



(12) 发明专利

(10) 授权公告号 CN 112084934 B

(45) 授权公告日 2024.03.15

(21) 申请号 202010934403.9

G06V 40/10 (2022.01)

(22) 申请日 2020.09.08

G06V 10/764 (2022.01)

(65) 同一申请的已公布的文献号

G06V 10/774 (2022.01)

申请公布号 CN 112084934 A

G06N 3/0464 (2023.01)

G06N 3/084 (2023.01)

(43) 申请公布日 2020.12.15

G06V 10/30 (2022.01)

(73) 专利权人 浙江工业大学

(56) 对比文件

地址 310006 浙江省杭州市下城区朝晖六区潮王路18号

CN 109543549 A, 2019.03.29

CN 110059620 A, 2019.07.26

(72) 发明人 邱飞岳 孔德伟 章国道 王丽萍 陈宏 郭海东 姜弼君

WO 2017133009 A1, 2017.08.10

审查员 狄希

(74) 专利代理机构 杭州浙科专利事务所(普通合伙) 33213

专利代理人 周红芳 朱盈盈

(51) Int. Cl.

G06V 40/20 (2022.01)

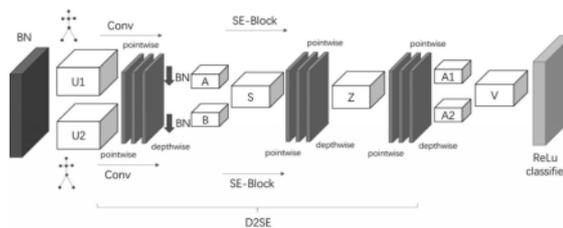
权利要求书1页 说明书6页 附图2页

(54) 发明名称

基于骨骼数据双通道深度可分离卷积的行为识别方法

(57) 摘要

基于骨骼数据双通道深度可分离卷积的行为识别方法,属于人体姿态行为识别技术领域。它包括以下步骤:一、获取人体行为姿态关节骨骼点数据;二、对骨骼点数据处理提取行为空间特征;三、构建D2SE双通道深度可分离卷积层,在时间维度上提取行为时间特征;四、将图卷积上的空间信息和D2SE网络层上的时间信息叠加提取姿态行为的时空信息;步骤五、使用ReLu函数获取骨骼动作分类。本发明使用图卷积神经网络GCN和D2SE网络层,对人体姿态行为骨骼数据使用空间上的图卷积来提取空间信息,本发明基于双通道,在提高基于深度分离卷积架构性能的同时不会引入额外的复杂度,同时能使卷积层的参数明显缩减。



1. 基于骨骼数据双通道深度可分离卷积的行为识别方法,其特征包括以下步骤:

步骤一、获取人体行为姿态关节骨骼点数据;

步骤二、使用类邻接矩阵策略对骨骼点数据处理提取行为空间特征;

步骤三、将Depth层和Point层嵌入卷积层中,构建D2SE双通道深度可分离卷积层,在时间维度上提取行为时间特征;

具体如下:

使用双向SE通道对输入数据进行卷积操作,SE层嵌入使用点卷积Depthwise和深度Pointwise卷积处理;

构建DS层网络,将DS层作为SE-Block的卷积层读入数据;

Depthwise卷积层将DS层接受参数作为输入;

对DS卷积后的Features进行ReLU激活得到特征图;

步骤四、将图卷积上的空间信息和D2SE网络层上的时间信息叠加提取姿态行为的时空信息;

步骤五、使用ReLU函数获取骨骼动作分类。

2. 根据权利要求1所述的基于骨骼数据双通道深度可分离卷积的行为识别方法,其特征包括以下步骤一,通过Azure for Kinect获取人体关节骨骼点数据,对获取的骨骼点数据进行预处理,具体构建步骤如下:

2.1、从导出的文件中获取姿态骨骼数据序列,去除噪声数据;

2.2、将去噪后的数据构建图数据。

3. 根据权利要求2所述的基于骨骼数据双通道深度可分离卷积的行为识别方法,其特征包括以下步骤二,将步骤一中预处理后的骨骼点数据使用类邻接矩阵策略转换成一张伪图像矩阵来提取行为空间特征。

4. 根据权利要求1所述的基于骨骼数据双通道深度可分离卷积的行为识别方法,其特征包括以下步骤三,构建D2SE双通道深度可分离卷积层,在时间维度上提取行为时间特征,所述D2SE网络包括双通道SE网络层及Conv层两个部分,所述双通道SE网络层对输入数据进行卷积操作,Conv层使用depthwise和pointwise卷积组合。

5. 根据权利要求1所述的基于骨骼数据双通道深度可分离卷积的行为识别方法,其特征包括以下步骤四,构建D2SE-GCN双通道深度可分离卷积网络将图卷积上的空间信息和D2SE网络层上的时间信息叠加提取姿态行为的时空信息,采用随机梯度下降SGD优化网络参数,预设初始学习率、权重衰减、训练和测试样本batch_size、训练epoch及warm_up预热学习率。

6. 根据权利要求5所述的基于骨骼数据双通道深度可分离卷积的行为识别方法,其特征包括以下步骤五,所述学习率的更新包括模型预热阶段、正式训练阶段,使用预热学习率能加快模型收敛速度。

7. 根据权利要求5所述的基于骨骼数据双通道深度可分离卷积的行为识别方法,其特征包括以下步骤六,将测试样本输入已训练好的双通道深度可分离卷积D2SE-GCN模型中,输出向量中最大值对应的姿态行为类别即可判断该测试样本的动作类型,实现姿态行为的准确识别。

基于骨骼数据双通道深度可分离卷积的行为识别方法

技术领域

[0001] 本发明属于人体姿态行为识别技术领域,具体涉及一种基于骨骼点数据双通道深度可分离卷积的姿态行为识别方法。

背景技术

[0002] 人体动作识别是近年来CV领域热门研究的一个方向,骨骼点动作识别是人体动作识别中的一个分支,旨在识别由骨骼点数据随时间变化构成的骨骼序列,人体动作识别的另外一个分支是处理RGB视频序列。

[0003] 基于RGB视频序列由于其时间序列性,一般使用GRU、3D卷积、LSTM变种等方式处理;但是由于RGB数据会放大光照、颜色以及遮盖物等因素的影响,进而导致模型鲁棒性不如骨骼数据拟合的模型。

[0004] 基于骨骼数据的模型,由于其骨骼和关节相互连接,和图结构的数据非常契合,加上其对颜色和光照、遮挡物的强鲁棒性,因此近年来基于骨骼数据的模型大部分是基于图卷积网络。

[0005] 然而针对特定场景中,Azure for Kinect设备开发应用程序所依赖运行设备的性能不足,以及对数据处理的时效性和准确率考虑,因此综合考虑设计一种基于骨骼点数据的姿态行为识别方法显得尤为重要。

发明内容

[0006] 针对现有技术中存在的上述问题,本发明的目的在于提供一种基于骨骼点坐标的姿态行为识别方法,通过在图卷积网络中嵌入D2SE块来处理骨骼数据,相比于传统的图卷积识别方法,能够降低模型训练参数和训练时间的同时保证动作识别的准确率。

[0007] 本发明提供如下技术方案:基于骨骼数据双通道深度可分离卷积的行为识别方法,其特征在于包括以下步骤:

[0008] 步骤一、获取人体行为姿态关节骨骼点数据;

[0009] 步骤二、使用类邻接矩阵策略对骨骼点数据处理提取行为空间特征;

[0010] 步骤三、将Depth层和Point层嵌入卷积层中,构建D2SE双通道深度可分离卷积层,在时间维度上提取行为时间特征;

[0011] 步骤四、将图卷积上的空间信息和D2SE网络层上的时间信息叠加提取姿态行为的时空信息;

[0012] 步骤五、使用ReLu函数获取骨骼动作分类。

[0013] 所述的基于骨骼数据双通道深度可分离卷积的行为识别方法,其特征在于所述步骤一中,通过Azure for Kinect获取人体关节骨骼点数据,对获取的骨骼点数据进行预处理,具体构建步骤如下:

[0014] 2.1、从导出的文件中获取姿态骨骼数据序列,去除噪声数据;

[0015] 2.2、将去噪后的数据构建图数据。

[0016] 所述的基于骨骼数据双通道深度可分离卷积的行为识别方法,其特征在于所述步骤二中,将步骤一中预处理后的骨骼点数据使用类邻接矩阵策略转换成一张伪图像矩阵来提取行为空间特征。

[0017] 所述的基于骨骼数据双通道深度可分离卷积的行为识别方法,其特征在于所述步骤三中,采用构建基于DSE网络的姿态行为识别方法提取时间特征,所述DSE网络包括双通道SE网络层及Conv层两个部分,所述双通道SE网络层对输入数据进行卷积操作,Conv层使用depthwise和pointwise卷积组合。

[0018] 所述的基于骨骼数据双通道深度可分离卷积的行为识别方法,其特征在于所述步骤四中,构建D2SE-GCN双通道深度可分离卷积网络将图卷积上的空间信息和D2SE网络层上的时间信息叠加提取姿态行为的时空信息,采用随机梯度下降SGD优化网络参数,预设初始学习率、权重衰减、训练和测试样本batch_size、训练epoch及warm_up预热学习率。

[0019] 所述的基于骨骼数据双通道深度可分离卷积的行为识别方法,其特征在于所述学习率的更新包括模型预热阶段、正式训练阶段,使用预热学习率能加快模型收敛速度。

[0020] 所述的基于骨骼数据双通道深度可分离卷积的行为识别方法,其特征在于所述步骤四中,将测试样本输入已训练好的双通道深度可分离卷积D2SE-GCN模型中,输出向量中最大值对应的姿态行为类别即可判断该测试样本的动作类型,实现姿态行为的准确识别。

[0021] 通过采用上述技术,与现有技术相比,本发明的有益效果如下:

[0022] 1) 本发明使用图卷积神经网络GCN和D2SE网络层,对人体姿态行为骨骼数据使用空间上的图卷积来提取空间信息,通过前后相邻帧的常规卷积来提取时间信息,最后通过二者的叠加来提取时空信息;

[0023] 2) 本发明是基于双通道,普通的SE块通过网络根据loss学习特征权重,使得有效的FeatureMap权重和无效的FeatureMap自适应,但是在网络的计算过程中不可避免的增加了参数和计算量,深度可分离卷积层网络架构隐式依赖提供一种更高效的划分方法,在提高基于深度分离卷积架构性能的同时不会引入额外的复杂度,同时能使卷积层的参数明显缩减,而双通道的模型在输入特征适应不同卷积核,多视野域卷积能使得SPLIT获取多个卷积核的不同感受野,通过线性变换操作、归一化、得到通道特征,和原始特征加乘,然后经过FUSE聚合来自多个路径的信息,从而获得全局选择性权重表现,双通道SE-Block能对读入的相邻帧骨骼数据时间维度信息自适应其卷积核的权重,使得模型的准确率得到更有效的提升。

附图说明

[0024] 图1为本发明的D2SE模型的网络层示意图;

[0025] 图2为本发明的Azure for Kinect骨骼点数据采集示意图;

[0026] 图3为本发明的人体骨骼数据具体分布示意图。

具体实施方式

[0027] 为了使本发明的目的、技术方案及优点更加清楚明白,以下结合说明书附图及实施例,对本发明进行进一步详细说明。应当理解,此处所描述的具体实施例仅仅用以解释本发明,并不用于限定本发明。

[0028] 相反,本发明涵盖任何由权利要求定义的在本发明的精髓和范围上做的替代、修改、等效方法以及方案。进一步,为了使公众对本发明有更好的了解,在下文对本发明的细节描述中,详尽描述了一些特定的细节部分。对本领域技术人员来说没有这些细节部分的描述也可以完全理解本发明。

[0029] 请参阅图1-3,一种基于骨骼数据双通道深度可分离卷积的行为识别方法,所述方法包括以下步骤:

[0030] 步骤一:使用Azure for Kinect获取人体行为姿态数据;

[0031] 步骤二:使用类邻接矩阵策略对骨骼点数据处理提取行为空间特征;

[0032] 步骤三:将Depth层和Point层嵌入卷积层中,构建D2SE双通道深度可分离卷积层,在时间维度上提取行为时间特征;

[0033] 步骤四:将图卷积上的空间信息和D2SE网络层上的时间信息叠加提取姿态行为的时空信息。

[0034] 步骤五:使用ReLu函数获取骨骼动作分类。

[0035] 通过Azure for Kinect获取人体32个关节骨骼点数据,如图3所示,对获取的骨骼点数据进行预处理,具体为:

[0036] 从文件中获取姿态骨骼序列,去除噪声数据,获取的骨骼点为32个身体关节数据,关节层次结构按照从人体中心向四肢的流向分布,每个连接点将父关节与子关节链接起来。

[0037] 获取的数据为连续的骨骼点帧数据,维度为(T、V、C),T表示时间,V表示关节点、C表示骨骼点坐标空间位置(X、Y、Z)。

[0038] 将处理后的骨骼点数据使用类邻接矩阵策略转换成一张伪图像矩阵来提取行为空间特征,具体操作为:首先构建一个图矩阵,骨骼点即为图矩阵的节点N,骨骼点之间的连接即为图矩阵的边S,对于每一个骨骼点,邻节点个数不固定,通过定义一个映射函数L将参数和邻节点对应,可用如下函数操作表示。

$$[0039] \quad f_{\text{out}}(v_i) = \sum_{v_j \in B_i} \frac{1}{Z_{ij}} f_{\text{in}}(N_j) \cdot w(l_i(v_j))$$

[0040] 其中f是输入输出的特征张量,w是权重,N是节点,I表示节点N于W权重之间映射,Z是归一化操作。

[0041] 建立D2SE层网络,使用D2SE卷积提取相邻帧上的时间信息,具体操作如下:

[0042] D2SE层网络分两部分,一个是SE-Block单元,本实例中使用双向SE通道对输入数据进行卷积操作,SE层嵌入使用点卷积Depthwise和深度Pointwise卷积处理,具体为:

[0043] 引入SE-Block单元:输入为骨骼点的channel数据,构建BLOCK_LIST块,步长为1,使用两个Branch,Branch全局参数r为16,L为32(r为压缩因子,L为d的最小值,d为卷积操作的输出特征,C为输入特征的channel),如下所示:

[0044] $d = \max(C/r, L)$

[0045] 构建DS层网络,将DS层作为SE-Block的卷积层读入数据。

[0046] DS层操作以及参数设置如下:

[0047] Pointwise卷积层(卷积核1*1,步长0) -> BN层归一化 -> Pointwise卷积层 -> BN归一

化层->Depthwise卷积层。

[0048] Pointwise卷积层卷积核kernel_size为1*1,步长stride为1,padding为0,dilation膨胀系数为1,卷积组groups为1,偏置项bias为false。

[0049] Depthwise卷积层将DS层接受参数作为输入。

[0050] 对于第一个Branch,执行一次DS卷积操作,DS参数为,3*3卷积核,步长为1,dilation为1。

[0051] 对DS卷积后的Features进行ReLU激活得到特征图A,如图1所示。

[0052] 对于第二个Branch,执行一次DS卷积操作,DS参数为5*5卷积核,步长为2,dilation为2;对DS卷积后的Features进行ReLU激活得到特征图B,如图1所示。

[0053] 将特征图A和特征图B使用SUM进行融合操作得到特征图U,使用自适应平均池化AdaptiveAvgPool_2d降采样,输出1*1尺寸的特征图S,如图1所示,H、W为输入特征图的高和宽,特征图S可表达为如下:

$$[0054] \quad s_c = F_{gp}(U_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W U_c(i, j)$$

[0055] 全连接层FC读入特征图S,同样使用两次DS卷积,参数为1*1的卷积核,步长为1,对卷积后的Features使用ReLU激活,得到特征图Z, δ 为ReLU函数,B为归一化函数,Ws为DS的输出特征,F(fc)为全连接函数。

$$[0056] \quad z = F_{fc}(s) = \delta(B(Ws))$$

[0057] 将全连接后的特征图Z再进行两次DS卷积操作,参数为1*1的卷积核,步长为1,padding为0。

[0058] 通过线性变换还原全连接层FC压缩状态,输出为两个channel,分别进行Softmax归一化操作,此时每个channel对应自身的价值,表示channel的重要程度,也即SEBlock能够自适应特征图感受野权重,选择最优的权重,将两个channel乘上原始对应的特征图得到新的特征图A1,A2,如图1所示。

[0059] 重复之前的SUM融合操作,得到输出特征V。

[0060] 原始特征图X经过DS卷积、融合FUSE、池化POOL、全连接FC、DS卷积、归一化BN、融合后得到特征图V,也即特征图V融合了多个感受野的信息。

[0061] 特征图V使用最大池化降采样,通过ReLU激活后进行分类。

[0062] 使用Azure for Kinect获取的数据训练D2SE-GCN网络,采用随机梯度下降SGD优化网络参数,初始学习率base_lr为0.1,权重衰减weight_decay为0.0001,训练和测试样本的batch_size设置为64,训练epoch设置为80,warm_up预热学习率设置为5个epoch。

[0063] 其中,学习率更新分为两个阶段,模型预热阶段和正式训练阶段,使用预热学习率能使得模型收敛速度变得更快,具体为:

[0064] 模型训练预热阶段,学习率更新为:

$$[0065] \quad lr = \frac{epoch + 1}{warm_up_epoch} * base_lr (epoch < warm_up_epoch)$$

[0066] 模型正式训练阶段,学习率更新为:

$$[0067] \quad lr = 0.1^{SUM(f(step))} \times base_lr$$

[0068] 模型使用交叉熵作为损失函数,在每batch_size样本被训练后,初始化梯度为0,反向传播求得梯度,更新模型参数。

[0069] 训练完所有epoch后,将测试集输入模型预测,使用top-1作为测试集预测结果,以此衡量模型的性能。

[0070] 将测试样本输入已训练好的双通道深度可分离卷积D2SE-GCN模型中,输出向量中最大值对应的姿态行为类别即可判断该测试样本的动作类型,实现姿态行为的准确识别。

[0071] 为了与公开模型对比性能,将D2SE-GCN与主流深度学习模型做对比,如表1-姿态行为识别Top1对比图所示,基于双通道深度可分离卷积的动作行为识别准确率较其他模型在NUT的X-Sub和X-View数据上表现更好。

[0072] 基于双通道的性能优势在于,普通的SE块通过网络根据loss学习特征权重,使得有效的FeatureMap权重和无效的FeatureMap自适应,但是在网络的计算过程中不可避免的增加了参数和计算量,深度可分离卷积层网络架构隐式依赖提供一种更高效的划分方法,在提高基于深度分离卷积架构性能的同时不会引入额外的复杂度,同时能使卷积层的参数明显缩减,而双通道的模型在输入特征适应不同卷积核,多视野域卷积能使得FUSE和SPLIT的操作后全局感受野自适应调整。

NTU-RGB-D	Top1	
	X-Sub(%)	X-View(%)
ST-GCN	81.5	88.3
SE-ST-GCN	82.3	87.5
ST-LSTM	69.2	77.7
STA-LSTM	73.4	81.2
VA-LSTM	79.2	87.7
Two-Stream	66.8	72.6
3DCNN		
TCN	74.3	83.1
D2SE-GCN	86.3	92.6

[0073]

[0074] 表1

[0075] 以上所述仅为本发明的较佳实施例而已,并不用以限制本发明,凡在本发明的精

神和原则之内所作的任何修改、等同替换和改进等,均应包含在本发明的保护范围之内。

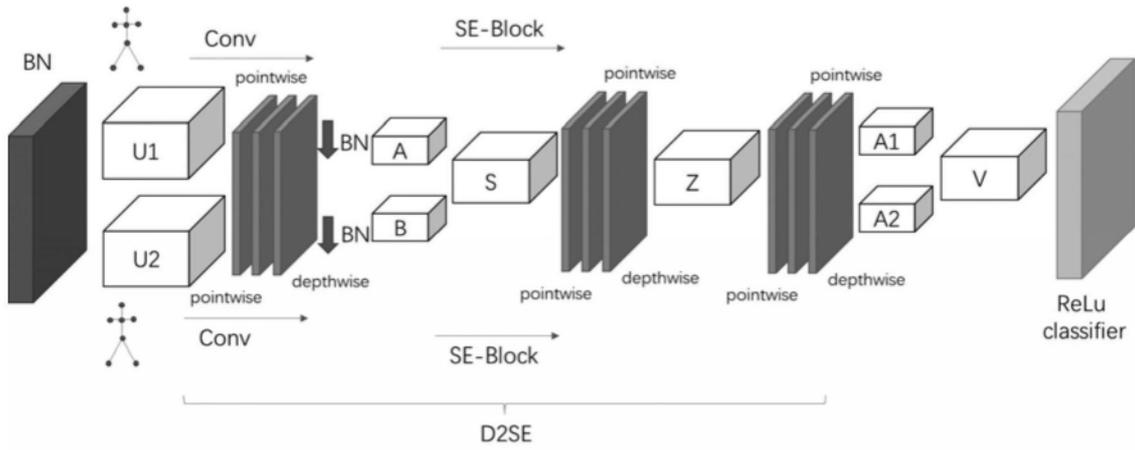


图1



图2

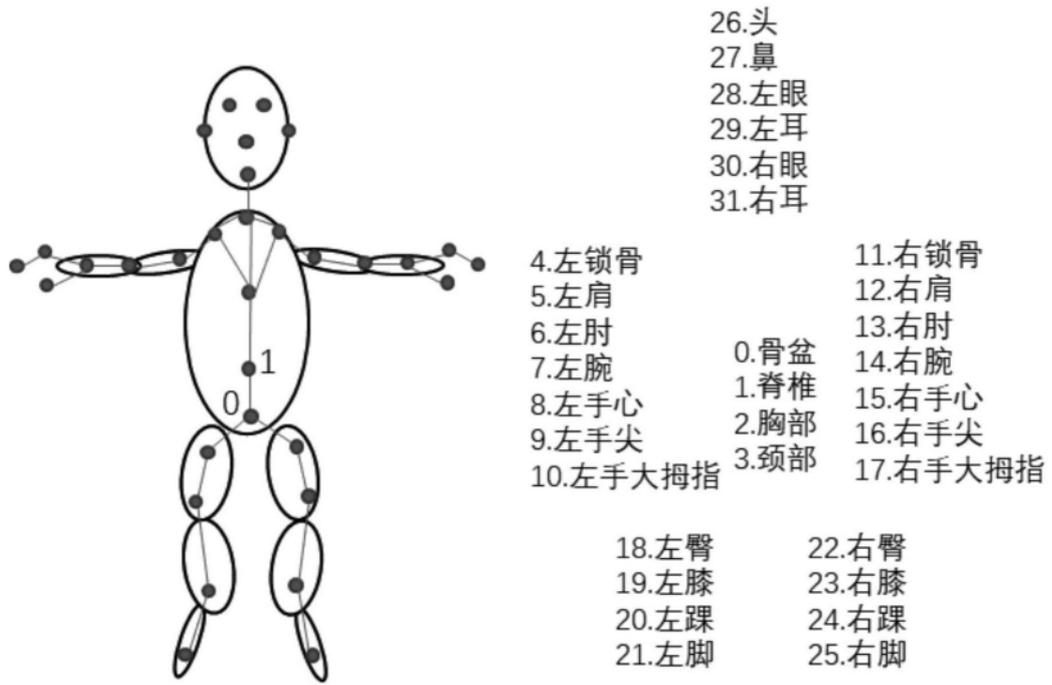


图3