

(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

(43) 国际公布日
2022 年 3 月 31 日 (31.03.2022)



(10) 国际公布号
WO 2022/062388 A1

- (51) 国际专利分类号:
G06K 9/00 (2006.01)
- (21) 国际申请号: PCT/CN2021/089924
- (22) 国际申请日: 2021 年 4 月 26 日 (26.04.2021)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
202011026282.4 2020 年 9 月 25 日 (25.09.2020) CN
- (71) 申请人: 苏州浪潮智能科技有限公司
(**INSPUR SUZHOU INTELLIGENT TECHNOLOGY CO., LTD.**) [CN/CN]; 中国江苏省苏州市吴中区吴中经济开发区郭巷街道官浦路 1 号 9 幢, Jiangsu 215000 (CN)。

(72) 发明人: 董刚(**DONG, Gang**); 中国江苏省苏州市吴中区郭巷街道官浦路 1 号 9 幢, Jiangsu 215000

(CN)。 赵雅倩(**ZHAO, Yaqian**); 中国江苏省苏州市吴中区郭巷街道官浦路 1 号 9 幢, Jiangsu 215000 (CN)。 李仁刚(**LI, Rengang**); 中国江苏省苏州市吴中区郭巷街道官浦路 1 号 9 幢, Jiangsu 215000 (CN)。 杨宏斌(**YANG, Hongbin**); 中国江苏省苏州市吴中区郭巷街道官浦路 1 号 9 幢, Jiangsu 215000 (CN)。 刘海威(**LIU, Haiwei**); 中国江苏省苏州市吴中区郭巷街道官浦路 1 号 9 幢, Jiangsu 215000 (CN)。 蒋东东(**JIANG, Dongdong**); 中国江苏省苏州市吴中区郭巷街道官浦路 1 号 9 幢, Jiangsu 215000 (CN)。

(74) 代理人: 北京集佳知识产权代理有限公司
(**UNITALEN ATTORNEYS AT LAW**); 中国北京市朝阳区建国门外大街 22 号赛特广场 7 层, Beijing 100004 (CN)。

(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,

(54) **Title:** VIDEO DATA PROCESSING METHOD AND SYSTEM, AND RELEVANT ASSEMBLIES

(54) 发明名称: 一种视频数据处理方法、系统及相关组件

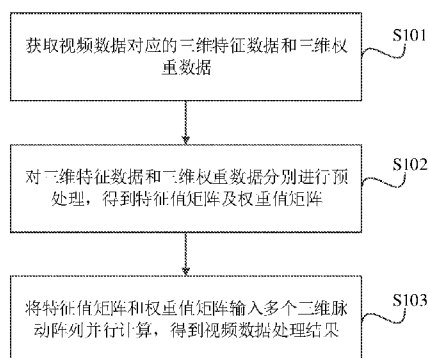


图 1

- S101 Obtain three-dimensional feature data and three-dimensional weight data corresponding to video data
- S102 Separately preprocess the three-dimensional feature data and the three-dimensional weight data to obtain a feature value matrix and a weight value matrix
- S103 Input the feature value matrix and the weight value matrix into a plurality of three-dimensional systolic arrays for parallel computing to obtain a video data processing result

(57) **Abstract:** A video data processing method, comprising: obtaining three-dimensional feature data and three-dimensional weight data corresponding to video data (S101); separately preprocessing the three-dimensional feature data and the three-dimensional weight data to obtain a feature value matrix and a weight value matrix (S102); and inputting the feature value matrix and the weight value matrix into a plurality of three-dimensional systolic arrays for parallel computing to obtain a video data processing result (S103). The present method can fully expand the degree of parallelism of computation, and a four-dimensional systolic computation architecture is constructed by using multiple three-dimensional systolic arrays, so as to perform parallel computing on a three-dimensional feature value matrix and a three-dimensional weight value matrix, thereby shortening the computation time of three-dimensional convolution, and improving the video data processing efficiency. A three-dimensional convolution parallel computing system, an electronic device, and a computer readable storage medium have the foregoing beneficial effects.

(57) **摘要:** 一种视频数据处理方法, 包括: 获取视频数据对应的三维特征数据和三维权重数据(S101); 对三维特征数据和三维权重数据分别进行预处理, 得到特征值矩阵及权重值矩阵(S102); 将特征值矩阵和权重值矩阵输入多个三维脉动阵列并行计算, 得到视频数据处理结果(S103)。本方法能够充分扩展计算的并行度, 采用多重三维脉动阵列构造四维脉动计算架构对三维特征值矩阵和三维权重值矩阵进行并行计算, 缩短了三维卷积的计算时间, 提高了视频数据处理效率。一种三维卷积并行计算系统、电子设备及计算机可读存储介质, 具有以上有益效果。

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, IT, 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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, 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))。

一种视频数据处理方法、系统及相关组件

本申请要求于 2020 年 09 月 25 日提交中国专利局、申请号为 202011026282.4、发明名称为“一种视频数据处理方法、系统及相关组件”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。

5 技术领域

本申请涉及视频数据处理领域，特别涉及一种视频数据处理方法、系统及相关组件。

背景技术

10 视频特征提取是视频数据处理的一个基本环节，几乎所有视频分析和处理的过程，都需要先对视频特征进行提取。三维 CNN（Convolutional Neural Networks，卷积神经网络）以其能更好的捕获视频中的时间和空间的特征信息，在视频分类、动作识别等领域发挥着巨大的优势，三维卷积是三维 CNN 中的主要计算步骤，通过三维卷积可以对视频数据进行分类
15 或从中提取特征，目前对于三维卷积计算的方法基本都是采用降低维度，将三维数据转化映射为二维数据甚至是一维数据，进行局部的并行计算，由于计算量巨大，运行速度非常缓慢，导致视频数据处理效率低。

因此，如何提供一种解决上述技术问题的方案是本领域技术人员目前需要解决的问题。

20

发明内容

本申请的目的是提供一种视频数据处理方法、系统、电子设备及计算机可读存储介质，能够充分扩展计算的并行度，采用多重三维脉动阵列构造四维脉动计算架构对三维特征值矩阵和三维权重值矩阵进行并行计算，
25 缩短了三维卷积的计算时间，提高了视频数据处理效率。

为解决上述技术问题，本申请提供了一种视频数据处理方法，包括：
获取视频数据对应的三维特征数据和三维权重数据；

对所述三维特征数据和所述三维权重数据分别进行预处理，得到特征

值矩阵及权重值矩阵；

将所述特征值矩阵和所述权重值矩阵输入多个三维脉动阵列并行计算，得到视频数据处理结果。

优选的，对所述三维特征数据进行预处理得到特征值矩阵的过程具体

5 包括：

对所述三维特征数据按卷积核尺寸进行切割，得到多个特征数据组，按预设映射关系将各个所述特征数据组转换成各自对应的二维矩阵；

根据所有所述二维矩阵得到特征值矩阵。

优选的，对所述三维权重数据进行预处理得到权重值矩阵的过程具体

10 包括：

将所述三维权重数据按所述预设映射关系进行重排，得到权重值矩阵。

优选的，所述将所述特征值矩阵和所述权重值矩阵输入多个三维脉动阵列并行计算，得到视频数据处理结果的过程具体包括：

15 通过第 i 个三维脉动阵列根据对应的目标中间值对第 i 个输入通道中的所述特征值矩阵及所述权重值矩阵执行计算操作，得到第 i 个计算结果， $i=1, 2, \dots, C_{in}$ ；

根据第 C_{in} 个计算结果得到视频数据处理结果；

其中，当 $i=1$ 时，所述目标中间值为 0，当 $1 < i \leq C_{in}$ 时，所述目标中间值为第 $i-1$ 个计算结果。

20 优选的，所述通过第 i 个三维脉动阵列根据对应的目标中间值对第 i 个输入通道中的所述特征值矩阵及所述权重值矩阵执行计算操作，得到第 i 个计算结果的过程具体包括：

将第 i 个输入通道中的所述特征值矩阵对应的 C_{out} 个权重值矩阵分别存储至第 i 个三维脉动阵列的 C_{out} 个计算单元中， C_{out} 为输出通道个数；

25 将所述第 i 个输入通道中特征值矩阵对应的各个子特征值矩阵按第一预设时间周期依次输入到所述第 i 个三维脉动阵列中；

通过每个所述计算单元根据目标中间值、接收到的特征值矩阵及存储的权重值矩阵执行计算操作，得到该计算单元对应的子计算结果；

通过所有所述子计算结果，得到第 i 个计算结果。

优选的，将所述第 i 个输入通道中特征值矩阵对应的每个子特征值矩阵输入到所述第 i 个三维脉动阵列中的过程包括：

将所述第 i 个输入通道中特征值矩阵对应的每个子特征值矩阵的第 r 行的 q 个特征值按第二预设时间周期分别输入到所述第 i 个三维脉动阵列的 C_{out} 个计算单元的第 r 行的 q 个 PE 中，所述子特征值矩阵的大小为 $p \times q$ ， p 、 q 均为正整数， $r=1, 2, \dots, p-1$ ；

其中，所述子特征值矩阵中第 $r+1$ 行的 q 个特征值输入到第 j 个计算单元与所述子特征值矩阵中第 r 行的 q 个特征值输入到所述第 j 个计算单元的时间间隔为所述第二预设时间周期，其中， $j=1, 2, \dots, C_{out}$ 。

10 优选的，所述通过每个所述计算单元根据目标中间值、接收到的特征值矩阵及存储的权重值矩阵执行计算操作，得到该计算单元对应的子计算结果的过程包括：

通过每个所述计算单元的第 r 行的 q 个 PE 按第一关系式执行计算操作，得到每个 PE 的计算结果；

15 所述第一关系式为 $h_{rw} = t_{rw} \times q_{rw} + c_{rw}$ ， h_w 为第 r 行、第 w 个 PE 的计算结果， t_w 为第 r 行、第 w 个 PE 接收到的特征值， q_w 为第 r 行、第 w 个 PE 的权重值， c_w 为第 r 行、第 w 个 PE 对应的目标中间值， $w=1, 2, \dots, q$ ；

根据相同列的所有所述 PE 的计算结果的和得到该计算单元的子计算结果。

20 优选的，所述根据第 C_{in} 个计算结果得到视频数据处理结果的过程具体包括：

获取第 C_{in} 个三维脉动阵列中每个所述计算单元的输出结果；

根据 C_{out} 个所述计算单元的输出结果得到视频数据处理结果。

25 优选的，所述获取第 C_{in} 个三维脉动阵列中每个所述计算单元的输出结果的过程具体包括：

通过第二关系式获取第 C_{in} 个三维脉动阵列中每个所述计算单元的输出结果，所述第二关系式为 $H = \sum_{w=1}^q \left(\sum_{r=1}^p h_{rw} \right)$ 。

为解决上述技术问题，本申请还提供了一种视频数据处理系统，包括：获取模块，用于获取视频数据对应的三维特征数据和三维权重数据；

预处理模块，用于对所述三维特征数据和所述三维权重数据分别进行预处理，得到特征值矩阵及权重值矩阵；

计算模块，用于将所述特征值矩阵和所述权重值矩阵输入多个三维脉动阵列并行计算，得到视频数据处理结果。

5 为解决上述技术问题，本申请还提供了一种电子设备，包括：

存储器，用于存储计算机程序；

处理器，用于执行所述计算机程序时实现如上文任意一项所述的视频数据处理方法的步骤。

10 为解决上述技术问题，本申请还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上存储有计算机程序，所述计算机程序被处理器执行时实现如上文任意一项所述的视频数据处理方法的步骤。

本申请提供了一种视频数据处理方法，对视频数据的三维特征值和三维权重值采取先降维再升维的预处理，在可行的条件下，充分扩展计算的并行度，采用多重三维脉动阵列构造四维脉动计算架构对特征值矩阵和权重值矩阵进行并行计算，缩短了三维卷积的计算时间，提高了视频数据处理效率。本申请还提供了一种三维卷积并行计算系统、电子设备及计算机可读存储介质，具有和上述三维卷积并行计算方法相同的有益效果。

15

附图说明

20 为了更清楚地说明本申请实施例，下面将对实施例中所需要使用的附图做简单的介绍，显而易见地，下面描述中的附图仅仅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 为本申请所提供的一种视频数据处理方法的步骤流程图；

25 图 2 为本申请所提供的一种映射关系示意图；

图 3 为本申请所提供的一种三维特征值重排示意图；

图 4 为本申请所提供的一种三维权重值重排示意图；

图 5 为本申请所提供的一种四维脉动阵列架构示意图；

图 6 为本申请所提供的一种三维脉动阵列计算模块的数据传输示意

图；

图7为本申请所提供的一种视频数据处理系统的结构示意图。

具体实施方式

5 本申请的核心是提供一种视频数据处理方法、系统、电子设备及计算机可读存储介质，能够充分扩展计算的并行度，采用多重三维脉动阵列构造四维脉动计算架构对三维特征值矩阵和三维权重值矩阵进行并行计算，缩短了三维卷积的计算时间，提高了视频数据处理效率。

10 为使本申请实施例的目的、技术方案和优点更加清楚，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。

15 为便于理解本申请所提供的一种基于三维卷积的视频数据处理方法，下面对卷积层的结构参数进行介绍，卷积层的结构参数主要有以下几种：

卷积核尺寸（KerCinel Size）：定义了卷积操作的感受野，在三维卷积中，通常设置为3，即卷积核大小为 $3 \times 3 \times 3$ 。

20 步幅（Stride）：定义了卷积核遍历图像时的步幅大小，其默认值通常设置为1，也可将步幅设置为2后对图像进行下采样，这种方式与最大池化类似。

边界扩充（Padding）：定义了网络层处理样本边界的方式，当卷积核大于1且不进行边界扩充，输出尺寸将相应缩小，当卷积核以标准方式进行边界扩充，则输出数据的空间尺寸将与输入相等。

25 假设输入数据的大小为 $a1 \times a2 \times a3$ ，输入通道个数为 Cin ，卷积核大小为 f ，即一个卷积核中包含了 $f \times f \times f$ 个权重值，输出通道个数为 $Cout$ ，因此权重值的总数目为 $f \times f \times f \times Cin \times Cout$ 。

基于上述情况，三维卷积最终的输出尺寸为

$$\left(\frac{a1-f+2\times padding}{Stride}+1\right)\times\left(\frac{a2-f+2\times padding}{Stride}+1\right)\times\left(\frac{a3-f+2\times padding}{Stride}+1\right)\times Cout$$

，该公式对于一维卷积、二维卷积仍然有效，只要调整输入数据的维度。

下面对本申请所提供的一种视频数据处理方法进行详细说明。

5 请参照图 1，图 1 为本申请所提供的一种视频数据处理方法的步骤流程图，该视频数据处理方法包括：

S101：获取视频数据对应的三维特征数据和三维权重数据；

首先需要说明的是，三维卷积的输入数据是由 Cin 个三维特征值矩阵和若干个（ $Cin\times Cout$ 个）三维权重值矩阵组成的，其中，单个三维特征值
10 矩阵尺寸为 $a1\times a2\times a3$ ，单个三维权重值矩阵尺寸为 $f\times f\times f$ ，因此，本实施例所提出的一种基于三维卷积的视频数据处理方法需要预先获取三维卷积的输入数据，即视频数据对应的三维特征数据和三维权重数据，以便后续对视频数据进行卷积运算。本实施例中的视频数据可以为安防监控录像对应的视频数据、也可以为自动驾驶过程中对应采集的视频数据、还可以为
15 流媒体在线视频对应的视频数据，本申请对于视频数据的应用领域不作具体的限定。

本步骤中，可以按预设获取周期获取视频数据对应的三维特征数据和三维权重数据，也可以在接收到获取指令后获取视频数据对应的三维特征数据和三维权重数据，本实施例在此对获取视频数据对应的三维特征数据
20 和三维权重数据的触发条件不作限定。

S102：对三维特征数据和三维权重数据分别进行预处理，得到特征值矩阵及权重值矩阵；

具体的，在获取到三维特征数据和三维权重数据后，先对二者进行降维预处理，使其符合三维脉动阵列的规模以及时序上的要求。作为一种优
25 选的实施例，对三维特征数据进行预处理得到特征值矩阵的过程具体包括：对三维特征数据按卷积核尺寸进行切割，得到多个特征数据组，然后按预

设映射关系将各个特征数据组转换成各自对应的二维矩阵，根据所有二维矩阵得到特征值矩阵；作为一种优选的实施例，对三维权重数据进行预处理得到权重值矩阵的过程具体包括：将三维权重数据按预设映射关系进行重排，得到权重值矩阵。

- 5 具体的，参照图 3 所示，对三维特征数据按卷积核尺寸分割，得到三维特征值矩阵 1、三维特征值矩阵 2 及三维特征值矩阵 3，按如图 2 所示的特定映射关系将三维特征值矩阵 1 组成二维矩阵 1'，按如图 2 所示的特定映射关系将三维特征值矩阵 2 组成二维矩阵 2'，按如图 2 所示的特定映射关系将三维特征值矩阵 3 组成二维矩阵 3'，将二维矩阵 1'，二维矩阵 2' 和二维矩阵 3' 按照按 a_1 、 a_2 、 a_3 的遍历顺序组合成新的特征值矩阵。其中，图 2 是以 $f=3$ 为例的根据数据循环的方式确定的映射关系，便于在硬件电路中实现。

- 具体的，对于三维卷积计算所需的三维权重值的预处理参照图 4 所示，将同一个输入通道对应的 C_{out} 个卷积核分别按图 2 所示的对应关系展开，
15 整合成一个新的权重值矩阵，具体的，按照如图 2 所示的特定映射关系将三维权重值矩阵 Z1 组成二维矩阵 Z1'，按照如图 2 所示的特定映射关系将三维权重值矩阵 Z2 组成二维矩阵 Z2'，按照如图 2 所示的特定映射关系将三维权重值矩阵 Z3 组成二维矩阵 Z3'，根据二维矩阵 Z1'、二维矩阵 Z2' 及二维矩阵 Z3' 得到新的权重值矩阵。

- 20 具体的，以 $f=3$ 为例，首先从第一个输入通道的三维特征数据中选取第一组 $3 \times 3 \times 3$ 个特征数据作为第一特征数据组，并将其按照图 2 所示的映射关系进行重新排序，得到第一特征数据组对应的二维矩阵，这里设为 t_{1-1-1} 、 t_{1-1-2} 、...、 t_{1-1-9} 、 t_{1-2-1} 、 t_{1-2-2} 、...、 t_{1-2-9} 、 t_{1-3-1} 、 t_{1-3-2} 、...、 t_{1-3-9} 共 27 个特征数据，其中， t_{1-1-1} 、 t_{1-1-2} 、...、 t_{1-1-9} 代表第一行中的
25 的九个数值， t_{1-2-1} 、 t_{1-2-2} 、...、 t_{1-2-9} 代表第二行中的九个数值， t_{1-3-1} 、 t_{1-3-2} 、...、 t_{1-3-9} 代表第三行中的九个数值，同理，可从第一个输入通道的三维特征数据中选取第二组 $3 \times 3 \times 3$ 个特征数据作为第二特征数据组，并将其按照图 2 所示的映射关系进行重新排序，得到第二特征数据组对应的二维矩阵，以此类推。第一个输入通道所对应的三维权重数据也要按照图

2 所示的映射关系进行重新排序得到多个权重值矩阵，特征值矩阵和权重值矩阵采用相同的数据结构，以便后续进行并行计算，提高并行运算效率。

S103：将特征值矩阵和权重值矩阵输入多个三维脉动阵列并行计算，得到视频数据处理结果。

5 具体的，当对视频数据的三维特征数据和三维权重数据都进行了如S102所描述的预处理之后，输入到采用多重三维脉动阵列构造四维脉动计算架构对特征值矩阵和权重值矩阵进行并行计算，从而得到三维卷积计算结果，将三维卷积计算结果作为视频数据处理结果，这里的视频数据处理结果可以是分类结果、特征提取结果等。通过扩展三维卷积计算的并行度，
10 提高了运算效率，尤其是针对视频处理方面对于实时性要求较高的特点，可以支持各种对实时三维图像的人工智能处理。

可见，本实施例中，对视频数据的三维特征值和三维权重值采取先降维再升维的预处理，在可行的条件下，充分扩展计算的并行度，采用多重三维脉动阵列构造四维脉动计算架构对特征值矩阵和权重值矩阵进行并行
15 计算，缩短了三维卷积的计算时间，提高了视频数据处理效率。

在上述实施例的基础上：

作为一种优选的实施例，将特征值矩阵和权重值矩阵输入多个三维脉动阵列并行计算，得到视频数据处理结果的过程具体包括：

20 通过第 i 个三维脉动阵列根据对应的目标中间值对第 i 个输入通道中的特征值矩阵及权重值矩阵执行计算操作，得到第 i 个计算结果， $i=1, 2, \dots, C_{in}$ ；

根据第 C_{in} 个计算结果得到视频数据处理结果；

其中，当 $i=1$ 时，目标中间值为0，当 $1 < i \leq C_{in}$ 时，目标中间值为第 $i-1$
25 个计算结果。

可以理解的是，一个输入通道对应一个三维脉动阵列，本实施例中，参照图5所示，存在多个三维脉动阵列，其中，每个三维脉动阵列均用于实现其对应的输入通道中的三维数据的卷积计算，由多个三维脉动阵列可以构造四维并行运算架构。在进行卷积计算时，需要用到中间值、权重值

和特征值，采用本实施例的方案可以将特征值和中间值按照特定的方向进行脉动传播，参照图 6 所示，实线箭头 Y 代表特征值的传输方向，虚线箭头 V 代表中间值的传输方向，形成立体网状的流水线操作。

可以理解的是，第一个三维脉动阵列对应的目标中间值为 0，第二个三维脉动阵列对应的目标中间值为第一个三维脉动阵列的计算结果，第三个三维脉动阵列对应的目标中间值为第二个三维脉动阵列的计算结果，以此类推，得到最后一个三维脉动阵列（即第 C_{in} 个三维脉动阵列）的计算结果后，根据该计算结果得到视频数据处理结果。

作为一种优选的实施例，通过第 i 个三维脉动阵列根据对应的目标中间值对第 i 个输入通道中的特征值矩阵及权重值矩阵执行计算操作，得到第 i 个计算结果的过程具体包括：

将第 i 个输入通道中的特征值矩阵对应的 C_{out} 个权重值矩阵分别存储至第 i 个三维脉动阵列的 C_{out} 个计算单元中， C_{out} 为输出通道个数；

将第 i 个输入通道中特征值矩阵对应的各个子特征值矩阵按第一预设时间周期依次输入到第 i 个三维脉动阵列中；

通过每个计算单元根据目标中间值、接收到的特征值矩阵及存储的权重值矩阵执行计算操作，得到该计算单元对应的子计算结果；

通过所有子计算结果，得到第 i 个计算结果。

具体的，假设 $f=3$ ，以第一个输入通道对应的特征值矩阵和权重值矩阵为例对三维脉动阵列执行的计算操作进行说明，其余输入通道，同理。

将第一个输入通道所对应的第一个三维权重数据矩阵按照图 2 所示的排列关系进行重新排序，预存至图 4 所示的三维脉动阵列的左斜条纹区域（第一计算单元）中，这里设为 $q1-1-1$ 、 $q1-1-2$ 、...、 $q1-1-9$ 、 $q1-2-1$ 、 $q1-2-2$ 、...、 $q1-2-9$ 、 $q1-3-1$ 、 $q1-3-2$ 、...、 $q1-3-9$ 共 27 个权重数据。第一个输入通道所对应的第二个三维权重数据矩阵按照图 2 所示的排列关系进行重新排序，并预存至图 4 所示的三维脉动阵列的右斜条纹区域（第二计算单元）中，该第一个输入通道所对应的第 C_{out} 个三维权重数据矩阵按照图 2 所示的排列关系进行重新排序，并预存至图 4 所示的三维脉动阵列的横条条纹区域（第 C_{out} 计算单元）中。对其余输入通道的三维数据的处理，同理。可以

理解的是，所有的三维权重数据经预处理后预存到三维脉动阵列中，在并行计算过程中，不发生变化。

具体的，根据每个输入通道中的特征值矩阵可以得到多个子特征值矩阵，每个子特征值矩阵大小为 $p \times q$ ，将各个子特征值矩阵按第一预设时间周期输入到三维脉动阵列中，如该输入通道中的特征值矩阵包括 3 个子特征值矩阵，则在第一个第一预设时间周期将第一个子特征值矩阵输入到三维脉动阵列中，在第二个第一预设时间周期将第二个子特征值矩阵输入到三维脉动阵列中，在第三个第一预设时间周期将第三个子特征值矩阵输入到三维脉动阵列中。

进一步的，不同输入通道中的特征值矩阵可以间隔预设时间段输入到各自对应的三维脉动阵列中，对每个输入通道中的特征值矩阵的处理方案，参照上文所述。

每个计算单元根据目标中间值、接收到的特征值矩阵及存储的权重值矩阵执行计算操作，得到该计算单元对应的子计算结果，该三维脉动阵列中的所有计算单元的子计算结果构成该三维脉动阵列的计算结果。

作为一种优选的实施例，将第 i 个输入通道中特征值矩阵对应的每个子特征值矩阵输入到第 i 个三维脉动阵列中的过程包括：

将第 i 个输入通道中特征值矩阵对应的每个子特征值矩阵的第 r 行的 q 个特征值按第二预设时间周期分别输入到第 i 个三维脉动阵列的 $Cout$ 个计算单元的第 r 行的 q 个 PE 中，子特征值矩阵的大小为 $p \times q$ ， p 、 q 均为正整数， $r=1, 2, \dots, p-1$ ；

其中，子特征值矩阵中第 $r+1$ 行的 q 个特征值输入到第 j 个计算单元与子特征值矩阵中第 r 行的 q 个特征值输入到第 j 个计算单元的时间间隔为第二预设时间周期，其中， $j=1, 2, \dots, Cout$ 。

具体的，本实施例对各个子特征值矩阵输入到三维脉动阵列的过程进行限定，假设 $f=3$ ，可以理解的是，在经过上述 S102 的预处理后，三维脉动阵列中的每一计算单元均包括 3×9 个 PE (Processing Elements, 处理单元)，每个 PE 会以预定的步骤和它相邻的 PE 进行数据传输。下面以其中一个子特征值矩阵的输入过程进行说明，子特征值矩阵的大小为 $p \times q$ ，

本实施例中, $p=3$, $q=9$, 参照图 4 所示, 在第一个第二预设周期, 将子特征矩阵第一行的 9 个特征数据一一对应输入到第一计算单元的第一行的 9 个 PE 中, 在第二个第二预设周期, 将子特征矩阵第一行的 9 个特征数据一一对应输入到第二计算单元的第一行的 9 个 PE 中, 在第三个第二预设周期, 将子特征矩阵第一行的 9 个特征数据一一对应输入到第三计算单元的第一行的 9 个 PE 中, 以此类推。进一步的, 为提高计算并行度, 在第二个第二预设周期, 在将子特征矩阵第一行的 9 个特征数据一一对应输入到第二计算单元的第一行的 9 个 PE 中的同时, 将子特征矩阵第二行的 9 个特征数据一一对应输入到第一计算单元的第二行的 9 个 PE 中。

5 作为一种优选的实施例通过每个计算单元根据目标中间值、接收到的特征值矩阵及存储的权重值矩阵执行计算操作, 得到该计算单元对应的子计算结果的过程包括:

通过每个计算单元的第 r 行的 q 个 PE 按第一关系式执行计算操作, 得到每个 PE 的计算结果;

15 第一关系式为 $h_{rw} = t_{rw} \times q_{rw} + c_{rw}$, h_w 为第 r 行、第 w 个 PE 的计算结果, t_w 为第 r 行、第 w 个 PE 接收到的特征值, q_w 为第 r 行、第 w 个 PE 的权重值, c_w 为第 r 行、第 w 个 PE 对应的目标中间值, $w=1, 2, \dots, q$;

根据相同列的所有 PE 的计算结果的和得到该计算单元的子计算结果。作为一种优选的实施例, 根据第 Cin 个计算结果得到视频数据处理结果的过程具体包括:

20

获取第 Cin 个三维脉动阵列中每个计算单元的输出结果;

根据 $Cout$ 个计算单元的输出结果得到视频数据处理结果。

作为一种优选的实施例, 获取第 Cin 个三维脉动阵列中每个计算单元的输出结果的过程具体包括:

25 通过第二关系式获取第 Cin 个三维脉动阵列中每个计算单元的输出结果, 第二关系式为 $H = \sum_{w=1}^q \left(\sum_{r=1}^p h_{rw} \right)$ 。

具体的, 下面对三维脉动阵列的具体计算过程进行解释说明:

首先从第一个输入通道的特征值矩阵中选取第一子特征值矩阵, 包括

t1-1-1、t1-1-2、...、t1-1-9、t1-2-1、t1-2-2、...、t1-2-9、t1-3-1、t1-3-2、...、t1-3-9 共 27 个特征数据，从第一个输入通道的特征值矩阵中选取第二子特征值矩阵，包括 t2-1-1、t2-1-2、...、t2-1-9、t2-2-1、t2-2-2、...、t2-2-9、t2-3-1、t2-3-2、...、t2-3-9 共 27 个特征数据，从第一个输入通道的特征值

5 矩阵中选取第三子特征值矩阵，包括 t3-1-1、t3-1-2、...、t3-1-9、t3-2-1、t3-2-2、...、t3-2-9、t3-3-1、t3-3-2、...、t3-3-9 共 27 个特征数据，从第一个输入通道的特征值矩阵中选取第四子特征值矩阵，包括 t4-1-1、t4-1-2、...、t4-1-9、t4-2-1、t4-2-2、...、t4-2-9、t4-3-1、t4-3-2、...、t4-3-9 共 27 个特征数据，以此类推，在第二个输入通道的特征值矩阵中选取第一子特征值

10 矩阵，包括 t'1-1-1、t'1-1-2、...、t'1-1-9、t'1-2-1、t'1-2-2、...、t'1-2-9、t'1-3-1、t'1-3-2、...、t'1-3-9 共 27 个特征数据。

在第一个预设时间周期内，特征数据 t1-1-1、t1-1-2、...、t1-1-9 送到图 4 所示左斜条纹区域的最上一行的 9 个 PE 中，分别与权重值 q1-1-1、q1-1-2、...、q1-1-9 对应相乘，即 t1-1-1 和 q1-1-1 相乘，t1-1-2 和 q1-1-2

15 相乘，以此类推，分别得到的中间值设为 m1-1-1、m1-1-2、...、m1-1-9，分别作为下一行各个 PE 的目标中间值。

在第二个预设时间周期内，特征数据 t1-2-1、t1-2-2、...、t1-2-9 送到图 4 所示左斜条纹的中间一行的 9 个 PE 中，与权重值 q1-2-1、q1-2-2、...、q1-2-9 对应相乘，并与输入的中间值 m1-1-1、m1-1-2、...、m1-1-9 相加，

20 即 t1-2-1 和 q1-2-1 相乘并加上 m1-1-1，t1-2-2 和 q1-2-2 相乘并加上 m1-1-2，以此类推，得到的中间值设为 m1-2-1、m1-2-2、...、m1-2-9。同时，特征数据的 t2-1-1、t2-1-2、...、t2-1-9 送到图 4 所示左斜条纹的最上一行的 9 个 PE 中，与 q1-1-1、q1-1-2、...、q1-1-9 对应相乘，即 t2-1-1 和 q1-1-1 相乘，t2-1-2 和 q1-1-2 相乘，以此类推，得到的中间值设为 m2-1-1、m2-1-2、...、

25 m2-1-9。同时，特征数据 t1-1-1、t1-1-2、...、t1-1-9 送到图 4 所示右斜条纹的最上一行的 9 个 PE 中，与 q2-1-1、q2-1-2、...、q2-1-9 对应相乘，即 t1-1-1 和 q2-1-1 相乘，t1-1-2 和 q2-1-2 相乘，以此类推，得到的中间值设为 n1-1-1、n1-1-2、...、n1-1-9。

在第三个预设时间周期内，特征数据 t1-3-1、t1-3-2、...、t1-3-9 送到

图 4 所示左斜条纹的最下一行的 9 个 PE 中, 与 $q1-3-1$ 、 $q1-3-2$ 、...、 $q1-3-9$ 对应相乘, 并与输入的中间值 $m1-2-1$ 、 $m1-2-2$ 、...、 $m1-2-9$ 相加, 即 $t1-3-1$ 和 $q1-3-1$ 相乘并加上 $m1-2-1$, $t1-3-2$ 和 $q1-3-2$ 相乘并加上 $m1-2-2$, 以此类推, 得到的中间值设为 $m1-3-1$ 、 $m1-3-2$ 、...、 $m1-3-9$ 。同时, 特征数据 $t2-2-1$ 、 $t2-2-2$ 、...、 $t2-2-9$ 送到图 4 所示左斜条纹的中间一行的 9 个 PE 中, 与 $q1-2-1$ 、 $q1-2-2$ 、...、 $q1-2-9$ 对应相乘, 并与输入的中间值 $m2-1-1$ 、 $m2-1-2$ 、...、 $m2-1-9$ 相加, 即 $t2-2-1$ 和 $q1-2-1$ 相乘并加上 $m2-1-1$, $t2-2-2$ 和 $q1-2-2$ 相乘并加上 $m2-1-2$, 以此类推, 得到的中间值设为 $m2-2-1$ 、 $m2-2-2$ 、...、 $m2-2-9$ 。同时, 特征数据 $t3-1-1$ 、 $t3-1-2$ 、...、 $t3-1-9$ 送到图 4 所示左斜条纹的最上一行的 9 个 PE 中, 与 $q1-1-1$ 、 $q1-1-2$ 、...、 $q1-1-9$ 对应相乘, 即 $t3-1-1$ 和 $q1-1-1$ 相乘, $t3-1-2$ 和 $q1-1-2$ 相乘, 以此类推, 得到的中间值设为 $m3-1-1$ 、 $m3-1-2$ 、...、 $m3-1-9$ 。同时, 特征数据 $t1-2-1$ 、 $t1-2-2$ 、...、 $t1-2-9$ 送到图 4 所示右斜条纹的中间一行的 9 个 PE 中, 与 $q2-2-1$ 、 $q2-2-2$ 、...、 $q2-2-9$ 对应相乘, 并与输入的中间值 $n1-1-1$ 、 $n1-1-2$ 、...、 $n1-1-9$ 相加, 即 $t1-2-1$ 和 $q2-2-1$ 相乘并加上 $n1-1-1$, $t1-2-2$ 和 $q2-2-2$ 相乘并加上 $n1-1-2$, 以此类推, 得到的中间值设为 $n1-2-1$ 、 $n1-2-2$ 、...、 $n1-2-9$ 。同时, 特征数据 $t2-1-1$ 、 $t2-1-2$ 、...、 $t2-1-9$ 送到图 4 所示右斜条纹的最上一行的 9 个 PE 中, 与 $q2-1-1$ 、 $q2-1-2$ 、...、 $q2-1-9$ 对应相乘, 即 $t2-1-1$ 和 $q2-1-1$ 相乘, $t2-1-2$ 和 $q2-1-2$ 相乘, 以此类推, 得到的中间值设为 $n2-1-1$ 、 $n2-1-2$ 、...、 $n2-1-9$ 。同时, 特征数据 $t1-1-1$ 、 $t1-1-2$ 、...、 $t1-1-9$ 送到图 4 所示右斜条纹后面一组 (图 4 中为了显示效果, 没有画出) 的最上一行的 9 个 PE 中, 与 $q3-1-1$ 、 $q3-1-2$ 、...、 $q3-1-9$ 对应相乘, 即 $t1-1-1$ 和 $q3-1-1$ 相乘, $t1-1-2$ 和 $q3-1-2$ 相乘, 以此类推, 得到的中间值设为 $o1-1-1$ 、 $o1-1-2$ 、...、 $o1-1-9$ 。

在第四个预设时间周期内, 特征数据 $t1-1-1$ 、 $t1-1-2$ 、...、 $t1-1-9$ 送到图 4 所示右斜条纹后面第二组 (图 4 中为了显示效果, 没有画出) 的最上一行的 9 个 PE 中, 与 $q4-1-1$ 、 $q4-1-2$ 、...、 $q4-1-9$ 对应相乘, 即 $t1-1-1$ 和 $q4-1-1$ 相乘, $t1-1-2$ 和 $q4-1-2$ 相乘, 以此类推, 得到的中间值设为 $k1-1-1$ 、 $k1-1-2$ 、...、 $k1-1-9$ 。同时, 特征数据 $t1-2-1$ 、 $t1-2-2$ 、...、 $t1-2-9$ 送到图 4 所示右斜条纹后面一组的中间一行的 9 个 PE 中, 与 $q3-2-1$ 、 $q3-2-2$ 、...、

$q3-2-9$ 对应相乘，并与输入的中间值 $o1-1-1$ 、 $o1-1-2$ 、...、 $o1-1-9$ 相加，即 $t1-2-1$ 和 $q3-2-1$ 相乘并加上 $o1-1-1$ ， $t1-2-2$ 和 $q3-2-2$ 相乘并加上 $o1-1-2$ ，以此类推，得到的中间值设为 $o1-2-1$ 、 $o1-2-2$ 、...、 $o1-2-9$ 。同时，特征数据 $t1-3-1$ 、 $t1-3-2$ 、...、 $t1-3-9$ 送到图 4 所示右斜条纹的最下一行的 9 个 PE 中，

5 与 $q2-3-1$ 、 $q2-3-2$ 、...、 $q2-3-9$ 对应相乘，并与输入的中间值 $n1-2-1$ 、 $n1-2-2$ 、...、 $n1-2-9$ 相加，即 $t1-3-1$ 和 $q1-3-1$ 相乘并加上 $m1-2-1$ ， $t1-3-2$ 和 $q1-3-2$ 相乘并加上 $m1-2-2$ ，以此类推，得到的中间值设为 $m1-3-1$ 、 $m1-3-2$ 、...、 $m1-3-9$ 。同时，特征数据 $t2-1-1$ 、 $t2-1-2$ 、...、 $t2-1-9$ 送到图 4 所示右斜条纹后面一组（图 4 中为了显示效果，没有画出）的最上一行的

10 9 个 PE 中，与 $q3-1-1$ 、 $q3-1-2$ 、...、 $q3-1-9$ 对应相乘，即 $t2-1-1$ 和 $q3-1-1$ 相乘， $t2-1-2$ 和 $q3-1-2$ 相乘，以此类推，得到的中间值设为 $o2-1-1$ 、 $o2-1-2$ 、...、 $o2-1-9$ 。同时，特征数据 $t2-2-1$ 、 $t2-2-2$ 、...、 $t2-2-9$ 送到图 4 所示右斜条纹的中间一行的 9 个 PE 中，与 $q2-2-1$ 、 $q2-2-2$ 、...、 $q2-2-9$ 对应相乘，并与输入的中间值 $n2-1-1$ 、 $n2-1-2$ 、...、 $n2-1-9$ 相加，即 $t2-2-1$ 和 $q2-2-1$ 相乘

15 并加上 $n2-1-1$ ， $t2-2-2$ 和 $q2-2-2$ 相乘并加上 $n2-1-2$ ，以此类推，得到的中间值设为 $n2-2-1$ 、 $n2-2-2$ 、...、 $n2-2-9$ 。同时，特征数据 $t2-3-1$ 、 $t2-3-2$ 、...、 $t2-3-9$ 送到图 4 所示左斜条纹的最下一行的 9 个 PE 中，与 $q1-3-1$ 、 $q1-3-2$ 、...、 $q1-3-9$ 对应相乘，并与输入的中间值 $m2-2-1$ 、 $m2-2-2$ 、...、 $m2-2-9$ 相加，即 $t2-3-1$ 和 $q1-3-1$ 相乘并加上 $m2-2-1$ ， $t2-3-2$ 和 $q1-3-2$ 相乘并加上 $m2-2-2$ ，

20 以此类推，得到的中间值设为 $m2-3-1$ 、 $m2-3-2$ 、...、 $m2-3-9$ 。同时，特征数据 $t3-1-1$ 、 $t3-1-2$ 、...、 $t3-1-9$ 送到图 4 所示右斜条纹的最上一行的 9 个 PE 中，与 $q2-1-1$ 、 $q2-1-2$ 、...、 $q2-1-9$ 对应相乘，即 $t3-1-1$ 和 $q2-1-1$ 相乘， $t3-1-2$ 和 $q2-1-2$ 相乘，以此类推，得到的中间值设为 $n3-1-1$ 、 $n3-1-2$ 、...、 $n3-1-9$ 。同时，特征数据 $t3-2-1$ 、 $t3-2-2$ 、...、 $t3-2-9$ 送到图 4 所示左斜条纹的

25 中间一行的 9 个 PE 中，与 $q1-2-1$ 、 $q1-2-2$ 、...、 $q1-2-9$ 对应相乘，并与输入的中间值 $m3-1-1$ 、 $m3-1-2$ 、...、 $m3-1-9$ 相加，即 $t3-2-1$ 和 $q1-2-1$ 相乘并加上 $m3-1-1$ ， $t3-2-2$ 和 $q1-2-2$ 相乘并加上 $m3-1-2$ ，以此类推，得到的中间值设为 $m3-2-1$ 、 $m3-2-2$ 、...、 $m3-2-9$ 。同时，特征数据 $t4-1-1$ 、 $t4-1-2$ 、...、 $t4-1-9$ 送到图 4 所示左斜条纹的最上一行的 9 个 PE 中，与 $q1-1-1$ 、 $q1-1-2$ 、...、

q1-1-9 对应相乘，即 t4-1-1 和 q1-1-1 相乘，t4-1-2 和 q1-1-2 相乘，以此类推，得到的中间值设为 m4-1-1、m4-1-2、...、m4-1-9。同时，第二个输入通道的特征数据 t'1-1-1、t'1-1-2、...、t'1-1-9 送到图 5 所示输入通道 2 特征值所对的左斜条纹的最上一行的 9 个 PE 中，与 q'1-1-1、q'1-1-2、...、q'1-1-9 对应相乘，并与输入的中间值 m1-3-1、m1-3-2、...、m1-3-9 相加即 t'1-1-1 和 q'1-1-1 相乘并加上 m1-3-1，t'1-1-2 和 q1-1-2 相乘并加上 m1-3-2，以此类推，得到的中间值设为 m'1-1-1、m'1-1-2、...、m'1-1-9，后面的计算步骤基本就是上面的叠加和重复。

由于是随着时间周期逐步分批启动各级运算单元，特征数据将在这些预算单元内水平传送，如图 6 中的实线箭头 Y 所示。各级乘累加的中间值在多组三维运算阵列之中进行传递，如图 6 中的纵向虚线箭头 V 所示。当 Cin 组的三维脉动阵列结果都纵向累加到了一起，就还需要额外将这个排列中的 9 列的结果再进行累加成一个数值，也就是由 $3 \times 9 \times Cin$ 乘积相加得来，得到的结果是一个输出通道中的一个数值，然后重复上述的过程，直到完成全部的计算。

这里再做进一步的扩展，由于深度学习网络中往往输入通道和输出通道数目比较大，达到 1000 以上的规模，因此完全串行计算很慢。这里就重复上面的三维计算阵列，实现多路并行计算。但多个三维计算阵列中的权重值是互不相同的，而且这些不同的三维计算阵列得到的结果需要再次累加才是最终的结果。

请参照图 7，图 7 为本申请所提供的一种视频数据处理系统的结构示意图，该视频数据处理系统包括：

获取模块 11，用于获取视频数据对应的三维特征数据和三维权重数据；

预处理模块 12，用于对三维特征数据和三维权重数据分别进行预处理，得到特征值矩阵及权重值矩阵；

计算模块 13，用于将特征值矩阵和权重值矩阵输入多个三维脉动阵列

并行计算，得到视频数据处理结果。

可见，本实施例中，对视频数据的三维特征值和三维权重值采取先降维再升维的预处理，在可行的条件下，充分扩展计算的并行度，采用多重三维脉动阵列构造四维脉动计算架构对特征值矩阵和权重值矩阵进行并行计算，缩短了三维卷积的计算时间，提高了视频数据处理效率。

作为一种优选的实施例，预处理模块 12 包括：

第一预处理单元，用于对所述三维特征数据按卷积核尺寸进行切割，得到多个子特征值矩阵，按预设映射关系将各个所述子特征值矩阵转换成各自对应的子特征值矩阵，根据所有所述子特征值矩阵得到特征值矩阵。

10 作为一种优选的实施例，预处理模块 12 还包括：

第二预处理单元，用于将三维权重数据按预设映射关系进行重排，得到权重值矩阵。

作为一种优选的实施例，计算模块 3 具体用于：

15 通过第 i 个三维脉动阵列根据对应的目标中间值对第 i 个输入通道中的特征值矩阵及权重值矩阵执行计算操作，得到第 i 个计算结果， $i=1, 2, \dots, C_{in}$ ；

根据第 C_{in} 个计算结果得到视频数据处理结果；

其中，当 $i=1$ 时，目标中间值为 0，当 $1 < i \leq C_{in}$ 时，目标中间值为第 $i-1$ 个计算结果。

20 作为一种优选的实施例，通过第 i 个三维脉动阵列根据对应的目标中间值对第 i 个输入通道中的特征值矩阵及权重值矩阵执行计算操作，得到第 i 个计算结果的过程具体包括：

将第 i 个输入通道中的特征值矩阵对应的 C_{out} 个权重值矩阵分别存储至第 i 个三维脉动阵列的 C_{out} 个计算单元中， C_{out} 为输出通道个数；

25 将第 i 个输入通道中特征值矩阵对应的各个子特征值矩阵按第一预设时间周期依次输入到第 i 个三维脉动阵列中；

通过每个计算单元根据目标中间值、接收到的特征值矩阵及存储的权重值矩阵执行计算操作，得到该计算单元对应的子计算结果；

通过所有子计算结果，得到第 i 个计算结果。

作为一种优选的实施例，将第 i 个输入通道中特征值矩阵对应的每个子特征值矩阵输入到第 i 个三维脉动阵列中的过程包括：

- 将第 i 个输入通道中特征值矩阵对应的每个子特征值矩阵的第 r 行的 q 个特征值按第二预设时间周期分别输入到第 i 个三维脉动阵列的 $Cout$ 个计算单元的第 r 行的 q 个 PE 中，子特征值矩阵的大小为 $p \times q$ ， p 、 q 均为正整数， $r=1, 2, \dots, p-1$ ；

其中，子特征值矩阵中第 $r+1$ 行的 q 个特征值输入到第 j 个计算单元与子特征值矩阵中第 r 行的 q 个特征值输入到第 j 个计算单元的时间间隔为第二预设时间周期，其中， $j=1, 2, \dots, Cout$ 。

- 10 作为一种优选的实施例通过每个计算单元根据目标中间值、接收到的特征值矩阵及存储的权重值矩阵执行计算操作，得到该计算单元对应的子计算结果的过程包括：

通过每个计算单元的第 r 行的 q 个 PE 按第一关系式执行计算操作，得到每个 PE 的计算结果；

- 15 第一关系式为 $h_{rw} = t_{rw} \times q_{rw} + c_{rw}$ ， h_w 为第 r 行、第 w 个 PE 的计算结果， t_w 为第 r 行、第 w 个 PE 接收到的特征值， q_w 为第 r 行、第 w 个 PE 的权重值， c_w 为第 r 行、第 w 个 PE 对应的目标中间值， $w=1, 2, \dots, q$ ；

根据相同列的所有 PE 的计算结果的和得到该计算单元的子计算结果。

- 20 作为一种优选的实施例，根据第 Cin 个计算结果得到视频数据处理结果的过程具体包括：

获取第 Cin 个三维脉动阵列中每个计算单元的输出结果；

根据 $Cout$ 个计算单元的输出结果得到视频数据处理结果。

作为一种优选的实施例，获取第 Cin 个三维脉动阵列中每个计算单元的输出结果的过程具体包括：

- 25 通过第二关系式获取第 Cin 个三维脉动阵列中每个计算单元的输出结果，第二关系式为 $H = \sum_{w=1}^q \left(\sum_{r=1}^p h_{rw} \right)$ 。

另一方面，本申请还提供了一种电子设备，包括：

存储器，用于存储计算机程序；

处理器，用于执行所述计算机程序时实现如上文任意一个实施例所描述的视频数据处理方法的步骤。

在本申请实施例中，该处理器，可以为中央处理器（Central Processing Unit, CPU），特定应用集成电路，数字信号处理器，现成可编程门阵列或者其他可编程逻辑器件等。

对于本申请所提供的一种电子设备的介绍请参照上述实施例，本申请在此不再赘述。

本申请所提供的一种电子设备具有和上述视频数据处理方法相同的有益效果。

10

另一方面，本申请还提供了一种计算机可读存储介质，计算机可读存储介质上存储有计算机程序，计算机程序被处理器执行时实现如上文任意一个实施例所描述的视频数据处理方法的步骤。

对于本申请所提供的一种计算机可读存储介质的介绍请参照上述实施例，本申请在此不再赘述。

本申请所提供的一种计算机可读存储介质具有和上述视频数据处理方法相同的有益效果。

还需要说明的是，在本说明书中，诸如第一和第二等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来，而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且，术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、方法、物品或者设备不仅包括那些要素，而且还包括没有明确列出的其他要素，或者是还包括为这种过程、方法、物品或者设备所固有的要素。在没有更多限制的状况下，由语句“包括一个...”限定的要素，并不排除在包括所述要素的过程、方法、物品或者设备中还存在另外的相同要素。

对所公开的实施例的上述说明，使本领域专业技术人员能够实现或使用本申请。对这些实施例的多种修改对本领域的专业技术人员来说将是显

而易见的，本文中所定义的一般原理可以在不脱离本申请的精神或范围的情况下，在其他实施例中实现。因此，本申请将不会被限制于本文所示的这些实施例，而是要符合与本文所公开的原理和新颖特点相一致的最宽的范围。

权 利 要 求

1、一种视频数据处理方法，其特征在于，包括：

获取视频数据对应的三维特征数据和三维权重数据；

5 对所述三维特征数据和所述三维权重数据分别进行预处理，得到特征值矩阵及权重值矩阵；

将所述特征值矩阵和所述权重值矩阵输入多个三维脉动阵列并行计算，得到视频数据处理结果。

2、根据权利要求1所述的视频数据处理方法，其特征在于，对所述三维特征数据进行预处理得到特征值矩阵的过程具体包括：

10 对所述三维特征数据按卷积核尺寸进行切割，得到多个特征数据组，按预设映射关系将各个所述特征数据组转换成各自对应的二维矩阵；

根据所有所述二维矩阵得到特征值矩阵。

3、根据权利要求2所述的视频数据处理方法，其特征在于，对所述三维权重数据进行预处理得到权重值矩阵的过程具体包括：

15 将所述三维权重数据按所述预设映射关系进行重排，得到权重值矩阵。

4、根据权利要求1-3任意一项所述的视频数据处理方法，其特征在于，所述将所述特征值矩阵和所述权重值矩阵输入多个三维脉动阵列并行计算，得到视频数据处理结果的过程具体包括：

20 通过第 i 个三维脉动阵列根据对应的目标中间值对第 i 个输入通道中的所述特征值矩阵及所述权重值矩阵执行计算操作，得到第 i 个计算结果， $i=1, 2, \dots, Cin$ ；

根据第 Cin 个计算结果得到视频数据处理结果；

其中，当 $i=1$ 时，所述目标中间值为0，当 $1 < i \leq Cin$ 时，所述目标中间值为第 $i-1$ 个计算结果。

25 5、根据权利要求4所述的视频数据处理方法，其特征在于，所述通过第 i 个三维脉动阵列根据对应的目标中间值对第 i 个输入通道中的所述特征值矩阵及所述权重值矩阵执行计算操作，得到第 i 个计算结果的过程具体包括：

将第 i 个输入通道中的所述特征值矩阵对应的 $Cout$ 个权重值矩阵分别

存储至第 i 个三维脉动阵列的 C_{out} 个计算单元中, C_{out} 为输出通道个数;

将所述第 i 个输入通道中特征值矩阵对应的各个子特征值矩阵按第一预设时间周期依次输入到所述第 i 个三维脉动阵列中;

通过每个所述计算单元根据目标中间值、接收到的特征值矩阵及存储的权重值矩阵执行计算操作, 得到该计算单元对应的子计算结果;

通过所有所述子计算结果, 得到第 i 个计算结果。

6、根据权利要求 5 所述的视频数据处理方法, 其特征在于, 将所述第 i 个输入通道中特征值矩阵对应的每个子特征值矩阵输入到所述第 i 个三维脉动阵列中的过程包括:

10 将所述第 i 个输入通道中特征值矩阵对应的每个子特征值矩阵的第 r 行的 q 个特征值按第二预设时间周期分别输入到所述第 i 个三维脉动阵列的 C_{out} 个计算单元的第 r 行的 q 个 PE 中, 所述子特征值矩阵的大小为 $p \times q$, p 、 q 均为正整数, $r=1, 2, \dots, p-1$;

15 其中, 所述子特征值矩阵中第 $r+1$ 行的 q 个特征值输入到第 j 个计算单元与所述子特征值矩阵中第 r 行的 q 个特征值输入到所述第 j 个计算单元的时间间隔为所述第二预设时间周期, 其中, $j=1, 2, \dots, C_{out}$ 。

7、根据权利要求 6 所述的视频数据处理方法, 其特征在于, 所述通过每个所述计算单元根据目标中间值、接收到的特征值矩阵及存储的权重值矩阵执行计算操作, 得到该计算单元对应的子计算结果的过程包括:

20 通过每个所述计算单元的第 r 行的 q 个 PE 按第一关系式执行计算操作, 得到每个 PE 的计算结果;

所述第一关系式为 $h_{rw} = t_{rw} \times q_{rw} + c_{rw}$, h_w 为第 r 行、第 w 个 PE 的计算结果, t_w 为第 r 行、第 w 个 PE 接收到的特征值, q_w 为第 r 行、第 w 个 PE 的权重值, c_w 为第 r 行、第 w 个 PE 对应的目标中间值, $w=1, 2, \dots, q$;

25 根据相同列的所有所述 PE 的计算结果的和得到该计算单元的子计算结果。

8、根据权利要求 7 所述的视频数据处理方法, 其特征在于, 所述根据第 C_{in} 个计算结果得到视频数据处理结果的过程具体包括:

获取第 C_{in} 个三维脉动阵列中每个所述计算单元的输出结果;

根据 C_{out} 个所述计算单元的输出结果得到视频数据处理结果。

9、根据权利要求 8 所述的视频数据处理方法，其特征在于，所述获取第 C_{in} 个三维脉动阵列中每个所述计算单元的输出结果的过程具体包括：

通过第二关系式获取第 C_{in} 个三维脉动阵列中每个所述计算单元的输

5 出结果，所述第二关系式为 $H = \sum_{w=1}^q \left(\sum_{r=1}^p h_{rw} \right)$ 。

10、一种视频数据处理系统，其特征在于，包括：

获取模块，用于获取视频数据对应的三维特征数据和三维权重数据；

预处理模块，用于对所述三维特征数据和所述三维权重数据分别进行预处理，得到特征值矩阵及权重值矩阵；

10 计算模块，用于将所述特征值矩阵和所述权重值矩阵输入多个三维脉动阵列并行计算，得到视频数据处理结果。

11、一种电子设备，其特征在于，包括：

存储器，用于存储计算机程序；

15 处理器，用于执行所述计算机程序时实现如权利要求 1-9 任意一项所述的视频数据处理方法的步骤。

12、一种计算机可读存储介质，其特征在于，所述计算机可读存储介质上存储有计算机程序，所述计算机程序被处理器执行时实现如权利要求 1-9 任意一项所述的视频数据处理方法的步骤。

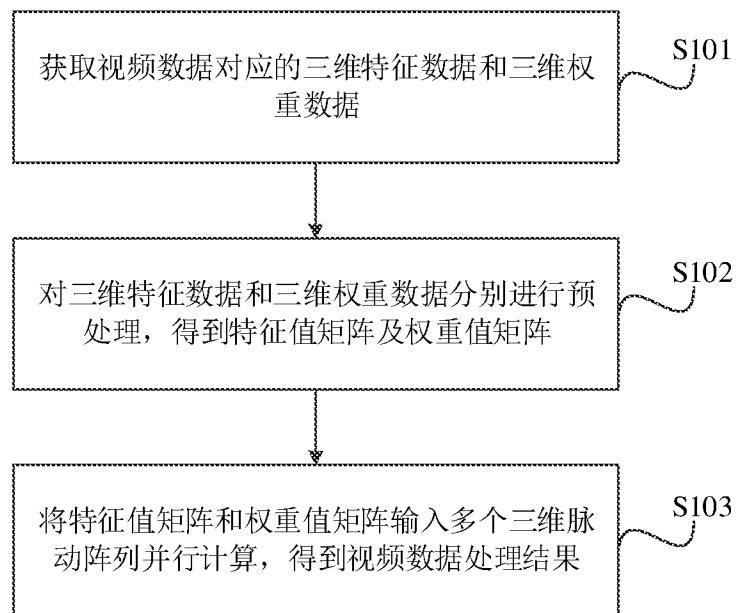


图 1

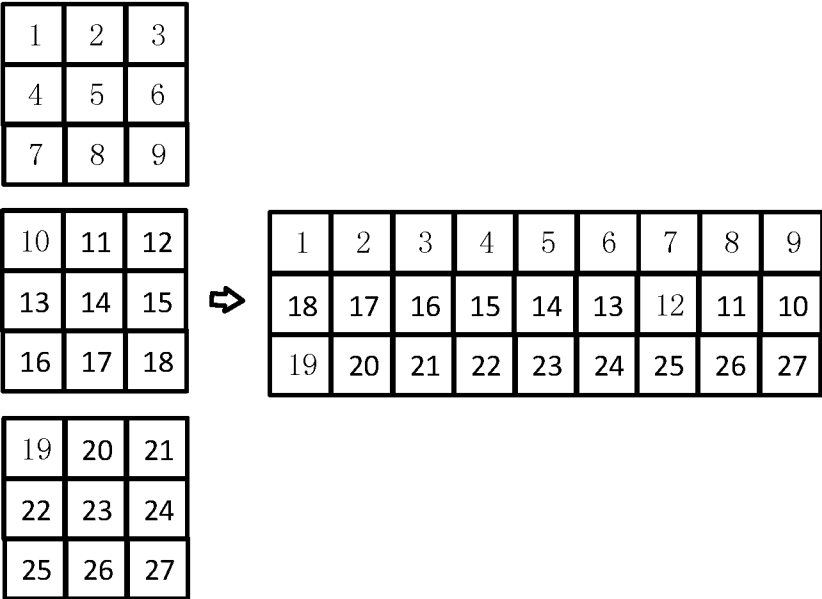


图 2

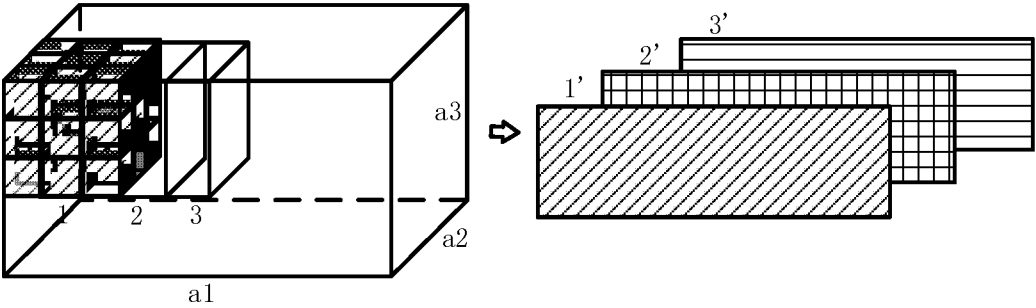


图 3

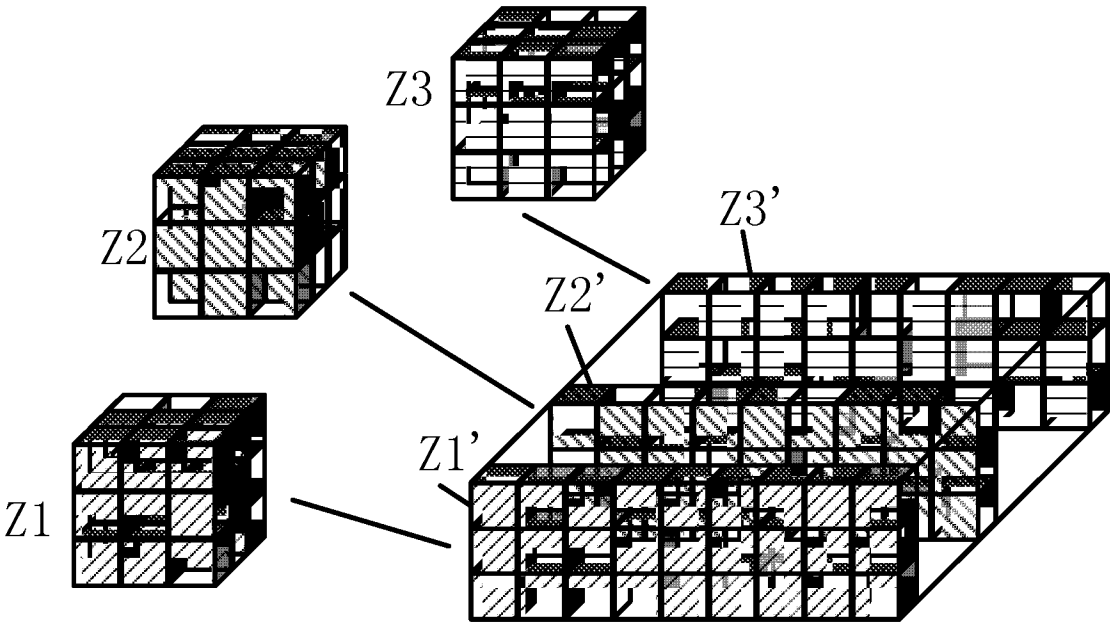


图 4

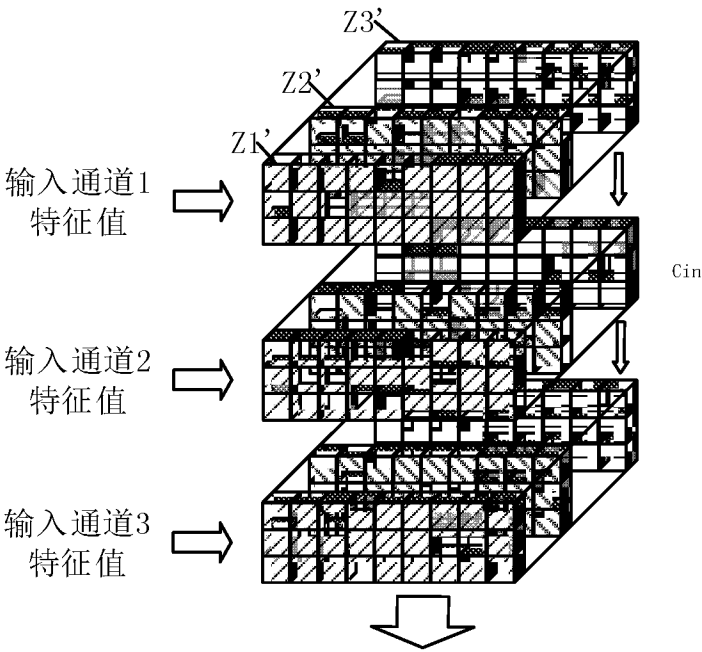


图 5

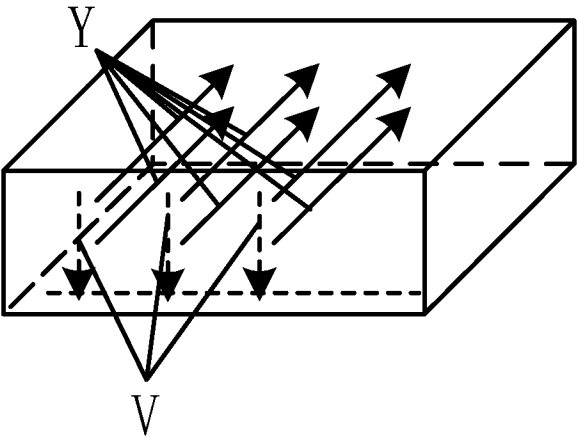


图 6

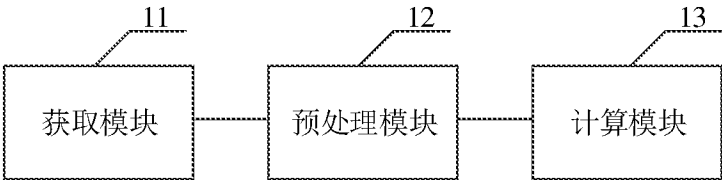


图 7

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2021/089924

A. CLASSIFICATION OF SUBJECT MATTER

G06K 9/00(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)

G06K G06N

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)

WPI, EPODOC, CNPAT, CNKI, IEEE: 视频, 三维, 特征, 权重, 矩阵, 并行, 计算, video, 3d, character, feature, weight, matrix, parallel, compute

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 112016522 A (SUZHOU INSPUR INTELLIGENT TECHNOLOGY CO., LTD.) 01 December 2020 (2020-12-01) claims 1-12	1-12
Y	CN 110785778 A (SHENZHEN DAJIANG INNOVATION TECHNOLOGY CO., LTD.) 11 February 2020 (2020-02-11) description, paragraphs [0024]-[0188], and figures 1-23	1-5, 10-12
Y	CN 111695672 A (BAIDU USA L.L.C.) 22 September 2020 (2020-09-22) description, paragraphs 0018-0080, figures 1-6	1-5, 10-12
A	US 2019065903 A1 (NORTHROP GRUMMAN SYSTEMS CORPORATION) 28 February 2019 (2019-02-28) entire document	1-12



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

12 July 2021

Date of mailing of the international search report

26 July 2021

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

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2021/089924

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	112016522	A	01 December 2020	None			
CN	110785778	A	11 February 2020	WO	2020034079	A1	20 February 2020
CN	111695672	A	22 September 2020	US	2020293866	A1	17 September 2020
US	2019065903	A1	28 February 2019	JP	2020532008	A	05 November 2020
				EP	3673417	A1	01 July 2020
				WO	2019040214	A1	28 February 2019

国际检索报告

国际申请号

PCT/CN2021/089924

A. 主题的分类

G06K 9/00(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)

G06K G06N

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))

WPI, EPDOC, CNPAT, CNKI, IEEE: 视频, 三维, 特征, 权重, 矩阵, 并行, 计算, video, 3d, character, feature, weight, matrix, parallel, compute

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
PX	CN 112016522 A (苏州浪潮智能科技有限公司) 2020年 12月 1日 (2020 - 12 - 01) 权利要求1-12	1-12
Y	CN 110785778 A (深圳市大疆创新科技有限公司) 2020年 2月 11日 (2020 - 02 - 11) 说明书第[0024]-[0188]段、附图1-23	1-5、10-12
Y	CN 111695672 A (百度美国有限责任公司) 2020年 9月 22日 (2020 - 09 - 22) 说明书第0018-0080段、附图1-6	1-5、10-12
A	US 2019065903 A1 (NORTHROP GRUMMAN SYSTEMS CORPORATION) 2019年 2月 28日 (2019 - 02 - 28) 全文	1-12

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 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

2021年 7月 12日

国际检索报告邮寄日期

2021年 7月 26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

授权官员

鞠博

电话号码 86-(10)-53961534

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2021/089924

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	112016522	A	2020年 12月 1日	无			
CN	110785778	A	2020年 2月 11日	WO	2020034079	A1	2020年 2月 20日
CN	111695672	A	2020年 9月 22日	US	2020293866	A1	2020年 9月 17日
US	2019065903	A1	2019年 2月 28日	JP	2020532008	A	2020年 11月 5日
				EP	3673417	A1	2020年 7月 1日
				WO	2019040214	A1	2019年 2月 28日