



(12)发明专利申请

(10)申请公布号 CN 109558564 A

(43)申请公布日 2019. 04. 02

(21)申请号 201811455785.6

(22)申请日 2018.11.30

(71)申请人 上海寒武纪信息科技有限公司

地址 200120 上海市浦东新区同汇路168号
B座6层

(72)发明人 不公告发明人

(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

代理人 刘新宇

(51)Int.Cl.

G06F 17/15(2006.01)

G06F 9/30(2006.01)

G06N 3/04(2006.01)

G06N 3/06(2006.01)

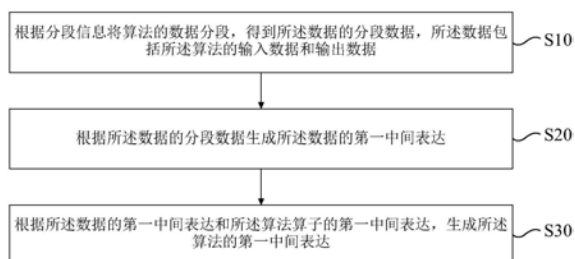
权利要求书2页 说明书16页 附图3页

(54)发明名称

运算方法、装置及相关产品

(57)摘要

本公开涉及运算方法、装置及相关产品,所述方法包括:根据分段信息将卷积算法的数据分段,得到所述数据的分段数据,所述数据包括所述卷积算法的输入数据和输出数据;根据所述数据的分段数据确定所述数据的第一中间表达;根据所述数据的第一中间表达和所述卷积算法算子的第一中间表达,生成所述卷积算法的第一中间表达。本公开实施例通过数据的第一中间表达和算子的第一中间表达得到算法的第一中间表达,不同的算法在不同的片上系统实现时,不需要设置特定的接口,提高了算法和片上系统之间的兼容性,降低了算法开发的难度。



1. 一种运算方法,其特征在于,所述方法包括:

根据分段信息将卷积算法的数据分段,得到所述数据的数据分段,所述数据包括所述卷积算法的输入数据和输出数据;

根据所述数据的数据分段确定所述数据的第一中间表达;

根据所述数据的第一中间表达和所述卷积算法算子的第一中间表达,生成所述卷积算法的第一中间表达。

2. 根据权利要求1所述的方法,其特征在于,所述数据为多维数据,根据分段信息将卷积算法的数据分段,得到所述数据的数据分段,包括:

根据分段信息将所述数据各维度数据分别进行分段,得到所述数据各维度数据的数据分段;

根据所述数据的数据分段的各维度数据,得到所述数据的数据分段。

3. 根据权利要求1所述的方法,其特征在于,所述方法还包括:

根据所述卷积算法的第一中间表达生成所述卷积算法的第一可执行指令。

4. 根据权利要求1所述的方法,其特征在于,所述分段信息根据片上缓存的大小确定。

5. 根据权利要求1所述的方法,其特征在于,所述输入数据包括输入神经元、权值、偏置,所述输出数据包括卷积输出神经元。

6. 根据权利要求5所述的方法,其特征在于,所述输入神经元的各维度数据包括:特征输入通道、输入特征图高度和输入特征图宽度;

所述卷积输出神经元的各维度数据包括:卷积通道、卷积特征图高度和卷积特征图宽度;

所述权值的各维度数据包括:特征输入通道、卷积通道、特征图高度卷积核和特征图宽度卷积核;

所述偏置的各维度数据包括:卷积通道。

7. 一种运算装置,其特征在于,所述装置包括:

数据分段获取模块,用于根据分段信息将卷积算法的数据分段,得到所述数据的数据分段,所述数据包括所述卷积算法的输入数据和输出数据;

数据中间表达确定模块,用于根据所述数据的数据分段确定所述数据的第一中间表达;

算法中间表达确定模块,用于根据所述数据的第一中间表达和所述卷积算法算子的第一中间表达,生成所述卷积算法的第一中间表达。

8. 根据权利要求7所述的装置,其特征在于,所述数据为多维数据,所述数据分段获取模块,包括:

数据分段获取子模块,用于根据分段信息将所述数据各维度数据分别进行分段,得到所述数据各维度数据的数据分段;

数据分段获取子模块,用于根据所述数据的数据分段的各维度数据,得到所述数据的数据分段。

9. 根据权利要求7所述的装置,其特征在于,所述装置还包括:

第一可执行指令获取模块,用于根据所述卷积算法的第一中间表达生成所述卷积算法的第一可执行指令。

10. 根据权利要求7所述的装置,其特征在于,所述分段信息根据片上缓存的大小确定。

11. 根据权利要求7所述的装置,其特征在于,所述输入数据包括输入神经元、权值、偏置,所述输出数据包括卷积输出神经元。

12. 根据权利要求11所述的装置,其特征在于,所述输入神经元的各维度数据包括:特征输入通道、输入特征图高度和输入特征图宽度;

所述卷积输出神经元的各维度数据包括:卷积通道、卷积特征图高度和卷积特征图宽度;

所述权值的各维度数据包括:特征输入通道、卷积通道、特征图高度卷积核和特征图宽度卷积核;

所述偏置的各维度数据包括:卷积通道。

13. 一种神经网络运算装置,其特征在于,所述神经网络运算装置包括一个或多个如权利要求7-12任一项所述的运算装置,所述神经网络运算装置用于完成设定的神经网络运算。

14. 一种组合运算装置,其特征在于,所述组合运算装置包括一个或多个如权利要求13任一项所述的神经网络运算装置,通用互联接口和其他处理装置;

所述神经网络运算装置与所述其他处理装置进行交互,共同完成用户指定的计算操作。

15. 一种神经网络芯片,其特征在于,所述神经网络芯片包括:

如权利要求7-12任一项所述的运算装置;或

如权利要求13所述的神经网络运算装置;或

如权利要求14所述的组合处理装置。

16. 一种电子设备,其特征在于,所述电子设备包括:

如权利要求7-12任一项所述的运算装置;或

如权利要求13所述的神经网络运算装置;或

如权利要求14所述的组合处理装置;或

如权利要求15所述的神经网络芯片。

运算方法、装置及相关产品

技术领域

[0001] 本公开涉及信息处理技术领域,尤其涉及一种运算方法、装置及相关产品。

背景技术

[0002] 不同的片上系统可以设置不同的语言表达,即设置不同的可用于在片上系统执行的代码。而神经网络算法自身的复杂度高,一个算法可能需要拆分后在片上系统执行。用于实现神经网络算法的编程语言也复杂多变。当使用不同编程语言实现的神经网络算法,在不同类型的片上系统上执行时,需要为特定语言的算法和特定的片上系统设置特定的接口,使得神经网络算法在片上系统实现时的复杂度高。

发明内容

[0003] 有鉴于此,本公开提出了一种运算方法、装置及相关产品,用以降低神经网络算法在芯片上实现的开发复杂度。

[0004] 根据本公开的一方面,提供了一种运算方法,所述方法包括:

[0005] 根据分段信息将卷积算法的数据分段,得到所述数据的分段数据,所述数据包括所述卷积算法的输入数据和输出数据;

[0006] 根据所述数据的分段数据确定所述数据的第一中间表达;

[0007] 根据所述数据的第一中间表达和所述卷积算法算子的第一中间表达,生成所述卷积算法的第一中间表达。

[0008] 在一种可能的实现方式中,所述数据为多维数据,根据分段信息将卷积算法的数据分段,得到所述数据的分段数据,包括:

[0009] 根据分段信息将所述数据各维度数据分别进行分段,得到所述数据各维度数据的维度分段数据;

[0010] 根据所述数据的各维度数据的维度分段数据,得到所述数据的分段数据。

[0011] 在一种可能的实现方式中,所述方法还包括:

[0012] 根据所述卷积算法的第一中间表达生成所述卷积算法的第一可执行指令。

[0013] 在一种可能的实现方式中,所述分段信息根据片上缓存的大小确定。

[0014] 在一种可能的实现方式中,所述输入数据包括输入神经元、权值、偏置,所述输出数据包括卷积输出神经元。

[0015] 在一种可能的实现方式中,所述输入神经元的各维度数据包括:特征输入通道、输入特征图高度和输入特征图宽度;

[0016] 所述卷积输出神经元的各维度数据包括:卷积通道、卷积特征图高度和卷积特征图宽度;

[0017] 所述权值的各维度数据包括:特征输入通道、卷积通道、特征图高度卷积核和特征图宽度卷积核;

[0018] 所述偏置的各维度数据包括:卷积通道。

- [0019] 根据本公开的一方面,提供了一种运算装置,所述装置包括:
- [0020] 分段数据获取模块,用于根据分段信息将卷积算法的数据分段,得到所述数据的数据分段,所述数据包括所述卷积算法的输入数据和输出数据;
- [0021] 数据中间表达确定模块,用于根据所述数据的数据分段确定所述数据的第一中间表达;
- [0022] 算法中间表达确定模块,用于根据所述数据的第一中间表达和所述卷积算法算子的第一中间表达,生成所述卷积算法的第一中间表达。
- [0023] 在一种可能的实现方式中,所述数据为多维数据,所述分段数据获取模块,包括:
- [0024] 维度分段数据获取子模块,用于根据分段信息将所述数据各维度数据分别进行分段,得到所述数据各维度数据的维度分段数据;
- [0025] 分段数据获取子模块,用于根据所述数据的各维度数据的维度分段数据,得到所述数据的数据分段。
- [0026] 在一种可能的实现方式中,所述装置还包括:
- [0027] 第一可执行指令获取模块,用于根据所述卷积算法的第一中间表达生成所述卷积算法的第一可执行指令。
- [0028] 在一种可能的实现方式中,所述分段信息根据片上缓存的大小确定。
- [0029] 在一种可能的实现方式中,所述输入数据包括输入神经元、权值、偏置,所述输出数据包括卷积输出神经元。
- [0030] 在一种可能的实现方式中,所述输入神经元的各维度数据包括:特征输入通道、输入特征图高度和输入特征图宽度;
- [0031] 所述卷积输出神经元的各维度数据包括:卷积通道、卷积特征图高度和卷积特征图宽度;
- [0032] 所述权值的各维度数据包括:特征输入通道、卷积通道、特征图高度卷积核和特征图宽度卷积核;
- [0033] 所述偏置的各维度数据包括:卷积通道。
- [0034] 根据本公开的一方面,提供了一种神经网络运算装置,所述神经网络运算装置包括一个或多个上述任一项所述的运算装置,所述神经网络运算装置用于完成设定的神经网络运算。
- [0035] 根据本公开的一方面,提供了一种组合运算装置,所述组合运算装置包括一个或多个上述的神经网络运算装置,通用互联接口和其他处理装置;
- [0036] 所述神经网络运算装置与所述其他处理装置进行交互,共同完成用户指定的计算操作。
- [0037] 根据本公开的一方面,提供了一种神经网络芯片,所述神经网络芯片包括:
- [0038] 上述的运算装置;或
- [0039] 上述的神经网络运算装置;或
- [0040] 上述的组合处理装置。
- [0041] 根据本公开的一方面,提供了一种电子设备,所述电子设备包括:
- [0042] 上述任一项所述的运算装置;或
- [0043] 上述的神经网络运算装置;或

[0044] 上述的组合处理装置;或

[0045] 上述的神经网络芯片。

[0046] 在本公开实施例中,根据分段信息将算法的数据分段得到数据的数据分段,可以根据数据的数据分段确定数据的第一中间表达,可以根据数据的第一中间表达和算法算子的第一中间表达,生成算法的第一中间表达。通过数据的第一中间表达和算子的第一中间表达得到算法的第一中间表达,不同的算法在不同的片上系统实现时,不需要设置特定的接口,提高了算法和片上系统之间的兼容性,降低了算法开发的难度。

[0047] 在一些实施例中,所述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。

[0048] 在一些实施例中,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。

[0049] 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

[0050] 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。

[0051] 图1示出根据本公开一实施例的运算方法的流程图;

[0052] 图2示出根据本公开一实施例的运算方法的流程图;

[0053] 图3示出根据本公开一实施例的运算方法的流程图;

[0054] 图4示出根据本公开一实施例的运算装置的框图;

[0055] 图5示出根据本公开一实施例的组合处理装置的框图。

具体实施方式

[0056] 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。

[0057] 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。

[0058] 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。

[0059] 图1示出根据本公开一实施例的运算方法的流程图,如图1所示,所述运算方法包括:

[0060] 步骤S10,根据分段信息将算法的数据分段,得到所述数据的数据分段,所述数据包括所述算法的输入数据和输出数据。

[0061] 在一种可能的实现方式中,神经网络的算法可以包括多个操作粒度,例如细粒度操作(具有有限尺度的操作指令)、粗粒度操作(卷积操作)和网络级别的操作。不同操作粒度对应的算法对应的数据的大小不同。由于用于完成神经网络算法的片上系统的缓存资源有限。各操作粒度对应的算法的输入数据和输出数据,可以在片上缓存中分段存储。

[0062] 在一种可能的实现方式中,对于不同的输入数据和输出数据,可以根据数据自身的长度和片上缓存的大小,预设与各输入数据和输出数据对应的分段信息。各数据可以对应相同的分段信息,也可以对应不同的分段信息。可以根据分段信息将输入数据和输出数据进行分段后,得到输入数据的数据分段数据和输出数据的数据分段数据。

[0063] 在一种可能的实现方式中,所述分段信息根据片上缓存的大小确定。所述分段信息包括分段长度和/或分段数量。例如,可以根据片上缓存的大小,确定分段信息为段的长度为B字节。卷积算法的输入数据包括输入神经元,输入神经元的长度为A字节($A > B$),可以将输入神经元分为两段,一段为B字节,一段为(A-B)字节。则输入神经元对应的分段数据为输入神经元第一分段数据B字节和输入神经元第二分段数据(A-B)字节。

[0064] 步骤S20,根据所述数据的数据分段数据生成所述数据的第一中间表达。

[0065] 在一种可能的实现方式中,可以根据需求预设中间表达及预设中间代码。可以设置C语言的中间表达。中间表达的语言可以与算法和片上系统的语言均不同,也可以和算法或片上系统的语言相同。本公开对此不做限定。

[0066] 在一种可能的实现方式中,可以将各输入数据和各输出数据的数据分段数据,分别编写各输入数据和输出数据的第一中间表达。在执行算法时,各数据可以分段执行,可以依次提取各数据的数据分段数据执行,各数据的数据分段数据可以共用相同的第一中间表达。例如,可以根据输入神经元的第一分段数据和第二分段数据,确定神经元数据的第一中间表达为Neuron input。输入神经元的第一分段数据和输入神经元的第二分段数据可以共用第一中间表达Neuron input。

[0067] 步骤S30,根据所述数据的第一中间表达和所述算法算子的第一中间表达,生成所述算法的第一中间表达。

[0068] 在一种可能的实现方式中,可以预设各算法的算子的第一中间表达。例如,卷积算法的算子的第一中间表达为ConvForward,池化算法的第一中间表达为MaxPoolForward。

[0069] 在一种可能的实现方式中,可以将算法的各输入数据的第一中间表达、各输出数据的第一中间表达和算法算子的第一中间表达,按照设定的方式组合得到算法的第一中间表达。例如,卷积算法的数据可以包括输入神经元(第一中间表达为Neuron input)、权值(第一中间表达为weight)、偏置(第一中间表达为bias),输出数据可以包括卷积输出神经元(第一中间表达为Neuron input),卷积算法的算子为卷积算子(第一中间表达为ConvForward),则卷积算法的中间表达可以为ConvForward(Neuron input,Neuron input,weight,bias)。

[0070] 在本实施例中,根据分段信息将算法的数据分段得到数据的数据分段数据,可以根据数据的数据分段数据确定数据的第一中间表达,可以根据数据的第一中间表达和算法算子的第一中间表达,生成算法的第一中间表达。通过数据的第一中间表达和算子的第一中间表达得到算法的第一中间表达,不同的算法在不同的片上系统实现时,不需要设置特定的接口,提高了算法和片上系统之间的兼容性,降低了算法开发的难度。

[0071] 图2示出根据本公开一实施例的运算方法的流程图,如图2所示,所述数据为多维数据,所述运算方法中步骤S10包括:

[0072] 步骤S11,根据分段信息将各维度数据分别进行分段,得到各维度数据的维度分段数据。

[0073] 步骤S12,根据所述数据的各维度数据的维度分段数据,得到所述数据的分段数据。

[0074] 在一种可能的实现方式中,算法的输入数据和输出数据可以是多维数据。对于输入数据的各维度数据和输出数据的各维度数据可以分别进行分段。例如,卷积算法的输入数据中包括输入神经元Neuron input,输入神经元Neuron input的各维度数据包括特征输入通道input_channal、输入特征图高度input_spatial_H和输入特征图宽度input_spatial_W。可以对特征输入通道的数据、输入特征图高度的数据和输入特征图宽度的数据分别进行分段。

[0075] 在一种可能的实现方式中,各维度数据可以对应不同的分段信息。可以根据与各维度数据对应的分段信息,得到各维度数据的维度分段数据。各维度数据的维度分段数据的长度可以不等。例如,特征输入通道input_channal的长度为C,对应的分段长度为A1,则特征输入通道的维度分段数据可以为input_channal(C,A1)。

[0076] 在一种可能的实现方式中,可以将各维度数据的维度分段数据按照设定的方式组合得到数据的分段数据。例如,输入神经元的分段数据为Neuron input(input_channal, input_spatial_H, input_spatial_W)。

[0077] 在本实施例中,可以根据分段信息将各维度数据分别进行分段,得到各维度数据的维度分段数据。根据所述数据的各维度数据的维度分段数据,得到所述数据的分段数据。将数据按照不同的维度进行分段后,可以使得数据的中间表达的使用范围更广,鲁棒性更强。

[0078] 图3示出根据本公开一实施例的运算方法的流程图,如图3所示,所述运算方法中步骤S30包括:

[0079] 步骤S31,确定数据的第一中间表达在片上缓存中对应的存储空间。

[0080] 步骤S32,根据所述存储空间生成所述数据的第二中间表达。

[0081] 步骤S33,根据所述数据的第二中间表达和所述算法算子的第二中间表达,生成所述算法的第二中间表达。

[0082] 在一种可能的实现方式中,在不同运算的算法中,可以将不同的数据存储在不同的存储器中。例如,在神经网络算法中,可以将输入神经元的数据、输出神经元的数据和权值的数据分别存储在片上块存储器上。

[0083] 在一种可能的实现方式中,可以将各数据的第一中间表达在片上缓存中对应不同的存储空间。例如,可以将输入神经元的第一中间表达对应至片上缓存的地址1、将权值的第一中间表达对应至片上缓存的地址2、将偏置的第一中间表达对应至片上缓存的地址3、将输出神经元的第一中间表达对应至片上缓存的地址4。可以为不同的数据的第一中间表达在片上缓存中确定不同大小的存储空间。

[0084] 在一种可能的实现方式中,可以根据个数据的第一中间表达对应的片上缓存的存储空间的地址,生成所述数据的第二中间表达。例如,输入神经元的第二中间表达可以为

地址1、权值的第二中间表达可以为地址2、偏置的第二中间表达可以为地址3、输出神经元的第二中间表达可以为地址4。

[0085] 在一种可能的实现方式中,可以预设各算法算子的第二中间表达。例如,卷积算子的第二中间表达为LConvForwar,池化算子的第二中间表达为LMaxPoolForward。各算子的第二中间表达可以直接调用片上缓存中的数据进行相应的运算。可以根据各数据的第二中间表达和算子的第二中间表达生成算法的第二中间表达。例如,卷积算法的第二中间表达可以为LConvForwar (地址1、地址2、地址3、地址4)。

[0086] 在本实施例中,确定数据的第一中间表达在片上缓存中对应的存储空间,根据存储空间生成数据的第二中间表达,根据数据的第二中间表达和算法算子的第二中间表达,生成算法的第二中间表达。根据为数据分配的片上缓存中的存储空间得到算法的第二中间表达,在片上缓存中直接调用数据进行计算,能够提高算法的运算效率,以及提高算法中间表达的兼容性。

[0087] 在一种可能的实现方式中,所述运算方法还包括:

[0088] 根据所述算法的第一中间表达生成所述算法的第一可执行指令,或

[0089] 根据所述算法的第二中间表达生成所述算法的第二可执行指令。

[0090] 在一种可能的实现方式中,可以预设片上系统执行代码和算法的中间表达之间的转换库。例如转换库可以利用汇编语言实现。可以为算法的第一中间表达和第二中间表达设置不同的转换库。可以利用转换库将算法的第一中间表达转换为第一可执行指令,或利用转换库将算法的第二中间表达转换为第二可执行指令。

[0091] 在本实施例中,可以根据算法的第一中间表达生成所述算法的第一可执行指令,或根据所述算法的第二中间表达生成所述算法的第二可执行指令。根据算法的中间表达生成的可执行指令可以在片上系统执行,能够提高算法中间表达的适用范围。

[0092] 在一种可能的实现方式中,所述算法包括卷积算法、池化算法、矩阵乘算法和矩阵加算法中的其中一种或任意组合。

[0093] 在一种可能的实现方式中,神经网络模型等各种运算装置中,可以用到的算法包括卷积算法、池化算法、矩阵乘算法和矩阵加算法等各种算法。可以将卷积算法、池化算法、矩阵乘算法和矩阵加算法中的其中一种或任意组合,按照本公开实施例中的运算方法转换为可执行指令后供片上系统执行。本公开对算法的类型和内容不做限定。

[0094] 在本实施例中,算法包括卷积算法、池化算法、矩阵乘算法和矩阵加算法中的其中一种或任意组合。利用本公开实施例中的运算方法将算法转换为可执行指令后供片上系统执行,提高了算法和片上系统之间的兼容性,降低了算法开发的难度。

[0095] 实施例1:

[0096] 在一种可能的实现方式中,根据分段信息将卷积算法的数据分段,得到所述数据的分段数据,所述数据包括所述卷积算法的输入数据和输出数据;

[0097] 根据所述数据的分段数据确定所述数据的第一中间表达;

[0098] 确定数据的第一中间表达在片上缓存中对应的存储空间;

[0099] 根据所述存储空间生成所述数据的第二中间表达;

[0100] 根据所述数据的第二中间表达和所述卷积算法算子的第二中间表达,生成所述卷积算法的第二中间表达。

[0101] 在一种可能的实现方式中,所述卷积算法的输入数据包括输入神经元、权值、偏置,所述卷积算法的输出数据包括卷积输出神经元。

[0102] 在一种可能的实现方式中,所述输入神经元的各维度数据包括:特征输入通道、输入特征图高度和输入特征图宽度;

[0103] 所述卷积输出神经元的各维度数据包括:卷积通道、卷积特征图高度和卷积特征图宽度;

[0104] 所述权值的各维度数据包括:特征输入通道、卷积通道、特征图高度卷积核和特征图宽度卷积核;

[0105] 所述偏置的各维度数据包括:卷积通道。

[0106] 在一种可能的实现方式中,可以根据分段信息将卷积算法的输入数据分段,得到输入数据的分段数据。即可以根据分段信息将输入神经元、权值、偏置分别进行分段,得到输入神经元分段数据、权值分段数据和偏置分段数据。可以根据分段信息将卷积算法的输出数据分段,得到输出数据的分段数据。即可以根据分段信息将卷积输出神经元进行分段,得到卷积输出神经元分段数据。在分段信息中,针对不同的输入数据和输出数据,分段长度和分段数量可以不同,本公开对此不做限定。

[0107] 在一种可能的实现方式中,所述卷积算法的数据为多维数据,根据分段信息将卷积算法的数据分段,得到所述数据的分段数据,包括:

[0108] 根据分段信息将所述数据各维度数据分别进行分段,得到所述数据各维度数据的维度分段数据;

[0109] 根据所述数据的各维度数据的维度分段数据,得到所述数据的分段数据。

[0110] 在一种可能的实现方式中,卷积算法的输入数据和输出数据为多维数据。可以根据分段信息,将卷积算法的输入数据和输出数据的各维度数据分别进行分段,得到输入数据和输出数据的各维度数据的维度分段数据。输入神经元的各维度数据包括:特征输入通道、输入特征图高度和输入特征图宽度。可以根据分段信息将特征输入通道的数据进行分段,得到特征输入通道的维度分段信息。可以根据分段信息将输入特征图高度的数据进行分段,得到输入特征图高度的维度分段信息。可以根据分段信息将输入特征图宽度的数据进行分段,得到输入特征图宽度的维度分段信息。同理,可以根据分段信息将卷积输出神经元的各维度数据进行分段,得到卷积通道的维度分段信息、卷积特征图高度的维度分段信息和卷积特征图宽度的维度分段信息。根据分段信息将权值的各维度数据进行分段,得到特征输入通道的维度分段信息、卷积通道的维度分段信息、特征图高度卷积核的维度分段信息和特征图宽度卷积核的维度分段信息。根据分段信息将偏置的各维度数据进行分段,得到卷积通道的维度分段信息。

[0111] 在一种可能的实现方式中,可以根据需求预设中间表达及预设中间代码。可以设置C语言的中间表达。中间表达的语言可以与算法和片上系统的语言均不同,也可以和算法或片上系统的语言相同。本公开对此不做限定。

[0112] 在一种可能的实现方式中,可以将卷积算法的各输入数据的分段数据和各输出数据的数据的分段数据,分别编写各输入数据的第一中间表达和各输出数据的第一中间表达。在执行算法时,各数据可以分段执行,可以依次提取各数据的数据的分段数据执行。

[0113] 在一种可能的实现方式中,各数据的数据的分段数据可以共用相同的第一中间表达。例

如,可以根据输入神经元的第一分段数据和第二分段数据,确定神经元数据的第一中间表达为Neuron input。输入神经元的第一分段数据和输入神经元的第二分段数据可以共用第一中间表达Neuron input。

[0114] 在一种可能的实现方式中,根据所述数据的第一中间表达和卷积算法算子的第一中间表达,生成卷积算法的第一中间表达。

[0115] 在一种可能的实现方式中,可以预设各算法的算子的第一中间表达。例如,卷积算法的算子的第一中间表达为ConvForward。

[0116] 在一种可能的实现方式中,可以将卷积算法的各输入数据的第一中间表达、各输出数据的第一中间表达和卷积算法算子的第一中间表达,按照设定的方式组合得到算法的第一中间表达。例如,卷积算法的数据可以包括输入神经元(第一中间表达为Neuron input)、权值(第一中间表达为weight)、偏置(第一中间表达为bias),输出数据可以包括卷积输出神经元(第一中间表达为Neuron input),卷积算法的算子为卷积算子(第一中间表达为ConvForward),则卷积算法的中间表达可以为ConvForward(Neuron input,Neuron input,weight,bias)。

[0117] 在一种可能的实现方式中,可以根据所述卷积算法的第一中间表达生成所述卷积算法的第一可执行指令,或根据所述卷积算法的第二中间表达生成所述卷积算法的第二可执行指令。

[0118] 在一种可能的实现方式中,可以预设片上系统执行代码和卷积算法的中间表达之间的转换库。例如转换库可以利用汇编语言实现。可以为卷积算法的第一中间表达和第二中间表达设置不同的转换库。可以利用转换库将卷积算法的第一中间表达转换为第一可执行指令,或利用转换库将卷积算法的第二中间表达转换为第二可执行指令。

[0119] 在本实施例中,根据分段信息将卷积算法的数据分段得到数据的分段数据,可以根据数据的分段数据确定数据的第一中间表达,可以根据数据的第一中间表达和卷积算法算子的第一中间表达,生成卷积算法的第一中间表达。通过数据的第一中间表达和算子的第一中间表达得到卷积算法的第一中间表达,卷积算法在不同的片上系统实现时,不需要设置特定的接口,提高了卷积算法和片上系统之间的兼容性,降低了算法开发的难度。

[0120] 实施例2:

[0121] 在一种可能的实现方式中,根据分段信息将卷积算法的数据分段,得到所述数据的数据分段,所述数据包括所述卷积算法的输入数据和输出数据;

[0122] 根据所述数据的数据分段确定所述数据的第一中间表达;

[0123] 根据所述数据的第一中间表达和所述卷积算法算子的第一中间表达,生成所述卷积算法的第一中间表达。

[0124] 在一种可能的实现方式中,所述数据为多维数据,根据分段信息将卷积算法的数据分段,得到所述数据的数据分段,包括:

[0125] 根据分段信息将所述数据各维度数据分别进行分段,得到所述数据各维度数据的维度分段数据;

[0126] 根据所述数据的各维度数据的维度分段数据,得到所述数据的数据分段。

[0127] 在一种可能的实现方式中,所述方法还包括:

[0128] 根据所述卷积算法的第一中间表达生成所述卷积算法的第一可执行指令。

[0129] 本实施例与实施例1的不同之处在于,只根据卷积算法的数据的第一中间表达和卷积算法算子的第一中间表达,生成卷积算法的第一中间表达,并根据卷积算法的第一中间表达生成卷积算法的第一可执行指令。没有根据卷积算法的第一中间表达在片上缓存中对应的存储空间生成所述卷积算法的第二中间表达,也没有根据卷积算法的第二中间表达生成卷积算法的第二可执行指令。

[0130] 实施例3:

[0131] 在一种可能的实现方式中,可以根据分段信息将池化算法的数据分段,得到所述数据的分段数据,所述数据包括所述池化算法的输入数据和输出数据;

[0132] 根据所述数据的分段数据确定所述数据的第一中间表达;

[0133] 确定数据的第一中间表达在片上缓存中对应的存储空间;

[0134] 根据所述存储空间生成所述数据的第二中间表达;

[0135] 根据所述数据的第二中间表达和所述池化算法算子的第二中间表达,生成所述池化算法的第二中间表达。

[0136] 在一种可能的实现方式中,所述池化算法的数据为多维数据,根据分段信息将池化算法的数据分段,得到所述数据的分段数据,包括:

[0137] 根据分段信息将各维度数据分别进行分段,得到各维度数据的维度分段数据;

[0138] 根据所述数据的各维度数据的维度分段数据,得到所述数据的分段数据。

[0139] 在一种可能的实现方式中,所述池化算法的输入数据包括卷积输出神经元,所述池化算法的输出数据包括池化输出神经元。

[0140] 在一种可能的实现方式中,可以根据分段信息将池化算法的输入数据分段,得到池化算法输入数据的分段数据。即可以根据分段信息将卷积输出神经元进行分段,得到卷积输出神经元分段数据。可以根据分段信息将池化算法的输出数据分段,得到池化算法输出数据的分段数据。即可以根据分段信息将池化输出神经元进行分段,得到池化输出神经元分段数据。在分段信息中,针对池化算法的输入数据和输出数据,分段长度和分段数量可以不同,本公开对此不做限定。

[0141] 在一种可能的实现方式中,所述卷积输出神经元的各维度数据包括:卷积通道、卷积特征图高度和卷积特征图宽度;

[0142] 所述池化输出神经元的各维度数据包括:卷积通道、池化特征图高度和池化特征图宽度。

[0143] 在一种可能的实现方式中,可以根据分段信息将池化算法中的卷积输出神经元的各维度数据进行分段,得到各维度数据的维度分段数据。即可以根据分段信息将卷积通道、卷积特征图高度和卷积特征图宽度分别进行分段,得到卷积通道的维度分段数据、卷积特征图高度的维度分段数据和卷积特征图宽度的维度分段数据。可以根据分段信息将池化算法中的池化输出神经元的各维度数据进行分段,得到各维度数据的维度分段数据。即可以根据分段信息将卷积通道、池化特征图高度和池化特征图宽度进行分段,得到卷积通道的维度分段数据、池化特征图高度的维度分段数据和池化特征图宽度的维度分段数据。

[0144] 在一种可能的实现方式中,可以根据需求预设中间表达及预设中间代码。可以设置C语言的中间表达。中间表达的语言可以与算法和片上系统的语言均不同,也可以和算法或片上系统的语言相同。本公开对此不做限定。

[0145] 在一种可能的实现方式中,可以将池化算法的各输入数据和各输出数据的分段数据,分别编写各输入数据和输出数据的第一中间表达。在执行池化算法时,各数据可以分段执行,可以依次提取各数据的分段数据执行,各数据的分段数据可以共用相同的第一中间表达。

[0146] 在一种可能的实现方式中,所述方法还包括:

[0147] 根据所述池化算法的第一中间表达生成所述池化算法的第一可执行指令,或

[0148] 根据所述池化算法的第二中间表达生成所述池化算法的第二可执行指令。

[0149] 在一种可能的实现方式中,可以预设片上系统执行代码和池化算法的中间表达之间的转换库。例如转换库可以利用汇编语言实现。可以为池化算法的第一中间表达和第二中间表达设置不同的转换库。可以利用转换库将池化算法的第一中间表达转换为第一可执行指令,或利用转换库将池化算法的第二中间表达转换为第二可执行指令。

[0150] 在本实施例中,根据分段信息将池化算法的数据分段得到数据的分段数据,可以根据数据的分段数据确定数据的第一中间表达,可以根据数据的第一中间表达和池化算法算子的第一中间表达,生成池化算法的第一中间表达。通过数据的第一中间表达和算子的第一中间表达得到池化算法的第一中间表达,池化算法在不同的片上系统实现时,不需要设置特定的接口,提高了池化算法和片上系统之间的兼容性,降低了算法开发的难度。

[0151] 应用实例4:

[0152] 在一种可能的实现方式中,根据分段信息将池化算法的数据分段,得到所述数据的分段数据,所述数据包括所述池化算法的输入数据和输出数据;

[0153] 根据所述数据的分段数据确定所述数据的第一中间表达;

[0154] 根据所述数据的第一中间表达和所述池化算法算子的第一中间表达,生成所述池化算法的第一中间表达。

[0155] 在一种可能的实现方式中,所述数据为多维数据,根据分段信息将池化算法的数据分段,得到所述数据的分段数据,包括:

[0156] 根据分段信息将各维度数据分别进行分段,得到各维度数据的维度分段数据;

[0157] 根据所述数据的各维度数据的维度分段数据,得到所述数据的分段数据。

[0158] 在一种可能的实现方式中,所述方法还包括:

[0159] 根据所述池化算法的第一中间表达生成所述池化算法的第一可执行指令。

[0160] 本实施例与实施例3的不同之处在于,只根据池化算法的数据的第一中间表达和池化算法算子的第一中间表达,生成池化算法的第一中间表达,并根据池化算法的第一中间表达生成池化算法的第一可执行指令。没有根据池化算法的第一中间表达在片上缓存中对应的存储空间生成所述池化算法的第二中间表达,以及根据池化算法的第二中间表达生成池化算法的第二可执行指令。

[0161] 实施例5:

[0162] 在一种可能的实现方式中,根据分段信息将矩阵乘算法的数据分段,得到所述数据的分段数据,所述数据包括所述矩阵乘算法的输入数据和输出数据;

[0163] 根据所述数据的分段数据确定所述数据的第一中间表达;

[0164] 确定数据的第一中间表达在片上缓存中对应的存储空间;

[0165] 根据所述存储空间生成所述数据的第二中间表达;

[0166] 根据所述数据的第二中间表达和所述矩阵乘算法算子的第二中间表达,生成所述矩阵乘算法的第二中间表达。

[0167] 在一种可能的实现方式中,所述矩阵乘算法的输入数据包括N行C列的第一矩阵数据,所述矩阵乘算法的输出数据包括N行M列的第二矩阵数据,根据分段信息将算法的数据分段,得到所述数据的数据分段,包括:

[0168] 根据分段信息将第一矩阵数据和第二矩阵数据分别分为N段,得到第一矩阵数据的数据分段和第二矩阵数据的数据分段,其中,第一矩阵数据的数据分段每段长度为C,第二矩阵数据的数据分段每段长度为M。

[0169] 在一种可能的实现方式中,可以根据分段信息将矩阵乘算法的输入数据分段,得到矩阵乘算法输入数据的数据分段。即可以根据分段信息将第一矩阵数据进行分段,得到第一矩阵数据数据分段。可以根据分段信息将矩阵乘算法的输出数据分段,得到矩阵乘算法输出数据的数据分段。即可以根据分段信息将第二矩阵数据进行分段,得到第二矩阵数据数据分段。在分段信息中,针对矩阵乘算法的输入数据和输出数据,分段长度和分段数量可以不同,本公开对此不做限定。

[0170] 在一种可能的实现方式中,可以根据需求预设中间表达及预设中间代码。可以设置C语言的中间表达。中间表达的语言可以与算法和片上系统的语言均不同,也可以和算法或片上系统的语言相同。本公开对此不做限定。

[0171] 在一种可能的实现方式中,可以将矩阵乘算法的各输入数据和各输出数据的数据分段,分别编写各输入数据和输出数据的第一中间表达。在执行矩阵乘算法时,各数据可以分段执行,可以依次提取各数据的数据分段执行,各数据的数据分段可以共用相同的第一中间表达。

[0172] 在一种可能的实现方式中,所述方法还包括:

[0173] 据所述矩阵乘算法的第一中间表达生成所述矩阵乘算法的第一可执行指令,或

[0174] 根据所述矩阵乘算法的第二中间表达生成所述矩阵乘算法的第二可执行指令。

[0175] 在一种可能的实现方式中,可以预设片上系统执行代码和矩阵乘算法的中间表达之间的转换库。例如转换库可以利用汇编语言实现。可以为矩阵乘算法的第一中间表达和第二中间表达设置不同的转换库。可以利用转换库将矩阵乘算法的第一中间表达转换为第一可执行指令,或利用转换库将矩阵乘算法的第二中间表达转换为第二可执行指令。

[0176] 在本实施例中,根据分段信息将矩阵乘算法的数据分段得到数据的数据分段,可以根据数据的数据分段确定数据的第一中间表达,可以根据数据的第一中间表达和矩阵乘算法算子的第一中间表达,生成矩阵乘算法的第一中间表达。通过数据的第一中间表达和算子的第一中间表达得到矩阵乘算法的第一中间表达,矩阵乘算法在不同的片上系统实现时,不需要设置特定的接口,提高了矩阵乘算法和片上系统之间的兼容性,降低了算法开发的难度。

[0177] 实施例6:

[0178] 在一种可能的实现方式中,根据分段信息将矩阵乘算法的数据分段,得到所述数据的数据分段,所述数据包括所述矩阵乘算法的输入数据和输出数据;

[0179] 根据所述数据的数据分段确定所述数据的第一中间表达;

[0180] 根据所述数据的第一中间表达和所述矩阵乘算法算子的第一中间表达,生成所述

矩阵乘算法的第一中间表达。

[0181] 在一种可能的实现方式中,所述方法还包括:

[0182] 根据所述矩阵乘算法的第一中间表达生成所述矩阵乘算法的第一可执行指令。

[0183] 本实施例与实施例5的不同之处在于,只根据矩阵乘算法的数据的第一中间表达和矩阵乘算法算子的第一中间表达,生成矩阵乘算法的第一中间表达,并根据矩阵乘算法的第一中间表达生成矩阵乘算法的第一可执行指令。没有根据矩阵乘算法的第一中间表达在片上缓存中对应的存储空间生成所述矩阵乘算法的第二中间表达,以及根据矩阵乘算法的第二中间表达生成矩阵乘算法的第二可执行指令。

[0184] 实施例7:

[0185] 在一种可能的实现方式中,根据分段信息将矩阵加算法的数据分段,得到所述数据的分段数据,所述数据包括所述矩阵加算法的输入数据和输出数据;

[0186] 根据所述数据的分段数据确定所述数据的第一中间表达;

[0187] 确定数据的第一中间表达在片上缓存中对应的存储空间;

[0188] 根据所述存储空间生成所述数据的第二中间表达;

[0189] 根据所述数据的第二中间表达和所述矩阵加算法算子的第二中间表达,生成所述矩阵加算法的第二中间表达。

[0190] 在一种可能的实现方式中,所述矩阵加算法的输入数据包括N行C列的第三矩阵数据和N行C列的第四矩阵数据,根据分段信息将算法的数据分段,得到所述数据的分段数据,包括:

[0191] 根据分段信息将第三矩阵数据和第四矩阵数据分别分为N段,得到第三矩阵数据的分段数据和第四矩阵数据的分段数据,其中,第三矩阵数据的分段数据每段长度为C,第四矩阵数据的分段数据每段长度为C。

[0192] 在一种可能的实现方式中,可以根据分段信息将矩阵加算法的输入数据分段,得到矩阵加算法输入数据的分段数据。即可以根据分段信息将第三矩阵数据进行分段,得到第三矩阵数据分段数据。可以根据分段信息将矩阵加算法的输出数据分段,得到矩阵加算法输出数据的分段数据。即可以根据分段信息将第四矩阵数据进行分段,得到第四矩阵数据分段数据。在分段信息中,针对矩阵加算法的输入数据和输出数据,分段长度和分段数量可以不同,本公开对此不做限定。

[0193] 在一种可能的实现方式中,所述方法还包括:

[0194] 据所述矩阵加算法的第一中间表达生成所述矩阵加算法的第一可执行指令,或

[0195] 根据所述矩阵加算法的第二中间表达生成所述矩阵加算法的第二可执行指令。

[0196] 在一种可能的实现方式中,可以预设片上系统执行代码和矩阵加算法的中间表达之间的转换库。例如转换库可以利用汇编语言实现。可以为矩阵加算法的第一中间表达和第二中间表达设置不同的转换库。可以利用转换库将矩阵加算法的第一中间表达转换为第一可执行指令,或利用转换库将矩阵加算法的第二中间表达转换为第二可执行指令。

[0197] 在本实施例中,根据分段信息将矩阵加算法的数据分段得到数据的分段数据,可以根据数据的分段数据确定数据的第一中间表达,可以根据数据的第一中间表达和矩阵加算法算子的第一中间表达,生成矩阵加算法的第一中间表达。通过数据的第一中间表达和算子的第一中间表达得到矩阵加算法的第一中间表达,矩阵加算法在不同的片上系统实现

时,不需要设置特定的接口,提高了矩阵加算法和片上系统之间的兼容性,降低了算法开发的难度。

[0198] 实施例8:

[0199] 在一种可能的实现方式中,根据分段信息将矩阵加算法的数据分段,得到所述数据的分段数据,所述数据包括所述矩阵加算法的输入数据和输出数据;

[0200] 根据所述数据的分段数据确定所述数据的第一中间表达;

[0201] 根据所述数据的第一中间表达和所述矩阵加算法算子的第一中间表达,生成所述矩阵加算法的第一中间表达。

[0202] 在一种可能的实现方式中,所述方法还包括:

[0203] 根据所述矩阵加算法的第一中间表达生成所述矩阵加算法的第一可执行指令。

[0204] 本实施例与实施例7的不同之处在于,只根据矩阵加算法的数据的第一中间表达和矩阵加算法算子的第一中间表达,生成矩阵加算法的第一中间表达,并根据矩阵加算法的第一中间表达生成矩阵加算法的第一可执行指令。没有根据矩阵加算法的第一中间表达在片上缓存中对应的存储空间生成所述矩阵加算法的第二中间表达,以及根据矩阵加算法的第二中间表达生成矩阵加算法的第二可执行指令。

[0205] 图4示出根据本公开一实施例的运算装置的框图,如图4所示,所述运算装置包括:

[0206] 分段数据获取模块10,用于根据分段信息将算法的数据分段,得到所述数据的分段数据,所述数据包括所述算法的输入数据和输出数据;

[0207] 数据中间表达确定模块20,用于根据所述数据的分段数据生成所述数据的第一中间表达;

[0208] 算法中间表达确定模块30,用于根据所述数据的第一中间表达和所述算法算子的第一中间表达,生成所述算法的第一中间表达。

[0209] 在一种可能的实现方式中,所述数据为多维数据,所述分段数据获取模块,包括:

[0210] 维度分段数据获取子模块,用于根据分段信息将各维度数据分别进行分段,得到各维度数据的维度分段数据;

[0211] 分段数据获取子模块,用于根据所述数据的各维度数据的维度分段数据,得到所述数据的分段数据。

[0212] 在一种可能的实现方式中,所述算法中间表达确定模块,包括:

[0213] 存储空间确定子模块,用于确定数据的第一中间表达在片上缓存中对应的存储空间;

[0214] 第二中间表达确定子模块,用于根据所述存储空间生成所述数据的第二中间表达;

[0215] 算法中间表达确定子模块,用于根据所述数据的第二中间表达和所述算法算子的第二中间表达,生成所述算法的第二中间表达。

[0216] 在一种可能的实现方式中,所述装置还包括:

[0217] 第一可执行指令生成模块,用于根据所述算法的第一中间表达生成所述算法的第一可执行指令,或

[0218] 第二可执行指令生成模块,用于根据所述算法的第二中间表达生成所述算法的第二可执行指令。

[0219] 在一种可能的实现方式中,所述分段信息根据片上缓存的大小确定。

[0220] 在一种可能的实现方式中,所述算法包括卷积算法、池化算法、矩阵乘算法和矩阵加算法中的其中一种或任意组合。

[0221] 在一种可能的实现方式中,所述卷积算法的输入数据包括输入神经元、权值、偏置,所述卷积算法的输出数据包括卷积输出神经元。

[0222] 在一种可能的实现方式中,所述输入神经元的各维度数据包括:特征输入通道、输入特征图高度和输入特征图宽度;

[0223] 所述卷积输出神经元的各维度数据包括:卷积通道、卷积特征图高度和卷积特征图宽度;

[0224] 所述权值的各维度数据包括:特征输入通道、卷积通道、特征图高度卷积核和特征图宽度卷积核;

[0225] 所述偏置的各维度数据包括:卷积通道。

[0226] 在一种可能的实现方式中,所述池化算法的输入数据包括卷积输出神经元,所述池化算法的输出数据包括池化输出神经元。

[0227] 在一种可能的实现方式中,所述卷积输出神经元的各维度数据包括:卷积通道、卷积特征图高度和卷积特征图宽度;

[0228] 所述池化输出神经元的各维度数据包括:卷积通道、池化特征图高度和池化特征图宽度。

[0229] 在一种可能的实现方式中,所述矩阵乘算法的输入数据包括N行C列的第一矩阵数据,所述矩阵乘算法的输出数据包括N行M列的第二矩阵数据,所述分段数据获取模块,包括:

[0230] 矩阵乘分段数据获取子模块,用于根据分段信息将第一矩阵数据和第二矩阵数据分别分为N段,得到第一矩阵数据的分段数据和第二矩阵数据的分段数据,其中,第一矩阵数据的分段数据每段长度为C,第二矩阵数据的分段数据每段长度为M。

[0231] 在一种可能的实现方式中,所述矩阵加算法的输入数据包括N行C列的第三矩阵数据和N行C列的第四矩阵数据,所述分段数据获取模块,包括:

[0232] 矩阵加分段数据获取子模块,用于根据分段信息将第三矩阵数据和第四矩阵数据分别分为N段,得到第三矩阵数据的分段数据和第四矩阵数据的分段数据,其中,第三矩阵数据的分段数据每段长度为C,第四矩阵数据的分段数据每段长度为C。

[0233] 图5示出根据本公开一实施例的组合处理装置的框图,如图5所示,所述组合处理装置,包括上述的神经网络运算装置,通用互联接口,和其他处理装置。

[0234] 神经网络运算装置与其他处理装置进行交互,共同完成用户指定的操作。其他处理装置,包括中央处理器CPU、图形处理器GPU、神经网络处理器等通用/专用处理器中的一种或以上的处理器类型。其他处理装置所包括的处理器数量不做限制。其他处理装置作为神经网络运算装置与外部数据和控制的接口,包括数据搬运,完成对本神经网络运算装置的开启、停止等基本控制;其他处理装置也可以和神经网络运算装置协作共同完成运算任务。通用互联接口,用于在所述神经网络运算装置与其他处理装置间传输数据和控制指令。该神经网络运算装置从其他处理装置中获取所需的输入数据,写入神经网络运算装置片上的存储装置;可以从其他处理装置中获取控制指令,写入神经网络运算装置片上的控制缓

存;也可以读取神经网络运算装置的存储模块中的数据并传输给其他处理装置。

[0235] 组合处理装置还可以包括存储装置,存储装置分别与所述神经网络运算装置和所述其他处理装置连接。存储装置用于保存在所述神经网络运算装置和所述其他处理装置的数据,尤其适用于所需要运算的数据在本神经网络运算装置或其他处理装置的内部存储中无法全部保存的数据。

[0236] 该组合处理装置可以作为手机、机器人、无人机、视频监控设备等设备的SOC片上系统,有效降低控制部分的核心面积,提高处理速度,降低整体功耗。此情况时,该组合处理装置的通用互联接口与设备的某些部件相连接。某些部件譬如摄像头,显示器,鼠标,键盘,网卡,wifi接口。

[0237] 在一种可能的实现方式中,本公开还提供神经网络芯片,其包括了上述神经网络运算装置或组合处理装置。

[0238] 在一种可能的实现方式中,本公开还提供芯片封装结构,其包括了上述芯片。

[0239] 在一种可能的实现方式中,本公开还提供板卡,其包括了上述芯片封装结构。

[0240] 在一种可能的实现方式中,本公开还提供电子设备,其包括了上述板卡。

[0241] 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。

[0242] 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。

[0243] 需要说明的是,对于前述的各方法实施例,为了简单描述,故将其都表述为一系列的动作组合,但是本领域技术人员应该知悉,本公开并不受所描述的动作顺序的限制,因为依据本公开,某些步骤可以采用其他顺序或者同时进行。其次,本领域技术人员也应该知悉,说明书中所描述的实施例均属于可选实施例,所涉及的动作和模块并不一定是本公开所必须的。

[0244] 在上述实施例中,对各个实施例的描述都各有侧重,某个实施例中未详述的部分,可以参见其他实施例的相关描述。

[0245] 在本公开所提供的几个实施例中,应该理解到,所揭露的装置,可通过其它的方式实现。例如,以上所描述的装置实施例仅仅是示意性的,例如所述单元的划分,仅仅为一种逻辑功能划分,实际实现时可以有另外的划分方式,例如多个单元或组件可以结合或者可以集成到另一个系统,或一些特征可以忽略,或不执行。另一点,所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口,装置或单元的间接耦合或通信连接,可以是电性或其它的形式。

[0246] 所述作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元显示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

[0247] 另外,在本公开各个实施例中的各功能单元可以集成在一个处理单元中,也可以是各个单元单独物理存在,也可以两个或两个以上单元集成在一个单元中。上述集成的单

元既可以采用硬件的形式实现,也可以采用软件程序模块的形式实现。

[0248] 所述集成的单元如果以软件程序模块的形式实现并作为独立的产品销售或使用时,可以存储在一个计算机可读取存储器中。基于这样的理解,本公开的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的全部或部分可以以软件产品的形式体现出来,该计算机软件产品存储在一个存储器中,包括若干指令用以使得一台计算机设备(可为个人计算机、服务器或者网络设备等)执行本公开各个实施例所述方法的全部或部分步骤。而前述的存储器包括:U盘、只读存储器(ROM,Read-Only Memory)、随机存取存储器(RAM,Random Access Memory)、移动硬盘、磁碟或者光盘等各种可以存储程序代码的介质。

[0249] 本领域普通技术人员可以理解上述实施例的各种方法中的全部或部分步骤可以通过程序来指令相关的硬件来完成,该程序可以存储于一计算机可读取存储器中,存储器可以包括:闪存盘、只读存储器(英文:Read-Only Memory,简称:ROM)、随机存取器(英文:Random Access Memory,简称:RAM)、磁盘或光盘等。

[0250] 以上对本公开实施例进行了详细介绍,本文中应用了具体个例对本公开的原理及实施方式进行了阐述,以上实施例的说明只是用于帮助理解本公开的方法及其核心思想;同时,对于本领域的一般技术人员,依据本公开的思想,在具体实施方式及应用范围上均会有改变之处,综上所述,本说明书内容不应理解为对本公开的限制。

[0251] 这里参照根据本公开实施例的方法、装置(系统)和计算机程序产品的流程图和/或框图描述了本公开的各个方面。应当理解,流程图和/或框图的每个方框以及流程图和/或框图中各方框的组合,都可以由计算机可读程序指令实现。

[0252] 附图中的流程图和框图显示了根据本公开的多个实施例的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上,流程图或框图中的每个方框可以代表一个模块、程序段或指令的一部分,所述模块、程序段或指令的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。在有些作为替换的实现中,方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如,两个连续的方框实际上可以基本并行地执行,它们有时也可以按相反的顺序执行,这依所涉及的功能而定。也要注意的,框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合,可以用执行规定的功能或动作的专用的基于硬件的系统来实现,或者可以用专用硬件与计算机指令的组合来实现。

[0253] 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。

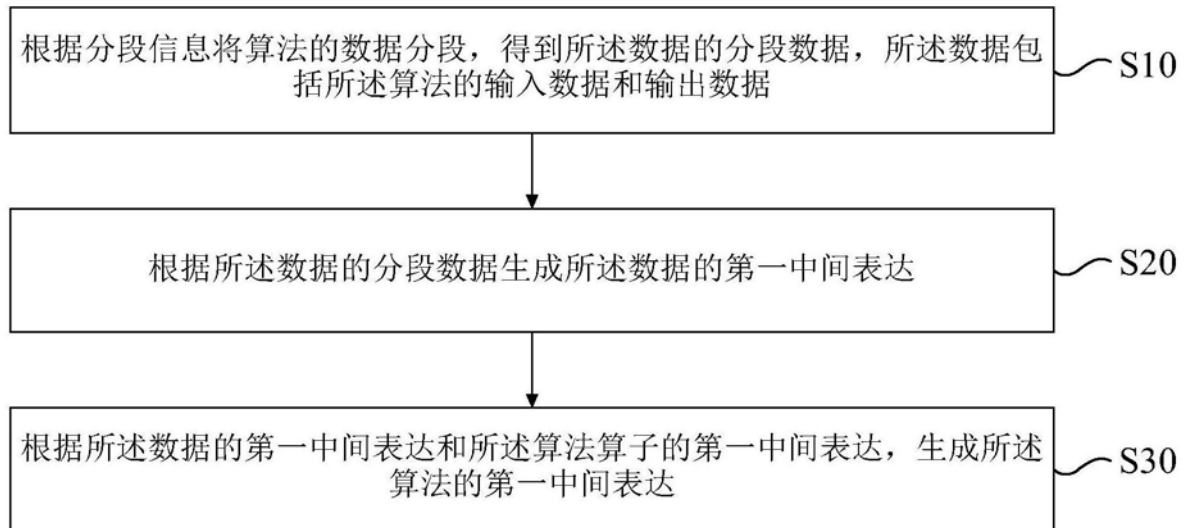


图1

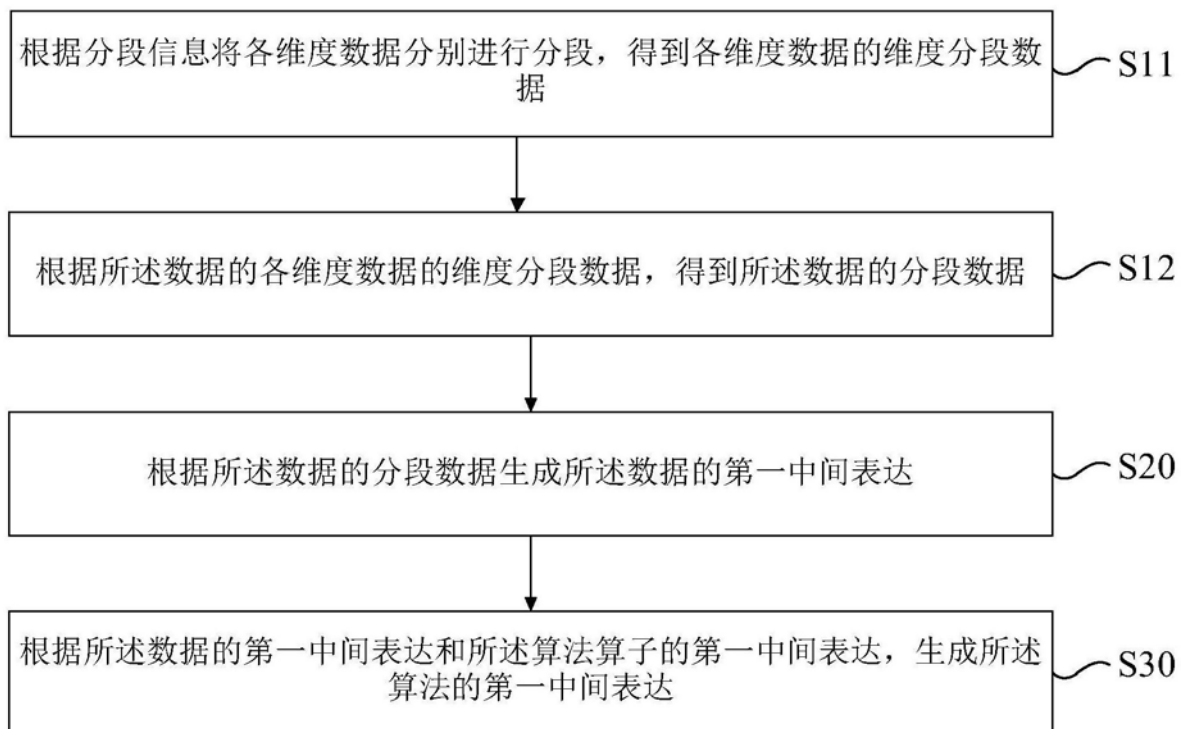


图2

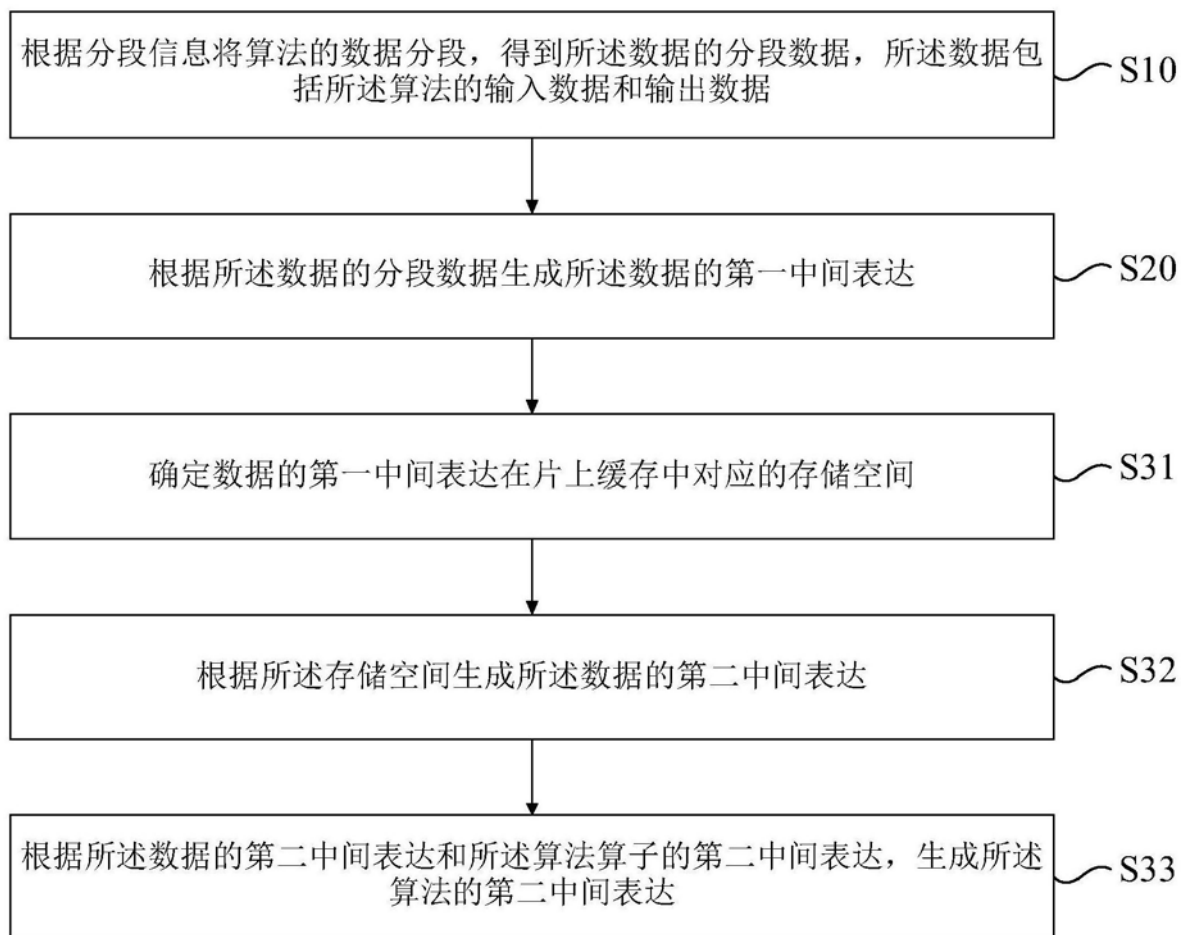


图3

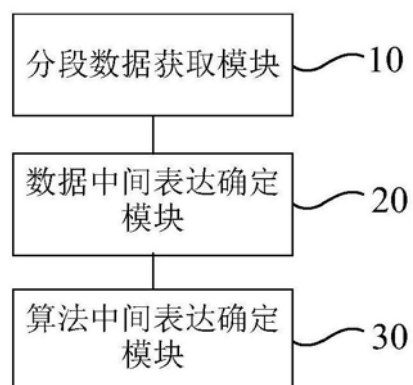


图4

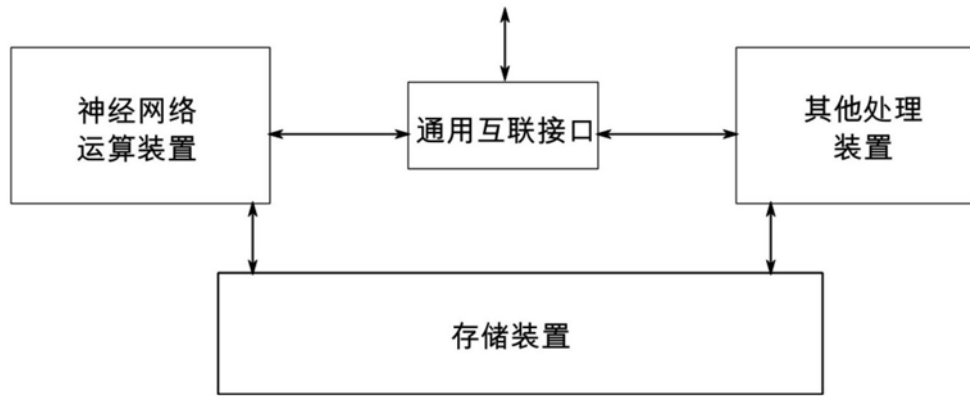


图5