在一种可能的实现方式中,在使用寄存器存储描述符的标识和内容时,可以使用寄存器的编号来表示描述符的标识。例如,寄存器的编号为0时,其存储的描述符的标识设置为0。当寄存器中的描述符有效时,可根据描述符所指示的张量数据的大小在缓存空间中分配一块区域用于存储该张量数据。

[0045] 在一种可能的实现方式中,描述符的标识及内容可存储在内部存储器,描述符所指示的张量数据可存储在外部存储器。例如,可以采用在片上存储描述符的标识及内容、在片下存储描述符所指示的张量数据的方式。

[0046] 在一种可能的实现方式中,与描述符的标识对应的数据存储空间的数据地址可以是固定地址。例如,可以为张量数据划分单独的数据存储空间,每个张量数据在数据存储空间的起始地址与描述符的标识一一对应。在这种情况下,执行单元根据描述符的内容即可确定与操作数对应的数据在数据存储空间中的数据地址,然后执行第一处理指令。

[0047] 在一种可能的实现方式中,在与描述符的标识对应的数据存储空间的数据地址为可变地址时,所述描述符还可用于指示N维的张量数据的地址,其中,所述描述符的内容还可包括表示张量数据的地址的至少一个地址参数。例如,张量数据为3维数据,在描述符指向该张量数据的地址时,描述符的内容可包括表示该张量数据的地址的一个地址参数,例如张量数据的起始物理地址,也可以包括该张量数据的地址的多个地址参数,例如张量数据的起始地址+地址偏移量,或张量数据基于各维度的地址参数。本领域技术人员可以根据实际需要地址参数进行设置,本公开对此不作限制。

[0048] 在一种可能的实现方式中,所述张量数据的地址参数包括所述描述符的数据基准点在所述张量数据的数据存储空间中的基准地址。其中,基准地址可根据数据基准点的变化而不同。本公开对数据基准点的选取不作限制。

[0049] 在一种可能的实现方式中,所述基准地址可包括所述数据存储空间的起始地址。在描述符的数据基准点是数据存储空间的第一个数据块时,描述符的基准地址即为数据存储空间的起始地址。在描述符的数据基准点是数据存储空间中第一个数据块以外的其他数据时,描述符的基准地址即为该数据块在数据存储空间中的地址。

[0050] 在一种可能的实现方式中,所述张量数据的形状参数包括以下至少一种:所述数据存储空间在N个维度方向的至少一个方向上的尺寸、所述存储区域在N个维度方向的至少一个方向上的尺寸、所述存储区域在N个维度方向的至少一个方向上的偏移量、处于N个维度方向的对角位置的至少两个顶点相对于所述数据基准点的位置、所述描述符所指示的张量数据的数据描述位置与数据地址之间的映射关系。其中,数据描述位置是描述符所指示的张量数据中的点或区域的映射位置,例如,张量数据为3维数据时,描述符可使用三维空间坐标(x,y,z)来表示该张量数据的形状,该张量数据的数据描述位置可以是使用三维空间坐标(x,y,z)表示的、该张量数据映射在三维空间中的点或区域的位置。

[0051] 应当理解,本领域技术人员可以根据实际情况选择表示张量数据的形状参数,本

公开对此不作限制。

[0052] 通过在数据存取过程中使用描述符,可建立数据之间的关联,从而降低数据存取的复杂度,提高指令处理效率。

[0053] 在一种可能的实现方式中,可根据所述描述符的数据基准点在所述张量数据的数据存储空间中的基准地址、所述数据存储空间中的N个维度方向的至少一个方向上的尺寸、所述存储区域在N个维度方向的至少一个方向上的尺寸和/或所述存储区域在N个维度方向的至少一个方向上的偏移量,确定所述张量数据的描述符的内容。图2示出根据本公开实施例的数据处理装置的数据存储空间的示意图。如图2所示,数据存储空间21采用行优先的方式存储了一个二维数据,可通过(x,y)来表示(其中,X轴水平向右,Y轴垂直向下),X轴方向上的尺寸(每行的尺寸)为ori_x(图中未示出),Y轴方向上的尺寸(总行数)为ori_y(图中未示出),数据存储空间21的起始地址PA_start(基准地址)为第一个数据块22的物理地址。数据块23是数据存储空间21中的部分数据,其在X轴方向上的偏移量25表示为offset_x,在Y轴方向上的偏移量24表示为offset_y,在X轴方向上的尺寸表示为size_x,在Y轴方向上的尺寸表示为size_y。

[0054] 在一种可能的实现方式中,使用描述符来定义数据块23时,描述符的数据基准点可使用数据存储空间21的第一个数据块,描述符的基准地址为数据存储空间21的起始地址PA_start,然后可以结合数据存储空间21在X轴的尺寸ori_x、在Y轴上的尺寸ori_y,以及数据块23在Y轴方向的偏移量offset_y、X轴方向上的偏移量offset_x、X轴方向上的尺寸size_x以及Y轴方向上的尺寸size_y来确定数据块23的描述符的内容。

[0055] 在一种可能的实现方式中,可以使用下述公式(1)来表示描述符的内容:

$$[0056] \begin{cases} \text{X方向: ori_x, offset_x, size_x} \\ \text{Y方向: ori_y, offset_y, size_y} \\ \text{PA_start} \end{cases} \quad (1)$$

[0057] 应当理解,虽然上述示例中,描述符的内容表示的是二维空间,但本领域技术人员可以根据实际情况对描述符的内容表示的具体维度进行设置,本公开对此不作限制。

[0058] 在一种可能的实现方式中,可根据所述描述符的数据基准点在所述数据存储空间中的基准地址、处于N个维度方向的对角位置的至少两个顶点相对于所述数据基准点的位置,确定所述张量数据的描述符的内容。

[0059] 举例来说,可以使用描述符的数据基准点在数据存储空间中的基准地址PA_base,以及对角位置的两个顶点相对于数据基准点的位置,确定出图2中数据块23的描述符的内容。首先,确定描述符的数据基准点以及其在数据存储空间中的基准地址PA_base,例如,可以在数据存储空间21中选取一个数据(例如,位置为(2,2)的数据)作为数据基准点,将该数据在数据存储空间中的物理地址作为基准地址PA_base;然后,确定数据块23的对角位置的至少两个顶点相对于数据基准点的位置,例如,使用左上至右下方向的对角位置顶点相对于数据基准点的位置,其中,左上角顶点的相对位置为(x_min,y_min),右下角顶点的相对位置为(x_max,y_max),然后可以根据基准地址PA_base、左上角顶点的相对位置(x_min,y_min)以及右下角顶点的相对位置(x_max,y_max)确定出数据块23的描述符的内容。

[0060] 在一种可能的实现方式中,可以使用下述公式(2)来表示描述符的内容:

$$[0061] \quad \begin{cases} \text{X方向: } x_{\min}, x_{\max} \\ \text{Y方向: } y_{\min}, y_{\max} \\ \text{PA_base} \end{cases} \quad (2)$$

[0062] 应当理解,虽然上述示例中使用左上角和右下角两个对角位置的顶点来确定描述符的内容,但本领域技术人员可以根据实际需要将对角位置的至少两个顶点的具体顶点进行设置,本公开对此不作限制。

[0063] 在一种可能的实现方式中,可根据所述描述符的数据基准点在所述数据存储空间中的基准地址,以及所述描述符所指示的张量数据的数据描述位置与数据地址之间的映射关系,确定所述张量数据的描述符的内容。其中,数据描述位置与数据地址之间的映射关系可以根据实际需要进行设定,例如,描述符所指示的张量数据为三维空间数据时,可是使用函数 $f(x, y, z)$ 来定义数据描述位置与数据地址之间的映射关系。

[0064] 在一种可能的实现方式中,可以使用下述公式(3)来表示描述符的内容:

$$[0065] \quad \begin{cases} f(x, y, z) \\ \text{PA_base} \end{cases} \quad (3)$$

[0066] 应当理解,本领域技术人员可以根据实际情况对数据描述位置与数据地址之间的映射关系进行设定,本公开对此不作限制。

[0067] 在一种可能的实现方式中,执行单元12接收到控制单元11发送的描述符的内容及第一处理指令时,可根据描述符的内容,确定与第一处理指令的操作数对应的数据在数据存储空间中的数据地址。其中,数据地址的计算由硬件自动完成,且描述符的内容的表示方式不同时,数据地址的计算方法也会不同。本公开对数据地址的具体计算方法不作限制。

[0068] 例如,操作数中描述符的内容是使用公式(1)表示的,描述符所指示的张量数据在数据存储空间中的偏移量分别为 $offset_x$ 和 $offset_y$,尺寸为 $size_x * size_y$,那么,该描述符所指示的张量数据在数据存储空间中的起始数据地址 $PA1_{(x,y)}$ 可以使用下述公式(4)来确定:

$$[0069] \quad PA1_{(x,y)} = PA_start + (offset_y - 1) * ori_x + offset_x \quad (4)$$

[0070] 根据上述公式(4)确定的数据起始地址 $PA1_{(x,y)}$,结合偏移量 $offset_x$ 和 $offset_y$,以及存储区域的尺寸 $size_x$ 和 $size_y$,可确定出描述符所指示的张量数据在数据存储空间中的存储区域。

[0071] 在一种可能的实现方式中,执行单元12接收到控制单元11发送的描述符的内容及第一处理指令,当操作数还包括针对描述符的数据描述位置时,可根据描述符的内容以及数据描述位置,确定操作数对应的数据在数据存储空间中的数据地址。通过这种方式,可以对描述符所指示的张量数据中的部分数据(例如一个或多个数据)进行处理。

[0072] 例如,操作数中描述符的内容是使用公式(1)表示的,描述符所指示的张量数据在数据存储空间中偏移量分别为 $offset_x$ 和 $offset_y$,尺寸为 $size_x * size_y$,操作数中包括的针对描述符的数据描述位置为 (x_q, y_q) ,那么,该描述符所指示的张量数据在数据存储空间中的数据地址 $PA2_{(x,y)}$ 可以使用下述公式(5)来确定:

$$[0073] \quad PA2_{(x,y)} = PA_start + (offset_y + y_q - 1) * ori_x + (offset_x + x_q) \quad (5)$$

[0074] 在一种可能的实现方式中,所述第一处理指令包括数据存取指令,所述操作数包

括源数据和目的数据，

[0075] 其中，所述控制单元11用于在所述源数据包括第一描述符的标识，且所述目的数据包括第二描述符的标识时，从所述描述符存储空间分别获取所述第一描述符的内容和所述第二描述符的内容；向所述执行单元12发送所述第一描述符的内容、所述第二描述符的内容及所述第一处理指令；

[0076] 其中，所述执行单元12用于根据接收到的所述第一描述符的内容和所述第二描述符的内容，分别获取所述源数据的第一数据地址和所述目的数据的第二数据地址；从所述第一数据地址读取数据并写入到所述第二数据地址。

[0077] 举例来说，在源数据和目的数据均包括描述符的标识的情况下，源数据可以是片下的一组待读取数据，其第一描述符的标识为1，目的数据是片上的一块存储空间，其第二描述符的标识为2。控制单元11可根据源数据中第一描述符的标识1以及目的数据中第二描述符的标识2，从描述符存储空间分别获取第一描述符的内容D1和第二描述符的内容D2，并将第一描述符的内容D1、第二描述符的内容D2以及第一处理指令发给执行单元12。其中，第一描述符的内容D1和第二描述符的内容可分别表示如下：

[0078] D1: $\begin{cases} \text{X方向: ori_x1, offset_x1, size_x1} \\ \text{Y方向: ori_y1, offset_y1, size_y1} \\ \text{PA_start1} \end{cases}$

[0079] D2: $\begin{cases} \text{X方向: ori_x2, offset_x2, size_x2} \\ \text{Y方向: ori_y2, offset_y2, size_y2} \\ \text{PA_start2} \end{cases}$

[0080] 执行单元12根据接收到的第一描述符的内容D1和第一描述符的内容D2，可分别获取源数据的起始物理地址PA3和目的数据的起始物理地址PA4，分别表示如下：

[0081] $PA3 = PA_start1 + (offset_y1 - 1) * ori_x1 + offset_x1$

[0082] $PA4 = PA_start2 + (offset_y2 - 1) * ori_x2 + offset_x2$

[0083] 执行单元12根据源数据的起始物理地址PA3和目的数据的起始物理地址PA4，以及第一描述符的内容D1和第二描述符的内容D2，可分别确定第一数据地址和第二数据地址，然后从第一数据地址读取数据并写入第二数据地址（可通过IO通路），从而完成将D1所指向的张量数据载入到D2所指向的存储空间，即，完成了数据从片下到片上的载入。应当理解，数据从片上到片下的传输，与上述过程类似，此处不再赘述。

[0084] 在一种可能的实现方式中，数据载入时可以设定载入数据的规模，其中，载入数据的规模，可根据使用需要或存储空间的大小进行设定。例如，根据使用需要，在数据载入时，可以设定载入D1中的前10个数据。本公开对载入数据的具体规模不作限制。

[0085] 通过这种方式可使用描述符完成数据的存取，描述符的内容存储在描述符存储空间，无需在每次存取时通过指令传入数据地址，从而可以提高处理器的数据存取效率。

[0086] 在一种可能的实现方式中，所述第一处理指令包括数据存取指令，所述操作数包括源数据和目的数据，

[0087] 其中，所述控制单元11用于在所述源数据包括第三描述符的标识时，从所述描述符存储空间获取所述第三描述符的内容；向所述执行单元12发送所述第三描述符的内容及

所述第一处理指令；

[0088] 其中,所述执行单元12用于根据接收到的所述第三描述符的内容,获取所述源数据的第三数据地址;从所述第三数据地址读取数据并写入到所述目的数据的数据地址。

[0089] 举例来说,在源数据和目的数据均包括描述符的标识的情况下,源数据可以是片下的一组待读取数据,其第一描述符的标识为3,目的数据是片上的一块存储空间。控制单元11可根据源数据中第三描述符的标识3,从描述符存储空间获取第三描述符的内容D3,并将第三描述符的内容D3以及第一处理指令发给执行单元12。其中,第三描述符的内容D3可表示如下:

$$[0090] \quad D3: \begin{cases} X\text{方向: } \text{ori_x3, offset_x3, size_x3} \\ Y\text{方向: } \text{ori_y3, offset_y3, size_y3} \\ \text{PA_start3} \end{cases}$$

[0091] 执行单元12根据接收到的第三描述符的内容D3,可获取源数据的起始物理地址PA5,表示如下:

$$[0092] \quad PA5 = PA_start3 + (\text{offset_y3} - 1) * \text{ori_x3} + \text{offset_x3}$$

[0093] 执行单元12可根据第一处理指令中目的数据的地址信息,计算目的数据的起始物理地址以及目标数据的数据地址。例如,目的数据的地址信息中包括起始地址、偏移量的通用寄存器编号,可以先访问通用寄存器获取偏移量,再将起始地址与偏移量相加,得到目的数据的起始物理地址;然后根据目标数据的尺寸,即可确定目的数据的数据地址。

[0094] 执行单元12根据源数据的起始物理地址PA5以及第三描述符的内容D3,可确定第三数据地址,然后从第三数据地址读取数据并写入到目的数据的数据地址(可通过I/O通路),从而完成将D3所指向的张量数据载入到目的数据的存储空间,即,完成了数据从片下到片上的载入。应当理解,数据从片上到片下的传输,与上述过程类似,此处不再赘述。

[0095] 在一种可能的实现方式中,所述第一处理指令包括数据存取指令,所述操作数包括源数据和目的数据,

[0096] 其中,所述控制单元11用于在所述目的数据包括第四描述符的标识时,从所述描述符存储空间获取所述第四描述符的内容;向所述执行单元12发送所述第四描述符的内容及所述第一处理指令;

[0097] 其中,所述执行单元12用于根据接收到的所述第四描述符的内容,获取所述目的数据的第四数据地址;从所述源数据的数据地址读取数据并写入到所述第四数据地址。

[0098] 举例来说,在目的数据包括描述符的标识的情况下,源数据可以是片下的一组待读取数据,目的数据是片上的一块存储空间,其第四描述符的标识为4。控制单元11可根据目标数据中第四描述符的标识4,从描述符存储空间获取第四描述符的内容D4,并将第四描述符的内容D4以及第一处理指令发给执行单元12。其中,第四描述符的内容D4可表示如下:

$$[0099] \quad D4: \begin{cases} X\text{方向: } \text{ori_x4, offset_x4, size_x4} \\ Y\text{方向: } \text{ori_y4, offset_y4, size_y4} \\ \text{PA_start4} \end{cases}$$

[0100] 执行单元12根据接收到的第四描述符的内容D4,可获取目标数据的起始物理地址PA6,表示如下:

[0101] $PA6 = PA_start4 + (offset_y4 - 1) * ori_x4 + offset_x4$

[0102] 执行单元12可根据第一处理指令中源数据的地址信息,计算源数据的起始物理地址以及源数据的数据地址,例如,源数据的地址信息中包括起始地址、偏移量的通用寄存器编号,可以先访问通用寄存器获取偏移量,再将起始地址与偏移量相加,得到源数据的起始物理地址;然后根据源数据的尺寸,即可确定源数据的数据地址。

[0103] 执行单元12根据目标数据的起始物理地址PA6以及第四描述符的内容D4,可确定第四数据地址,然后从源数据的数据地址读取数据并写入到第四数据地址(可通过I0通路),从而完成将源数据载入到D4所指向的存储空间,即,完成了数据从片下到片上的载入。应当理解,数据从片上到片下的传输,与上述过程类似,此处不再赘述。

[0104] 在一种可能的实现方式中,所述第一处理指令包括运算指令,其中,所述执行单元根据所述数据地址,执行与所述第一处理指令对应的数据处理可包括:从所述数据地址读取数据并执行与所述运算指令对应的运算。

[0105] 举例来说,在第一处理指令为运算指令,且运算指令操的作数包括描述符标识的情况下,控制单元11可根据描述符的标识,从描述符存储空间获取描述符的内容,并将描述符的内容及运算指令发送给执行单元12;执行单元12根据接收到的描述符的内容,确定待运算数据的数据地址,从该数据地址读取待运算的数据并执行与运算指令对应的运算。例如,运算指令为ADD指令,在其操作数包括描述符的标识时,控制单元11可根据描述符的标识获取描述符的内容,并将描述符的内容及ADD指令发送给执行单元12,执行单元12可根据描述符的内容确定出待运算数据的数据地址,并执行ADD运算。在操作数是片下数据时,执行单元12可根据数据地址,通过I0通路将数据载入到片上,然后从操作数在片上的数据地址中读取数据并执行ADD操作。

[0106] 通过这种方式,可以使用描述符完成数据的传输,提高处理器的数据存取效率。

[0107] 根据本公开实施例的数据处理装置,引入了能够描述数据的形状的描述符,使得第一处理指令运行过程中能够通过描述符来确定数据的地址,从硬件方面简化了指令生成方式,从而降低降低数据存取的复杂,并且提高处理器存取数据的效率。

[0108] 需要说明的是,尽管以上述实施例作为示例介绍了数据处理装置如上,但本领域技术人员能够理解,本公开应不限于此。事实上,用户完全可根据个人喜好和/或实际应用场景灵活设定各模块,只要符合本公开的技术方案即可。

[0109] 图3示出根据本公开实施例的数据处理方法的流程图。如图3所示,所述数据处理方法应用于处理器,所述方法包括:

[0110] 步骤S41,在解码后的第一处理指令的操作数包括描述符的标识时,根据所述描述符的标识,从描述符存储空间获取所述描述符的内容,所述描述符用于指示张量的形状;

[0111] 步骤S42,根据所述描述符的内容,确定与所述第一处理指令的操作数对应的数据在数据存储空间中的数据地址;

[0112] 步骤S43,根据所述数据地址,执行与所述第一处理指令对应的数据处理。

[0113] 在一种可能的实现方式中,步骤S42可包括:

[0114] 在所述操作数还包括针对所述描述符的数据描述位置时,根据所述描述符的内容以及所述数据描述位置,确定所述操作数对应的数据在所述数据存储空间中的数据地址。

[0115] 在一种可能的实现方式中,所述描述符用于指示N维的张量数据的形状,N为正整

数，

[0116] 其中，所述描述符的内容包括表示张量数据的形状的至少一个形状参数。

[0117] 在一种可能的实现方式中，所述描述符还用于指示N维的张量数据的地址，其中，所述描述符的内容还包括表示张量数据的地址的至少一个地址参数。

[0118] 在一种可能的实现方式中，所述张量数据的地址参数包括所述描述符的数据基准点在所述张量数据的数据存储空间中的基准地址；

[0119] 其中，所述张量数据的形状参数包括以下至少一种：

[0120] 所述数据存储空间在N个维度方向的至少一个方向上的尺寸、所述张量数据的存储区域在N个维度方向的至少一个方向上的尺寸、所述存储区域在N个维度方向的至少一个方向上的偏移量、处于N个维度方向的对角位置的至少两个顶点相对于所述数据基准点的位置、所述描述符所指示的张量数据的数据描述位置与数据地址之间的映射关系。

[0121] 在一种可能的实现方式中，所述基准地址包括所述数据存储空间的起始地址。

[0122] 在一种可能的实现方式中，所述第一处理指令包括数据存取指令，所述操作数包括源数据和目的数据，

[0123] 其中，在解码后的第一处理指令的操作数包括描述符的标识时，步骤S41可包括：在所述源数据包括第一描述符的标识，且所述目的数据包括第二描述符的标识时，从所述描述符存储空间分别获取所述第一描述符的内容和所述第二描述符的内容；

[0124] 其中，步骤S42可包括：根据所述第一描述符的内容和所述第二描述符的内容，分别获取所述源数据的第一数据地址和所述目的数据的第二数据地址；

[0125] 其中，步骤S43可包括：从所述第一数据地址读取数据并写入到所述第二数据地址。

[0126] 在一种可能的实现方式中，所述第一处理指令包括数据存取指令，所述操作数包括源数据和目的数据，

[0127] 其中，在解码后的第一处理指令的操作数包括描述符的标识时，步骤S41可包括：在所述源数据包括第三描述符的标识时，从所述描述符存储空间获取所述第三描述符的内容；

[0128] 其中，步骤S42可包括：根据所述第三描述符的内容，获取所述源数据的第三数据地址；

[0129] 其中，步骤S43可包括：从所述第三数据地址读取数据并写入到所述目的数据的数据地址。

[0130] 在一种可能的实现方式中，所述第一处理指令包括数据存取指令，所述操作数包括源数据和目的数据，

[0131] 其中，在解码后的第一处理指令的操作数包括描述符的标识时，步骤S41可包括：在所述目的数据包括第四描述符的标识时，从所述描述符存储空间获取所述第四描述符的内容；

[0132] 其中，步骤S42可包括：根据所述第四描述符的内容，获取所述目的数据的第四数据地址；

[0133] 其中，步骤S43可包括：从所述源数据的数据地址读取数据并写入到所述第四数据地址。

[0134] 在一种可能的实现方式中,所述第一处理指令包括运算指令,

[0135] 其中,步骤S43可包括:从所述数据地址读取数据并执行与所述运算指令对应的运算。

[0136] 在一种可能的实现方式中,所述方法还包括:

[0137] 对接收到的第一处理指令进行解码处理,得到解码后的第一处理指令,

[0138] 其中,解码后的第一处理指令包括操作码以及一个或多个操作数,所述操作码用于指示与所述第一处理指令对应的处理类型。

[0139] 在一种可能的实现方式中,所述描述符存储空间为所述处理器的内部存储器中的存储空间,所述数据存储空间为所述处理器的内部存储器或外部存储器中的存储空间。

[0140] 在一种可能的实现方式中,还公开了一种人工智能芯片,其包括了上述数据处理装置。

[0141] 在一种可能的实现方式中,还公开了一种板卡,其包括存储器件、接口装置和控制器件以及上述人工智能芯片;其中,所述人工智能芯片与所述存储器件、所述控制器件以及所述接口装置分别连接;所述存储器件,用于存储数据;所述接口装置,用于实现所述人工智能芯片与外部设备之间的数据传输;所述控制器件,用于对所述人工智能芯片的状态进行监控。

[0142] 图4示出根据本公开实施例的板卡的结构框图,参阅图4,上述板卡除了包括上述芯片389以外,还可以包括其他的配套部件,该配套部件包括但不限于:存储器件390、接口装置391和控制器件392;

[0143] 所述存储器件390与所述人工智能芯片通过总线连接,用于存储数据。所述存储器件可以包括多组存储单元393。每一组所述存储单元与所述人工智能芯片通过总线连接。可以理解,每一组所述存储单元可以是DDR SDRAM(英文:Double Data Rate SDRAM,双倍速率同步动态随机存储器)。

[0144] DDR不需要提高时钟频率就能加倍提高SDRAM的速度。DDR允许在时钟脉冲的上升沿和下降沿读出数据。DDR的速度是标准SDRAM的两倍。在一个实施例中,所述存储装置可以包括4组所述存储单元。每一组所述存储单元可以包括多个DDR4颗粒(芯片)。在一个实施例中,所述人工智能芯片内部可以包括4个72位DDR4控制器,上述72位DDR4控制器中64bit用于传输数据,8bit用于ECC校验。可以理解,当每一组所述存储单元中采用DDR4-3200颗粒时,数据传输的理论带宽可达到25600MB/s。

[0145] 在一个实施例中,每一组所述存储单元包括多个并联设置的双倍速率同步动态随机存储器。DDR在一个时钟周期内可以传输两次数据。在所述芯片中设置控制DDR的控制器,用于对每个所述存储单元的数据传输与数据存储的控制。

[0146] 所述接口装置与所述人工智能芯片电连接。所述接口装置用于实现所述人工智能芯片与外部设备(例如服务器或计算机)之间的数据传输。例如在一个实施例中,所述接口装置可以为标准PCIE接口。比如,待处理的数据由服务器通过标准PCIE接口传递至所述芯片,实现数据转移。优选的,当采用PCIE 3.0X 16接口传输时,理论带宽可达到16000MB/s。在另一个实施例中,所述接口装置还可以是其他的接口,本申请并不限制上述其他的接口的具体表现形式,所述接口单元能够实现转接功能即可。另外,所述人工智能芯片的计算结果仍由所述接口装置传送给外部设备(例如服务器)。

[0147] 所述控制器件与所述人工智能芯片电连接。所述控制器件用于对所述人工智能芯片的状态进行监控。具体的,所述人工智能芯片与所述控制器件可以通过SPI接口电连接。所述控制器件可以包括单片机(Micro Controller Unit,MCU)。如所述人工智能芯片可以包括多个处理芯片、多个处理核或多个处理电路,可以带动多个负载。因此,所述人工智能芯片可以处于多负载和轻负载等不同的工作状态。通过所述控制装置可以实现对所述人工智能芯片中多个处理芯片、多个处理和或多个处理电路的工作状态的调控。

[0148] 在一种可能的实现方式中,公开了一种电子设备,其包括了上述人工智能芯片。电子设备包括数据处理装置、机器人、电脑、打印机、扫描仪、平板电脑、智能终端、手机、行车记录仪、导航仪、传感器、摄像头、服务器、云端服务器、相机、摄像机、投影仪、手表、耳机、移动存储、可穿戴设备、交通工具、家用电器、和/或医疗设备。

[0149] 所述交通工具包括飞机、轮船和/或车辆;所述家用电器包括电视、空调、微波炉、冰箱、电饭煲、加湿器、洗衣机、电灯、燃气灶、油烟机;所述医疗设备包括核磁共振仪、B超仪和/或心电图仪。

[0150] A1、一种数据处理装置,所述装置包括控制单元以及执行单元,所述控制单元被配置为:

[0151] 在解码后的第一处理指令的操作数包括描述符的标识时,根据所述描述符的标识,从描述符存储空间获取所述描述符的内容,所述描述符用于指示张量的形状;

[0152] 向所述执行单元发送所述描述符的内容及所述第一处理指令,以使所述执行单元根据所述描述符的内容执行所述第一处理指令。

[0153] A2、根据权利要求A1所述的装置,所述执行单元被配置为:

[0154] 根据接收到的描述符的内容,确定与所述第一处理指令的操作数对应的数据在数据存储空间中的数据地址;

[0155] 根据所述数据地址,执行与所述第一处理指令对应的数据处理。

[0156] A3、根据权利要求A2所述的装置,根据接收到的描述符的内容,确定与所述第一处理指令的操作数对应的数据在数据存储空间中的数据地址,包括:

[0157] 在所述操作数还包括针对所述描述符的数据描述位置时,根据所述描述符的内容以及所述数据描述位置,确定所述操作数对应的数据在所述数据存储空间中的数据地址。

[0158] A4、根据权利要求A1-A3中任意一项所述的装置,所述描述符用于指示N维的张量数据的形状,N为正整数,

[0159] 其中,所述描述符的内容包括表示张量数据的形状的至少一个形状参数。

[0160] A5、根据权利要求A4所述的装置,所述描述符还用于指示N维的张量数据的地址,其中,所述描述符的内容还包括表示张量数据的地址的至少一个地址参数。

[0161] A6、根据权利要求A5所述的装置,所述张量数据的地址参数包括所述描述符的数据基准点在所述张量数据的数据存储空间中的基准地址;

[0162] 其中,所述张量数据的形状参数包括以下至少一种:

[0163] 所述数据存储空间在N个维度方向的至少一个方向上的尺寸、所述张量数据的存储区域在N个维度方向的至少一个方向上的尺寸、所述存储区域在N个维度方向的至少一个方向上的偏移量、处于N个维度方向的对角位置的至少两个顶点相对于所述数据基准点的位置、所述描述符所指示的张量数据的数据描述位置与数据地址之间的映射关系。

- [0164] A7、根据权利要求A6所述的装置,所述基准地址包括所述数据存储空间的起始地址。
- [0165] A8、根据权利要求A2-A7中任意一项所述的装置,所述第一处理指令包括数据存取指令,所述操作数包括源数据和目的数据,
- [0166] 其中,所述控制单元被配置为:
- [0167] 在所述源数据包括第一描述符的标识,且所述目的数据包括第二描述符的标识时,从所述描述符存储空间分别获取所述第一描述符的内容和所述第二描述符的内容;
- [0168] 向所述执行单元发送所述第一描述符的内容、所述第二描述符的内容及所述第一处理指令;
- [0169] 其中,所述执行单元被配置为:
- [0170] 根据接收到的所述第一描述符的内容和所述第二描述符的内容,分别获取所述源数据的第一数据地址和所述目的数据的第二数据地址;
- [0171] 从所述第一数据地址读取数据并写入到所述第二数据地址。
- [0172] A9、根据权利要求A2-A7中任意一项所述的装置,所述第一处理指令包括数据存取指令,所述操作数包括源数据和目的数据,
- [0173] 其中,所述控制单元被配置为:
- [0174] 在所述源数据包括第三描述符的标识时,从所述描述符存储空间获取所述第三描述符的内容;
- [0175] 向所述执行单元发送所述第三描述符的内容及所述第一处理指令;
- [0176] 其中,所述执行单元被配置为:
- [0177] 根据接收到的所述第三描述符的内容,获取所述源数据的第三数据地址;
- [0178] 从所述第三数据地址读取数据并写入到所述目的数据的数据地址。
- [0179] A10、根据权利要求A2-A7中任意一项所述的装置,所述第一处理指令包括数据存取指令,所述操作数包括源数据和目的数据,
- [0180] 其中,所述控制单元被配置为:
- [0181] 在所述目的数据包括第四描述符的标识时,从所述描述符存储空间获取所述第四描述符的内容;
- [0182] 向所述执行单元发送所述第四描述符的内容及所述第一处理指令,
- [0183] 其中,所述执行单元被配置为:
- [0184] 根据接收到的所述第四描述符的内容,获取所述目的数据的第四数据地址;
- [0185] 从所述源数据的数据地址读取数据并写入到所述第四数据地址。
- [0186] A11、根据权利要求A2所述的装置,所述第一处理指令包括运算指令,
- [0187] 其中,所述执行单元根据所述数据地址,执行与所述第一处理指令对应的数据处理,包括:
- [0188] 从所述数据地址读取数据并执行与所述运算指令对应的运算。
- [0189] A12、根据权利要求A1-A11中任意一项所述的装置,所述控制单元还被配置为:
- [0190] 对接收到的第一处理指令进行解码处理,得到解码后的第一处理指令,
- [0191] 其中,解码后的第一处理指令包括操作码以及一个或多个操作数,所述操作码用于指示与所述第一处理指令对应的处理类型。

[0192] A13、根据权利要求A1-A12中任意一项所述的装置，所述描述符存储空间为所述控制单元的内部存储器中的存储空间，所述数据存储空间为所述控制单元的内部存储器或与所述控制单元连接的外部存储器中的存储空间。

[0193] A14、一种数据处理方法，所述方法应用于处理器，所述方法包括：

[0194] 在解码后的第一处理指令的操作数包括描述符的标识时，根据所述描述符的标识，从描述符存储空间获取所述描述符的内容，所述描述符用于指示张量的形状；

[0195] 根据所述描述符的内容，确定与所述第一处理指令的操作数对应的数据在数据存储空间中的数据地址；

[0196] 根据所述数据地址，执行与所述第一处理指令对应的数据处理。

[0197] A15、根据权利要求A14所述的方法，根据所述描述符的内容，确定与所述第一处理指令的操作数对应的数据在数据存储空间中的数据地址，包括：

[0198] 在所述操作数还包括针对所述描述符的数据描述位置时，根据所述描述符的内容以及所述数据描述位置，确定所述操作数对应的数据在所述数据存储空间中的数据地址。

[0199] A16、根据权利要求A14或A15所述的方法，所述描述符用于指示N维的张量数据的形状，N为正整数，

[0200] 其中，所述描述符的内容包括表示张量数据的形状的至少一个形状参数。

[0201] A17、根据权利要求A16所述的方法，所述描述符还用于指示N维的张量数据的地址，其中，所述描述符的内容还包括表示张量数据的地址的至少一个地址参数。

[0202] A18、根据权利要求A17所述的方法，所述张量数据的地址参数包括所述描述符的数据基准点在所述张量数据的数据存储空间中的基准地址；

[0203] 其中，所述张量数据的形状参数包括以下至少一种：

[0204] 所述数据存储空间在N个维度方向的至少一个方向上的尺寸、所述张量数据的存储区域在N个维度方向的至少一个方向上的尺寸、所述存储区域在N个维度方向的至少一个方向上的偏移量、处于N个维度方向的对角位置的至少两个顶点相对于所述数据基准点的位置、所述描述符所指示的张量数据的数据描述位置与数据地址之间的映射关系。

[0205] A19、根据权利要求A18所述的方法，所述基准地址包括所述数据存储空间的起始地址。

[0206] A20、根据权利要求A14-A19中任意一项所述的方法，所述第一处理指令包括数据存取指令，所述操作数包括源数据和目的数据，

[0207] 其中，在解码后的第一处理指令的操作数包括描述符的标识时，根据所述描述符的标识，从描述符存储空间获取所述描述符的内容，包括：在所述源数据包括第一描述符的标识，且所述目的数据包括第二描述符的标识时，从所述描述符存储空间分别获取所述第一描述符的内容和所述第二描述符的内容；

[0208] 其中，根据所述描述符的内容，确定与所述第一处理指令的操作数对应的数据在数据存储空间中的数据地址，包括：根据所述第一描述符的内容和所述第二描述符的内容，分别获取所述源数据的第一数据地址和所述目的数据的第二数据地址；

[0209] 其中，根据所述数据地址，执行与所述第一处理指令对应的数据处理，包括：从所述第一数据地址读取数据并写入到所述第二数据地址。

[0210] A21、根据权利要求A14-A19中任意一项所述的方法，所述第一处理指令包括数据

存取指令,所述操作数包括源数据和目的数据,

[0211] 其中,在解码后的第一处理指令的操作数包括描述符的标识时,根据所述描述符的标识,从描述符存储空间获取所述描述符的内容,包括:在所述源数据包括第三描述符的标识时,从所述描述符存储空间获取所述第三描述符的内容;

[0212] 其中,根据所述描述符的内容,确定与所述第一处理指令的操作数对应的数据在数据存储空间中的数据地址,包括:根据所述第三描述符的内容,获取所述源数据的第三数据地址;

[0213] 其中,根据所述数据地址,执行与所述第一处理指令对应的数据处理,包括:从所述第三数据地址读取数据并写入到所述目的数据的数据地址。

[0214] A22、根据权利要求A14-A19中任意一项所述的方法,所述第一处理指令包括数据存取指令,所述操作数包括源数据和目的数据,

[0215] 其中,在解码后的第一处理指令的操作数包括描述符的标识时,根据所述描述符的标识,从描述符存储空间获取所述描述符的内容,包括:在所述目的数据包括第四描述符的标识时,从所述描述符存储空间获取所述第四描述符的内容;

[0216] 其中,根据所述描述符的内容,确定与所述第一处理指令的操作数对应的数据在数据存储空间中的数据地址,包括:

[0217] 根据所述第四描述符的内容,获取所述目的数据的第四数据地址;

[0218] 其中,根据所述数据地址,执行与所述第一处理指令对应的数据处理,包括:从所述源数据的数据地址读取数据并写入到所述第四数据地址。

[0219] A23、根据权利要求A14所述的方法,所述第一处理指令包括运算指令,

[0220] 其中,根据所述数据地址,执行与所述第一处理指令对应的数据处理,包括:

[0221] 从所述数据地址读取数据并执行与所述运算指令对应的运算。

[0222] A24、根据权利要求A14-A23中任意一项所述的方法,所述方法还包括:

[0223] 对接收到的第一处理指令进行解码处理,得到解码后的第一处理指令,

[0224] 其中,解码后的第一处理指令包括操作码以及一个或多个操作数,所述操作码用于指示与所述第一处理指令对应的处理类型。

[0225] A25、根据权利要求A14-A24中任意一项所述的方法,所述描述符存储空间为所述处理器的内部存储器中的存储空间,所述数据存储空间为所述处理器的内部存储器或外部存储器中的存储空间。

[0226] A26、一种人工智能芯片,所述芯片包括如权利要求A1-A13中任意一项所述的数据处理装置。

[0227] A27、一种电子设备,所述电子设备包括如权利要求A26所述的人工智能芯片。

[0228] A28、一种板卡,所述板卡包括:存储器件、接口装置和控制器件以及如权利要求A26所述的人工智能芯片;

[0229] 其中,所述人工智能芯片与所述存储器件、所述控制器件以及所述接口装置分别连接;

[0230] 所述存储器件,用于存储数据;

[0231] 所述接口装置,用于实现所述人工智能芯片与外部设备之间的数据传输;

[0232] 所述控制器件,用于对所述人工智能芯片的状态进行监控。

[0233] A29、根据权利要求A28所述的板卡,所述存储器件包括:多组存储单元,每一组所述存储单元与所述人工智能芯片通过总线连接,所述存储单元为:DDR SDRAM;

[0234] 所述芯片包括:DDR控制器,用于对每个所述存储单元的数据传输与数据存储的控制;

[0235] 所述接口装置为:标准PCIE接口。

[0236] 以上已经描述了本公开的各实施例,上述说明是示例性的,并非穷尽性的,并且也不限于所披露的各实施例。在不偏离所说明的各实施例的范围和精神的情况下,对于本技术领域的普通技术人员来说许多修改和变更都是显而易见的。本文中所用术语的选择,旨在最好地解释各实施例的原理、实际应用或对市场中的技术的改进,或者使本技术领域的其它普通技术人员能理解本文披露的各实施例。

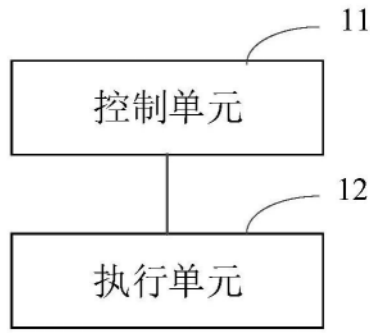


图1

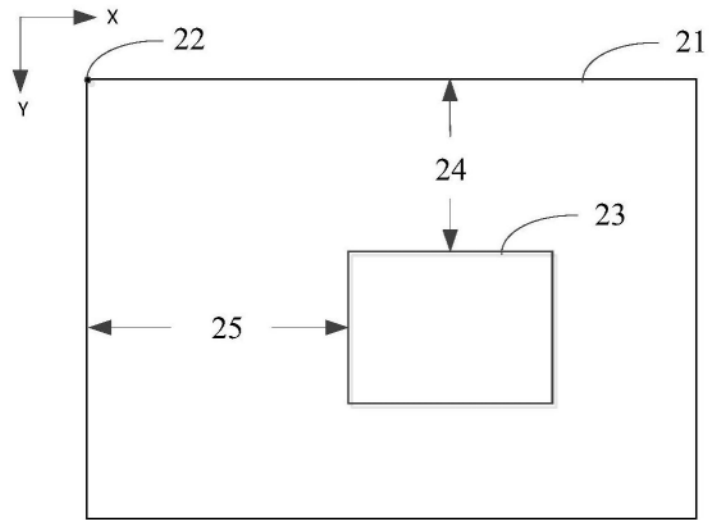


图2

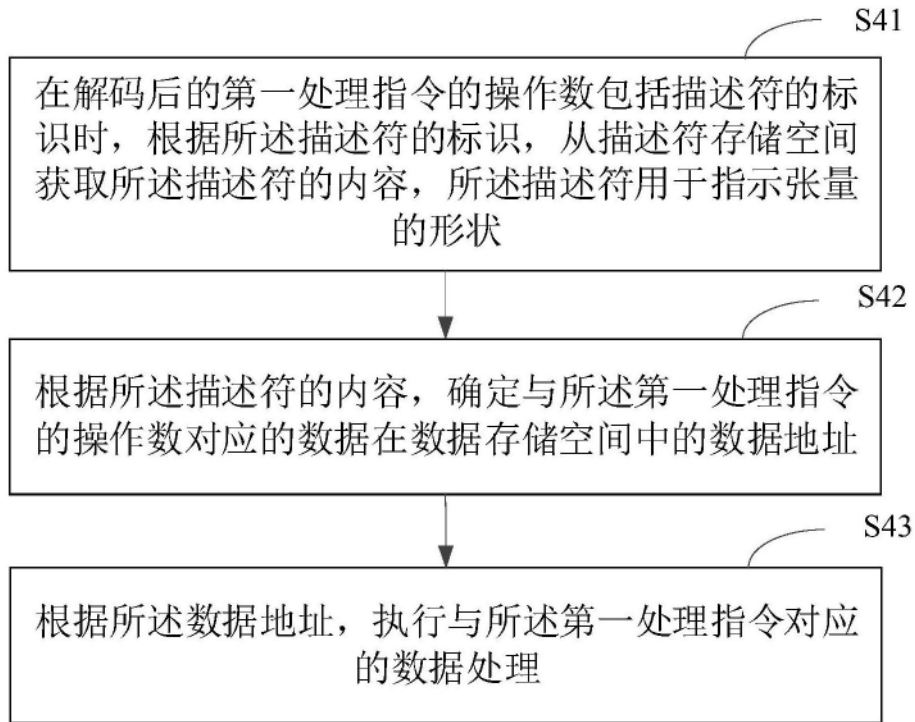


图3

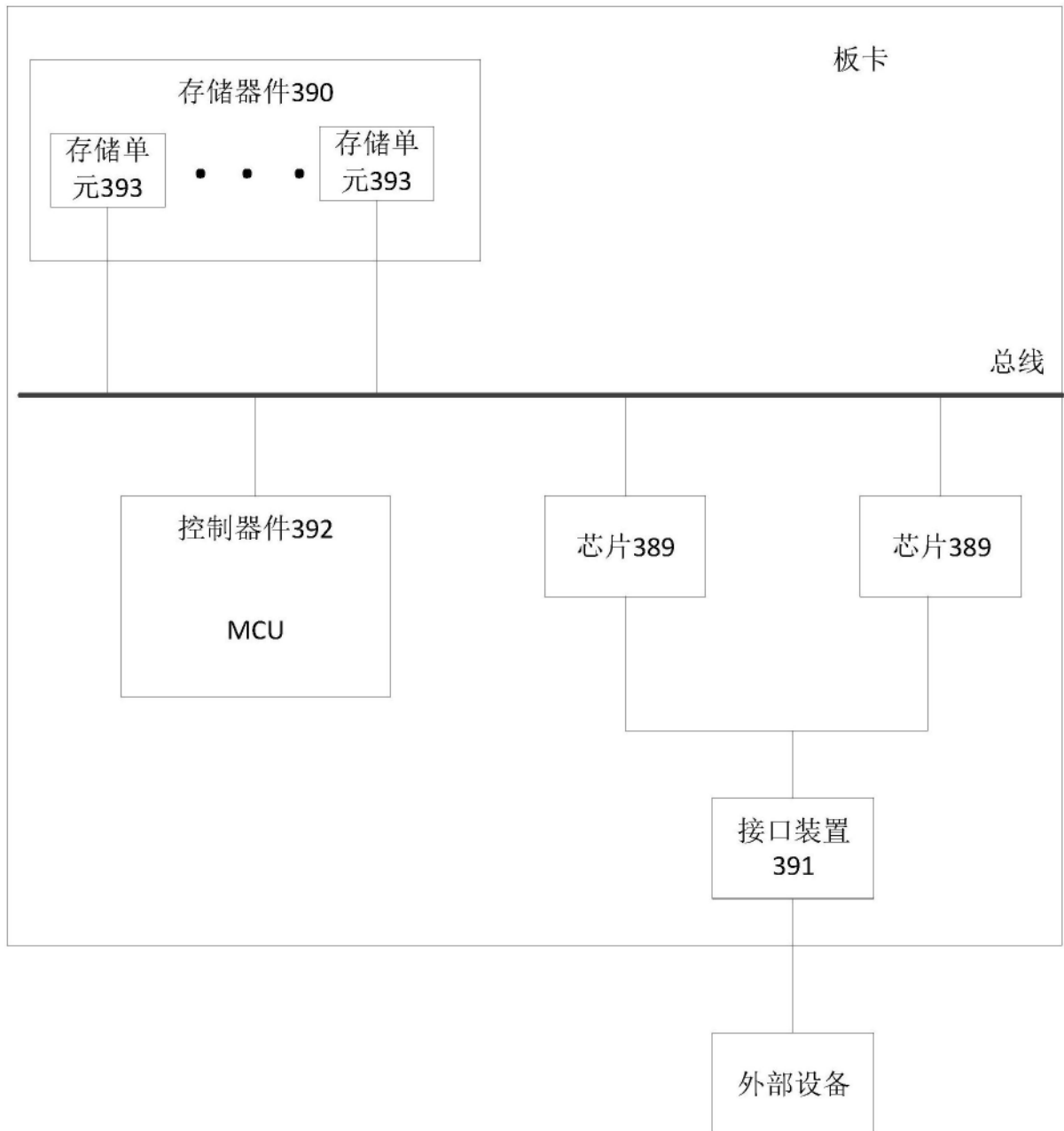


图4