(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 118781663 A
(43)申请公布日 2024.10.15

(21)申请号 202411264600.9

(22)申请日 2024.09.10

(71)申请人 泉州装备制造研究所
地址 362100 福建省泉州市台商投资区洛阳镇上浦村吉贝511号

(72)发明人 李琦铭 林清锋 李俊 谢银辉 吴锦滢

(74)专利代理机构 泉州三允知识产权代理事务所(普通合伙) 35265
专利代理师 安乔

(51)Int.Cl.
G06V 40/20(2022.01)
G06N 3/045(2023.01)
G06N 3/0464(2023.01)
G06V 10/764(2022.01)
G06V 10/80(2022.01)
G06V 10/82(2022.01)
G06V 20/40(2022.01)
G06V 20/52(2022.01)

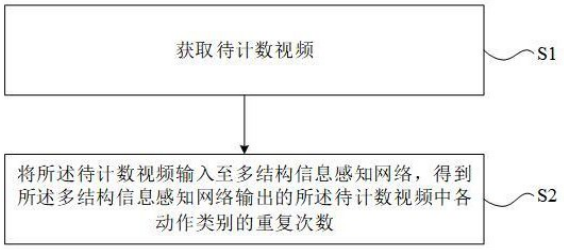
权利要求书3页 说明书15页 附图8页

(54)发明名称

基于多结构信息感知网络的重复动作计数方法及装置

(57)摘要

本发明涉及计算机视觉技术领域,提供一种基于多结构信息感知网络的重复动作计数方法及装置,采用的多结构信息感知网络包括结构信息提取模块、结构信息融合模块以及重复计数模块,通过结构信息提取模块提取待计数视频的每一视频帧中的结构信息进行辅助判别,可以提升多结构信息感知网络的性能。结构信息融合模块通过多重注意力机制自适应地捕获结构信息的相关性,通过多重卷积操作对结构信息的局部特征进行挖掘,关注局部细节变化,二者互补,保证各动作类别的重复次数的准确性。重复计数模块通过应用各动作类别对应的阈值,可以实现对待计数视频中的重复动作的准确计数。



CN 118781663 A

1.一种基于多结构信息感知网络的重复动作计数方法,其特征在于,包括:

获取待计数视频;

将所述待计数视频输入至多结构信息感知网络,得到所述多结构信息感知网络输出的所述待计数视频中各动作类别的重复次数;

其中,所述多结构信息感知网络包括结构信息提取模块、结构信息融合模块以及重复计数模块;

所述结构信息提取模块用于提取所述待计数视频的每一视频帧中的结构信息;所述结构信息包括各关节点的位置信息、指定关节点的角度信息和目标关节点对之间的距离信息;

所述结构信息融合模块用于基于多重注意力机制以及多重卷积操作,对所述结构信息进行融合,得到融合特征,并基于所述融合特征,得到所述待计数视频的每一视频帧中各动作类别的得分;

所述重复计数模块用于基于所述待计数视频的各视频帧中各动作类别的得分,应用所述各动作类别对应的阈值,对所述待计数视频中的重复动作进行计数。

2.根据权利要求1所述的基于多结构信息感知网络的重复动作计数方法,其特征在于,所述结构信息提取模块具体用于:

基于人体姿态追踪算法,提取所述待计数视频的每一视频帧中所述各关节点的位置信息;

基于所述待计数视频的每一视频帧中所述指定关节点及其相邻关节点的位置信息,确定所述待计数视频的每一视频帧中所述指定关节点的角度信息;

基于所述待计数视频的每一视频帧中所述目标关节点对的位置信息,确定所述待计数视频的每一视频帧中所述目标关节点对之间的距离信息。

3.根据权利要求1所述的基于多结构信息感知网络的重复动作计数方法,其特征在于,所述结构信息融合模块包括信息融合与嵌入模块、多重注意力模块、结构特征挖掘模块以及特征映射模块;

所述信息融合与嵌入模块用于将所述待计数视频的每一视频帧中的结构信息进行拼接,得到拼接结果,并将所述拼接结果嵌入至所述多重注意力模块的特征空间内,得到嵌入特征;

所述多重注意力模块用于基于多重注意力机制,将所述嵌入特征在全局上建立长距离依赖关系,得到全局特征;

所述结构特征挖掘模块用于基于多个卷积模块,对所述全局特征进行局部特征挖掘融合,得到所述融合特征;

所述特征映射模块用于基于全连接层,对所述融合特征进行分类预测,得到所述待计数视频的每一视频帧中各动作类别的得分。

4.根据权利要求3所述的基于多结构信息感知网络的重复动作计数方法,其特征在于,所述结构特征挖掘模块具体包括依次连接的第一卷积模块、第一拼接层、第二卷积模块、第二拼接层、第三卷积模块、第三拼接层、第四拼接层以及第四卷积模块;

所述第一卷积模块的输入端用于与所述多重注意力模块的输出端连接;所述第二拼接层的输入端还用于与所述第一拼接层的输出端连接;所述第三拼接层的输入端还用于与所

述多重注意力模块的输出端连接。

5.根据权利要求3所述的基于多结构信息感知网络的重复动作计数方法,其特征在于,所述多重注意力模块包括依次连接的第一注意力模块、第一叠加层、第二注意力模块、第二叠加层、第三注意力模块以及第三叠加层;

所述第一注意力模块的输入端、所述第一叠加层的输入端以及所述第三叠加层的输入端均用于与所述信息融合与嵌入模块的输出端连接;所述第二叠加层的输入端还用于与所述第一叠加层的输出端连接。

6.根据权利要求1所述的基于多结构信息感知网络的重复动作计数方法,其特征在于,所述重复计数模块具体用于:

对于任一动作类别,基于所述待计数视频的各视频帧中所述任一动作类别的得分,应用所述任一动作类别对应的第一阈值和第二阈值,确定所述第一阈值和所述第二阈值按顺序连续触发的次数,并将所述次数作为所述任一动作类别的重复次数。

7.根据权利要求1-6中任一项所述的基于多结构信息感知网络的重复动作计数方法,其特征在于,所述多结构信息感知网络基于如下步骤训练得到:

将视频样本中各视频帧样本输入至初始感知网络,得到所述初始感知网络中的结构信息融合模块输出的所述各视频帧样本的样本特征以及所述初始感知网络输出的所述各视频帧样本的样本动作类别;所述各视频帧样本包括锚点样本、正样本和负样本;

基于所述样本特征,计算所述锚点样本与所述正样本之间的第一特征距离以及所述锚点样本与所述负样本之间的第二特征距离,并基于所述第一特征距离以及所述第二特征距离,计算三重边界损失;

基于所述样本动作类别以及所述各视频帧样本携带的动作类别标签,计算二元交叉熵损失;

基于所述三重边界损失以及所述二元交叉熵损失,计算综合损失,并基于所述综合损失,对所述初始感知网络的结构参数进行迭代优化,得到所述多结构信息感知网络。

8.一种基于多结构信息感知网络的重复动作计数装置,其特征在于,包括:

视频获取模块,用于获取待计数视频;

重复动作计数模块,用于将所述待计数视频输入至多结构信息感知网络,得到所述多结构信息感知网络输出的所述待计数视频中各动作类别的重复次数;

其中,所述多结构信息感知网络包括结构信息提取模块、结构信息融合模块以及重复计数模块;

所述结构信息提取模块用于提取所述待计数视频的每一视频帧中的结构信息;所述结构信息包括各关节点的位置信息、指定关节点的角度信息和目标关节点对之间的距离信息;

所述结构信息融合模块用于基于多重注意力机制以及多重卷积操作,对所述结构信息进行融合,得到融合特征,并基于所述融合特征,得到所述待计数视频的每一视频帧中各动作类别的得分;

所述重复计数模块用于基于所述待计数视频的各视频帧中各动作类别的得分,应用所述各动作类别对应的阈值,对所述待计数视频中的重复动作进行计数。

9.一种电子设备,包括存储器、处理器及存储在所述存储器上并可在所述处理器上运

行的计算机程序,其特征在于,所述处理器执行所述程序时实现如权利要求1-7中任一项所述的基于多结构信息感知网络的重复动作计数方法。

10.一种非暂态计算机可读存储介质,其上存储有计算机程序,其特征在于,所述计算机程序被处理器执行时实现如权利要求1-7中任一项所述的基于多结构信息感知网络的重复动作计数方法。

基于多结构信息感知网络的重复动作计数方法及装置

技术领域

[0001] 本发明涉及计算机视觉技术领域,尤其涉及一种基于多结构信息感知网络的重复动作计数方法及装置。

背景技术

[0002] 随着人工智能技术的日益发展,视频分析领域也迎来了一场革命,其中之一是在视频重复动作计数方面的应用,重复动作计数是一种利用视频捕捉技术来计算特定动作重复次数的技术,该技术在评估运动员训练效果以及监测和判断其身体状况方面具有巨大潜力,同时也可以用于健身领域,帮助个人追踪进度并衡量他们的健身强度。

[0003] 现有的重复动作计数方法主要分为两类:传统方法和基于计算机视觉的方法。

[0004] 传统方法主要包括人工计数和传感器辅助计数。人工计数需要有专门的记录员,这种方法耗费人力,并且对于某些频率较快的动作进行精准计数往往难度较大,存在因反应延迟导致的计数误差,也可能产生由于记录员疲劳而导致计数错误的情况。传感器辅助计数方法一般通过在运动场地安装红外线传感器、压力传感器等,或者是让运动人员佩戴相应的传感器,然后对传感器的数据信息进行分析,进而实现重复动作计数,这种方法虽然准确度高,但是设备搭载复杂,不同的动作所使用的传感器也不尽相同,布置成本较高,此外佩戴传感器极有可能影响发挥或者造成安全事故。

[0005] 基于计算机视觉的方法可以克服基于传统方法的效率低及接触特性等缺点。该类方法以数据驱动的方式通过上下文感知或时间相关性建模解决上述问题,从而在通用场景中实现重复计数。然而,其计数精度远远不能满足体能测试场景下的实际应用需求。现有方法通过将视频的每一帧作为一个整体来关注全局空间信息,缺乏判别存在周期性运动的局部区域特征的能力,从而难以识别细粒度的局部周期性运动,进而导致重复计数误差大。

发明内容

[0006] 本发明提供一种基于多结构信息感知网络的重复动作计数方法及装置,用以解决现有技术中存在的缺陷。

[0007] 本发明提供一种基于多结构信息感知网络的重复动作计数方法,包括:

获取待计数视频;

将所述待计数视频输入至多结构信息感知网络,得到所述多结构信息感知网络输出的所述待计数视频中各动作类别的重复次数;

其中,所述多结构信息感知网络包括结构信息提取模块、结构信息融合模块以及重复计数模块;

所述结构信息提取模块用于提取所述待计数视频的每一视频帧中的结构信息;所述结构信息包括各关节点的位置信息、指定关节点的角度信息和目标关节点对之间的距离信息;

所述结构信息融合模块用于基于多重注意力机制以及多重卷积操作,对所述结构

信息进行融合,得到融管脚合特征,并基于所述融合特征,得到所述待计数视频的每一视频帧中各动作类别的得分;

所述重复计数模块用于基于所述待计数视频的各视频帧中各动作类别的得分,应用所述各动作类别对应的阈值,对所述待计数视频中的重复动作进行计数。

[0008] 本发明还提供一种基于多结构信息感知网络的重复动作计数装置,包括:

视频获取模块,用于获取待计数视频;

重复动作计数模块,用于将所述待计数视频输入至多结构信息感知网络,得到所述多结构信息感知网络输出的所述待计数视频中各动作类别的重复次数;

其中,所述多结构信息感知网络包括结构信息提取模块、结构信息融合模块以及重复计数模块;

所述结构信息提取模块用于提取所述待计数视频的每一视频帧中的结构信息;所述结构信息包括各关节点的位置信息、指定关节点的角度信息和目标关节点对之间的距离信息;

所述结构信息融合模块用于基于多重注意力机制以及多重卷积操作,对所述结构信息进行融合,得到融合特征,并基于所述融合特征,得到所述待计数视频的每一视频帧中各动作类别的得分;

所述重复计数模块用于基于所述待计数视频的各视频帧中各动作类别的得分,应用所述各动作类别对应的阈值,对所述待计数视频中的重复动作进行计数。

[0009] 本发明还提供一种电子设备,包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序,所述处理器执行所述程序时实现如上述任一种所述的XXXX方法。

[0010] 本发明还提供一种非暂态计算机可读存储介质,其上存储有计算机程序,该计算机程序被处理器执行时实现如上述任一种所述的基于多结构信息感知网络的重复动作计数方法。

[0011] 本发明还提供一种计算机程序产品,包括计算机程序,所述计算机程序被处理器执行时实现如上述任一种所述的基于多结构信息感知网络的重复动作计数方法。

[0012] 与现有技术相比,本发明具有如下有益效果:

本发明提供的基于多结构信息感知网络的重复动作计数方法及装置,首先获取待计数视频;然后将待计数视频输入至多结构信息感知网络,得到多结构信息感知网络输出的待计数视频中各动作类别的重复次数。该多结构信息感知网络包括结构信息提取模块、结构信息融合模块以及重复计数模块,通过结构信息提取模块提取待计数视频的每一视频帧中的结构信息进行辅助判别,可以提升多结构信息感知网络的性能。结构信息融合模块通过多重注意力机制自适应地捕获结构信息的相关性,通过多重卷积操作对结构信息的局部特征进行挖掘,关注局部细节变化,二者互补,保证各动作类别的重复次数的准确性。重复计数模块通过应用各动作类别对应的阈值,可以实现对待计数视频中的重复动作的准确计数。

附图说明

[0013] 为了更清楚地说明本发明或现有技术中的技术方案,下面将对实施例或现有技术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,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。

[0014] 图1是本发明提供的基于多结构信息感知网络的重复动作计数方法的流程示意图;

图2是本发明提供的基于多结构信息感知网络的重复动作计数方法中多结构信息感知网络的结构示意图;

图3是本发明提供的基于多结构信息感知网络的重复动作计数方法中待计数视频的每一视频帧中各关节点的示意图;

图4是本发明提供的基于多结构信息感知网络的重复动作计数方法中待计数视频的每一视频帧中指定关节点的角度信息示意图;

图5是本发明提供的基于多结构信息感知网络的重复动作计数方法中待计数视频的每一视频帧中目标关节点对之间的距离信息示意图;

图6是本发明提供的基于多结构信息感知网络的重复动作计数方法中多结构信息感知网络的多重注意力模块的结构示意图;

图7是本发明提供的基于多结构信息感知网络的重复动作计数方法中多结构信息感知网络的每个注意力模块的结构示意图;

图8是本发明提供的基于多结构信息感知网络的重复动作计数方法中多结构信息感知网络的结构特征挖掘模块的结构示意图;

图9是本发明提供的基于多结构信息感知网络的重复动作计数方法中各视频帧中各动作类别的得分示意图;

图10是本发明提供的基于多结构信息感知网络的重复动作计数方法中多结构信息感知网络的重复计数模块的结构示意图;

图11是本发明提供的基于多结构信息感知网络的重复动作计数装置的结构示意图;

图12是本发明提供的电子设备的结构示意图。

具体实施方式

[0015] 为使本发明的目的、技术方案和优点更加清楚,下面将结合本发明中的附图,对本发明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发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,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,都属于本发明保护的范围。

[0016] 图1为本发明实施例中提供的一种基于多结构信息感知网络的重复动作计数方法的流程示意图,如图1所示,该方法包括:

S1,获取待计数视频;

S2,将所述待计数视频输入至多结构信息感知网络,得到所述多结构信息感知网络输出的所述待计数视频中各动作类别的重复次数;

其中,所述多结构信息感知网络包括结构信息提取模块、结构信息融合模块以及重复计数模块;

所述结构信息提取模块用于提取所述待计数视频的每一视频帧中的结构信息;所述结构信息包括各关节点的位置信息、指定关节点的角度信息和目标关节点对之间的距离

信息;

所述结构信息融合模块(MIF-Module)用于基于多重注意力机制以及多重卷积操作,对所述结构信息进行融合,得到融合特征,并基于所述融合特征,得到所述待计数视频的每一视频帧中各动作类别的得分;

所述重复计数模块用于基于所述待计数视频的各视频帧中各动作类别的得分,应用所述各动作类别对应的阈值,对所述待计数视频中的重复动作进行计数。

[0017] 具体地,本发明实施例中提供的基于多结构信息感知网络的重复动作计数方法,其执行主体为基于多结构信息感知网络的重复动作计数装置,该装置可以配置于计算机内,该计算机可以为本地计算机或云计算机,本地计算机可以是电脑、平板等,此处不作具体限定。

[0018] 首先执行步骤S1,获取待计数视频,该待计数视频可以包括多个视频帧,每个视频帧中可以包含有一个或多个动作类别。

[0019] 然后执行步骤S2,将待计数视频输入至多结构信息感知网络(MIA-Net),由多结构信息感知网络输出的待计数视频中各动作类别的重复次数。

[0020] 如图2所示,多结构信息感知网络可以包括顺次连接的结构信息提取模块、结构信息融合模块以及重复计数模块。

[0021] 结构信息提取模块用于提取待计数视频的每一视频帧中的结构信息;结构信息包括各关节点的位置信息、指定关节点的角度信息和目标关节点对之间的距离信息。

[0022] 结构信息提取模块可以通过人体姿态追踪算法提取待计数视频的每一视频帧中各关节点的位置信息,各关节点如图3所示,可以包括头部关节、肩部关节、肘部关节、臀部关节、膝部关节和脚部关节。

[0023] 该人体姿态追踪算法可以是Blazepose算法。即有:

$$V = \{f_i\}^T \in R^{C\times H\times W\times T};$$

$$V \to P = \{p_i\}^T \in R^{D\times K\times T};$$

其中,V表示待计数视频,P表示待计数视频的关节点信息,关节点即显著部分特征点。f_i 表示待计数视频中的第i个视频帧,C 表示通道数,通常为三个通道,H 表示高度,W 表示宽度,T 表示帧数。p_i 表示第i个视频帧中的关节点信息。为了表示每个视频帧中的关节点信息,使用序列 $D\times K\times T$ 来表示。D 表示每个关节点的维度,一般是三维的,分别是二维的位置信息和一维的深度信息,K 表示关节点的个数。

[0024] 由于不同的动作类别会导致运动过程中关节点之间的角度信息和距离信息的变化不同,角度信息和距离信息等隐藏的结构信息有助于区分不同的动作类别。如图4所示,根据真实场景,左侧肘部角度θ11、左侧肩部角度θ12、左侧臀部角度θ13、左侧膝部角度θ14、右侧肘部角度θ21、右侧肩部角度θ22、右侧臀部角度θ23以及右侧膝部角度θ24的变化在大多数动作中变化最为明显,并且这些角度的变化对于具体动作具有很强的代表性。因此,结构信息提取模块可以计算指定关节点的角度信息作为辅助的结构信息。此处,指定关节点可以包括各关节点,也可以包括肘部关节点、肩部关节点、臀部关节点以及膝部关节点。

[0025] 指定关节点的角度信息可以通过待计数视频的每一视频帧中指定关节点及其相邻关节点的位置信息确定,计算指定关节点的角度信息的数学表达式包括:

$$\overrightarrow{BA} = (x_i - x_j, y_i - y_j, z_i - z_j);$$

$$\overrightarrow{BC} = (x_k - x_j, y_k - y_j, z_k - z_j);$$

$$\| \overrightarrow{BA} \| = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2};$$

$$\| \overrightarrow{BC} \| = \sqrt{(x_k - x_j)^2 + (y_k - y_j)^2 + (z_k - z_j)^2};$$

$$\theta = \arccos\left(\frac{\overrightarrow{BA} \cdot \overrightarrow{BC}}{\|\overrightarrow{BA}\| \|\overrightarrow{BC}\|}\right);$$

其中,B表示指定关节点,A、C分别为指定关节点B的相邻关节点,(x_i, y_i, z_i)为A的位置信息,(x_j, y_j, z_j)为B的位置信息,(x_k, y_k, z_k)为C的位置信息,$\overrightarrow{BA}$为A与B的连线矢量,$\overrightarrow{BC}$为B与C的连线矢量,θ为B的角度信息。

[0026] 如图5所示,在大多数动作中,左侧腕部关节点和肩部关节点之间的距离信息D_1、右侧腕部关节点和肩部关节点之间的距离信息D_2,左侧腕部关节点和臀部关节点之间的距离信息D_3、右侧腕部关节点和臀部关节点之间的距离信息D_4,左侧肩部关节点与膝部关节点之间的距离信息D_5、右侧肩部关节点与膝部关节点之间的距离信息D_6,左侧头部关节点和膝部关节点之间的距离信息D_7和右侧头部关节点和膝部关节点之间的距离信息D_8,这些关节点之间的距离信息对于重复动作的判断也起着关键作用。

[0027] 因此,结构信息提取模块将目标关节点对之间的距离信息作为另外的结构信息,其目的是增强并辅助坐标信息进行重复动作的判别。目标关节点对可以包括腕部关节点和肩部关节点、腕部关节点和臀部关节点、肩部关节点与膝部关节点、头部关节点和膝部关节点。

[0028] 目标关节点对之间的距离信息可以通过目标关节点对中各目标关节点的位置信息确定,计算目标关节点对之间的距离信息的数学表达式可以表示为:

$$Distance(E,F) = \sqrt{(x_m - x_n)^2 + (y_m - y_n)^2 + (z_m - z_n)^2}。$$

[0029] 其中,$Distance(E,F)$为目标关节点E与F构成的目标关节点对之间的距离信息,(x_m, y_m, z_m)为E的位置信息,(x_n, y_n, z_n)为F的位置信息。

[0030] 基于此,待计数视频的每一视频帧中各关节点的位置信息集合 $P_{coordinate}$ 、待计数视频的每一视频帧中指定关节点的角度信息集合 P_{angle} 以及待计数视频的每一视频帧中目标关节点对之间的距离信息集合 $P_{distance}$ 可以表示为:

$$P_{coordinate} = \{p_1, p_2, \ldots, p_n\}^T ;$$

$$P_{angle} = \{a_1, a_2, \ldots, a_m\}^T ;$$

$$P_{distance} = \{d_1, d_2, \ldots, d_t\}^T ;$$

其中, p_n , a_m , $d_t \in R^{D\times K}$ 分别表示待计数视频中某一视频帧中第n个关节点的位置信息、第m个指定关节点的角度信息、第t个目标关节点对之间的距离信息。 n , m , t 分别代表待计数视频中某一视频帧中关节点的个数、指定关节点的个数和目标关节点对的个数。

[0031] 结构信息融合模块可以利用多重注意力机制以及多重卷积操作,对结构信息进行融合,得到融合特征。通过多重注意力机制可以对结构信息在全局上建立长距离依赖关系,通过多重卷积操作可以捕捉结构信息中的局部细微变化,进而得到融合特征。

[0032] 结构信息融合模块还可以借助于全连接层,建立融合特征与动作类别之间的映射关系,将融合特征输入至全连接(Fully Connected,FC)层,得到待计数视频的每一视频帧中各动作类别的得分。

[0033] 重复计数模块可以利用待计数视频的各视频帧中各动作类别的得分,通过动作触发器,应用各动作类别对应的阈值,对待计数视频中的重复动作进行计数。其中,每个动作类别均可对应有进入阈值和退出阈值,进入阈值为动作类别的开始姿态的得分,退出阈值为动作类别的结束姿态的得分。

[0034] 当某个动作类别对应的进入阈值与退出阈值连续触发,则说明该动作类别产生一次。进而,可以统计对待计数视频中各动作类别对应的进入阈值与退出阈值的连续触发次数,得到各动作类别的重复次数。

[0035] 本发明实施例中提供的基于多结构信息感知网络的重复动作计数方法,首先获取待计数视频;然后将待计数视频输入至多结构信息感知网络,得到多结构信息感知网络输出的待计数视频中各动作类别的重复次数。该多结构信息感知网络包括结构信息提取模块、结构信息融合模块以及重复计数模块,通过结构信息提取模块提取待计数视频的每一视频帧中的结构信息进行辅助判别,可以提升多结构信息感知网络的性能。结构信息融合模块通过多重注意力机制自适应地捕获结构信息的相关性,通过多重卷积操作对结构信息的局部特征进行挖掘,关注局部细节变化,二者互补,保证各动作类别的重复次数的准确性。重复计数模块通过应用各动作类别对应的阈值,可以实现对待计数视频中的重复动作的准确计数。

[0036] 在上述实施例的基础上,所述结构信息融合模块包括信息融合与嵌入模块(IFE-Module)、多重注意力模块(MA-Module)、结构特征挖掘模块(SFM-Module)以及特征映射模

块;

所述信息融合与嵌入模块用于将所述待计数视频的每一视频帧中的结构信息进行拼接,得到拼接结果,并将所述拼接结果嵌入至所述多重注意力模块的特征空间内,得到嵌入特征;

所述多重注意力模块用于基于多重注意力机制,将所述嵌入特征在全局上建立长距离依赖关系,得到全局特征;

所述结构特征挖掘模块用于基于多个卷积模块,对所述全局特征进行局部特征挖掘融合,得到所述融合特征;

所述特征映射模块用于基于全连接层,对所述融合特征进行分类预测,得到所述待计数视频的每一视频帧中各动作类别的得分。

[0037] 具体地,为了获得多重结构信息,信息融合与嵌入模块首先将每一视频帧中的结构信息进行拼接,得到拼接结果$P_{out} \in R^{D \times K'}$,其中$K'$为拼接后的信息个数。拼接结果$P_{out}$可以表示为:

$$P_{out} = Concat\left(P_{coordinate}, P_{angle}, P_{distance}\right);$$

其中,Concat为拼接操作。

[0038] 随后,信息融合与嵌入模块使用嵌入层(Embedding)将拼接结果嵌入到多重注意力模块的特征空间内,得到高维的嵌入特征$E_{out} \in R^{D' \times K'}$,其中$D'$为嵌入后每个关节点的维度个数。此处,嵌入层可以包括批归一化层(Batch Norm,BN)和两个线性模块(LBR),该线性模块包括线性层(Linear)、批归一化层和激活层(ReLU)。

[0039] 嵌入特征E_{out}可以表示为:

$$E_{out} = LBR\left(LBR\left(BN\left(P_{out}\right)\right)\right)。$$

[0040] 嵌入特征E_{out}被传递到多重注意力模块中,通过多重注意力模块中自适应地更新每个特征的权重,学习每个特征之间的相关性,从而生成具有更高代表性和更多关键信息的全局特征。为了在不增加网络参数量的情况下减少信息的损失和学习更加复杂的特征表示,该模块通过残差结构将注意力模块的输出特征进行连接。

[0041] 如图6所示,多重注意力模块包括依次连接的第一注意力模块、第一叠加层、第二注意力模块、第二叠加层、第三注意力模块以及第三叠加层,第一注意力模块的输入端、第一叠加层的输入端以及第三叠加层的输入端均用于与信息融合与嵌入模块的输出端连接,用于输入嵌入特征E_{out};第二叠加层的输入端还用于与第一叠加层的输出端连接。最终,由第三叠加层的输出端输出全局特征$MA_{out} \in R^{D' \times K'}$:

$$MA_{out_1} = Attention(E_{out}) + E_{out};$$

$$MA_{out_2} = Attention\left(E_{out_1}\right) + MA_{out_1};$$

$$MA_{out} = Attention(E_{out_2}) + E_{out}$$;

其中，MA_{out_1} 为第一注意力模块的输出，MA_{out_2} 为第二注意力模块的输出，$Attention$ 为注意力机制操作。

[0042] 可以理解的是，如图7所示，每个注意力模块均可以将输入通过三个线性层生成查询向量W_Q、键向量W_K和值向量W_V，通过将键向量W_K与值向量W_V进行相乘，并将乘积结果通过归一化层(Softmax)得到归一化特征，通过将查询向量W_Q与归一化特征进行相乘，并将乘积结果与输入进行两次叠加，得到输出。

[0043] 为了更有效地从具有代表性的全局特征 MA_{out} 中提取最显著的局部特征，结构特征挖掘模块使用全局特征 MA_{out} 作为输入，并通过多个卷积模块进行新一轮的特征提取及拼接，目的是将多方面的信息进行融合，获得更多的显著细节信息，以提高显著特征的表示，得到融合特征 SFM_{out}。

[0044] 此处，如图8所示，结构特征挖掘模块具体包括依次连接的第一卷积模块、第一拼接层(Concat1)、第二卷积模块、第二拼接层(Concat2)、第三卷积模块、第三拼接层(Concat3)、第四拼接层(Concat4)以及第四卷积模块。通过结构特征挖掘模块可以减少有效信息的损失，同时学习更显著和更丰富的高级特征，为后续的特征映射提供最具代表性的输入。

[0045] 第一卷积模块的输入端用于与多重注意力模块的输出端连接，用于输入全局特征 MA_{out}；第二拼接层的输入端还用于与第一拼接层的输出端连接；第三拼接层的输入端还用于与多重注意力模块的输出端连接，也用于输入全局特征 MA_{out}，输出融合特征 SFM_{out}。

[0046] 第一卷积模块可以包括一个卷积块，第二卷积模块可以包括2个卷积块，第三卷积模块可以包括1个卷积块，第四卷积模块可以包括4个卷积块。此处，卷积块可以包括一个1×1的卷积层(Conv)、批归一化层(BetchNorm)和激活层(ReLU)，卷积块可以通过CBR表示。通过卷积块，可以逐步降低网络的维数，保持参数参数不增加。

[0047] 基于此，结构特征挖掘模块的操作可以通过如下公式表示：

$$SFM_1 = CBR(MA_{out})$$;

$$SFM_{1_1} = Concat1(SFM_1, SFM_1)$$;

$$SFM_2 = \mathrm{CBR}\left(CBR(SFM_{1_1})\right)$$;

$$SFM_{2_1} = Concat2(SFM_2, SFM_{1_1})$$;

$$SFM_3 = CBR(SFM_{2_1})$$;

$$SFM_{3_1} = Concat3\left(MA_{out_3}, SFM_3\right);$$

$$SFM_{3_2} = Concat4\left(SFM_{3_1}, SFM_{3_1}\right)。$$

[0048] $$SFM_{out} = CBR\left(CBR\left(\mathrm{CBR}\left(CBR\left(SFM_{3_2}\right)\right)\right)\right)。$$

[0049] 其中，SFM_1 为第一卷积模块的输出，SFM_{1_1} 为第一拼接层的输出，SFM_2 为第二卷积模块的输出，SFM_{2_1} 为第二拼接层的输出，SFM_3 为第三卷积模块的输出，SFM_{3_1} 为第三拼接层的输出，$SFM_{3_2} \in \boldsymbol{R}^{\boldsymbol{D}'' \times \boldsymbol{K}'}$ 为第四拼接层的输出，D'' 为经过多次特征提取和融合后的特征数量。

[0050] 可以理解的是，第四卷积模块可以构成分类解码器，用以输出融合特征SFM_{out}。

[0051] 特征映射模块用于基于全连接层，对融合特征进行分类预测，得到待计数视频的每一视频帧中各动作类别的得分。此处，特征映射模块可以包括展平层(flatten)和全连接层，通过展平层可以将融合特征转换为一维特征，通过全连接层，可以将一维特征转换为待计数视频的每一视频帧中各动作类别的得分。即有：

$$\hat{S} = FC\left(Flatten\left(SFM_{out}\right)\right)。$$

[0052] 其中，$\hat{S} \in R^{C \times T}$ 为某一动作类别的得分。

[0053] 如图9所示，为各视频帧中各动作类别的得分示意图。图9中，每一视频帧中各动作类别的得分可以通过长方体表示，长方体越高，表示得分越高。

[0054] 本发明实施例中，通过结构信息融合模块的具体结构，确定待计数视频的每一视频帧中各动作类别的得分，可以保证得分的准确性，由于后续对各动作类别的重复次数的准确记录。

[0055] 在上述实施例的基础上，所述重复计数模块具体用于：

对于任一动作类别，基于所述待计数视频的各视频帧中所述任一动作类别的得分，应用所述任一动作类别对应的第一阈值和第二阈值，确定所述第一阈值和所述第二阈值按顺序连续触发的次数，并将所述次数作为所述任一动作类别的重复次数。

[0056] 具体地，如图10所示，对于任一动作类别，可以遍历待计数视频中的所有视频帧，确定所有视频帧中该任一动作类别的得分，应用该任一动作类别对应的第一阈值和第二阈值，确定第一阈值和第二阈值按顺序连续触发的次数，并将该次数作为该任一动作类别的重复次数。

[0057] 此处，第一阈值可以小于第二阈值，第一阈值可以是该任一动作类别的退出阈值，为该任一动作类别的结束姿态的得分，第二阈值可以是该动作类别的进入阈值，为该任一动作类别的开始姿态的得分。图10中，横坐标为视频帧的序号，纵坐标为各视频帧中该任一动作类别的得分。从图10中可以看出，该任一动作类别的重复次数为7。

[0058] 本发明实施例中，通过引入第一阈值和第二阈值确定动作类别的重复次数，可以

简化计数流程,提高计数效率。

[0059] 在上述实施例的基础上,所述多结构信息感知网络基于如下步骤训练得到:

将视频样本中各视频帧样本输入至初始感知网络,得到所述初始感知网络中的结构信息融合模块得到的所述各视频帧样本的样本特征以及所述初始感知网络输出的所述各视频帧样本的样本动作类别;所述各视频帧样本包括锚点样本、正样本和负样本;

基于所述样本特征,计算所述锚点样本与所述正样本之间的第一特征距离以及所述锚点样本与所述负样本之间的第二特征距离,并基于所述第一特征距离以及所述第二特征距离,计算三重边界损失;

基于所述样本动作类别以及所述各视频帧样本携带的动作类别标签,计算二元交叉熵损失;

基于所述三重边界损失以及所述二元交叉熵损失,计算综合损失,并基于所述综合损失,对所述初始感知网络的结构参数进行迭代优化,得到所述多结构信息感知网络。

[0060] 具体地,在多结构信息感知网络的训练过程中,可以先将视频样本中各视频帧样本输入至初始感知网络,该初始感知网络与多结构信息感知网络的结构相同,不同的是初始感知网络的结构参数是初始化的,多结构信息感知网络的结构参数是通过训练过程优化后得到的。

[0061] 初始感知网络也包括结构信息提取模块、结构信息融合模块以及重复计数模块。在将视频样本中各视频帧样本输入至初始感知网络后,得到初始感知网络中初始融合模块输出的各视频帧样本的样本特征以及初始感知网络输出的各视频帧样本的样本动作类别。

[0062] 各视频帧样本中可以包括锚点样本、正样本和负样本。其中,锚点样本是指具有真实动作类别的真实样本,正样本是指与真实样本相同动作类别的样本,负样本是指与真实样本不同动作类别的样本。

[0063] 此后,利用样本特征,可以计算锚点样本与正样本之间的第一特征距离以及锚点样本与负样本之间的第二特征距离,并利用第一特征距离以及第二特征距离,计算三重边界损失(Triplet Margin Loss) L_{tri} 。

[0064] 即有:

$$L_{tri} = max(CS(a,p) - CS(a,n) + margin, 0)$$;

其中,a为锚点样本,p为正样本,n为负样本, $CS(a,p)$ 为第一特征距离, $CS(a,n)$ 为第二特征距离, $margin$ 为定值。CS 代表余弦相似度,用来度量特征之间的相似度。

[0065] 利用样本动作类别以及各视频帧样本携带的动作类别标签,计算二元交叉熵损失(Binary Cross Entropy Loss) L_{bce} ,即有:

$$L_{bce} = -\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{C}\sum_{j=1}^{C} loss(i,j)\right)$$;

$$loss(i,j) = y_{ij} log(p_{ij}) + (1 - y_{ij}) log(1 - p_{ij})$$;

其中, N 代表批次大小,其中每一帧构成一个批次,因此N为待计数视频中视频帧

的个数，C 代表类别的数量。y_{ij} 代表第i个视频帧中第j个动作类别标签，p_{ij} 是第i个视频帧中第j个样本动作类别。

[0066] 三重边界损失 L_{tri} 可以减少锚点样本与正样本之间的第一特征距离，同时增加锚点样本与负样本之间的第二特征距离。这样的目的是可以更好地区分每个动作类别，从而提高性能。二元交叉熵损失 L_{bce} 可以对每个动作类别进行二值分类。

[0067] 最后利用三重边界损失 L_{tri} 与二元交叉熵损失 L_{bce} 的加权求和，得到综合损失 L_{total}，即有：

$$L_{total} = L_{bce} + \alpha L_{tri};$$

其中，α 为是控制三重边界损失 L_{tri} 以及二元交叉熵损失 L_{bce} 的加权因子，这确保在网络训练过程中三重边界损失 L_{tri} 以及二元交叉熵损失 L_{bce} 的相对重要性在相同值得范围内。

[0068] 利用综合损失 L_{total} ，对初始感知网络的结构参数进行迭代优化，直至综合损失收敛或达到预设迭代次数，得到多结构信息感知网络。

[0069] 在训练过程中，本发明实施例中使用PyTorch-Lightning框架来训练多结构信息感知网络。该PyTorch-Lightning框架在正式开始训练之前执行一个训练步骤，监控批处理中综合损失的变化，以便自动选择初始最优学习率。此外，每一轮遍历(即epoch)完成后，会进行一次验证，如果连续6个epoch验证综合损失没有减少，则会自动调整学习率。此外，本发明实施例中，将优化器设置为Adam，并在NVIDIA PCIe A100 GPU上使用Triplet Margin Loss和BCELoss训练整体架构。

[0070] 与传统的视频级方法相比，姿态级方法关注人体关节坐标的变化，显著提高了性能。然而，姿态级方法忽略了在运动过程中存在于人体关节之间的隐藏结构信息。因此，本发明实施例中，将在姿态级方法的基础上，进一步研究在运动过程中对人体姿态变化具有显著性和可区分性的结构信息，并用这些重要的结构信息来补充简单的坐标信息。

[0071] 评估模型性能的主要指标是平均绝对误差(Mean Absolute Error，MAE)和离一误差(OBO)。MAE表示模型预测与实际情况之间的平均绝对误差。另一方面，OBO被定义为如果网络的预测值与真实值相差不超过1(通常小于或等于1)，则认为网络计数的预测值是正确的。它们可以定义如下：

$$MAE = \frac{1}{M}\sum_{i=1}^{M}\frac{|\tilde{c}_i - c_i|}{\tilde{c}_i};$$

$$OBO = \frac{1}{M}\sum_{i=1}^{M}[|\tilde{c}_i - c_i| \le 1];$$

其中，$\tilde{c}_i$ 为真实值，c_i 为预测值，N为视频样本的个数。

[0072] 如表1所示,本发明实施例中提供的多结构信息感知网络MIA-Net与RepCount-pose数据集上的一些常规方法进行了比较,最佳结果是MAE为0.203、OBO为0.592。与常规的视频级方法相比,MIA-Net将MAE降低了18.1%,将OBO提高了20.6%。

[0073] 此外,与姿态关键点的最新方法PoseRAC相比,MIA-Net将MAE降低了3.3%,将OBO提高了3.2%。实验结果表明,MIA-Net有效地学习了多结构信息之间的关系,并在融合特征与动作类别之间建立了良好的映射关系,提高了MIA-Net的性能。

[0074] 表1 MIA-Net算法与现有算法在RepCount-pose数据集上两个关键客观指标的对比

类型	方法	MAE↓	OBO↑
视频级方法	RepNet	0.995	0.013
	X3D	0.911	0.106
	Zhang et al.	0.879	0.155
	TANet	0.662	0.099
	Video Swin Transformer	0.576	0.132
	Huang et al.	0.527	0.159
	TransRAC	0.443	0.291
姿态关键点	PoseRAC	0.236	0.560
	MIA-Net	0.203	0.592

[0075] 如图11所示,在上述实施例的基础上,本发明实施例中提供了一种基于多结构信息感知网络的重复动作计数装置,包括:

视频获取模块111,用于获取待计数视频;

重复动作计数模块112,用于将所述待计数视频输入至多结构信息感知网络,得到所述多结构信息感知网络输出的所述待计数视频中各动作类别的重复次数;

其中,所述多结构信息感知网络包括结构信息提取模块、结构信息融合模块以及重复计数模块;

所述结构信息提取模块用于提取所述待计数视频的每一视频帧中的结构信息;所述结构信息包括各关节点的位置信息、指定关节点的角度信息和目标关节点对之间的距离信息;

所述结构信息融合模块用于基于多重注意力机制以及多重卷积操作,对所述结构信息进行融合,得到融合特征,并基于所述融合特征,得到所述待计数视频的每一视频帧中各动作类别的得分;

所述重复计数模块用于基于所述待计数视频的各视频帧中各动作类别的得分,应用所述各动作类别对应的阈值,对所述待计数视频中的重复动作进行计数。

[0076] 在上述实施例的基础上,所述结构信息提取模块具体用于:

基于人体姿态追踪算法,提取所述待计数视频的每一视频帧中所述各关节点的位置信息;

基于所述待计数视频的每一视频帧中所述指定关节点及其相邻关节点的位置信息,确定所述待计数视频的每一视频帧中所述指定关节点的角度信息;

基于所述待计数视频的每一视频帧中所述目标关节点对的位置信息,确定所述待计数视频的每一视频帧中所述目标关节点对之间的距离信息。

[0077] 在上述实施例的基础上,所述结构信息融合模块包括信息融合与嵌入模块、多重注意力模块、结构特征挖掘模块以及特征映射模块;

所述信息融合与嵌入模块用于将所述待计数视频的每一视频帧中的结构信息进行拼接,得到拼接结果,并将所述拼接结果嵌入至所述多重注意力模块的特征空间内,得到嵌入特征;

所述多重注意力模块用于基于多重注意力机制,将所述嵌入特征在全局上建立长距离依赖关系,得到全局特征;

所述结构特征挖掘模块用于基于多个卷积模块,对所述全局特征进行局部特征挖掘融合,得到所述融合特征;

所述特征映射模块用于基于全连接层,对所述融合特征进行分类预测,得到所述待计数视频的每一视频帧中各动作类别的得分。

[0078] 在上述实施例的基础上,所述结构特征挖掘模块具体包括依次连接的第一卷积模块、第一拼接层、第二卷积模块、第二拼接层、第三卷积模块、第三拼接层、第四拼接层以及第四卷积模块;

所述第一卷积模块的输入端用于与所述多重注意力模块的输出端连接;所述第二拼接层的输入端还用于与所述第一拼接层的输出端连接;所述第三拼接层的输入端还用于与所述多重注意力模块的输出端连接。

[0079] 在上述实施例的基础上,所述多重注意力模块包括依次连接的第一注意力模块、第一叠加层、第二注意力模块、第二叠加层、第三注意力模块以及第三叠加层;

所述第一注意力模块的输入端、所述第一叠加层的输入端以及所述第三叠加层的输入端均用于与所述信息融合与嵌入模块的输出端连接;所述第二叠加层的输入端还用于与所述第一叠加层的输出端连接。

[0080] 在上述实施例的基础上,所述重复计数模块具体用于:

对于任一动作类别,基于所述待计数视频的各视频帧中所述任一动作类别的得分,应用所述任一动作类别对应的第一阈值和第二阈值,确定所述第一阈值和所述第二阈值按顺序连续触发的次数,并将所述次数作为所述任一动作类别的重复次数。

[0081] 在上述实施例的基础上,还包括训练模块,用于:

将视频样本中各视频帧样本输入至初始感知网络,得到所述初始感知网络中的结构信息融合模块输出的所述各视频帧样本的样本特征以及所述初始感知网络输出的所述各视频帧样本的样本动作类别;所述各视频帧样本包括锚点样本、正样本和负样本;

基于所述样本特征,计算所述锚点样本与所述正样本之间的第一特征距离以及所述锚点样本与所述负样本之间的第二特征距离,并基于所述第一特征距离以及所述第二特征距离,计算三重边界损失;

基于所述样本动作类别以及所述各视频帧样本携带的动作类别标签,计算二元交叉熵损失;

基于所述三重边界损失以及所述二元交叉熵损失,计算综合损失,并基于所述综合损失,对所述初始感知网络的结构参数进行迭代优化,得到所述多结构信息感知网络。

[0082] 具体地,本发明实施例中提供的基于多结构信息感知网络的重复动作计数中各模块的作用与上述方法类实施例中各步骤的操作流程是一一对应的,实现的效果也是一致的,具体参见上述实施例,本发明实施例中对此不再赘述。

[0083] 图12示例了一种电子设备的实体结构示意图,如图12所示,该电子设备可以包括:处理器(Processor)121、通信接口(Communications Interface)122、存储器(Memory)123和通信总线124,其中,处理器121,通信接口122,存储器123通过通信总线124完成相互间的通信。处理器121可以调用存储器123中的逻辑指令,以执行上述各实施例中提供的基于多结构信息感知网络的重复动作计数方法。

[0084] 此外,上述的存储器123中的逻辑指令可以通过软件功能单元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介质中。基于这样的理解,本发明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介质中,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)执行本发明各个实施例所述方法的全部或部分步骤。而前述的存储介质包括:U盘、移动硬盘、只读存储器(Read-Only Memory,ROM)、随机存取存储器(Random Access Memory,RAM)、磁碟或者光盘等各种可以存储程序代码的介质。

[0085] 另一方面,本发明还提供一种计算机程序产品,所述计算机程序产品包括计算机程序,计算机程序可存储在非暂态计算机可读存储介质上,所述计算机程序被处理器执行时,计算机能够执行上述各实施例中提供的基于多结构信息感知网络的重复动作计数方法。

[0086] 又一方面,本发明还提供一种非暂态计算机可读存储介质,其上存储有计算机程序,该计算机程序被处理器执行时实现以执行上述各实施例中提供的基于多结构信息感知网络的重复动作计数方法。

[0087] 以上所描述的装置实施例仅仅是示意性的,其中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模块来实现本实施例方案的目的。本领域普通技术人员在不付出创造性的劳动的情况下,即可以理解并实施。

[0088] 通过以上的实施方式的描述,本领域的技术人员可以清楚地了解到各实施方式可借助软件加必需的通用硬件平台的方式来实现,当然也可以通过硬件。基于这样的理解,上述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计算机可读存储介质中,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)执行各个实施例或者实施例的某些部分所述的方法。

[0089] 最后应说明的是:以上实施例仅用以说明本发明的技术方案,而非对其限制;尽管参照前述实施例对本发明进行了详细的说明,本领域的普通技术人员应当理解:其依然可以对前述各实施例所记载的技术方案进行修改,或者对其中部分技术特征进行等同替换;

而这些修改或者替换,并不使相应技术方案的本质脱离本发明各实施例技术方案的精神和范围。

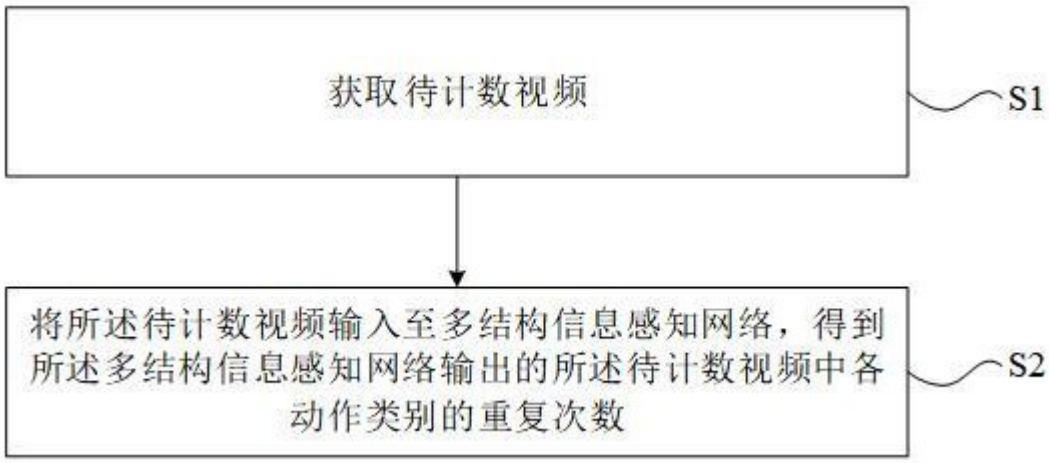

获取待计数视频
S1
将所述待计数视频输入至多结构信息感知网络，得到所述多结构信息感知网络输出的所述待计数视频中各动作类别的重复次数
S2

图 1

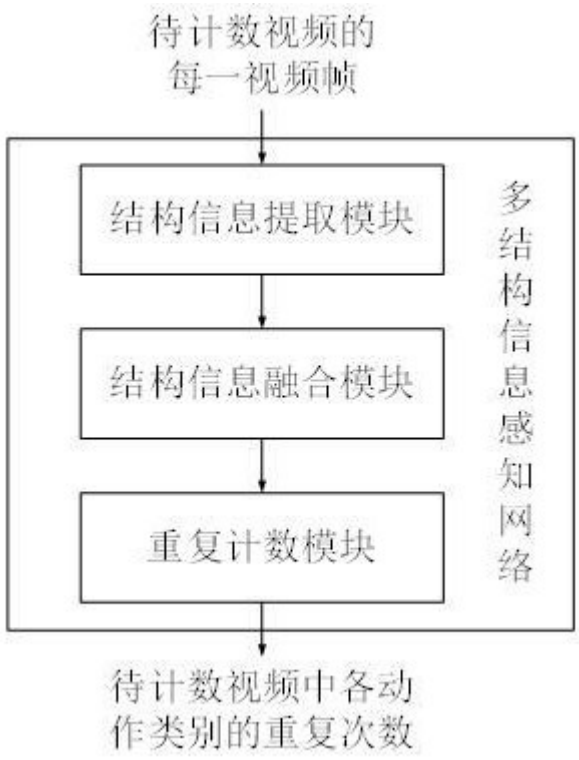

待计数视频的每一视频帧
结构信息提取模块
结构信息融合模块
重复计数模块
多结构信息感知网络
待计数视频中各动作类别的重复次数

图 2

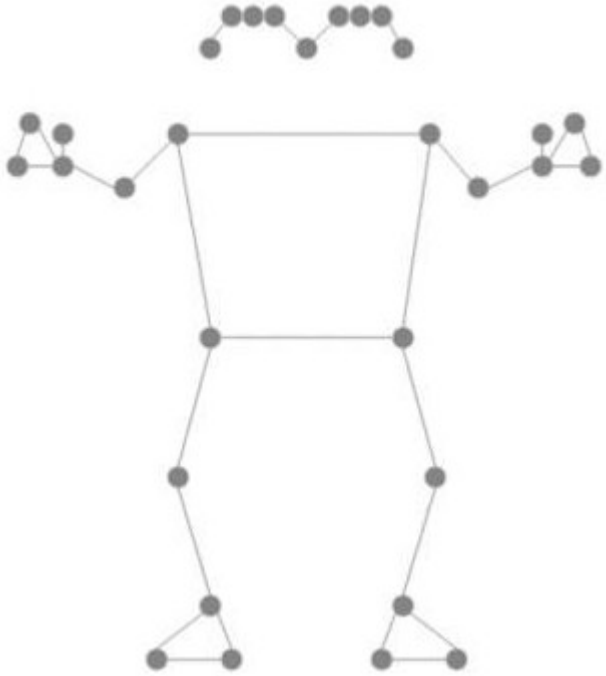


图 3

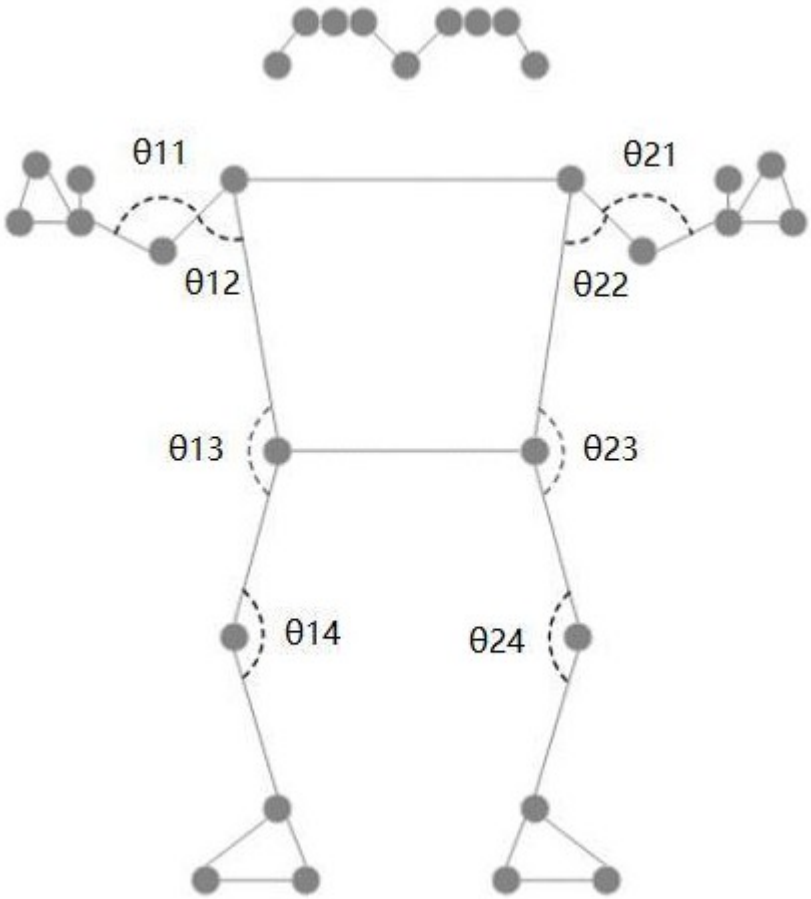


图 4

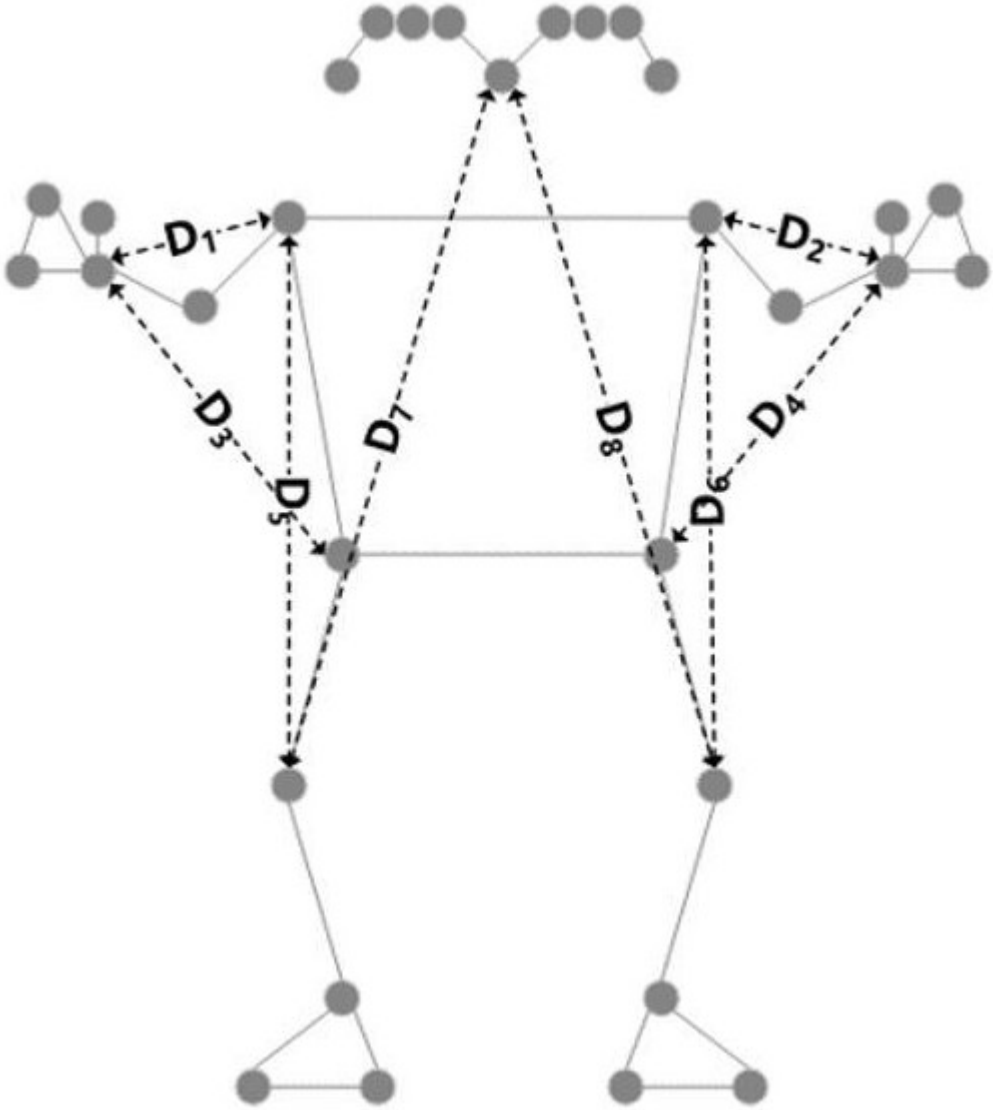

D1
D2
D3
D4
D5
D6
D7
D8

图 5

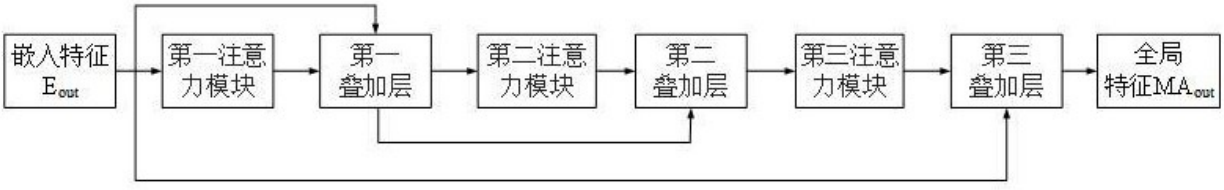

嵌入特征
Eout
第一注意力模块
第一叠加层
第二注意力模块
第二叠加层
第三注意力模块
第三叠加层
全局特征MAout

图 6

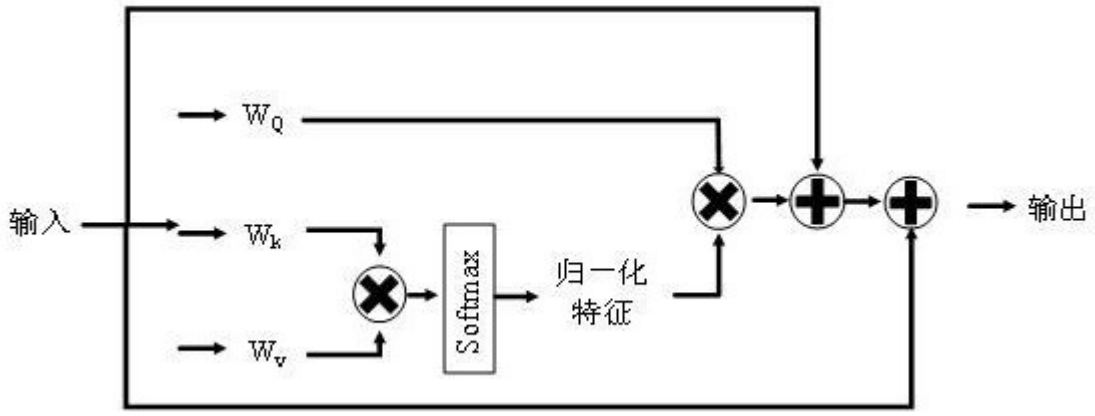


图 7

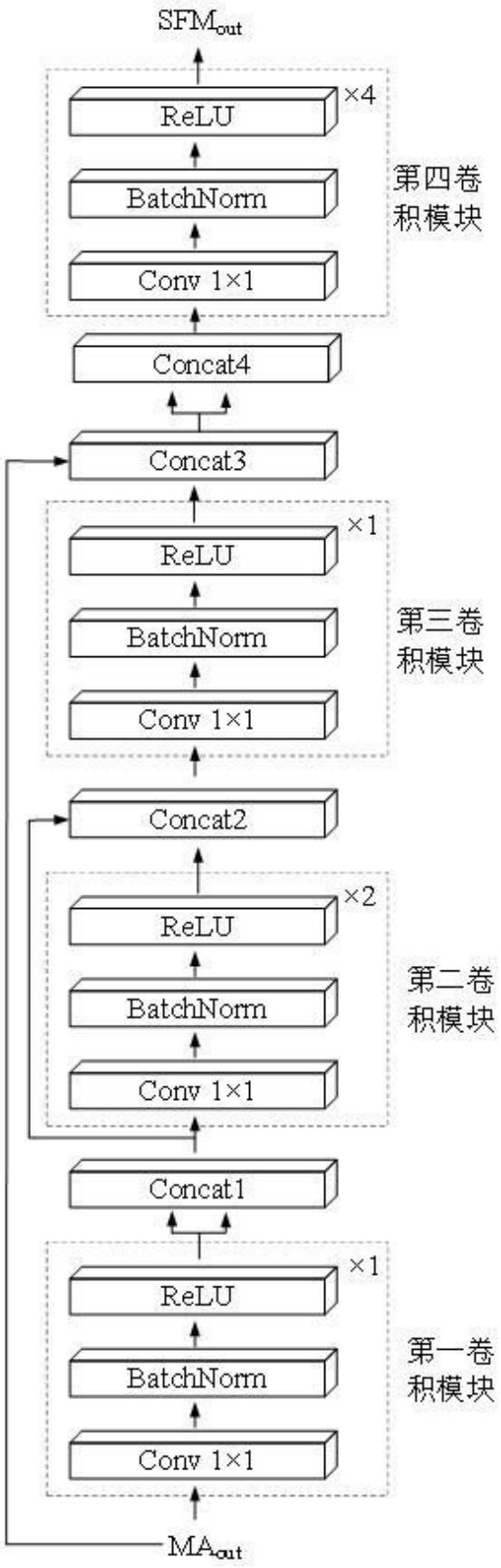

SFM$_{out}$
×4
ReLU
BatchNorm
Conv 1×1
第四卷积模块
Concat4
Concat3
×1
ReLU
BatchNorm
Conv 1×1
第三卷积模块
Concat2
×2
ReLU
BatchNorm
Conv 1×1
第二卷积模块
Concat1
×1
ReLU
BatchNorm
Conv 1×1
第一卷积模块
MA$_{out}$

图 8

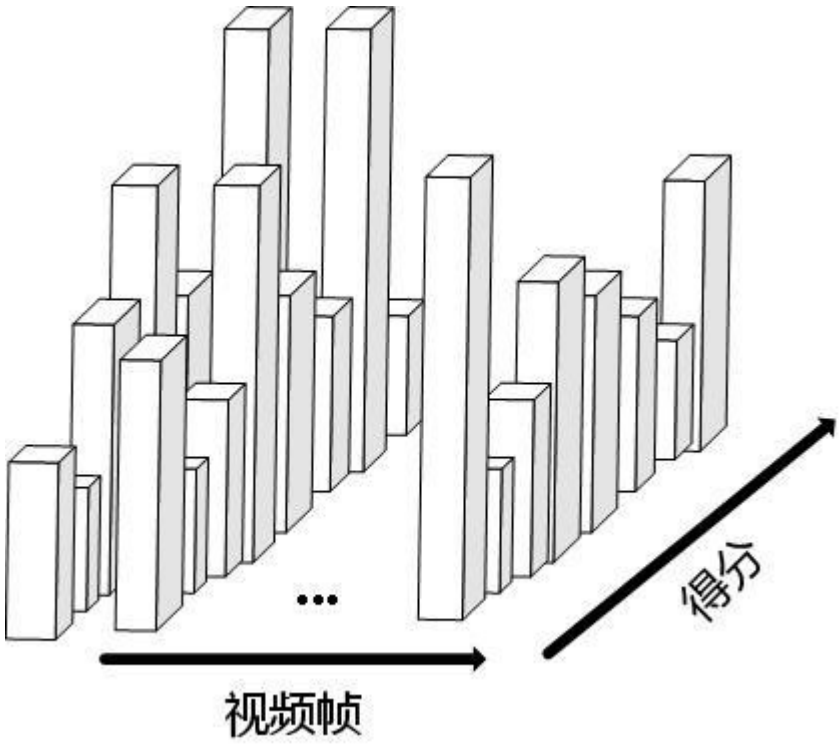


图 9

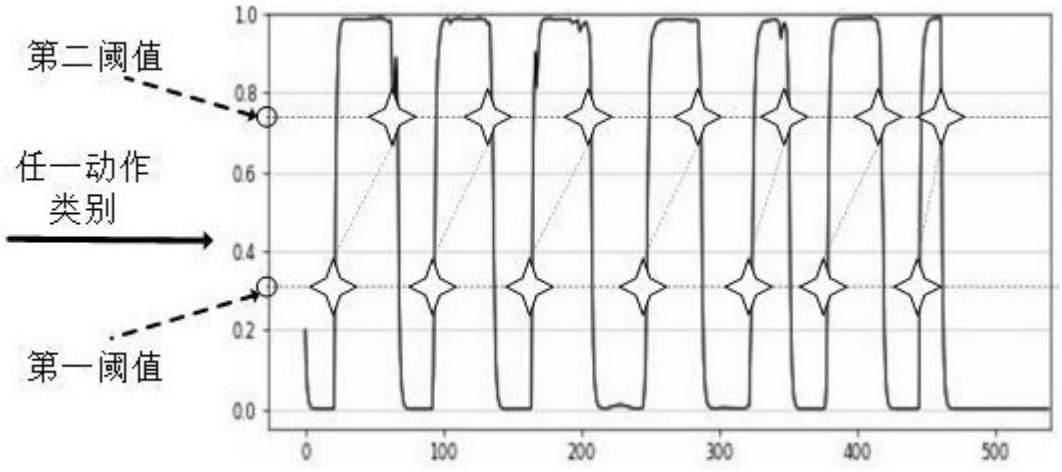


图 10

图 11

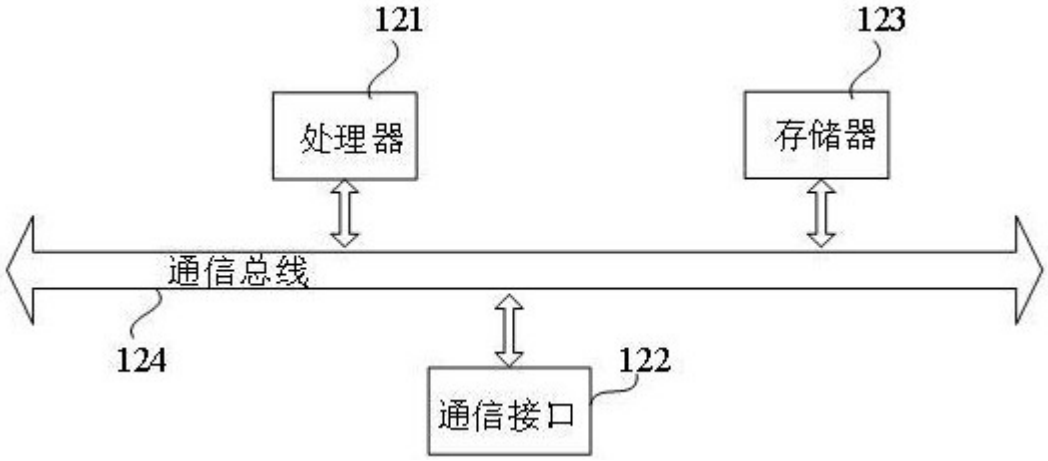


图 12