

(12) 发明专利



(10) 授权公告号 CN 113111760 B (45) 授权公告日 2023.05.02

- (21)申请号 202110372148.8
- (22)申请日 2021.04.07
- (65) 同一申请的已公布的文献号 申请公布号 CN 113111760 A
- (43) 申请公布日 2021.07.13
- (73)专利权人 同济大学 地址 200092 上海市杨浦区四平路1239号
- (72)发明人 刘成菊 党荣浩 陈启军 张恒
- (74) 专利代理机构 上海科盛知识产权代理有限 公司 31225

专利代理师 杨宏泰

(51) Int.Cl.

GO6V 40/20 (2022.01)

GO6V 10/774 (2022.01)

GO6V 10/82 (2022.01)

(54)发明名称

基于通道注意力的轻量化图卷积人体骨架 动作识别方法

(57) 摘要

本发明涉及一种基于通道注意力的轻量化 图卷积人体骨架动作识别方法,包括以下步骤: S1:获取视频影像中人体骨架的骨架序列信息; S2:添加关节点索引信息,并计算骨架关节点的 邻接矩阵:S3:利用邻接矩阵以及引入通道注意 力机制的残差GCN网络提取骨架序列的空间特 征,所述的邻接矩阵在使用时进行随机失活处 理;S4:添加帧索引信息,并进行最大池化处理; S5:利用一阶CNN网络提取骨架序列的时间特征; S6: 通过最大池化生成最终的特征序列,并得到 识别分类结果,与现有技术相比,本发明具有鲁 棒性、实时性和准确性高等优点。

Ю CN 113111760



GO6N 3/0464 (2023.01) GO6N 3/08 (2023.01)

(56) 对比文件

CN 108830157 A,2018.11.16 CN 110852182 A,2020.02.28

审查员 张诗浩

权利要求书2页 说明书6页 附图5页



1.一种基于通道注意力的轻量化图卷积人体骨架动作识别方法,其特征在于,包括以下步骤:

S1:获取视频影像中人体骨架的骨架序列信息;

S2:添加关节点索引信息,并计算骨架关节点的邻接矩阵;

S3:利用邻接矩阵以及引入通道注意力机制的残差GCN网络提取骨架序列的空间特征, 所述的邻接矩阵在使用时进行随机失活处理;

S4:添加帧索引信息,并进行最大池化处理;

S5:利用一阶CNN网络提取骨架序列的时间特征;

S6:通过最大池化生成最终的特征序列,并得到识别分类结果;

步骤S4具体为:

首先将帧索引信息通过独热编码加到每一帧的空间特征中,再利用空间最大池化的将 每一帧的关节信息进行聚合;

步骤S6具体为:

利用时间维度的最大池化生成最终的特征序列,再用全连接层和softmax层生成最终的分类结果。

2.根据权利要求1所述的一种基于通道注意力的轻量化图卷积人体骨架动作识别方法,其特征在于,步骤S1具体包括:

S11:对视频影像中的人体骨架建模并进行预处理,得到初始骨架序列信息;

S12:利用相邻帧之间作差的方式,获取初始骨架序列信息中的一阶信息、二阶信息和 三阶信息;

S13:在保证数据初始维度不变的情况下,将一阶信息、二阶信息和三阶信息融合相加, 得到最终的骨架序列信息。

3.根据权利要求1所述的一种基于通道注意力的轻量化图卷积人体骨架动作识别方法,其特征在于,步骤S2中,利用两个全连接层对骨架关节点之间相似性和亲和度的建模,通过内积计算得到邻接矩阵。

4.根据权利要求3所述的一种基于通道注意力的轻量化图卷积人体骨架动作识别方法,其特征在于,所述的邻接矩阵的计算式为:

 $G = \theta(z_{t,i}) \cdot \varphi(z_{t,j})^T$

其中,G为邻接矩阵, z_t 为第t帧的骨架序列信息, $\theta(z_{t,i})$ 为双层全连接层 θ 对第t帧骨架 序列信息 z_t 进行编码后的数据, $\varphi(z_{t,j})$ 为双层全连接层 φ 对第t帧骨架序列信息 z_t 进行编码 后的数据,所述的双层全连接层 θ 和双层全连接层 φ 为两个权重不共享的双层连接层。

5.根据权利要求1所述的一种基于通道注意力的轻量化图卷积人体骨架动作识别方法,其特征在于,所述的引入通道注意力机制的残差GCN网络为多层结构,每一层分别包括GCN网络模块、残差连接模块和通道注意力机制模块,所述的通道注意力机制模块接入于GCN网络模块的输出和残差连接模块之间,所述的GCN网络模块的输入包括邻接矩阵和融合了关节点索引信息的骨架序列信息。

6.根据权利要求5所述的一种基于通道注意力的轻量化图卷积人体骨架动作识别方法,其特征在于,所述的引入通道注意力机制的残差GCN网络的每一层中输入GCN网络模块

的邻接矩阵均采用前一个邻接矩阵经过随机失活之后的结果。

7.根据权利要求5所述的一种基于通道注意力的轻量化图卷积人体骨架动作识别方法,其特征在于,所述的通道注意力机制模块首先通过全局平局池化层聚合每个通道的信息,再通过利用Relu层进行非线性激活的两个全连接层,最后通过sigmoid层得到每个通道的权重值。

8.根据权利要求5或7所述的一种基于通道注意力的轻量化图卷积人体骨架动作识别 方法,其特征在于,所述的通道注意力机制模块用于获取每个通道的权重,其计算式如下:

 $Y_t = G_t Z_t W_v$

 $M_t = Attention(Y_t)$

 $Y_{t-out} = M_t Y_t$

其中, W_y 为可学习的参数, $M_t = \{m_t^1, m_t^2, ..., m_t^C\}$, m_t^c 表示第t时刻通道c的注意力权重, Y_t 为经过图卷积神经网络网络进行节点信息传递后的特征图, Y_{t-out} 为经过通道激活后的人体骨架特征图, G_t 为经随机失活后输入通道注意力机制模块的邻接矩阵, Z_t 为输入通道注意力机制模块的骨架序列信息。

9.根据权利要求1所述的一种基于通道注意力的轻量化图卷积人体骨架动作识别方法,其特征在于,步骤S5中,所述的一阶CNN网络首先利用一维的CNN模型进行局部时间特征聚合,然后进行批标准化和Relu激活,再利用dropout层防止过拟合,最后再利用1×1卷积将特征映射到高维空间。

基于通道注意力的轻量化图卷积人体骨架动作识别方法

技术领域

[0001] 本发明涉及计算机视觉和深度学习领域,尤其是涉及一种基于通道注意力的轻量 化图卷积人体骨架动作识别方法。

背景技术

[0002] 人体动作识别是计算机视觉领域的重要研究方向,有着广阔的应用场景和市场价值,比如:异常行为监控、用户行为分析等。骨架序列数据是一种抽象化的人体动作数据,利用3D坐标、关节索引和关节连接关系表示了人体各个关键部位的运动。早期的骨架动作识别方法中,大多采用手工提取的特征,利用特征映射等方式对数据进行处理整合。深度学习的方法兴起后,利用神经网络对骨架的时空信息进行建模渐渐的变为主流,相继出现了基于RNN、CNN、GCN等各种方法,并且效果明显优于基于手工特征的方法。

[0003] 基于RNN递归神经网络的方法主要是利用LSTM/GRU等模型建模骨架序列的时间动态变化。但这种方法只是将每一帧的关节三维坐标信息以某种顺序排列成一个向量输入到递归神经网络中,并对不同的关节进行区分,也就丢失了重要的空间信息。CNN卷积神经网络在图像处理领域愈发的火热,研究者通过将骨架信息编排成伪图片的形式,即将三维坐标(x,y,z)类比成图像中的三通道(R,G,B),将时间序列和关节索引类比成图像的长宽。这样,通过基本的二维CNN卷积操作就能够将时间和空间的信息聚合起来,提取时空的联合特征。但是这样的数据组织无法表示关节之间的拓扑结构和连接关系,关节之间的相互影响也难以进行有效的建模。

[0004] 用于骨架动作识别的GCN图卷积神经网络包括ST-GCN、AS-GCN和2S-AGCN。ST-GCN 是利用GCN方法解决动作识别问题的开端,此方法使用时空图卷积网络不断的聚合相邻关 节和相邻帧之间的特征,实现了对于人体关节自然连接关系的建模。AS-GCN对于ST-GCN中 的问题进行了改进,发现自然连接较远的关节之间也会有隐连接,如走路时手和脚的摆动。 所以AS-GCN将训练分为两路,一路侧重于探索结构之间的连接,另一路侧重于探索动作导 致的节点之间的隐连接。2S-AGCN中提出了自适应图卷积方法,将邻接矩阵分为三部分的叠 加:原始连接、训练出来的关节相关性、根据每个输入计算的关节相似性,更加灵活的对关 节之间的关系进行了建模。

[0005] 如今利用图卷积神经网络的骨架动作识别方法仍在不断的发展中,对于建模节点-骨架关系、时空联合特征以及轻量化网络等方面,仍需要进一步探索,以进一步提高骨架动作识别的鲁棒性、实时性和准确性。

发明内容

[0006] 本发明的目的就是为了克服上述现有技术存在的缺陷而提供一种鲁棒性好、实时 性强和准确性高的基于通道注意力的轻量化图卷积人体骨架动作识别方法。

[0007] 本发明的目的可以通过以下技术方案来实现:

[0008] 一种基于通道注意力的轻量化图卷积人体骨架动作识别方法,包括以下步骤:

[0009] S1:获取视频影像中人体骨架的骨架序列信息;

[0010] S2:添加关节点索引信息,并计算骨架关节点的邻接矩阵;

[0011] S3:利用邻接矩阵以及引入通道注意力机制的残差GCN网络提取骨架序列的空间 特征,所述的邻接矩阵在使用时进行随机失活处理;

[0012] S4:添加帧索引信息,并进行最大池化处理;

[0013] S5:利用一阶CNN网络提取骨架序列的时间特征;

[0014] S6:通过最大池化生成最终的特征序列,并得到识别分类结果。

[0015] 进一步地,步骤S1具体包括:

[0016] S11:对视频影像中的人体骨架建模并进行预处理,得到初始骨架序列信息;

[0017] S12:利用相邻帧之间作差的方式,获取初始骨架序列信息中的一阶信息、二阶信息和三阶信息;

[0018] S13:在保证数据初始维度不变的情况下,将一阶信息、二阶信息和三阶信息融合相加,得到最终的骨架序列信息。

[0019] 进一步地,步骤S2中,利用两个全连接层对骨架关节点之间相似性和亲和度的建模,通过内积计算得到邻接矩阵。

[0020] 更进一步地,所述的邻接矩阵的计算式为:

 $[0021] \quad G = \theta(z_{t,i}) \cdot \varphi(z_{t,j})^T$

[0022] 其中,G为邻接矩阵, z_t 为第t帧的骨架序列信息, $\theta(z_{t,i})$ 为双层全连接层 θ 对第t帧 骨架序列信息 z_t 进行编码后的数据, $\varphi(z_{t,j})$ 为双层全连接层 φ 对第t帧骨架序列信息 z_t 进 行编码后的数据,所述的双层全连接层 θ 和双层全连接层 φ 为两个权重不共享的双层连接 层。

[0023] 进一步地,所述的引入通道注意力机制的残差GCN网络为多层结构,每一层分别包括GCN网络模块、残差连接模块和通道注意力机制模块,所述的通道注意力机制模块接入于GCN网络模块的输出和残差连接模块之间,所述的GCN网络模块的输入包括邻接矩阵和融合了关节点索引信息的骨架序列信息。

[0024] 更进一步地,所述的引入通道注意力机制的残差GCN网络的每一层中输入GCN网络 模块的邻接矩阵均采用前一个邻接矩阵经过随机失活之后的结果。

[0025] 更进一步地,所述的通道注意力机制模块首先通过全局平局池化层聚合每个通道的信息,再通过利用Relu层进行非线性激活的两个全连接层,最后通过sigmoid层得到每个通道的权重值。

[0026] 更进一步地,所述的通道注意力机制模块用于获取每个通道的权重,其计算式如下:

 $\begin{bmatrix} 0027 \end{bmatrix} \qquad Y_{t} = G_{t}Z_{t}W_{v}$

[0028]
$$M_{+}$$
=Attention(Y_{+})

 $[0029] \quad Y_{t-out} = M_t Y_t$

[0030] 其中, W_y 为可学习的参数, $M_t = \{m_t^1, m_t^2, ..., m_t^C\}$, m_t^c 表示第t时刻通道c的注意 力权重, Y_t 为经过图卷积神经网络网络进行节点信息传递后的特征图, Y_{t-out} 为经过通道激 活后的人体骨架特征图, G_t 为经随机失活后输入通道注意力机制模块的邻接矩阵, Z_t 为输入

通道注意力机制模块的骨架序列信息。

[0031] 进一步地,步骤S4具体为:

[0032] 首先将帧索引信息通过独热编码加到每一帧的空间特征中,再利用空间最大池化的将每一帧的关节信息进行聚合;

[0033] 步骤S6具体为:

[0034] 利用时间维度的最大池化生成最终的特征序列,再用全连接层和softmax层生成最终的分类结果。

[0035] 进一步地,步骤S5中,所述的一阶CNN网络首先利用一维的CNN模型进行局部时间 特征聚合,然后进行批标准化和Relu激活,再利用dropout层防止过拟合,最后再利用1×1 卷积将特征映射到高维空间。

[0036] 与现有技术相比,本发明具有以下优点:

[0037] 1)本发明在图卷积网络中引入了通道注意力机制,将节点通道之间的相互依赖性关系进行了建模,对GCN输出的特征在通道方向进行自适应的重新校准,提高了整个网络对于节点特征通道信息挖掘的深度,从而提高了整体的识别效果,和识别稳定性;

[0038] 2)本发明对邻接矩阵进行了随机失活处理,降低了关节之间的强依赖关系,也使 得训练过程中邻接矩阵生成的训练和后面空间特征提取的训练有效的分离,消除了特定关 节之间的强依赖关系,减弱了关节间关系建模的过拟合情况,让生成邻接矩阵的网络拥有 更强的泛用性,使得在识别数据集外的骨架视频序列时能够有更好的鲁棒性;

[0039] 3)本发明将空间特征提取和时间特征提取分离开来,有效的避免了时空图卷积庞 大的计算量,也使得特征提取的难度降低,能够使用更加简单,参数量更小的方法实现特征 的聚合,较小的计算量也使得这种算法在移动端和可穿戴设备上的部署可行性更强;

[0040] 4)本发明在空间特征提取和时间特征提取前分别引入了关节点索引信息和时间 序列索引信息,使得空间上的骨架连接和时间上的骨架运动都拥有方向上的意义,提高识 别分类的效果;

[0041] 5)本发明将初始骨架序列信息分成一阶信息、二阶信息和三阶信息的三通道信号,并将三通道信号融合相加,在保证数据初始维度不变的情况下,得到更加丰富的骨架序 列信息,提高识别分类的效果。

附图说明

[0042] 图1是本发明算法的流程图;

[0043] 图2是本发明算法的空间特征提取,即引入通道注意力机制的残差GCN网络(CA-GCN)的结构示意图;

[0044] 图3是本发明算法的时间特征提取,即一阶CNN网络(first-order CNN)的结构示意图;

[0045] 图4是本发明算法的通道注意力机制模块(CA-Block)的结构示意图;

[0046] 图5是NTU-RGBD数据集的骨架结构图;

[0047] 图6是本发明动作识别方法与其他现有动作识别方法在NTU-RGBD60数据集上的效果对比示意图。

具体实施方式

[0048] 下面结合附图和具体实施例对本发明进行详细说明。显然,所描述的实施例是本 发明的一部分实施例,而不是全部实施例。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的前提下所获得的所有其他实施例,都应属于本发明保护的范围。

[0049] 实施例

[0050] 如图1所示,本发明公开了一种基于通道注意力的轻量化图卷积人体骨架动作识别方法,包括以下步骤:

[0051] S1:获取视频影像中人体骨架的骨架序列信息,具体包括:

[0052] S11:对视频影像中的人体骨架建模并进行预处理,得到初始骨架序列信息;

[0053] S12:利用相邻帧之间作差的方式,获取初始骨架序列信息中的一阶信息、二阶信息和三阶信息;

[0054] S13:在保证数据初始维度不变的情况下,将一阶信息、二阶信息和三阶信息融合相加,得到最终的骨架序列信息。

[0055] 其中,步骤S11中通过openpose等姿态估计算法对视频影像中的人体骨架建模;初始骨架序列信息中的一阶信息、二阶信息和三阶信息分别对应为位置、速度和加速度的三通道信号,三通道相加融合能够得到更加丰富的骨架序列信息。

[0056] S2:添加关节点索引信息,并计算骨架关节点的邻接矩阵,具体为:利用两个的FC 层建模初始骨架点之间的相似性和亲和度,从而计算出可学习的邻接矩阵;

[0057] S3:利用邻接矩阵以及引入通道注意力机制的残差GCN网络提取骨架序列的空间特征;

[0058] 其中,邻接矩阵在使用时进行随机失活处理,对其矩阵中的信息进行一定概率的随机dropout,消除了特定关节之间的强依赖关系,减弱了关节间关系建模的过拟合情况; 通道注意力机制接到图卷积输出和残差连接之间,建模关节通道之间的相互依赖性,并自 适应的校准通道之间的特征响应;

[0059] S4:添加帧索引信息,并进行空间最大池化处理,利用空间最大池化聚合关节信息;

[0060] S5:利用一阶CNN网络提取骨架序列的时间特征,具体为:利用一维CNN对时间序列进行特征提取,再用1×1卷积将特征映射到高维空间;

[0061] S6:通过最大池化生成最终的特征序列,并得到识别分类结果,具体为用时间维度上的最大池化层聚合时间信息生成最终的特征序列。

[0062] 本发明中空间特征的提取和时间特征的提取是分开进行的,首先利用GCN提取空间信息,再利用一维CNN提取时间信息,两者串联进行,且在空间特征提取和时间特征提取前分别引入了关节点索引信息和时间序列索引信息,使得空间上的骨架连接和时间上的骨架运动都拥有方向上的意义。

[0063] 如图1所示,下面给出本发明的具体实施过程:

[0064] 步骤1:通过openpose等姿态估计算法对视频影像中的人体骨架进行建模。

[0065] 本步骤中,采用NTU-RGBD数据集作为训练数据,其骨架数据如图5所示,有25个关节点,在测试时使用openpose方法生成25个节点骨架数据。

[0066] 步骤2:对骨架信息进行标准化预处理,并根据受试者和相机角度划分训练集和测

试集。

[0067] 本步骤中,首先从skeleton文件中读取原始的骨架数据,以一个视频骨架序列为一组,并按照.pk1的格式储存原始数据。之后剔除出现视频过短、骨架横纵比失调、骨架数 过多等问题的骨架序列,并储存为.pk1格式。最后将每一帧的骨架图居中,并按照不同受试 者和不同摄像头角度这两种方式划分测试集和训练集。

[0068] 步骤3:利用相邻帧之间做差获得骨架的高阶信息,将骨架的一阶、二阶和三阶信息融合并添加骨架索引信息。

[0069] 本步骤中,先利用相邻帧之间做差生成骨架序列的速度和加速度信息,为了保证 序列长度相同在做差后均将初始端的数字复制一次。之后利用两层的全连接层对位置、速 度和加速度信息分别进行编码,再通过编码后的直接相加进行信息的融合。

[0070] 步骤4:利用融合后的节点信息计算骨架点的邻接矩阵。

[0071] 本步骤中,使用两个不同的双层全连接层 θ 和 φ 对融合了关节点索引信息的骨架序列信息Z(T×J×C_n)进行编码,分别得到 $\theta(z_{t,i})$ 和 $\varphi(z_{t,j})$,其中T为帧数,J为关节点数,C_n为维度,为了求得关节之间的连接关系,即邻接矩阵,我们将这两个编码后的数据进行内积 $\theta(z_{t,i}) \cdot \varphi(z_{t,j})^{T}$ 。得到后面图卷积过程中使用的邻接矩阵G,即邻接矩阵的计算式为:

 $[0072] \quad G = \theta(z_{t,i}) \cdot \varphi(z_{t,i})^T$

[0073] 其中,G为邻接矩阵,z_t为第t帧的骨架序列信息,θ(z_{t,i})为双层全连接层θ对第t帧 骨架序列信息z_t进行编码后的数据,**φ(z_{t,j})**为双层全连接层**φ**对第t帧骨架序列信息z_t进行 编码后的数据,所述的双层全连接层θ和双层全连接层**φ**为两个权重不共享的双层连接层。 [0074] 步骤5:利用邻接矩阵和串联了通道注意力机制的残差GCN网络提取骨架序列的空

间特征。

[0075] 本步骤中,如图2所示,利用通道注意力机制和GCN结合的方式实现的空间特征的 提取,引入通道注意力机制的残差GCN网络(CA-GCN)为多层结构,每一层分别包括GCN网络 模块、残差连接模块和通道注意力机制模块,通道注意力机制模块接入于GCN网络模块的输 出和残差连接模块之间,GCN网络模块的输入包括邻接矩阵和融合了关节点索引信息的骨 架序列信息。

[0076] 首先,为了减少计算量,将每个GCN网络模块的邻接矩阵都采用之前计算出来的邻接矩阵G经过随机失活之后的结果,即将邻接矩阵中的元素按照一定的概率随机置零,再利用Y_t=G_tZ_tW_y公式在骨架图之间进行信息传递,其中W_y是可学习的参数。之后再通过如图4所示的通道注意力机制模块探索通道之间的依赖关系,并求得每个通道的权重,公式如下:

[0077] $M_t = Attention(Y_t)$

 $[0078] \quad Y_{t-out} = M_t Y_t$

[0079] 其中,W_y为可学习的参数, $M_t = \{m_t^1, m_t^2, ..., m_t^c\}$, m_t^c 表示第t时刻通道c的注意 力权重,Y_t为经过图卷积神经网络网络进行节点信息传递后的特征图,Y_{t-out}为经过通道激 活后的人体骨架特征图,G_t为经随机失活后输入通道注意力机制模块的邻接矩阵,Z_t为未经 过图卷积信息聚合的骨架序列信息。

[0080] 通道注意力机制模块先通过一个全局平局池化聚合每个通道的信息,再通过用

Relu层进行非线性激活的两个全连接层和最后的sigmoid层得到每个通道的权重值。

[0081] 步骤6:添加帧索引信息并对骨架点信息进行最大池化。

[0082] 本步骤中,首先将帧索引信息进行独热编码加到每一帧的空间特征中,再利用最 大池化的方法将每一帧的关节信息进行聚合,得到一个T×1×C。维度的向量。

[0083] 步骤7:利用一阶的CNN网络提取骨架序列时间特征。

[0084] 本步骤中,如图3所示,一阶的CNN网络首先利用一维的CNN对骨架序列的局部时间 特征进行聚合,卷积核大小为(1,3),padding=(0,1)。之后进行归一化和Relu激活,再利用 dropout层防止过拟合。后一个CNN是将学习到的时间特征映射到更高维的核空间,从而增 强特征向量的表达能力。

[0085] 步骤8:对帧的时间信息进行最大池化,并通过全连接层输出分类类别。

[0086] 本步骤中,利用时间维度的最大池化生成最终的特征序列,再用全连接层和 softmax生成最终的分类结果。

[0087] 本发明将CNN领域中的通道注意力机制引入到了GCN中来,并提高了对于骨架动作 识别任务中空间信息提取的能力,让通道之间的相互依赖关系得以被建模。并且将CNN网络 中的dropout随机神经元失活的思想引入到了邻接矩阵的应用中,使得每一次GCN都运用的 不同的邻接矩阵进行计算,让骨架之间的耦合关系更加的灵活不至于过拟合,增强了对于 不同数据集的鲁棒性。通过时空信息提取网络串联的方式,降低了整体的计算量和参数量, 使用有针对性的定制化方式对时空信息进行处理是提高信息提取效率的关键。本算法在 NTU-RGBD60数据集上以不同受试者为数据集划分依据的情况下,测试准确为88.565%,参 数量为839464。如图6所示,相比于ST-GCN、AS-GCN等先进的图网络方法,我们方法的参数量 和精度都有着较大的优势。

[0088] 以上所述,仅为本发明的具体实施方式,但本发明的保护范围并不局限于此,任何 熟悉本技术领域的工作人员在本发明揭露的技术范围内,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发明的保护范围之内。因此,本发明的保护范围应以权利 要求的保护范围为准。











图5



图6