

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局



(10) 国际公布号

WO 2020/199476 A1

(43) 国际公布日

2020 年 10 月 8 日 (08.10.2020)

(51) 国际专利分类号:

G06N 3/04 (2006.01)

(21) 国际申请号:

PCT/CN2019/103137

(22) 国际申请日:

2019 年 8 月 28 日 (28.08.2019)

(25) 申请语言:

中文

(26) 公布语言:

中文

(30) 优先权:

201910268881.8 2019年4月4日 (04.04.2019) CN

(71) 申请人: 平安科技(深圳)有限公司(PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) [CN/CN];

中国广东省深圳福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518057 (CN)。

(72) 发明人: 郭跃超(GUO, Yuechao); 中国广东省深圳福田区福田街道福安社区益田路 5033 号

平安金融中心 23 楼, Guangdong 518057 (CN)。

高鹏(GAO, Peng); 中国广东省深圳福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518057 (CN)。 谢国彤(XIE, Guotong); 中国广东省深圳福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518057 (CN)。 唐义君(TANG, Yijun); 中国广东省深圳福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518057 (CN)。 张萌(ZHANG, Meng); 中国广东省深圳福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518057 (CN)。

(74) 代理人: 深圳市力道知识产权代理事务所(普通合伙)(SHENZHEN LIDAO INTELLECTUAL PROPERTY AGENCY (GENERAL PARTNERSHIP)); 中国广东省深圳市福田区园岭街道八卦一路 50 号鹏基商务时空大厦 1717-1719, Guangdong 518000 (CN)。

(54) Title: NEURAL NETWORK ACCELERATION METHOD AND APPARATUS BASED ON PULSATION ARRAY, AND COMPUTER DEVICE AND STORAGE MEDIUM

(54) 发明名称: 基于脉动阵列的神经网络加速方法、装置、计算机设备及存储介质

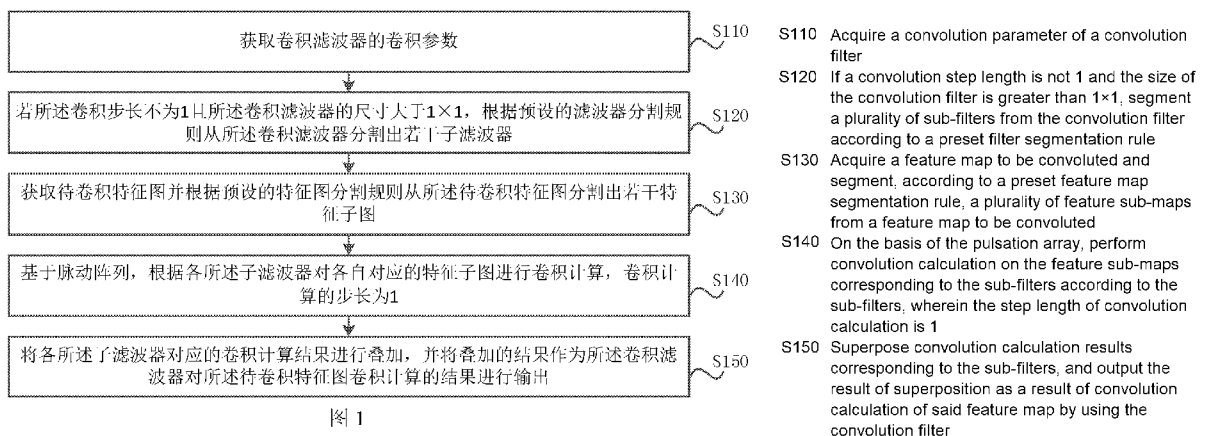


图 1

(57) Abstract: A neural network acceleration method and apparatus based on a pulsation array, and a computer device and a storage medium. The method comprises: acquiring a convolution parameter of a convolution filter; segmenting a plurality of sub-filters from the convolution filter according to a preset filter segmentation rule; segmenting, according to a preset feature map segmentation rule, a plurality of feature sub-maps from a feature map to be convoluted; on the basis of the pulsation array, performing convolution calculation on the corresponding feature sub-maps according to the sub-filters; and superposing convolution calculation results corresponding to the sub-filters.

(57) 摘要: 一种基于脉动阵列的神经网络加速方法、装置、计算机设备及存储介质, 该方法包括: 获取卷积滤波器的参数; 根据滤波器分割规则从卷积滤波器分割出若干子滤波器; 根据特征图分割规则从待卷积特征图分割出若干特征子图; 基于脉动阵列, 根据各子滤波器对对应的特征子图进行卷积计算; 将各子滤波器对应的卷积计算结果进行叠加。

(81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告 (条约第21条(3))。

说明书

基于脉动阵列的神经网络加速方法、装置、计算机设备及存储介质

本申请要求于 2019 年 4 月 4 日提交中国专利局、申请号为 201910268881.8、发明名称为“基于脉动阵列的神经网络加速方法、装置、计算机设备及存储介质”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本申请涉及神经网络技术领域，尤其涉及一种基于脉动阵列的神经网络加速方法、装置、计算机设备及存储介质。

背景技术

常用的神经网络最主要的部分就是卷积的计算，卷积计算中常常会碰到卷积滤波器不等于 1 的情况，在这种情况下，主流的一些神经网络计算库，例如 CUDNN（NVIDIA 的深度网络计算库）在计算这种卷积的时候会显著的变慢。一些深度学习加速器例如现场可编程门阵列（Field-Programmable Gate Array, FPGA），专用网络处理器（network process units, NPU）等在卷积部分通常会用脉动阵列的结构实现，此种结构对于卷积滤波器的卷积滤波器不等于 1 的情况非常不友好。

现有技术一般通过先计算卷积步长为 1 时的卷积结果，然后下采样丢弃不需要的卷积结果以得到特定卷积步长的特征图，这样做显然会浪费计算和调度资源，同样会使得卷积计算变慢。

发明内容

本申请实施例提供一种基于脉动阵列的神经网络加速方法、装置、计算机设备及存储介质，能够较佳地解决步长不为 1 的卷积计算会浪费脉动阵列计算和调度资源的问题。

第一方面，本申请提供了一种基于脉动阵列的神经网络加速方法，所述方法包括：

获取卷积滤波器的卷积参数，所述卷积参数包括卷积步长和所述卷积滤波器的尺寸；

若所述卷积步长不为 1 且所述卷积滤波器的尺寸大于 1×1 ，根据预设的滤波器分割规则从所述卷积滤波器分割出若干子滤波器，各所述子滤波器的尺寸小于所述卷积滤波器的尺寸；

获取待卷积特征图并根据预设的特征图分割规则从所述待卷积特征图分割出若干特征子图，所述若干特征子图与所述若干子滤波器一一对应；

基于脉动阵列，根据各所述子滤波器对各自对应的特征子图进行卷积计算，卷积计算的步长为 1；

将各所述子滤波器对应的卷积计算结果进行叠加，并将叠加的结果作为所述卷积滤波器对所述待卷积特征图卷积计算的结果进行输出。

第二方面，本申请提供了一种基于脉动阵列的神经网络加速装置，所述装置包括：

卷积参数获取模块，用于获取卷积滤波器的卷积参数，所述卷积参数包括卷积步长和所述卷积滤波器的尺寸；

滤波器分割模块，用于若所述卷积步长不为 1 且所述卷积滤波器的尺寸大于 1×1 ，根据预设的滤波器分割规则从所述卷积滤波器分割出若干子滤波器，各所述子滤波器的尺寸小于所述卷积滤波器的尺寸；

特征图分割模块，用于获取待卷积特征图并根据预设的特征图分割规则从所述待卷积特征图分割出若干特征子图，所述若干特征子图与所述若干子滤波器一一对应；

卷积模块，用于基于脉动阵列，根据各所述子滤波器对各自对应的特征子图进行卷积计算，卷积计算的步长为 1；

叠加模块，用于将各所述子滤波器对应的卷积计算结果进行叠加，并将叠加的结果作为

说明书

所述卷积滤波器对所述待卷积特征图卷积计算的结果进行输出。

第三方面，本申请提供了一种计算机设备，所述计算机设备包括存储器和处理器；所述存储器用于存储计算机程序；所述处理器，用于执行所述计算机程序并在执行所述计算机程序时实现上述的基于脉动阵列的神经网络加速方法。

第四方面，本申请提供了一种计算机可读存储介质，所述计算机可读存储介质存储有计算机程序，若所述计算机程序被处理器执行，实现上述的基于脉动阵列的神经网络加速方法。

本申请公开了一种基于脉动阵列的神经网络加速方法、装置、设备及存储介质，通过在卷积步长不为 1 时根据预设的滤波器分割规则从卷积滤波器分割出若干子滤波器，以及根据预设的特征图分割规则从待卷积特征图分割出若干特征子图实现可以以卷积步长为 1 执行卷积计算，各子滤波器对应的卷积计算结果进行叠加后的叠加结果与根据原卷积滤波器对待卷积特征图执行的卷积步长不为 1 的卷积计算结果相同，即分割操作前后的两种卷积计算是等价的；但是由于分割操作后卷积步长为 1，可以更充分的利用脉动阵列的计算能力。

附图说明

为了更清楚地说明本申请实施例的技术方案，下面将对实施例描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图是本申请的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

图 1 为本申请一实施例的基于脉动阵列的神经网络加速方法的流程示意图；

图 2 为本申请另一实施例的基于脉动阵列的神经网络加速方法的流程示意图；

图 3 为卷积步长为 2 且卷积滤波器尺寸为 2×2 时分割卷积滤波器的示意图；

图 4 为卷积步长为 2 且卷积滤波器尺寸为 2×2 时分割待卷积特征图的示意图；

图 5 为图 1 中分割待卷积特征图一实施方式的子流程示意图；

图 6 为分割补零后的待卷积特征图的示意图；

图 7 为脉动阵列的结构示意图；

图 8 为脉动阵列进行卷积计算的示意图；

图 9 为图 1 中基于脉动阵列进行卷积计算的子流程示意图；

图 10 为本申请再一实施例的基于脉动阵列的神经网络加速方法的流程示意图；

图 11 为卷积步长为 2 且卷积滤波器尺寸为 3×3 时分割卷积滤波器的示意图；

图 12 为卷积步长为 2 且卷积滤波器尺寸为 3×3 时分割卷积滤波器的子流程示意图；

图 13 为卷积步长为 2 且卷积滤波器尺寸为 3×3 时分割待卷积特征图的示意图；

图 14 为本申请又一实施例的基于脉动阵列的神经网络加速方法的流程示意图；

图 15 为卷积步长为 3 且卷积滤波器尺寸为 3×3 时分割卷积滤波器的示意图；

图 16 为卷积步长为 3 且卷积滤波器尺寸为 3×3 时分割待卷积特征图的子流程示意图；

图 17 为卷积步长为 3 且卷积滤波器尺寸为 3×3 时分割待卷积特征图的示意图；

图 18 为根据神经网络加速方法对深度卷积神经网络下采样的拓扑结构进行等价变换的示意图；

图 19 为本申请实施例的基于脉动阵列的神经网络加速装置的结构示意图；

图 20 为本申请另一实施例的基于脉动阵列的神经网络加速装置的结构示意图；

图 21 为本申请一实施例提供的一种计算机设备的结构示意图。

具体实施方式

下面将结合本申请实施例中的附图，对本申请实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例是本申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本申请保护的范围。

说明书

附图中所示的流程图仅是示例说明，不是必须包括所有的内容和操作/步骤，也不是必须按所描述的顺序执行。例如，有的操作/步骤还可以分解、组合或部分合并，因此实际执行的顺序有可能根据实际情况改变。另外，虽然在装置示意图中进行了功能模块的划分，但是在某些情况下，可以以不同于装置示意图中的模块划分。

本申请的实施例提供了一种基于脉动阵列的神经网络加速方法、装置、设备及存储介质。其中，该基于脉动阵列的神经网络加速方法可以应用于终端或服务器中，以实现加速基于脉动阵列的神经网络的训练或推理。

下面结合附图，对本申请的一些实施方式作详细说明。在不冲突的情况下，下述的实施例及实施例中的特征可以相互组合。

请参阅图 1，图 1 是本申请的实施例提供的一种基于脉动阵列的神经网络加速方法的流程示意图。

如图 1 所示，基于脉动阵列的神经网络加速方法包括以下步骤：

步骤 S110、获取卷积滤波器的卷积参数。

其中，所述卷积参数包括卷积步长和所述卷积滤波器的尺寸。

滤波器 filter，又称内核 kernel、特征检测器 (feature detector)，在输入的图像或特征图上滑动滤波器并计算点乘即为卷积操作，卷积操作的输出的矩阵叫做卷积特征 (Convolved Feature)、激活图 (Activation Map) 或者特征图 (Feature Map)。

示例性的，在神经网络中的神经元进行卷积操作前，先获取预先存储的或者初始化的卷积滤波器的卷积参数。

在本实施例中，卷积参数包括卷积步长 stride 和所述卷积滤波器的尺寸，即卷积滤波器的高度 h 和宽度 w；在另一些实施例中，卷积参数还包括输入通道数和/或输出通道数；其中输入通道数 in depth 是由待卷积特征图的通道数所决定的，输出通道数 out depth 等于卷积滤波器的个数，可以决定卷积结束后输出特征图的通道数。

基于脉动阵列的神经网络加速方法可以用于输入通道数等于 1 或大于 1 的场景，也可用于输出通道数等于 1 或大于 1 的场景。

步骤 S120、若所述卷积步长不为 1 且所述卷积滤波器的尺寸大于 1×1 ，根据预设的滤波器分割规则从所述卷积滤波器分割出若干子滤波器。

其中，各所述子滤波器的尺寸小于所述卷积滤波器的尺寸。

一些深度学习加速器例如 FPGA，专用 NPU 等在卷积部分通常会用脉动阵列的结构实现，但是这种结构对于卷积步长不等于 1 的情况非常不友好；本实施例将尺寸大于 1×1 卷积滤波器分割为若干子滤波器，以使得各子滤波器分别以等于 1 的卷积步长执行卷积操作，从而充分利用脉动阵列结构的性能。

在一些实施例中，如图 2 和图 3 所示，步骤 S120 若所述卷积步长不为 1 且所述卷积滤波器的尺寸大于 1×1 ，根据预设的滤波器分割规则从所述卷积滤波器分割出若干子滤波器，具体包括：

步骤 S121、若所述卷积步长为 2 且所述卷积滤波器的尺寸为 2×2 ，从所述卷积滤波器分割出 4 个子滤波器，各所述子滤波器的尺寸为 1×1 。

如图 3 所示，某卷积操作对应的卷积参数为 [6 6 2 2]，即输入通道数 in depth 为 6，输出通道数 out depth 等于 6，6 个卷积滤波器 Kernel Tensor 的尺寸均为 2×2 。

如图 3 所示，将每个 2×2 的卷积滤波器分别分割为 4 个 1×1 的子滤波器。以第一个卷积滤波器分割的 4 个子滤波器为例，其中第一个子滤波器包括所述卷积滤波器奇数行奇数列的权值 w1，第二个子滤波器包括所述卷积滤波器奇数行偶数列的权值 w2，第三个子滤波器包括所述卷积滤波器偶数行奇数列的权值 w3，第四个子滤波器包括所述卷积滤波器偶数行偶数列的权值 w4。

具体的，将卷积滤波器第一行第一列的权值分配给第一个 1×1 的子滤波器，将卷积滤波

说明书

器第一行第二列的权值分配给第二个 1×1 的子滤波器, 将卷积滤波器第二行第一列的权值分配给第三个 1×1 的子滤波器, 将卷积滤波器第二行第二列的权值分配给第四个 1×1 的子滤波器。

步骤 S130、获取待卷积特征图并根据预设的特征图分割规则从所述待卷积特征图分割出若干特征子图。

其中, 所述若干特征子图与所述若干子滤波器一一对应。

示例性的, 待卷积特征图的通道数可以等于 1 或大于 1, 待卷积特征图的通道数可以决定相应卷积滤波器卷积参数中的输入通道数。

在一些实施例中, 如图 2 和图 4 所示, 若所述卷积步长为 2 且所述卷积滤波器的尺寸为 2×2 , 步骤 S130 获取待卷积特征图并根据预设的特征图分割规则从所述待卷积特征图分割出若干特征子图, 具体包括:

步骤 S1311、将所述待卷积特征图奇数行奇数列的数值分配至第一个特征子图的相应位置。

步骤 S1312、将所述待卷积特征图奇数行偶数列的数值分配至第二个特征子图的相应位置。

步骤 S1313、将所述待卷积特征图偶数行奇数列的数值分配至第三个特征子图的相应位置。

步骤 S1314、将所述待卷积特征图偶数行偶数列的数值分配至第四个特征子图的相应位置。

示例性的, 待卷积特征图中同一行的数值在各特征子图中也位于同一行, 待卷积特征图中同一列的数值在各特征子图中也位于同一列。

如图 4 所示, 获取到的卷积特征图 input Tensor 是 $[1\ 6\ 4\ 4]$ 的卷积特征图, 该待卷积特征图的通道数为 6, 宽和高均为 4。根据预设的分割特征从所述待卷积特征图分割出了 4 个特征子图。

在一些实施例中, 如图 5 所示, 步骤 S130 获取待卷积特征图并根据预设的特征图分割规则从所述待卷积特征图分割出若干特征子图具体包括:

步骤 S131、获取待卷积特征图。

示例性的, 获取的卷积特征图如图 6 所示。

步骤 S132、若获取的待卷积特征图的长或宽不是所述卷积步长的整数倍, 对所述待卷积特征图的预设位置进行补零以使补零后的待卷积特征图的长或宽是所述卷积步长的整数倍。

在本实施例中, 获取的卷积特征图的长、宽均为 3, 示例性的在该卷积特征图的右侧和下方补零 padding, 使得补零后的待卷积特征图的长、宽为 4。

步骤 S133、根据预设的特征图分割规则从补零后的待卷积特征图分割出若干特征子图。

如图 6 所示, 根据上述步骤 S1311-步骤 S1314 从补零后的待卷积特征图分割出了 4 个特征子图。

在一些实施例中, 待卷积特征图的结构为 NCHW, 如 $[1\ 6\ 4\ 4]$, N 代表数量, C 代表通道数 channel, H 代表高度, W 代表宽度; 因此实例数 batchsize=1, 通道数 channel=6, 待卷积特征图的高度 H=4, 待卷积特征图的宽度 W=4 时, 即待卷积张量的数量为 1, 有 6 个通道, 各通道均为一个待卷积特征图。可以根据先分割、卷积同一数量编号下不同通道的待卷积特征图, 然后分割、卷积下一数量编号下不同通道的待卷积特征图。

步骤 S140、基于脉动阵列, 根据各所述子滤波器对各自对应的特征子图进行卷积计算, 卷积计算的步长为 1。

脉动阵列 (Systolic Array) 核心概念就是让数据在运算单元的阵列中进行流动, 减少访存的次数, 并且使得结构更加规整, 布线更加统一, 提高频率。

在一些实施例中, 如图 7 所示, 脉动阵列 Systolic Array 包括 $L \times L$ 个处理单元 PE, 所述

说明书

脉动阵列连接于权值寄存器 filter buffer、输入寄存器 in buffer 和输出寄存器 out buffer。每行处理单元 PE 的左侧、每列处理单元 PE 的上侧都设有先进先出寄存器 FIFO。滤波器的权值通过先进先出寄存器 FIFO 存储并传输给同一行的所有处理单元 PE，第一行和第一列的处理单元 PE 接收来自输入寄存器中待卷积特征图的数据，并且第一行和第一列的处理单元 PE 均向各自右下角的处理单元 PE 传输来自待卷积特征图的数据。这样的设计最大化了数据的复用。

示例性的，如图 8 所示，脉动阵列根据一个 3×3 滤波器 W 对一个 5×5 的特征图 X 进行二维卷积。

假设滤波器 W 和特征图 X 有以下形式：

$$W = \begin{bmatrix} w1 \\ w2 \\ w3 \end{bmatrix}, \quad X = \begin{bmatrix} x1 \\ x2 \\ x3 \\ x4 \\ x5 \end{bmatrix}$$

其中， w_i 和 x_j 分别代表滤波器 W 和特征图 X 的某一行数据，则最后一行的三个处理单元 PE 输出三行卷积结果：

$$Y = W * X = \begin{bmatrix} w1 \times x1 + w2 \times x2 + w3 \times x3 \\ w1 \times x2 + w2 \times x3 + w3 \times x4 \\ w1 \times x3 + w2 \times x4 + w3 \times x5 \end{bmatrix}$$

其中 * 表示一维卷积计算。

在一些实施例中，如图 9 所示，步骤 S140 基于脉动阵列，根据各所述子滤波器对各自对应的特征子图进行卷积计算，具体包括：

步骤 S141、将所述子滤波器的权值加载至连接于所述脉动阵列的权值寄存器中。

示例性的，将子滤波器的权值加载至权值寄存器 filter buffer 中，子滤波器的权值通过先进先出寄存器 FIFO 存储并传输给同一行的处理单元 PE。

步骤 S142、将所述子滤波器对应的特征子图加载至连接于所述脉动阵列的输入寄存器中。

示例性的，将与子滤波器对应的特征子图加载至输入寄存器 in buffer 中，脉动阵列第一行和第一列的处理单元 PE 接收来自输入寄存器 in buffer 中特征子图的数据。

步骤 S143、获取所述脉动阵列卷积计算后的输出结果。

示例性的，脉动阵列第一行和第一列的处理单元 PE 均向各自右下角的处理单元 PE 传输来自特征子图的数据；最后一行的处理单元 PE 输出所述子滤波器对对应的特征子图进行卷积计算，卷积步长为 1 的卷积结果。

如图 3 所示，第一个子滤波器的权值为 $w1$ ，对图 4 中与其对应的第一个特征子图进行卷积计算；第二个子滤波器的权值为 $w2$ ，对第二个特征子图进行卷积计算；第三个子滤波器的权值为 $w3$ ，对第三个特征子图进行卷积计算；第四个子滤波器的权值为 $w4$ ，对第四个特征子图进行卷积计算。第一至第四个子滤波器对应的卷积计算的结果如下：

$$\begin{bmatrix} w1 \times p11 & w1 \times p13 \\ w1 \times p31 & w1 \times p33 \end{bmatrix}, \begin{bmatrix} w2 \times p12 & w2 \times p14 \\ w2 \times p32 & w2 \times p34 \end{bmatrix}, \begin{bmatrix} w3 \times p21 & w3 \times p23 \\ w3 \times p41 & w3 \times p43 \end{bmatrix}, \begin{bmatrix} w4 \times p22 & w4 \times p24 \\ w4 \times p42 & w4 \times p44 \end{bmatrix}。$$

步骤 S150、将各所述子滤波器对应的卷积计算结果进行叠加，并将叠加的结果作为所述卷积滤波器对所述待卷积特征图卷积计算的结果进行输出。

示例性的，将 4 个子滤波器对应的卷积计算结果进行叠加，得到：

说明书

$$\begin{bmatrix} w1 \times p11 + w2 \times p12 + w3 \times p21 + w4 \times p22 & w1 \times p13 + w2 \times p14 + w3 \times p23 + w4 \times p24 \\ w1 \times p31 + w2 \times p32 + w3 \times p41 + w4 \times p42 & w1 \times p33 + w2 \times p34 + w3 \times p41 + w4 \times p44 \end{bmatrix}。$$

如果直接根据图 3 左侧的卷积滤波器对图 4 左侧的卷积特征图以卷积步长为 2 进行卷积计算，卷积计算的结果为：

$$\begin{bmatrix} w1 \times p11 + w2 \times p12 + w3 \times p21 + w4 \times p22 & w1 \times p13 + w2 \times p14 + w3 \times p23 + w4 \times p24 \\ w1 \times p31 + w2 \times p32 + w3 \times p41 + w4 \times p42 & w1 \times p33 + w2 \times p34 + w3 \times p41 + w4 \times p44 \end{bmatrix}。$$

因此，本实施例的基于脉动阵列的神经网络加速方法，通过在卷积步长不为 1 时根据预设的滤波器分割规则从所述卷积滤波器分割出若干子滤波器，以及根据预设的特征图分割规则从所述待卷积特征图分割出若干特征子图实现可以以卷积步长为 1 执行卷积计算，各子滤波器对应的卷积计算结果进行叠加后的叠加结果与根据原卷积滤波器对待卷积特征图执行的卷积步长不为 1 的卷积计算结果相同，即分割操作前后的两种卷积计算是等价的；因此叠加的结果可以作为所述卷积滤波器对所述待卷积特征图卷积计算的结果进行输出，以进行后续的处理如再一次卷积、池化、分类等；但是由于分割操作后卷积步长为 1，可以更充分的利用脉动阵列的计算能力。

示例性的，如图 10 和图 11 所示，步骤 S120 若所述卷积步长不为 1 且所述卷积滤波器的尺寸大于 1×1 ，根据预设的滤波器分割规则从所述卷积滤波器分割出若干子滤波器，具体包括：

步骤 S122、若所述卷积步长为 2 且所述卷积滤波器的尺寸为 3×3 ，从所述卷积滤波器分割出 4 个子滤波器，各所述子滤波器的尺寸为 2×2 。

其中第一个子滤波器包括所述卷积滤波器奇数行奇数列的权值，第二个子滤波器包括所述卷积滤波器奇数行偶数列的权值，第三个子滤波器包括所述卷积滤波器偶数行奇数列的权值，第四个子滤波器包括所述卷积滤波器偶数行偶数列的权值。

在一些实施例中，卷积滤波器 kernel (filter) 的尺寸无法整除卷积步长 (stride)，可以通过在卷积滤波器的预设位置进行补零以使补零后的卷积滤波器的长或宽是所述卷积步长的整数倍。在本实施例中，卷积滤波器的尺寸为 3×3 ，卷积步长为 2，卷积滤波器的尺寸无法整除卷积步长，可以通过补零操作以使补零后的卷积滤波器的长或宽是所述卷积步长的整数倍，从而使卷积滤波器可以根据预设的滤波器分割规则分割出若干子滤波器。

具体的，如图 11 和图 12 所示，所述若所述卷积步长为 2 且所述卷积滤波器的尺寸为 3×3 ，从所述卷积滤波器分割出 4 个子滤波器，各所述子滤波器的尺寸为 2×2 ，具体包括：

步骤 S11、将所述卷积滤波器奇数行奇数列的权值分配至第一个子滤波器。

步骤 S12、将所述卷积滤波器奇数行偶数列的权值分配至第二个子滤波器的第一列，并以 0 填充所述第二个子滤波器的第二列。

步骤 S13、将所述卷积滤波器偶数行奇数列的权值分配至第三个子滤波器的第一行，并以 0 填充所述第三个子滤波器的第二行。

步骤 S14、将所述卷积滤波器偶数行偶数列的权值分配至第四个子滤波器的第一行第一列，并以 0 填充所述第四个子滤波器的其余位置。

在本实施例中，如图 10 和图 13 所示，若所述卷积步长为 2 且所述卷积滤波器的尺寸为 3×3 ，步骤 S130 获取待卷积特征图并根据预设的特征图分割规则从所述待卷积特征图分割出若干特征子图，具体包括：

步骤 S1321、将所述待卷积特征图奇数行奇数列的数值分配至第一个特征子图的相应位置。

步骤 S1322、将所述待卷积特征图奇数行偶数列的数值分配至第二个特征子图的相应位置。

说明书

步骤 S1323、将所述待卷积特征图偶数行奇数列的数值分配至第三个特征子图的相应位置。

步骤 S1324、将所述待卷积特征图偶数行偶数列的数值分配至第四个特征子图的相应位置。

卷积步长为 2 且所述卷积滤波器的尺寸为 3×3 时, 根据预设的特征图分割规则从所述待卷积特征图分割出 4 个特征子图; 如果待卷积特征图的通道数为 1, 则分割后的通道数为 4。

在一些实施例中, 如图 14 和图 15 所示, 步骤 S120 若所述卷积步长不为 1 且所述卷积滤波器的尺寸大于 1×1 , 根据预设的滤波器分割规则从所述卷积滤波器分割出若干子滤波器, 具体包括:

步骤 S123、若所述卷积步长为 3 且所述卷积滤波器的尺寸为 3×3 , 从所述卷积滤波器分割出 9 个子滤波器, 各所述子滤波器的尺寸为 1×1 且分别包括所述卷积滤波器 9 个权值中的一个。

示例性的, 将卷积滤波器第一行第一列的权值分配给第一个 1×1 的子滤波器, 将卷积滤波器第一行第二列的权值分配给第二个 1×1 的子滤波器, 将卷积滤波器第一行第三列的权值分配给第三个 1×1 的子滤波器, 将卷积滤波器第二行第一列的权值分配给第四个 1×1 的子滤波器, 将卷积滤波器第二行第二列的权值分配给第五个 1×1 的子滤波器, 以此类推。

在本实施例中, 如图 16 和图 17 所示, 若所述卷积步长为 3 且所述卷积滤波器的尺寸为 3×3 , 步骤 S130 获取待卷积特征图并根据预设的特征图分割规则从所述待卷积特征图分割出若干特征子图, 具体包括:

步骤 S1331、将所述待卷积特征图第 $3n+1$ 行第 $3n+1$ 列的数值分配至第一个特征子图的相应位置。其中 n 为自然数。

步骤 S1332、将所述待卷积特征图第 $3n+1$ 行第 $3n+2$ 列的数值分配至第二个特征子图的相应位置。

步骤 S1333、将所述待卷积特征图第 $3n+1$ 行第 $3n+3$ 列的数值分配至第三个特征子图的相应位置。

步骤 S1334、将所述待卷积特征图第 $3n+2$ 行第 $3n+1$ 列的数值分配至第四个特征子图的相应位置。

步骤 S1335、将所述待卷积特征图第 $3n+2$ 行第 $3n+2$ 列的数值分配至第五个特征子图的相应位置。

步骤 S1336、将所述待卷积特征图第 $3n+2$ 行第 $3n+3$ 列的数值分配至第六个特征子图的相应位置。

步骤 S1337、将所述待卷积特征图第 $3n+3$ 行第 $3n+1$ 列的数值分配至第七个特征子图的相应位置。

步骤 S1338、将所述待卷积特征图第 $3n+3$ 行第 $3n+2$ 列的数值分配至第八个特征子图的相应位置。

步骤 S1339、将所述待卷积特征图第 $3n+3$ 行第 $3n+3$ 列的数值分配至第九个特征子图的相应位置。

示例性的, 如图 17 所示, 获取到的待卷积特征图的长宽均为 8, 不是所述卷积步长, 即 3 的整数倍, 则对所述待卷积特征图的预设位置进行补零以使补零后的待卷积特征图的长或宽是所述卷积步长的整数倍; 然后根据预设的特征图分割规则从补零后的待卷积特征图分割出 9 个特征子图。

本申请的基于脉动阵列的神经网络加速方法通过在卷积步长不为 1 时根据预设的滤波器分割规则从卷积滤波器分割出若干子滤波器, 以及根据预设的特征图分割规则从待卷积特征图分割出若干特征子图实现可以以卷积步长为 1 执行卷积计算。可以很好的与一些特殊的专用深度网络加速器例如 FPGA, NPU 等底层都会采用的脉动阵列 (Systolic Array) 的结构适

说明书

配，可以节省计算资源，而且这种分割方法本身就是一种特殊的计算逻辑，可以集成进入各种深度学习的框架中。本申请提供的分割变换方法，并不影响深度网络本身的前传和后传路径，而且因为节省了计算资源，实际上提升了训练和推理的速度。

在一些实施例中，如图 18 所示为根据本申请的基于脉动阵列的神经网络加速方法对传统的深度卷积神经网络 ResNet50 的下采样的拓扑结构进行分割变换的示意图；箭头左侧为传统的深度卷积神经网络 ResNet50 的下采样的拓扑结构部分的简化模型，箭头右侧为经过分割变换等价变换后的计算拓扑结构。

相较于传统的深度卷积神经网络的计算图结构，经过分割变换等价变换后的计算拓扑结构如下的优势：1.省去了传统 ResNet50 在左侧的两个 1×1 的映射卷积，减少计算资源。2.传统 ResNet50 右侧的残差分量部分，可以转换为直接恒等映射 (Identity Mapping)，有利于残差的传播。本申请的基于脉动阵列的神经网络加速方法可以应用在很多网络模型中，例如 densNet，或者 shakeshake 网络等等，只要存于在下采样部分的网络，都可以采用本申请提供的神经网络加速方法做变换后，再进行计算训练等。

请参阅图 19，图 19 是本申请一实施例提供的基于脉动阵列的神经网络加速装置的结构示意图，该基于脉动阵列的神经网络加速装置可以配置于服务器中，用于执行前述的基于脉动阵列的神经网络加速方法。

如图 19 所示，该基于脉动阵列的神经网络加速装置，包括：

卷积参数获取模块 110，用于获取卷积滤波器的卷积参数，所述卷积参数包括卷积步长和所述卷积滤波器的尺寸。

滤波器分割模块 120，用于若所述卷积步长不为 1 且所述卷积滤波器的尺寸大于 1×1 ，根据预设的滤波器分割规则从所述卷积滤波器分割出若干子滤波器，各所述子滤波器的尺寸小于所述卷积滤波器的尺寸。

特征图分割模块 130，用于获取待卷积特征图并根据预设的特征图分割规则从所述待卷积特征图分割出若干特征子图，所述若干特征子图与所述若干子滤波器一一对应。

卷积模块 140，用于基于脉动阵列，根据各所述子滤波器对各自对应的特征子图进行卷积计算，卷积计算的步长为 1。

叠加模块 150，用于将各所述子滤波器对应的卷积计算结果进行叠加，并将叠加的结果作为所述卷积滤波器对所述待卷积特征图卷积计算的结果进行输出。

在一些实施例中，如图 20 所示，特征图分割模块 130 包括：

特征图获取子模块 131，用于获取待卷积特征图。

补零子模块 132，用于若获取的待卷积特征图的长或宽不是所述卷积步长的整数倍，对所述待卷积特征图的预设位置进行补零以使补零后的待卷积特征图的长或宽是所述卷积步长的整数倍。

特征图分割子模块 133，用于根据预设的特征图分割规则从补零后的待卷积特征图分割出若干特征子图。

在一些实施例中，如图 20 所示，卷积模块 140 包括：

权值加载子模块 141，用于将所述子滤波器的权值加载至连接于所述脉动阵列的权值寄存器中；

子图加载子模块 142，用于将所述子滤波器对应的特征子图加载至连接于所述脉动阵列的输入寄存器中。

输出子模块 143，用于获取所述脉动阵列卷积计算后的输出结果。

在一些实施例中，如图 20 所示，滤波器分割模块 120 包括第一滤波器分割子模块 121，用于若所述卷积步长为 2 且所述卷积滤波器的尺寸为 2×2 ，从所述卷积滤波器分割出 4 个子滤波器，各所述子滤波器的尺寸为 1×1 ；其中第一个子滤波器包括所述卷积滤波器奇数行奇数列的权值，第二个子滤波器包括所述卷积滤波器奇数行偶数列的权值，第三个子滤波器包

说明书

括所述卷积滤波器偶数行奇数列的权值，第四个子滤波器包括所述卷积滤波器偶数行偶数列的权值。

特征图分割模块 130 包括第一特征图分割子模块 1301，用于若所述卷积步长为 2 且所述卷积滤波器的尺寸为 2×2 ，将所述待卷积特征图奇数行奇数列的数值分配至第一个特征子图的相应位置，将所述待卷积特征图奇数行偶数列的数值分配至第二个特征子图的相应位置，将所述待卷积特征图偶数行奇数列的数值分配至第三个特征子图的相应位置，将所述待卷积特征图偶数行偶数列的数值分配至第四个特征子图的相应位置。

在一些实施例中，如图 20 所示，滤波器分割模块 120 包括第二滤波器分割子模块 122，用于若所述卷积步长为 2 且所述卷积滤波器的尺寸为 3×3 ，从所述卷积滤波器分割出 4 个子滤波器，各所述子滤波器的尺寸为 2×2 ；其中第一个子滤波器包括所述卷积滤波器奇数行奇数列的权值，第二个子滤波器包括所述卷积滤波器奇数行偶数列的权值，第三个子滤波器包括所述卷积滤波器偶数行奇数列的权值，第四个子滤波器包括所述卷积滤波器偶数行偶数列的权值。

特征图分割模块 130 包括第二特征图分割子模块 1302，用于若所述卷积步长为 2 且所述卷积滤波器的尺寸为 3×3 ，将所述待卷积特征图奇数行奇数列的数值分配至第一个特征子图的相应位置，将所述待卷积特征图奇数行偶数列的数值分配至第二个特征子图的相应位置，将所述待卷积特征图偶数行奇数列的数值分配至第三个特征子图的相应位置，将所述待卷积特征图偶数行偶数列的数值分配至第四个特征子图的相应位置。

在一些实施例中，如图 20 所示，滤波器分割模块 120 包括第三滤波器分割子模块 123，用于若所述卷积步长为 3 且所述卷积滤波器的尺寸为 3×3 ，从所述卷积滤波器分割出 9 个子滤波器，各所述子滤波器的尺寸为 1×1 且分别包括所述卷积滤波器 9 个权值中的一个。

特征图分割模块 130 包括第三特征图分割子模块 1303，用于若所述卷积步长为 3 且所述卷积滤波器的尺寸为 3×3 ，将所述待卷积特征图第 $3n+1$ 行第 $3n+1$ 列的数值分配至第一个特征子图的相应位置，将所述待卷积特征图第 $3n+1$ 行第 $3n+2$ 列的数值分配至第二个特征子图的相应位置，将所述待卷积特征图第 $3n+1$ 行第 $3n+3$ 列的数值分配至第三个特征子图的相应位置，将所述待卷积特征图第 $3n+2$ 行第 $3n+1$ 列的数值分配至第四个特征子图的相应位置，将所述待卷积特征图第 $3n+2$ 行第 $3n+2$ 列的数值分配至第五个特征子图的相应位置，将所述待卷积特征图第 $3n+2$ 行第 $3n+3$ 列的数值分配至第六个特征子图的相应位置，将所述待卷积特征图第 $3n+3$ 行第 $3n+1$ 列的数值分配至第七个特征子图的相应位置，将所述待卷积特征图第 $3n+3$ 行第 $3n+2$ 列的数值分配至第八个特征子图的相应位置，将所述待卷积特征图第 $3n+3$ 行第 $3n+3$ 列的数值分配至第九个特征子图的相应位置，其中 n 为自然数。

需要说明的是，所属领域的技术人员可以清楚地了解到，为了描述的方便和简洁，上述描述的装置和各模块、单元的具体工作过程，可以参考前述方法实施例中的对应过程，在此不再赘述。

本申请的方法、装置可用于众多通用或专用的计算系统环境或配置中。例如：个人计算机、服务器计算机、手持设备或便携式设备、平板型设备、多处理器系统、基于微处理器的系统、机顶盒、可编程的消费电子设备、网络 PC、小型计算机、大型计算机、包括以上任何系统或设备的分布式计算环境等等。

示例性的，上述的方法、装置可以实现为一种计算机程序的形式，该计算机程序可以在如图 21 所示的计算机设备上运行。

请参阅图 21，图 21 是本申请实施例提供的一种计算机设备的结构示意图。该计算机设备可以是服务器或终端。

参阅图 21，该计算机设备包括通过系统总线连接的处理器、存储器和网络接口，其中，存储器可以包括非易失性存储介质和内存。

非易失性存储介质可存储操作系统和计算机程序。该计算机程序包括程序指令，该程序

说明书

指令被执行时,可使得处理器执行任意一种基于脉动阵列的神经网络加速方法。

处理器用于提供计算和控制能力,支撑整个计算机设备的运行。

内存储器为非易失性存储介质中的计算机程序的运行提供环境,该计算机程序被处理器执行时,可使得处理器执行任意一种基于脉动阵列的神经网络加速方法。

该网络接口用于进行网络通信,如发送分配的任务等。本领域技术人员可以理解,该计算机设备的结构,仅仅是与本申请方案相关的部分结构的框图,并不构成对本申请方案所应用于其上的计算机设备的限定,具体的计算机设备可以包括比图中所示更多或更少的部件,或者组合某些部件,或者具有不同的部件布置。

应当理解的是,处理器可以是中央处理单元(Central Processing Unit, CPU),该处理器还可以是其他通用处理器、数字信号处理器(Digital Signal Processor, DSP)、专用集成电路(Application Specific Integrated Circuit, ASIC)、现场可编程门阵列(Field-Programmable Gate Array, FPGA)或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。其中,通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。

其中,在一个实施例中,所述处理器用于运行存储在存储器中的计算机程序,以实现前述基于脉动阵列的神经网络加速方法的步骤。

示例性的,所述处理器用于运行存储在存储器中的计算机程序,以实现如下步骤:

获取卷积滤波器的卷积参数,所述卷积参数包括卷积步长和所述卷积滤波器的尺寸;

若所述卷积步长不为1且所述卷积滤波器的尺寸大于 1×1 ,根据预设的滤波器分割规则从所述卷积滤波器分割出若干子滤波器,各所述子滤波器的尺寸小于所述卷积滤波器的尺寸;

获取待卷积特征图并根据预设的特征图分割规则从所述待卷积特征图分割出若干特征子图,所述若干特征子图与所述若干子滤波器一一对应;

基于脉动阵列,根据各所述子滤波器对各自对应的特征子图进行卷积计算,卷积计算的步长为1;

将各所述子滤波器对应的卷积计算结果进行叠加,并将叠加的结果作为所述卷积滤波器对所述待卷积特征图卷积计算的结果进行输出。

通过以上的实施方式的描述可知,本领域的技术人员可以清楚地了解到本申请可借助软件加必需的通用硬件平台的方式来实现。基于这样的理解,本申请的技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来,该计算机软件产品可以存储在存储介质中,如ROM/RAM、磁碟、光盘等,包括若干指令用以使得一台计算机设备(可以是个人计算机,服务器,或者网络设备等)执行本申请各个实施例或者实施例的某些部分所述的方法,如:

一种计算机可读存储介质,所述计算机可读存储介质存储有计算机程序,所述计算机程序中包括程序指令,所述处理器执行所述程序指令,实现本申请实施例提供的任一项基于脉动阵列的神经网络加速方法。

其中,所述计算机可读存储介质可以是前述实施例所述的计算机设备的内部存储单元,例如所述计算机设备的硬盘或内存。所述计算机可读存储介质也可以是所述计算机设备的外部存储设备,例如所述计算机设备上配备的插接式硬盘,智能存储卡(Smart Media Card, SMC),安全数字(Secure Digital, SD)卡,闪存卡(Flash Card)等。

以上所述,仅为本申请的具体实施方式,但本申请的保护范围并不局限于此,任何熟悉本技术领域的技术人员在本申请揭露的技术范围内,可轻易想到各种等效的修改或替换,这些修改或替换都应涵盖在本申请的保护范围之内。因此,本申请的保护范围应以权利要求的保护范围为准。

权 利 要 求 书

1. 一种基于脉动阵列的神经网络加速方法, 包括:

获取卷积滤波器的卷积参数, 所述卷积参数包括卷积步长和所述卷积滤波器的尺寸;

若所述卷积步长不为 1 且所述卷积滤波器的尺寸大于 1×1 , 根据预设的滤波器分割规则从所述卷积滤波器分割出若干子滤波器, 各所述子滤波器的尺寸小于所述卷积滤波器的尺寸;

获取待卷积特征图并根据预设的特征图分割规则从所述待卷积特征图分割出若干特征子图, 所述若干特征子图与所述若干子滤波器一一对应;

基于脉动阵列, 根据各所述子滤波器对各自对应的特征子图进行卷积计算, 卷积计算的步长为 1;

将各所述子滤波器对应的卷积计算结果进行叠加, 并将叠加的结果作为所述卷积滤波器对所述待卷积特征图卷积计算的结果进行输出。

2. 如权利要求 1 所述的神经网络加速方法, 其中, 所述获取待卷积特征图并根据预设的特征图分割规则从所述待卷积特征图分割出若干特征子图, 具体包括:

获取待卷积特征图;

若获取的待卷积特征图的长或宽不是所述卷积步长的整数倍, 对所述待卷积特征图的预设位置进行补零以使补零后的待卷积特征图的长或宽是所述卷积步长的整数倍;

根据预设的特征图分割规则从补零后的待卷积特征图分割出若干特征子图。

3. 如权利要求 2 所述的神经网络加速方法, 其中, 所述若所述卷积步长不为 1 且所述卷积滤波器的尺寸大于 1, 根据预设的滤波器分割规则从所述卷积滤波器分割出若干子滤波器, 具体包括:

若所述卷积步长为 2 且所述卷积滤波器的尺寸为 2×2 , 从所述卷积滤波器分割出 4 个子滤波器, 各所述子滤波器的尺寸为 1×1 ; 其中第一个子滤波器包括所述卷积滤波器奇数行奇数列的权值, 第二个子滤波器包括所述卷积滤波器奇数行偶数列的权值, 第三个子滤波器包括所述卷积滤波器偶数行奇数列的权值, 第四个子滤波器包括所述卷积滤波器偶数行偶数列的权值;

若所述卷积步长为 2 且所述卷积滤波器的尺寸为 2×2 , 所述获取待卷积特征图并根据预设的特征图分割规则从所述待卷积特征图分割出若干特征子图, 具体包括:

将所述待卷积特征图奇数行奇数列的数值分配至第一个特征子图的相应位置, 将所述待卷积特征图奇数行偶数列的数值分配至第二个特征子图的相应位置, 将所述待卷积特征图偶数行奇数列的数值分配至第三个特征子图的相应位置, 将所述待卷积特征图偶数行偶数列的数值分配至第四个特征子图的相应位置。

4. 如权利要求 2 所述的神经网络加速方法, 其中, 所述若所述卷积步长不为 1 且所述卷积滤波器的尺寸大于 1, 根据预设的滤波器分割规则从所述卷积滤波器分割出若干子滤波器, 具体包括:

若所述卷积步长为 2 且所述卷积滤波器的尺寸为 3×3 , 从所述卷积滤波器分割出 4 个子滤波器, 各所述子滤波器的尺寸为 2×2 ; 其中第一个子滤波器包括所述卷积滤波器奇数行奇数列的权值, 第二个子滤波器包括所述卷积滤波器奇数行偶数列的权值, 第三个子滤波器包括所述卷积滤波器偶数行奇数列的权值, 第四个子滤波器包括所述卷积滤波器偶数行偶数列的权值;

若所述卷积步长为 2 且所述卷积滤波器的尺寸为 3×3 , 所述获取待卷积特征图并根据预设的特征图分割规则从所述待卷积特征图分割出若干特征子图, 具体包括:

将所述待卷积特征图奇数行奇数列的数值分配至第一个特征子图的相应位置, 将所述待卷积特征图奇数行偶数列的数值分配至第二个特征子图的相应位置, 将所述待卷积特征图偶数行奇数列的数值分配至第三个特征子图的相应位置, 将所述待卷积特征图偶数行偶数列的数值分配至第四个特征子图的相应位置。

5. 如权利要求 4 所述的神经网络加速方法, 其中, 所述若所述卷积步长为 2 且所述卷积

权 利 要 求 书

滤波器的尺寸为 3×3 ，从所述卷积滤波器分割出 4 个子滤波器，各所述子滤波器的尺寸为 2×2 ，具体包括：

将所述卷积滤波器奇数行奇数列的权值分配至第一个子滤波器；

将所述卷积滤波器奇数行偶数列的权值分配至第二个子滤波器的第一列，并以 0 填充所述第二个子滤波器的第二列；

将所述卷积滤波器偶数行奇数列的权值分配至第三个子滤波器的第一行，并以 0 填充所述第三个子滤波器的第二行；

将所述卷积滤波器偶数行偶数列的权值分配至第四个子滤波器的第一行第一列，并以 0 填充所述第四个子滤波器的其余位置。

6. 如权利要求 2 所述的神经网络加速方法，其中，所述若所述卷积步长不为 1 且所述卷积滤波器的尺寸大于 1，根据预设的滤波器分割规则从所述卷积滤波器分割出若干子滤波器，具体包括：

若所述卷积步长为 3 且所述卷积滤波器的尺寸为 3×3 ，从所述卷积滤波器分割出 9 个子滤波器，各所述子滤波器的尺寸为 1×1 且分别包括所述卷积滤波器 9 个权值中的一个；

若所述卷积步长为 3 且所述卷积滤波器的尺寸为 3×3 ，所述获取待卷积特征图并根据预设的特征图分割规则从所述待卷积特征图分割出若干特征子图，具体包括：

将所述待卷积特征图第 $3n+1$ 行第 $3n+1$ 列的数值分配至第一个特征子图的相应位置，将所述待卷积特征图第 $3n+1$ 行第 $3n+2$ 列的数值分配至第二个特征子图的相应位置，将所述待卷积特征图第 $3n+1$ 行第 $3n+3$ 列的数值分配至第三个特征子图的相应位置，将所述待卷积特征图第 $3n+2$ 行第 $3n+1$ 列的数值分配至第四个特征子图的相应位置，将所述待卷积特征图第 $3n+2$ 行第 $3n+2$ 列的数值分配至第五个特征子图的相应位置，将所述待卷积特征图第 $3n+2$ 行第 $3n+3$ 列的数值分配至第六个特征子图的相应位置，将所述待卷积特征图第 $3n+3$ 行第 $3n+1$ 列的数值分配至第七个特征子图的相应位置，将所述待卷积特征图第 $3n+3$ 行第 $3n+2$ 列的数值分配至第八个特征子图的相应位置，将所述待卷积特征图第 $3n+3$ 行第 $3n+3$ 列的数值分配至第九个特征子图的相应位置，其中 n 为自然数。

7. 如权利要求 1-6 中任一项所述的神经网络加速方法，其中：所述基于脉动阵列，根据各所述子滤波器对各自对应的特征子图进行卷积计算，具体包括：

将所述子滤波器的权值加载至连接于所述脉动阵列的权值寄存器中；

将所述子滤波器对应的特征子图加载至连接于所述脉动阵列的输入寄存器中；

获取所述脉动阵列卷积计算后的输出结果。

8. 一种基于脉动阵列的神经网络加速装置，包括：

卷积参数获取模块，用于获取卷积滤波器的卷积参数，所述卷积参数包括卷积步长和所述卷积滤波器的尺寸；

滤波器分割模块，用于若所述卷积步长不为 1 且所述卷积滤波器的尺寸大于 1×1 ，根据预设的滤波器分割规则从所述卷积滤波器分割出若干子滤波器，各所述子滤波器的尺寸小于所述卷积滤波器的尺寸；

特征图分割模块，用于获取待卷积特征图并根据预设的特征图分割规则从所述待卷积特征图分割出若干特征子图，所述若干特征子图与所述若干子滤波器一一对应；

卷积模块，用于基于脉动阵列，根据各所述子滤波器对各自对应的特征子图进行卷积计算，卷积计算的步长为 1；

叠加模块，用于将各所述子滤波器对应的卷积计算结果进行叠加，并将叠加的结果作为所述卷积滤波器对所述待卷积特征图卷积计算的结果进行输出。

9. 一种计算机设备，所述计算机设备包括存储器和处理器；

所述存储器用于存储计算机程序；

所述处理器，用于执行所述计算机程序并在执行所述计算机程序时实现如下步骤：

权 利 要 求 书

获取卷积滤波器的卷积参数，所述卷积参数包括卷积步长和所述卷积滤波器的尺寸；

若所述卷积步长不为 1 且所述卷积滤波器的尺寸大于 1×1 ，根据预设的滤波器分割规则从所述卷积滤波器分割出若干子滤波器，各所述子滤波器的尺寸小于所述卷积滤波器的尺寸；

获取待卷积特征图并根据预设的特征图分割规则从所述待卷积特征图分割出若干特征子图，所述若干特征子图与所述若干子滤波器一一对应；

基于脉动阵列，根据各所述子滤波器对各自对应的特征子图进行卷积计算，卷积计算的步长为 1；

将各所述子滤波器对应的卷积计算结果进行叠加，并将叠加的结果作为所述卷积滤波器对所述待卷积特征图卷积计算的结果进行输出。

10. 如权利要求 9 所述的计算机设备，其中，所述处理器在实现所述获取待卷积特征图并根据预设的特征图分割规则从所述待卷积特征图分割出若干特征子图时，用于实现如下步骤：

获取待卷积特征图；

若获取的待卷积特征图的长或宽不是所述卷积步长的整数倍，对所述待卷积特征图的预设位置进行补零以使补零后的待卷积特征图的长或宽是所述卷积步长的整数倍；

根据预设的特征图分割规则从补零后的待卷积特征图分割出若干特征子图。

11. 如权利要求 10 所述的计算机设备，其中，所述处理器在实现所述若所述卷积步长不为 1 且所述卷积滤波器的尺寸大于 1，根据预设的滤波器分割规则从所述卷积滤波器分割出若干子滤波器时，用于实现如下步骤：

若所述卷积步长为 2 且所述卷积滤波器的尺寸为 2×2 ，从所述卷积滤波器分割出 4 个子滤波器，各所述子滤波器的尺寸为 1×1 ；其中第一个子滤波器包括所述卷积滤波器奇数行奇数列的权值，第二个子滤波器包括所述卷积滤波器奇数行偶数列的权值，第三个子滤波器包括所述卷积滤波器偶数行奇数列的权值，第四个子滤波器包括所述卷积滤波器偶数行偶数列的权值；

若所述卷积步长为 2 且所述卷积滤波器的尺寸为 2×2 ，所述获取待卷积特征图并根据预设的特征图分割规则从所述待卷积特征图分割出若干特征子图，具体包括：

将所述待卷积特征图奇数行奇数列的数值分配至第一个特征子图的相应位置，将所述待卷积特征图奇数行偶数列的数值分配至第二个特征子图的相应位置，将所述待卷积特征图偶数行奇数列的数值分配至第三个特征子图的相应位置，将所述待卷积特征图偶数行偶数列的数值分配至第四个特征子图的相应位置。

12. 如权利要求 10 所述的计算机设备，其中，所述处理器在实现所述若所述卷积步长不为 1 且所述卷积滤波器的尺寸大于 1，根据预设的滤波器分割规则从所述卷积滤波器分割出若干子滤波器时，用于实现如下步骤：

若所述卷积步长为 2 且所述卷积滤波器的尺寸为 3×3 ，从所述卷积滤波器分割出 4 个子滤波器，各所述子滤波器的尺寸为 2×2 ；其中第一个子滤波器包括所述卷积滤波器奇数行奇数列的权值，第二个子滤波器包括所述卷积滤波器奇数行偶数列的权值，第三个子滤波器包括所述卷积滤波器偶数行奇数列的权值，第四个子滤波器包括所述卷积滤波器偶数行偶数列的权值；

若所述卷积步长为 2 且所述卷积滤波器的尺寸为 3×3 ，所述获取待卷积特征图并根据预设的特征图分割规则从所述待卷积特征图分割出若干特征子图，具体包括：

将所述待卷积特征图奇数行奇数列的数值分配至第一个特征子图的相应位置，将所述待卷积特征图奇数行偶数列的数值分配至第二个特征子图的相应位置，将所述待卷积特征图偶数行奇数列的数值分配至第三个特征子图的相应位置，将所述待卷积特征图偶数行偶数列的数值分配至第四个特征子图的相应位置。

13. 如权利要求 10 所述的计算机设备，其中，所述处理器在实现所述若所述卷积步长不

权 利 要 求 书

为 1 且所述卷积滤波器的尺寸大于 1, 根据预设的滤波器分割规则从所述卷积滤波器分割出若干子滤波器时, 用于实现如下步骤:

若所述卷积步长为 3 且所述卷积滤波器的尺寸为 3×3 , 从所述卷积滤波器分割出 9 个子滤波器, 各所述子滤波器的尺寸为 1×1 且分别包括所述卷积滤波器 9 个权值中的一个;

若所述卷积步长为 3 且所述卷积滤波器的尺寸为 3×3 , 所述获取待卷积特征图并根据预设的特征图分割规则从所述待卷积特征图分割出若干特征子图, 具体包括:

将所述待卷积特征图第 $3n+1$ 行第 $3n+1$ 列的数值分配至第一个特征子图的相应位置, 将所述待卷积特征图第 $3n+1$ 行第 $3n+2$ 列的数值分配至第二个特征子图的相应位置, 将所述待卷积特征图第 $3n+1$ 行第 $3n+3$ 列的数值分配至第三个特征子图的相应位置, 将所述待卷积特征图第 $3n+2$ 行第 $3n+1$ 列的数值分配至第四个特征子图的相应位置, 将所述待卷积特征图第 $3n+2$ 行第 $3n+2$ 列的数值分配至第五个特征子图的相应位置, 将所述待卷积特征图第 $3n+2$ 行第 $3n+3$ 列的数值分配至第六个特征子图的相应位置, 将所述待卷积特征图第 $3n+3$ 行第 $3n+1$ 列的数值分配至第七个特征子图的相应位置, 将所述待卷积特征图第 $3n+3$ 行第 $3n+2$ 列的数值分配至第八个特征子图的相应位置, 将所述待卷积特征图第 $3n+3$ 行第 $3n+3$ 列的数值分配至第九个特征子图的相应位置, 其中 n 为自然数。

14. 如权利要求 9-13 中任一项所述的计算机设备, 其中: 所述处理器在实现所述基于脉动阵列, 根据各所述子滤波器对各自对应的特征子图进行卷积计算时, 用于实现如下步骤:

将所述子滤波器的权值加载至连接于所述脉动阵列的权值寄存器中;

将所述子滤波器对应的特征子图加载至连接于所述脉动阵列的输入寄存器中;

获取所述脉动阵列卷积计算后的输出结果。

15. 一种计算机可读存储介质, 所述计算机可读存储介质存储有计算机程序, 若所述计算机程序被处理器执行, 实现如下步骤:

获取卷积滤波器的卷积参数, 所述卷积参数包括卷积步长和所述卷积滤波器的尺寸;

若所述卷积步长不为 1 且所述卷积滤波器的尺寸大于 1×1 , 根据预设的滤波器分割规则从所述卷积滤波器分割出若干子滤波器, 各所述子滤波器的尺寸小于所述卷积滤波器的尺寸;

获取待卷积特征图并根据预设的特征图分割规则从所述待卷积特征图分割出若干特征子图, 所述若干特征子图与所述若干子滤波器一一对应;

基于脉动阵列, 根据各所述子滤波器对各自对应的特征子图进行卷积计算, 卷积计算的步长为 1;

将各所述子滤波器对应的卷积计算结果进行叠加, 并将叠加的结果作为所述卷积滤波器对所述待卷积特征图卷积计算的结果进行输出。

16. 如权利要求 15 所述的存储介质, 其中, 所述处理器在实现所述获取待卷积特征图并根据预设的特征图分割规则从所述待卷积特征图分割出若干特征子图时, 用于实现如下步骤:

获取待卷积特征图;

若获取的待卷积特征图的长或宽不是所述卷积步长的整数倍, 对所述待卷积特征图的预设位置进行补零以使补零后的待卷积特征图的长或宽是所述卷积步长的整数倍;

根据预设的特征图分割规则从补零后的待卷积特征图分割出若干特征子图。

17. 如权利要求 16 所述的存储介质, 其中, 所述处理器在实现所述若所述卷积步长不为 1 且所述卷积滤波器的尺寸大于 1, 根据预设的滤波器分割规则从所述卷积滤波器分割出若干子滤波器时, 用于实现如下步骤:

若所述卷积步长为 2 且所述卷积滤波器的尺寸为 2×2 , 从所述卷积滤波器分割出 4 个子滤波器, 各所述子滤波器的尺寸为 1×1 ; 其中第一个子滤波器包括所述卷积滤波器奇数行奇数列的权值, 第二个子滤波器包括所述卷积滤波器奇数行偶数列的权值, 第三个子滤波器包括所述卷积滤波器偶数行奇数列的权值, 第四个子滤波器包括所述卷积滤波器偶数行偶数列的权值;

权 利 要 求 书

若所述卷积步长为 2 且所述卷积滤波器的尺寸为 2×2 ，所述获取待卷积特征图并根据预设的特征图分割规则从所述待卷积特征图分割出若干特征子图，具体包括：

将所述待卷积特征图奇数行奇数列的数值分配至第一个特征子图的相应位置，将所述待卷积特征图奇数行偶数列的数值分配至第二个特征子图的相应位置，将所述待卷积特征图偶数行奇数列的数值分配至第三个特征子图的相应位置，将所述待卷积特征图偶数行偶数列的数值分配至第四个特征子图的相应位置。

18. 如权利要求 16 所述的存储介质，其中，所述处理器在实现所述若所述卷积步长不为 1 且所述卷积滤波器的尺寸大于 1，根据预设的滤波器分割规则从所述卷积滤波器分割出若干子滤波器时，用于实现如下步骤：

若所述卷积步长为 2 且所述卷积滤波器的尺寸为 3×3 ，从所述卷积滤波器分割出 4 个子滤波器，各所述子滤波器的尺寸为 2×2 ；其中第一个子滤波器包括所述卷积滤波器奇数行奇数列的权值，第二个子滤波器包括所述卷积滤波器奇数行偶数列的权值，第三个子滤波器包括所述卷积滤波器偶数行奇数列的权值，第四个子滤波器包括所述卷积滤波器偶数行偶数列的权值；

若所述卷积步长为 2 且所述卷积滤波器的尺寸为 3×3 ，所述获取待卷积特征图并根据预设的特征图分割规则从所述待卷积特征图分割出若干特征子图，具体包括：

将所述待卷积特征图奇数行奇数列的数值分配至第一个特征子图的相应位置，将所述待卷积特征图奇数行偶数列的数值分配至第二个特征子图的相应位置，将所述待卷积特征图偶数行奇数列的数值分配至第三个特征子图的相应位置，将所述待卷积特征图偶数行偶数列的数值分配至第四个特征子图的相应位置。

19. 如权利要求 16 所述的存储介质，其中，所述处理器在实现所述若所述卷积步长不为 1 且所述卷积滤波器的尺寸大于 1，根据预设的滤波器分割规则从所述卷积滤波器分割出若干子滤波器时，用于实现如下步骤：

若所述卷积步长为 3 且所述卷积滤波器的尺寸为 3×3 ，从所述卷积滤波器分割出 9 个子滤波器，各所述子滤波器的尺寸为 1×1 且分别包括所述卷积滤波器 9 个权值中的一个；

若所述卷积步长为 3 且所述卷积滤波器的尺寸为 3×3 ，所述获取待卷积特征图并根据预设的特征图分割规则从所述待卷积特征图分割出若干特征子图，具体包括：

将所述待卷积特征图第 $3n+1$ 行第 $3n+1$ 列的数值分配至第一个特征子图的相应位置，将所述待卷积特征图第 $3n+1$ 行第 $3n+2$ 列的数值分配至第二个特征子图的相应位置，将所述待卷积特征图第 $3n+1$ 行第 $3n+3$ 列的数值分配至第三个特征子图的相应位置，将所述待卷积特征图第 $3n+2$ 行第 $3n+1$ 列的数值分配至第四个特征子图的相应位置，将所述待卷积特征图第 $3n+2$ 行第 $3n+2$ 列的数值分配至第五个特征子图的相应位置，将所述待卷积特征图第 $3n+2$ 行第 $3n+3$ 列的数值分配至第六个特征子图的相应位置，将所述待卷积特征图第 $3n+3$ 行第 $3n+1$ 列的数值分配至第七个特征子图的相应位置，将所述待卷积特征图第 $3n+3$ 行第 $3n+2$ 列的数值分配至第八个特征子图的相应位置，将所述待卷积特征图第 $3n+3$ 行第 $3n+3$ 列的数值分配至第九个特征子图的相应位置，其中 n 为自然数。

20. 如权利要求 15-19 中任一项所述的存储介质，其中：所述处理器在实现所述基于脉动阵列，根据各所述子滤波器对各自对应的特征子图进行卷积计算时，用于实现如下步骤：

将所述子滤波器的权值加载至连接于所述脉动阵列的权值寄存器中；

将所述子滤波器对应的特征子图加载至连接于所述脉动阵列的输入寄存器中；

获取所述脉动阵列卷积计算后的输出结果。

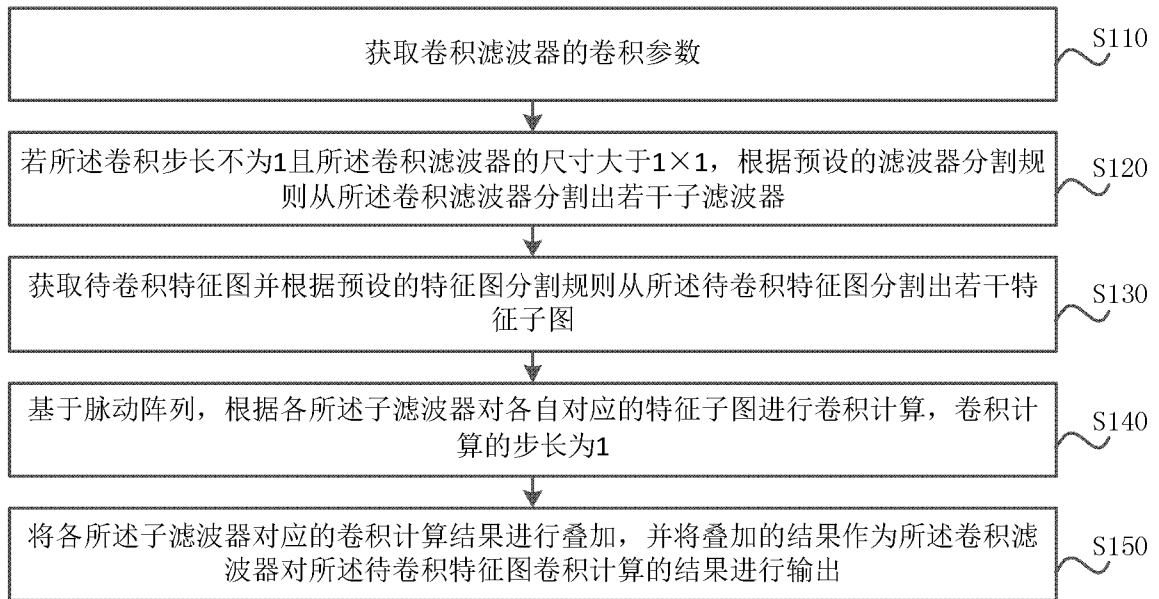


图 1

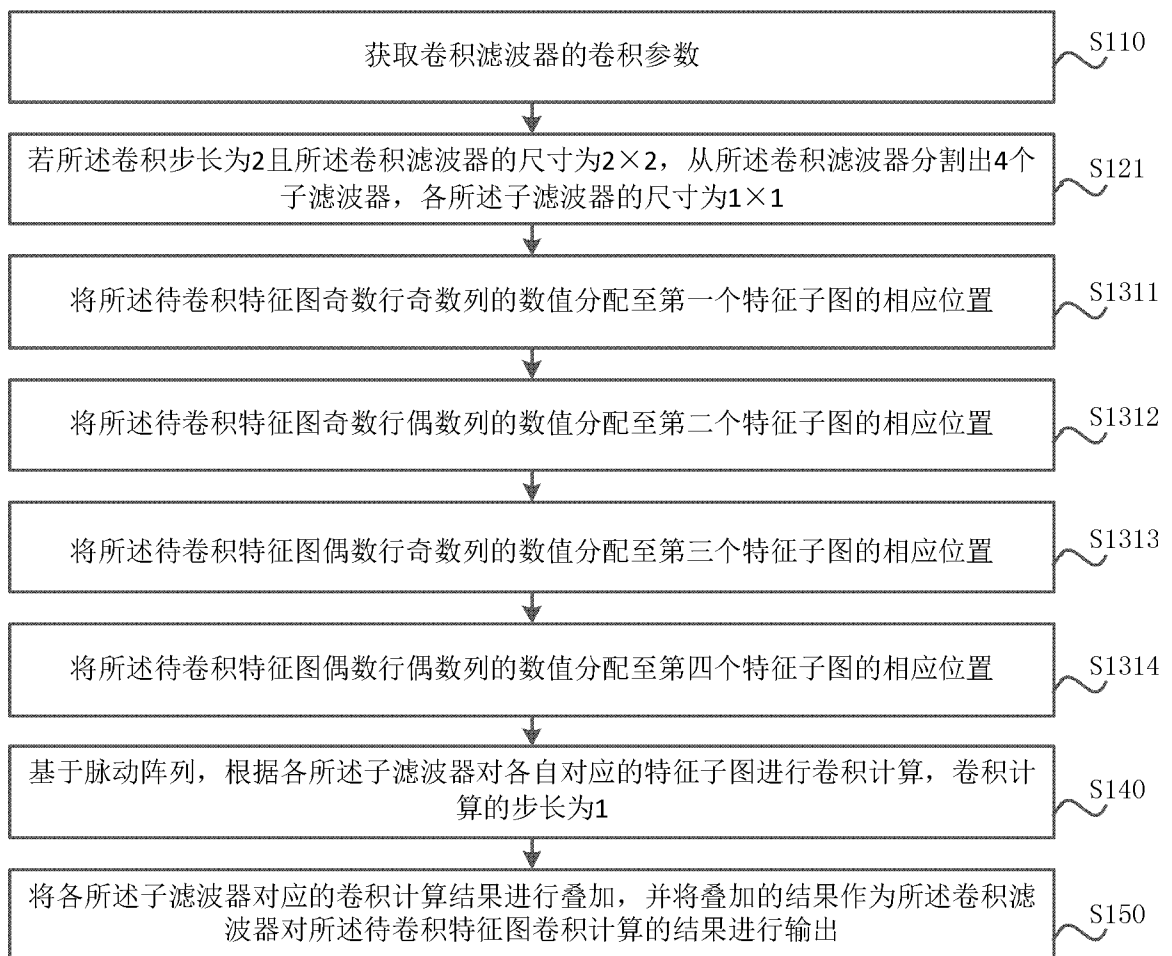
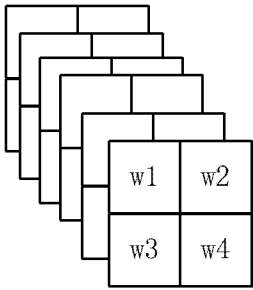


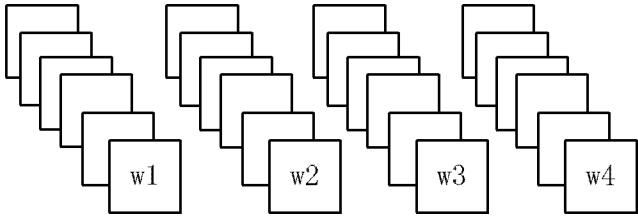
图 2

卷积滤波器[6 6 2 2]



卷积步长 = 2

卷积滤波器[6 24 1 1]



卷积步长 = 1

图 3

输入张量[1 24 2 2]

输入张量[1 6 4 4]

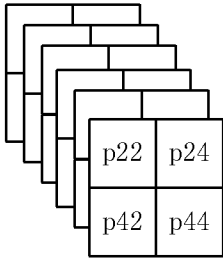
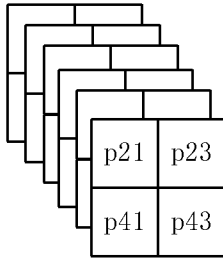
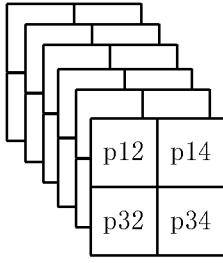
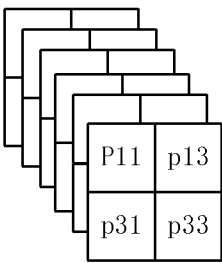
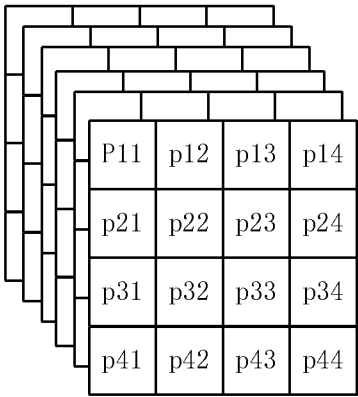


图 4

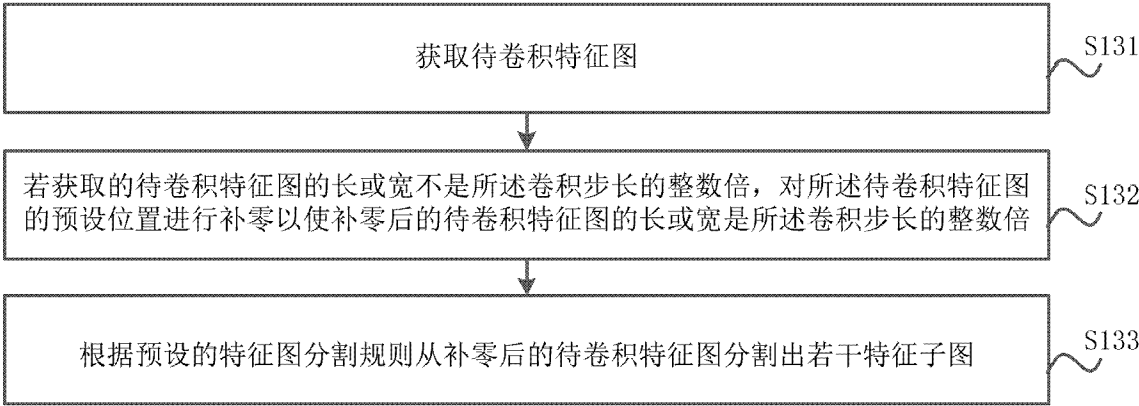


图 5

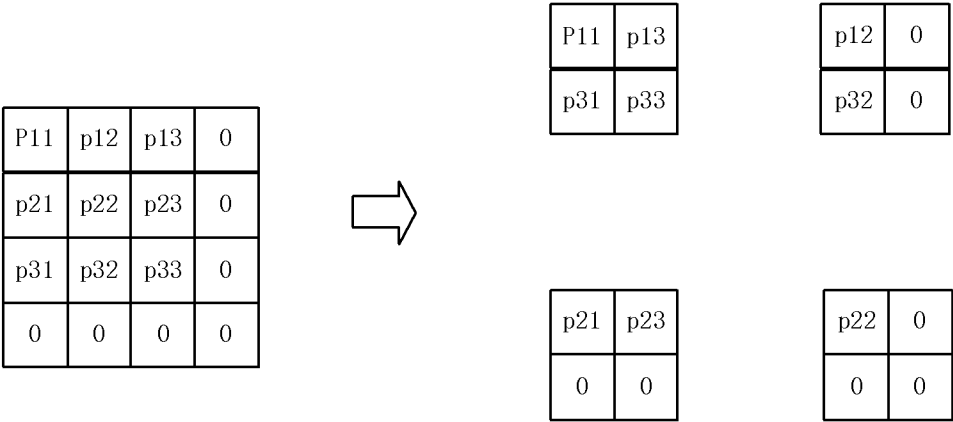


图 6

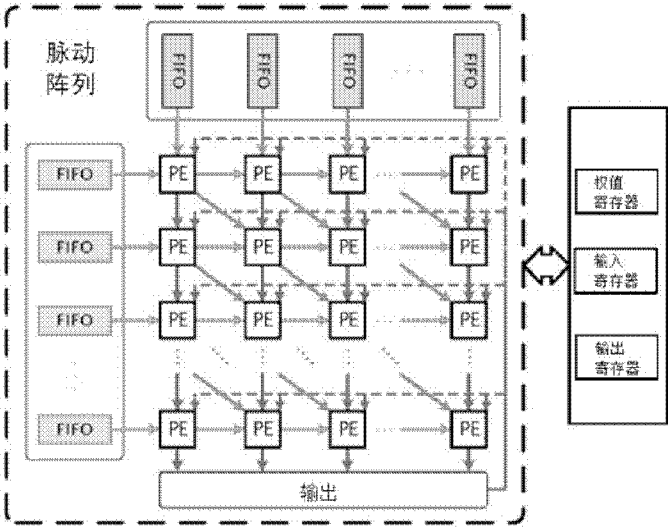


图 7

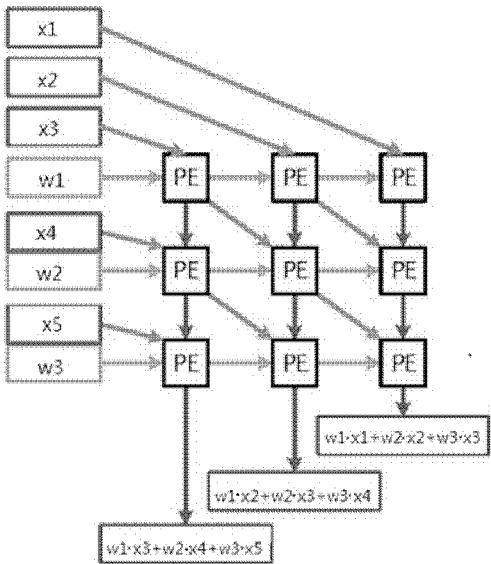
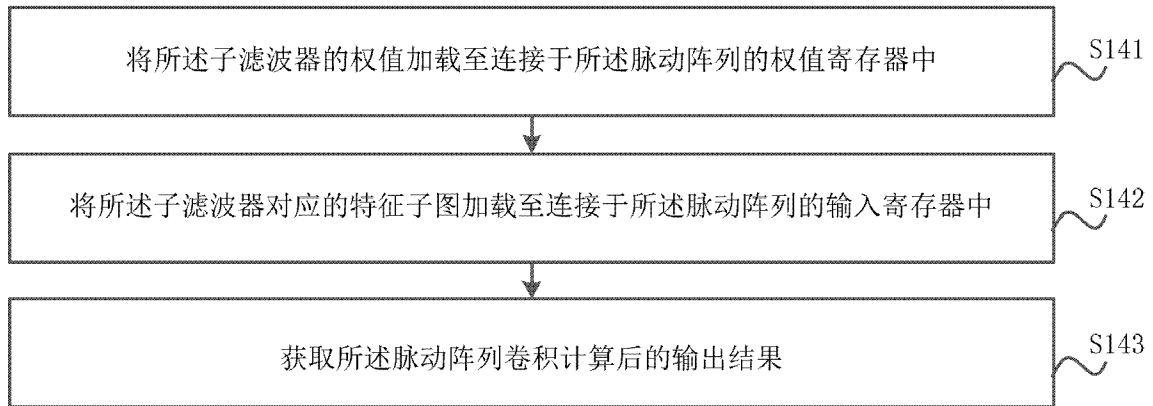


图 8



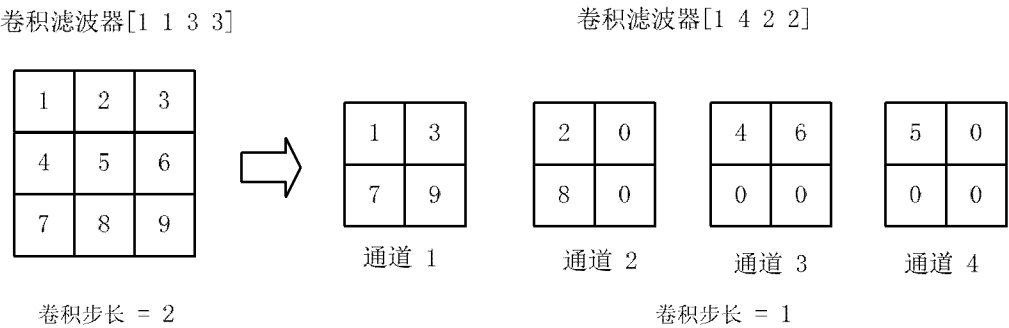


图 11

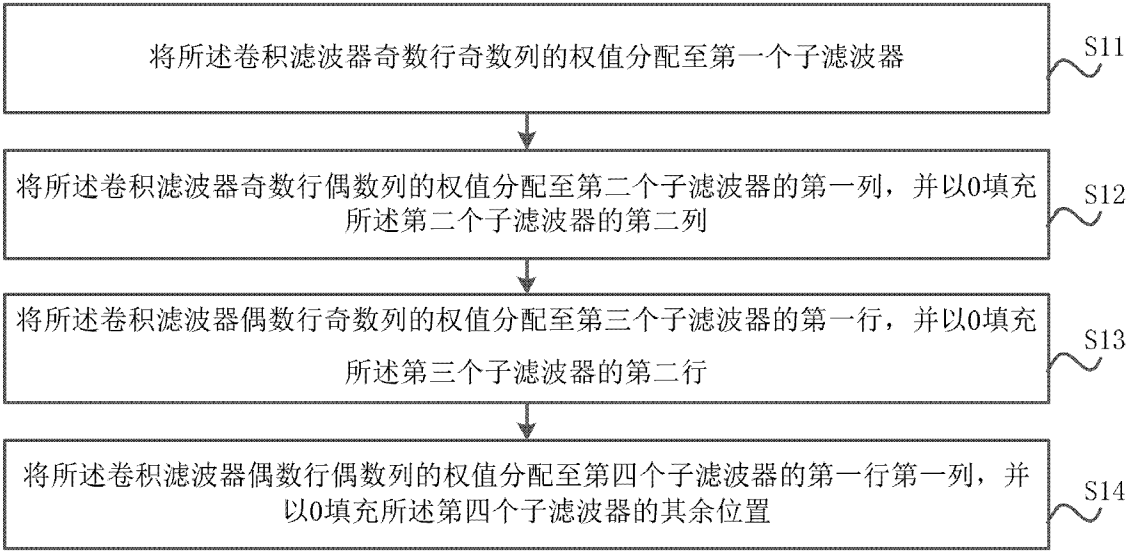


图 12

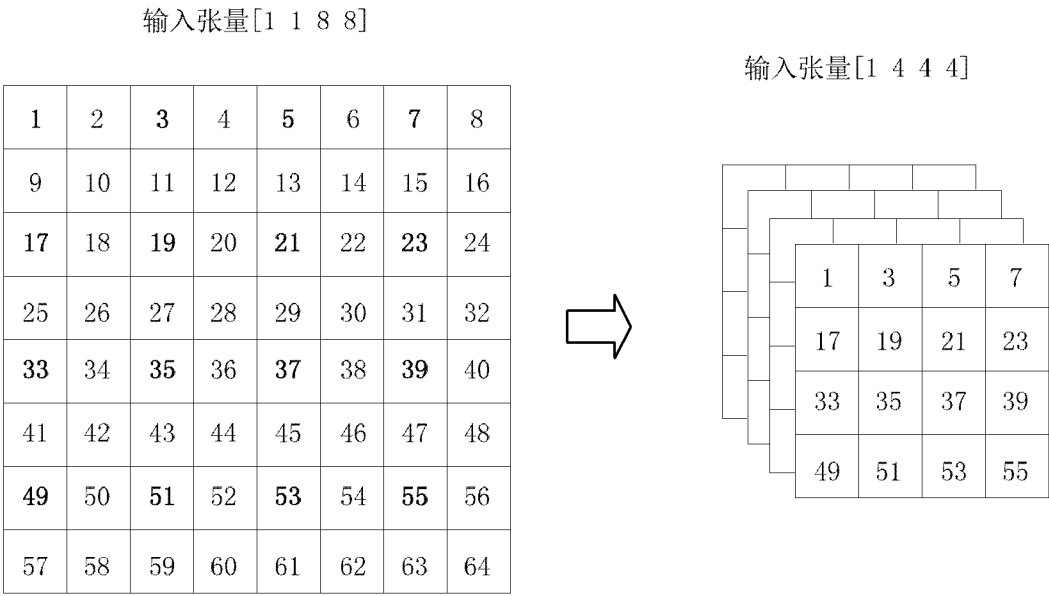


图 13

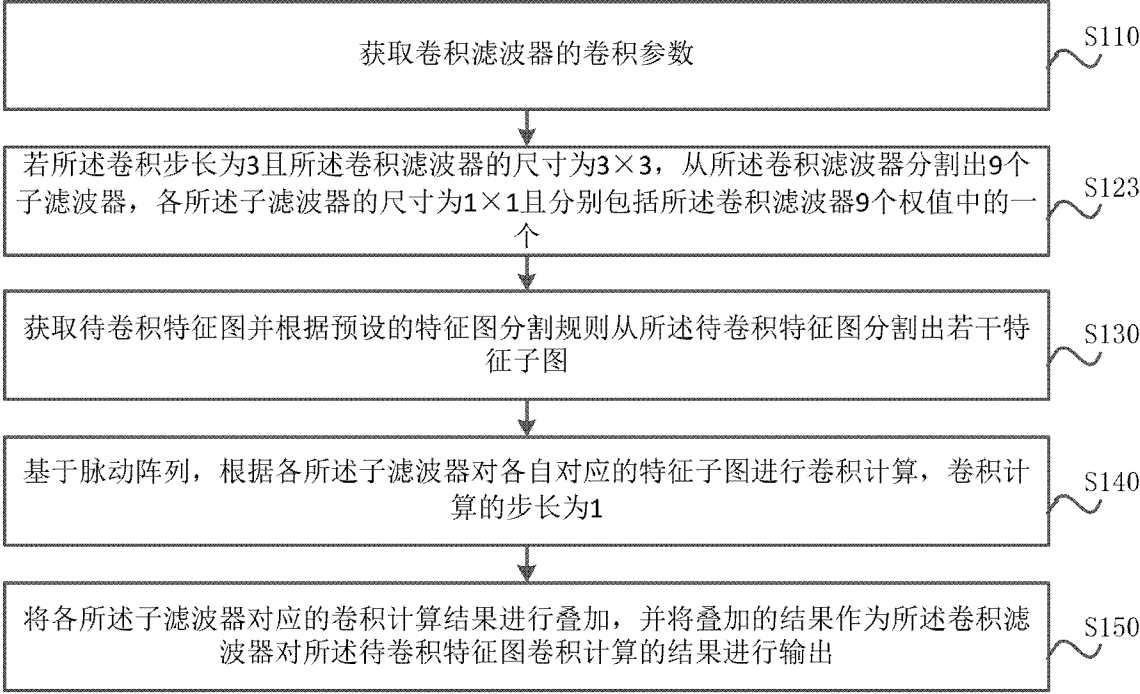


图 14

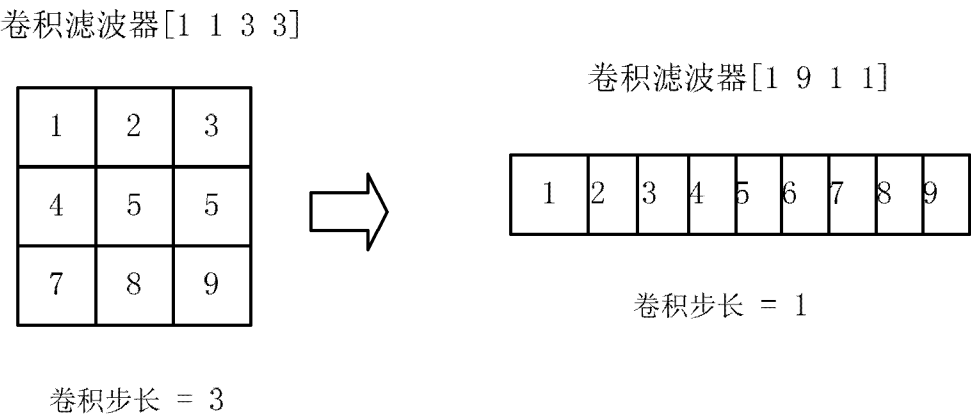


图 15

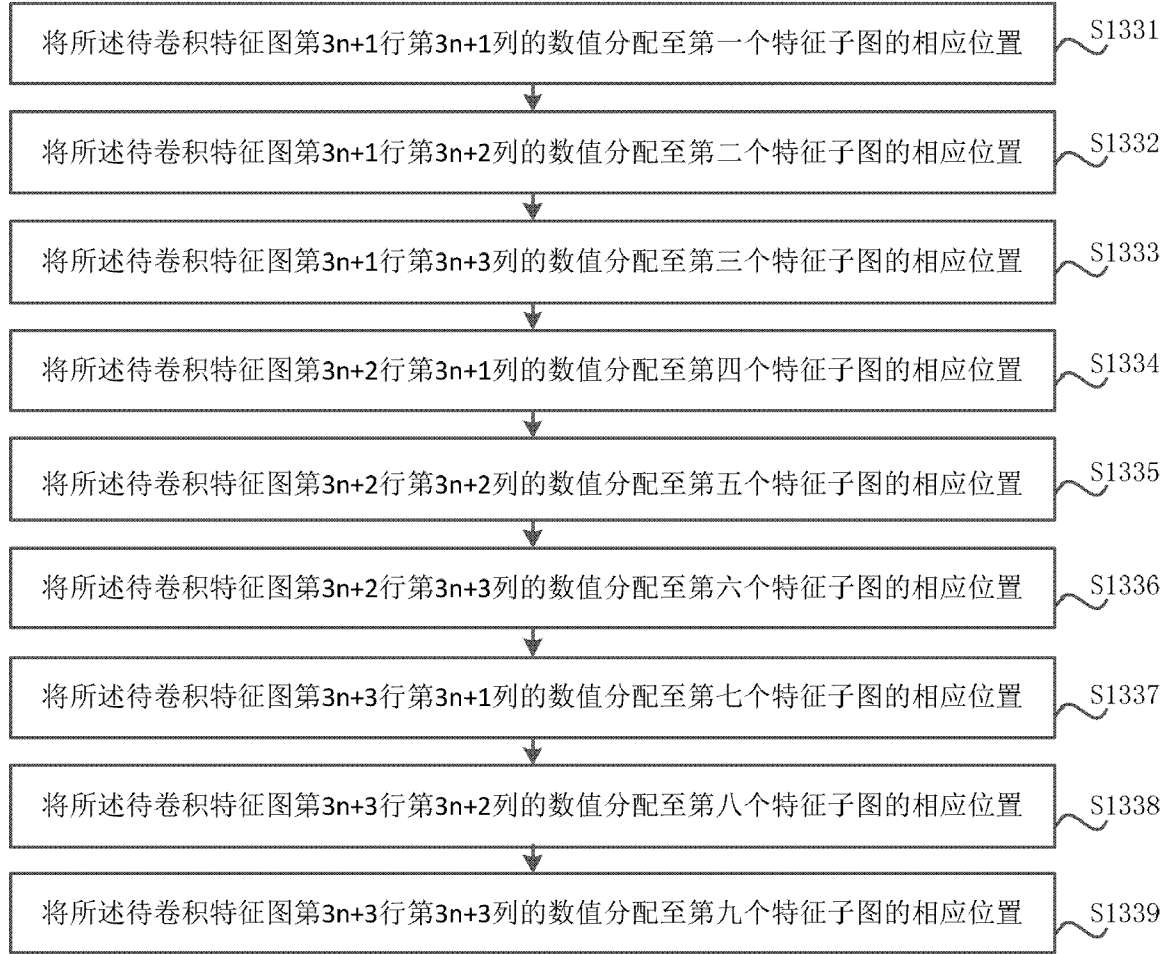


图 16

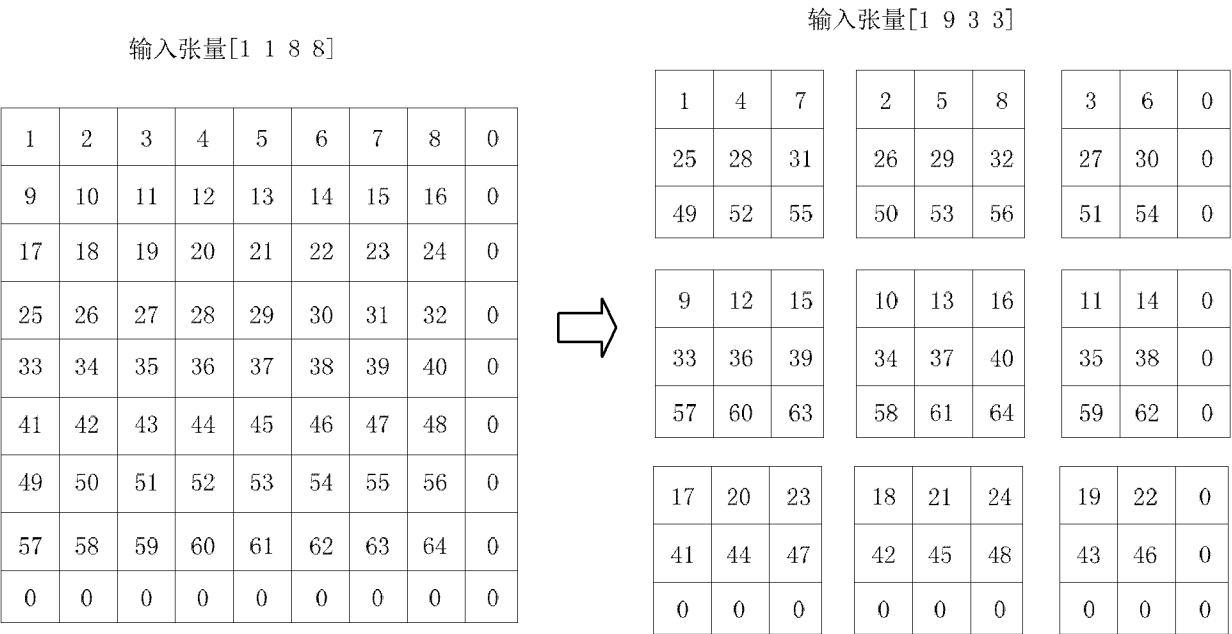


图 17

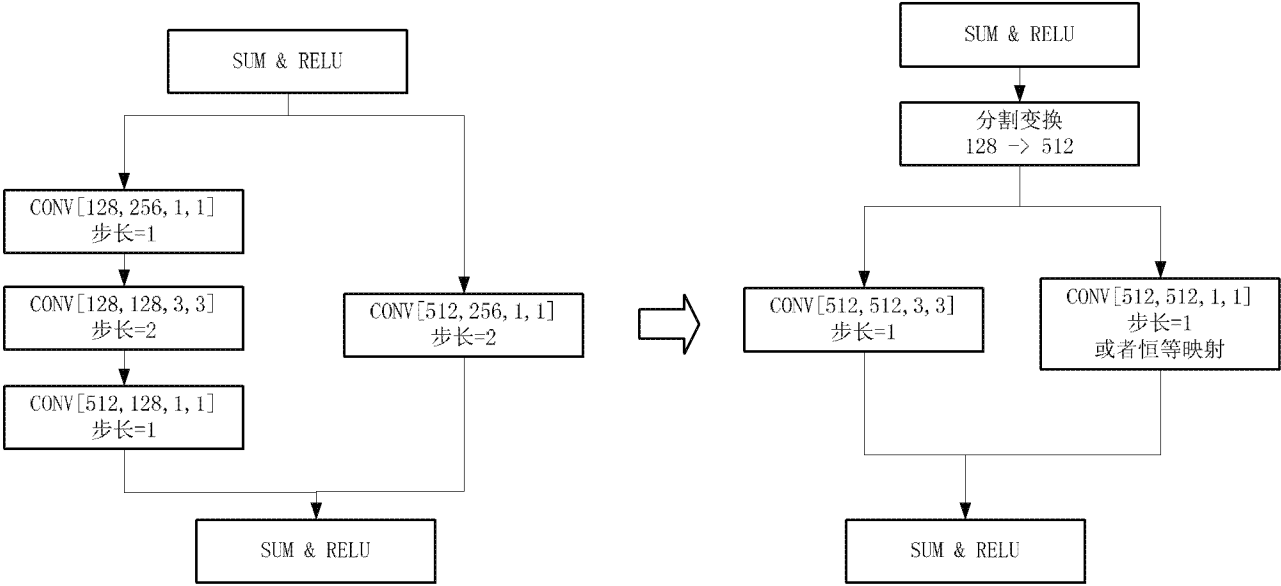


图 18

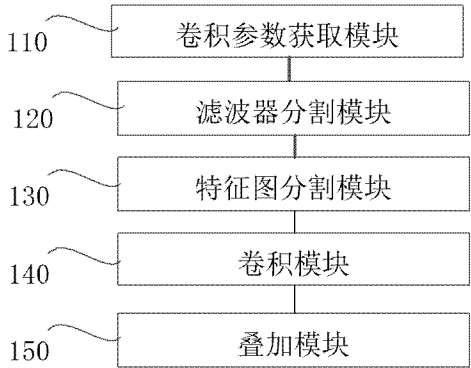


图 19

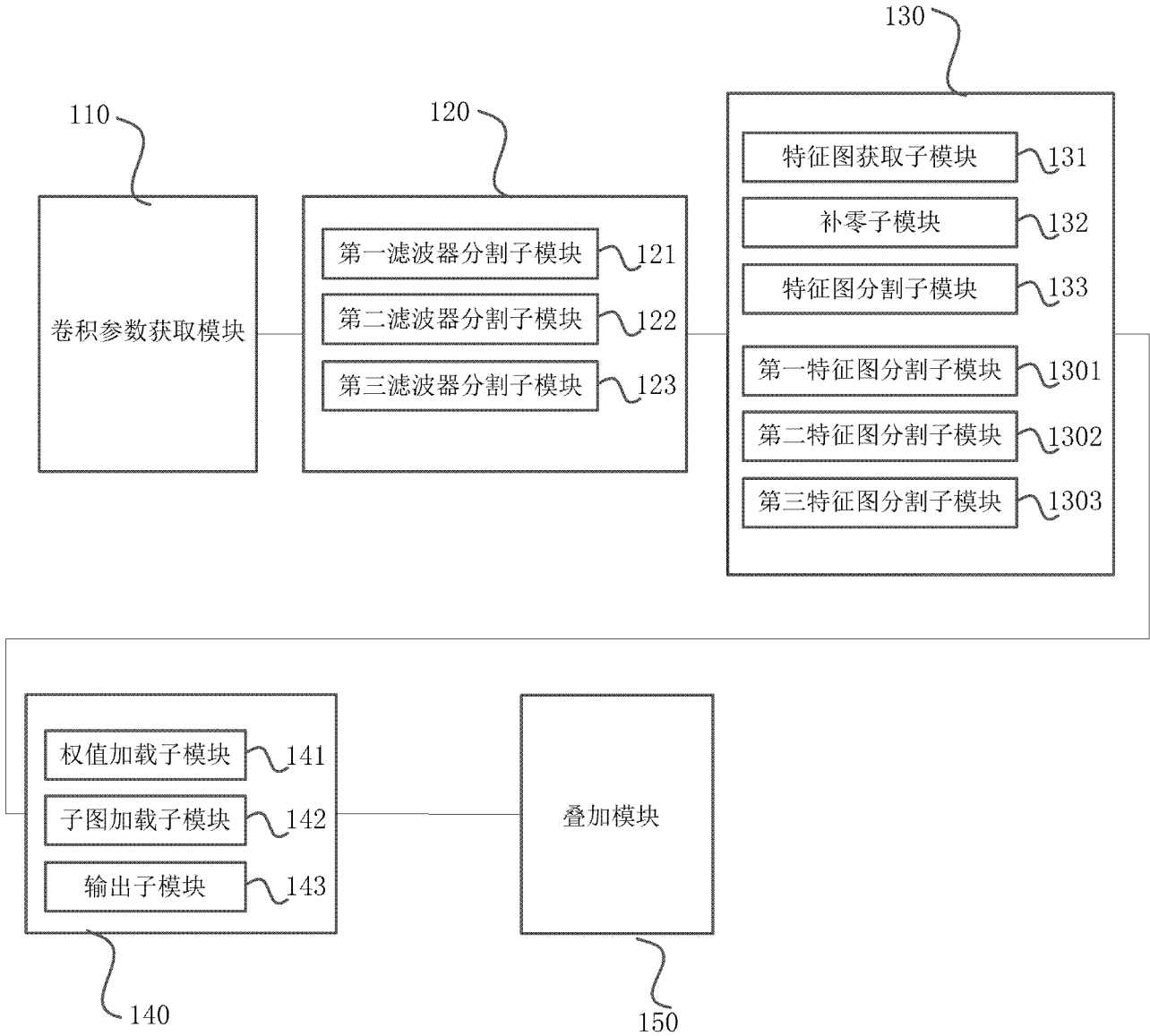


图 20

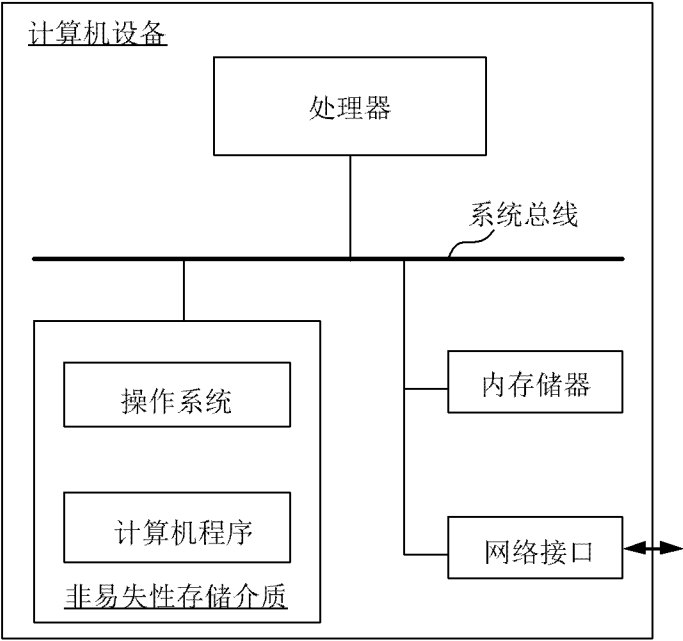


图 21

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/103137

A. CLASSIFICATION OF SUBJECT MATTER

G06N 3/04(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06N; G06K; G06T

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT, EPODOC, WPI, CNKI, IEEE: 神经网络, 卷积, 点乘, 加速, 分割, 脉动阵列, 划分, 分解, 细分, 子, 内核, 核函数, 卷积核, 尺寸, 高度, 宽度, 步长, 步进, 滑动; CNN, convol+, accelerat+, divide, sub, kerner, size, height, width, stride, step, slide

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 110135556 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 16 August 2019 (2019-08-16) claims 1-10, and description, paragraphs [0004]-[0179]	1-20
X	CN 108875908 A (SAMSUNG ELECTRONICS CO., LTD.) 23 November 2018 (2018-11-23) description, paragraphs [0004]-[0050], and figures 1-6	1-20
A	KR 20190035445 A (SAMSUNG ELECTRONICS CO., LTD.) 03 April 2019 (2019-04-03) entire document	1-20
A	CN 109934339 A (SOUTHEAST UNIVERSITY et al.) 25 June 2019 (2019-06-25) entire document	1-20
A	CN 107885700 A (NATIONAL UNIVERSITY OF DEFENSE TECHNOLOGY) 06 April 2018 (2018-04-06) entire document	1-20
A	CN 108268931 A (HUAWEI TECHNOLOGIES CO., LTD.) 10 July 2018 (2018-07-10) entire document	1-20



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

25 December 2019

Date of mailing of the international search report

14 January 2020

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Authorized officer

Facsimile No. (86-10)62019451

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2019/103137

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	110135556	A	16 August 2019	None			
CN	108875908	A	23 November 2018	KR	20180125884	A	26 November 2018
				US	2018336462	A1	22 November 2018
KR	20190035445	A	03 April 2019	WO	2019066183	A1	04 April 2019
CN	109934339	A	25 June 2019	None			
CN	107885700	A	06 April 2018	None			
CN	108268931	A	10 July 2018	None			

国际检索报告

国际申请号

PCT/CN2019/103137

A. 主题的分类 G06N 3/04 (2006.01) i 按照国际专利分类 (IPC) 或者同时按照国家分类和 IPC 两种分类																							
B. 检索领域 检索的最低限度文献 (标明分类系统和分类号) G06N; G06K; G06T 包含在检索领域中的除最低限度文献以外的检索文献 在国际检索时查阅的电子数据库 (数据库的名称, 和使用的检索词 (如使用)) CNPAT, EPDOC, WPI, CNKI, IEEE: 神经网络, 卷积, 点乘, 加速, 分割, 脉动阵列, 划分, 分解, 细分, 子, 内核, 核函数, 卷积核, 尺寸, 高度, 宽度, 步长, 步进, 滑动; CNN, convol+, accelerat+, divide, sub, kerner, size, height, width, stride, step, slide																							
C. 相关文件 <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>PX</td> <td>CN 110135556 A (平安科技深圳有限公司) 2019年 8月 16日 (2019 - 08 - 16) 权利要求1-10, 说明书第[0004]-[0179]段</td> <td>1-20</td> </tr> <tr> <td>X</td> <td>CN 108875908 A (三星电子株式会社) 2018年 11月 23日 (2018 - 11 - 23) 说明书第[0004]-[0050]段、图1-6</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>KR 20190035445 A (SAMSUNG ELECTRONICS CO., LTD.) 2019年 4月 3日 (2019 - 04 - 03) 全文</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>CN 109934339 A (东南大学 等) 2019年 6月 25日 (2019 - 06 - 25) 全文</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>CN 107885700 A (中国人民解放军国防科技大学) 2018年 4月 6日 (2018 - 04 - 06) 全文</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>CN 108268931 A (华为技术有限公司) 2018年 7月 10日 (2018 - 07 - 10) 全文</td> <td>1-20</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	PX	CN 110135556 A (平安科技深圳有限公司) 2019年 8月 16日 (2019 - 08 - 16) 权利要求1-10, 说明书第[0004]-[0179]段	1-20	X	CN 108875908 A (三星电子株式会社) 2018年 11月 23日 (2018 - 11 - 23) 说明书第[0004]-[0050]段、图1-6	1-20	A	KR 20190035445 A (SAMSUNG ELECTRONICS CO., LTD.) 2019年 4月 3日 (2019 - 04 - 03) 全文	1-20	A	CN 109934339 A (东南大学 等) 2019年 6月 25日 (2019 - 06 - 25) 全文	1-20	A	CN 107885700 A (中国人民解放军国防科技大学) 2018年 4月 6日 (2018 - 04 - 06) 全文	1-20	A	CN 108268931 A (华为技术有限公司) 2018年 7月 10日 (2018 - 07 - 10) 全文	1-20
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求																					
PX	CN 110135556 A (平安科技深圳有限公司) 2019年 8月 16日 (2019 - 08 - 16) 权利要求1-10, 说明书第[0004]-[0179]段	1-20																					
X	CN 108875908 A (三星电子株式会社) 2018年 11月 23日 (2018 - 11 - 23) 说明书第[0004]-[0050]段、图1-6	1-20																					
A	KR 20190035445 A (SAMSUNG ELECTRONICS CO., LTD.) 2019年 4月 3日 (2019 - 04 - 03) 全文	1-20																					
A	CN 109934339 A (东南大学 等) 2019年 6月 25日 (2019 - 06 - 25) 全文	1-20																					
A	CN 107885700 A (中国人民解放军国防科技大学) 2018年 4月 6日 (2018 - 04 - 06) 全文	1-20																					
A	CN 108268931 A (华为技术有限公司) 2018年 7月 10日 (2018 - 07 - 10) 全文	1-20																					
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。																							
<table border="0"> <tr> <td> * 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件 </td> <td> “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件 </td> </tr> </table>			* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件	“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件																			
* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件	“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件																						
国际检索实际完成的日期 2019年 12月 25日		国际检索报告邮寄日期 2020年 1月 14日																					
ISA/CN的名称和邮寄地址 中国国家知识产权局 (ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451		受权官员 董洪梅 电话号码 86-(10)-53961528																					

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2019/103137

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	110135556	A	2019年 8月 16日	无			
CN	108875908	A	2018年 11月 23日	KR	20180125884	A	2018年 11月 26日
				US	2018336462	A1	2018年 11月 22日
KR	20190035445	A	2019年 4月 3日	WO	2019066183	A1	2019年 4月 4日
CN	109934339	A	2019年 6月 25日	无			
CN	107885700	A	2018年 4月 6日	无			
CN	108268931	A	2018年 7月 10日	无			