Journal of System Simulation

Volume 36 | Issue 3 Article 3

3-15-2024

Human Action Recognition Based on Skeleton Edge Information Under Projection Subspace

Benyue Su
School of Computer and Information, Anqing Normal University, Anqing 246133, China; School of Mathematics and Computer,Tongling University, Tongling 244061, China, subenyue@sohu.com

Peng Zhang
School of Computer and Information, Anqing Normal University, Anqing 246133, China; School of Mathematics and Computer,Tongling University, Tongling 244061, China

Bangguo Zhu
School of Computer and Information, Anqing Normal University, Anqing 246133, China; School of Mathematics and Computer,Tongling University, Tongling 244061, China

Mengjuan Guo
School of Computer and Information, Anqing Normal University, Anqing 246133, China; School of Mathematics and Computer,Tongling University, Tongling 244061, China

See next page for additional authors

Human Action Recognition Based on Skeleton Edge Information Under Projection Subspace

Abstract

Abstract: In recent years, human action recognition based on skeleton data has received a lot of attention in the fields of computer vision and human-computer interaction. Most of the existing methods focus on modeling the skeleton points in the original 3D coordinate space. However, skeleton points ignore the physical chain structure of the human body itself, which makes it difficult to portray the local correlation of human motion. In addition, due to the diversity of camera views, it is difficult to explore the comprehensive representation of actions in different views under the original point-based 3D space. In view of this, this paper proposed an action recognition method based on skeleton edge information in the projection subspace. The method defined skeleton edge information combined with the body's own connection for capturing the spatial characteristics of the action. The direction and size information of skeleton edge motion was introduced on the basis of the skeleton edge information for capturing the temporal characteristics of the action. The 2D projection subspace was used for action characterization under different subspace perspectives. A suitable feature fusion strategy was explored, and the above features were extracted comprehensively through the improved CNN framework. Experimental results on two challenging large datasets NTU-RGB+D 60 (evaluation metrics are cross-subject and cross-view) and NTU-RGB+D 120 (evaluation metrics are cross-subject and cross-set) show that compared with the benchmark method, the proposed method improves the accuracy under the four metrics by 3.2%, 2.4%, 3.1%, and 5.8%, respectively.

Keywords

skeleton data, skeleton edges, edge direction, edge size, projection subspace

Authors

Benyue Su, Peng Zhang, Bangguo Zhu, Mengjuan Guo, and Min Sheng

Recommended Citation

Su Benyue, Zhang Peng, Zhu Bangguo, et al. Human Action Recognition Based on Skeleton Edge Information Under Projection Subspace[J]. Journal of System Simulation, 2024, 36(3): 555-563.

第 36 卷第 3 期 系统仿真学报© Vol. 36 No. 3
2024 年 3 月 Journal of System Simulation Mar. 2024

投影子空间下基于骨骼边信息的人体动作识别

苏本跃[1,2]，张鹏[1,2]，朱邦国[1,2]，郭梦娟[1,2]，盛敏[3]
(1. 安庆师范大学 计算机与信息学院，安徽 安庆 246133；2. 铜陵学院 数学与计算机学院，安徽 铜陵 244061；
3. 安庆师范大学 数理学院，安徽 安庆 246133)

摘要：近年来，基于骨骼数据的人体动作识别在计算机视觉、人机交互等领域受到了广泛的关注。现有的方法大多关注于在原始的3D坐标空间下对骨骼点进行建模。然而，骨骼点忽略了人体自身的物理链状结构，很难刻画人体运动的局部相关性；此外，由于相机视角的多样性，在原始的基于点的3D空间下难以探索动作在不同视角下的综合表征。鉴于此，*提出了一种投影子空间下基于骨骼边信息的动作识别方法。定义了结合人体自身连接的骨骼边信息，用于捕获动作的空间特性；在骨骼边信息的基础上引入了骨骼边运动的方向与大小信息，用于获取动作的时间特性；采用2D投影子空间的方式在不同的子空间视角下进行动作表征；探索了合适的特征融合策略，通过改进的CNN框架对上述特征进行综合提取*。在2个具有挑战性的大型数据集NTU-RGB+D 60(评价指标为cross-subject与cross-view)和NTU-RGB+D 120(评价指标为cross-subject与cross-set)上的实验结果表明，相比基准方法，所提方法在4个指标下精度分别提升了3.2%、2.4%、3.1%和5.8%。

关键词：骨骼数据；骨骼边；边方向；边大小；投影子空间
中图分类号：TP391.4 文献标志码：A 文章编号：1004-731X(2024)03-0555-09
DOI: 10.16182/j.issn1004731x.joss.22-1234
引用格式: 苏本跃, 张鹏, 朱邦国, 等. 投影子空间下基于骨骼边信息的人体动作识别[J]. 系统仿真学报, 2024, 36(3): 555-563.
Reference format: Su Benyue, Zhang Peng, Zhu Bangguo, et al. Human Action Recognition Based on Skeleton Edge Information Under Projection Subspace[J]. Journal of System Simulation, 2024, 36(3): 555-563.

Human Action Recognition Based on Skeleton Edge Information Under Projection Subspace

Su Benyue[1,2], *Zhang Peng*[1,2], *Zhu Bangguo*[1,2], *Guo Mengjuan*[1,2], *Sheng Min*[3]
(1. School of Computer and Information, Anqing Normal University, Anqing 246133, China; 2. School of Mathematics and Computer, Tongling University, Tongling 244061, China; 3. School of Mathematics and Physics, Anqing Normal University, Anqing 246133, China)

Abstract: In recent years, human action recognition based on skeleton data has received a lot of attention in the fields of computer vision and human-computer interaction. Most of the existing methods focus on modeling the skeleton points in the original 3D coordinate space. However, skeleton points ignore the physical chain structure of the human body itself, which makes it difficult to portray the local correlation of human motion. In addition, due to the diversity of camera views, it is difficult to explore the comprehensive representation of actions in different views under the original point-based 3D space. *In view of this, this paper proposed an action recognition method based on skeleton edge information in the projection subspace. The method defined skeleton edge information combined with the body's own connection for capturing the spatial characteristics of the action. The direction and size information of*

收稿日期：2022-10-17 修回日期：2023-01-31
基金项目：安徽省领军人才团队项目(皖教秘人［2019］16号)；安庆师范大学与铜陵学院联合培养研究生科研创新基金(tlaqsflhy2)
第一作者：苏本跃(1971-)，男，教授，博士，研究方向为机器学习与模式识别、图形图像处理。E-mail： subenyue@sohu.com

skeleton edge motion was introduced on the basis of the skeleton edge information for capturing the temporal characteristics of the action. The 2D projection subspace was used for action characterization under different subspace perspectives. A suitable feature fusion strategy was explored, and the above features were extracted comprehensively through the improved CNN framework. Experimental results on two challenging large datasets NTU-RGB+D 60 (evaluation metrics are cross-subject and cross-view) and NTU-RGB+D 120 (evaluation metrics are cross-subject and cross-set) show that compared with the benchmark method, the proposed method improves the accuracy under the four metrics by 3.2%, 2.4%, 3.1%, and 5.8%, respectively.

Keywords: skeleton data; skeleton edges; edge direction; edge size; projection subspace

0 引言

人体动作识别是指赋予计算机以“人的智能”，使得计算机能智能地识别出人的动作类型。随着数字图像处理和智能硬件制造技术的飞速发展，人体动作识别已广泛应用于人机交互、虚拟现实、工业系统、医疗保健和康复等领域[1]。与RGB视频数据和深度数据相比，骨骼数据具有简洁性、易于存储和受光照影响较小等优点[2]，在人体动作识别等方面受到了广泛的关注。

深度学习方法在基于骨骼数据的动作识别中的应用十分广泛，其结构大致分为三类，即循环神经网络(recurrent neural network, RNN)、图卷积网络(graph convolutional network, GCN)和卷积神经网络(convolutional neural network, CNN)[3]。其中，RNN架构将骨骼数据按时序处理为长向量的形式，仅仅关注于在时间上对动作进行建模;为了更好的获取动作的空间信息，GCN将骨骼数据根据关节点生成骨骼边表示的拓扑结构，通过学习节点之间的关系来提取动作特征，然而其复杂的图形结构会给卷积核的构建带来困难。CNN通过将关节和时间分别编码为类似图像的行和列的方式，将骨骼数据视为“伪图像”的形式来进行动作识别。相较于RNN和GCN，CNN可以用更小的代价同时建模动作的时间和空间特征。

在基于CNN的方法中，对骨骼数据的建模方式非常重要。一些研究者通过将骨骼数据的x，y，z坐标对应为图像的r，g，b通道的方式[4]，利用图像识别的方法进行动作识别。也有些研究者通过定义骨骼关节点轨迹、距离等高级特征来更具体的表征动作。然而这些方法大多关注于关节点自身的特征构建，很少关注人体骨架自身的物理结构，刻画人体的局部结构特性。

在现实场景中，通常在一个相机的不同位置获取一个动作，不同位置的同一个动作可能会产生不同的识别结果，动作的视角问题也同样值得关注[5]。为了解决视角的一致性问题，有些研究者们通过将不同视角下的坐标进行旋转规整到一个视图下[6]，然而在规整动作需要对齐相关位置，可能会导致动作的混淆。也有些研究者通过增加不同位置的相机以丰富不同的视图信息来获取动作的更精确表达，但是在实际应用中需要付出非常昂贵的代价[7]。

综合以上考虑，本文提出一种投影子空间下基于骨骼边信息的人体动作识别方法(subspace edge CNN, SE-CNN)。首先定义了骨骼边及骨骼边的运动信息(边运动的大小和方向)分别刻画人体运动的时空特性，其中骨骼边可以结合人体自身的物理连接，在空间上建模人体局部坐标结构；骨骼边的大小和方向能够在时间上探索边的运动变化。然后，通过构建骨骼边的投影子空间的方式来解决不同视角下动作识别问题。最后通过改进的CNN框架来对动作进行最终的分类。

1 方法

本节对CNN的骨骼动作识别流程进行介绍，阐述本文提出的投影子空间下基于骨骼边信息框

DOI: 10.16182/j.issn1004731x.joss.22-1234

架，包括骨骼边信息及其运动方向和大小信息的定义、投影子空间的构建和SE-CNN网络的具体结构。具体的网络框架见图1，网络输入包含边信息、边运动的方向和大小信息(经过投影后分别表示为包含3个“32×24×2”的矩阵，依次表示为时间帧、边和二维坐标)，分别经过坐标级、边级和视图级的特征提取，经过全连接层输出最终的结果。图1中，*xy*，*xz*，*yz*分别表示3D原始坐标*xyz*在2D空间上的投影。

1.1 基于CNN的骨骼数据动作识别

CNN骨骼动作识别流程图如图2所示：首先将骨骼序列表达为“伪图像”的形式，然后再通过相关模型提取特征，最后经过全连接层得到最终的分类结果。以基于关节点的建模为例，将骨骼序列的关节点和时间表达为伪图像的行和列，关节点的坐标维度视为特征图通道的方式进行输入构建。具体地，有$Input=R^{C\times T\times V\times M}$，其中，$C$为通道，即关节点的$x,y,z$坐标维度，大小为3；$T$为时间帧，一般有统一的帧数处理；$V$为关节维度，表示不同的关节点，数目为25；$M$为人数，在单人动作中为1，在双人动作中为2。

1.2 基于骨骼边的时空特征构建

1.2.1 骨骼边

与传统的基于关节点自身定义的特征不同的是，本文采用相邻关节点之间形成的骨骼边，并在此基础上进行空间特征的构建。如图3所示，本文将人体运动中较稳定的脊柱底部点“1”作为中心点，结合人体物理结构，在中心点向外的方向趋势上构建关节点之间连接形成骨骼边向量，总数为24。特别地，由于CNN需要构建伪图像模型，假设图像的行代表关节维度(此处为骨骼边)，本文将相同部位的骨骼边分别按照躯干、左上肢、右上肢、左下肢和右下肢的顺序依次进行行排列，不同的部位在图3中有不同的颜色。基于骨骼边的特征构建，可以充分利用人体的自然连接，在空间上捕捉人体动作的变化。

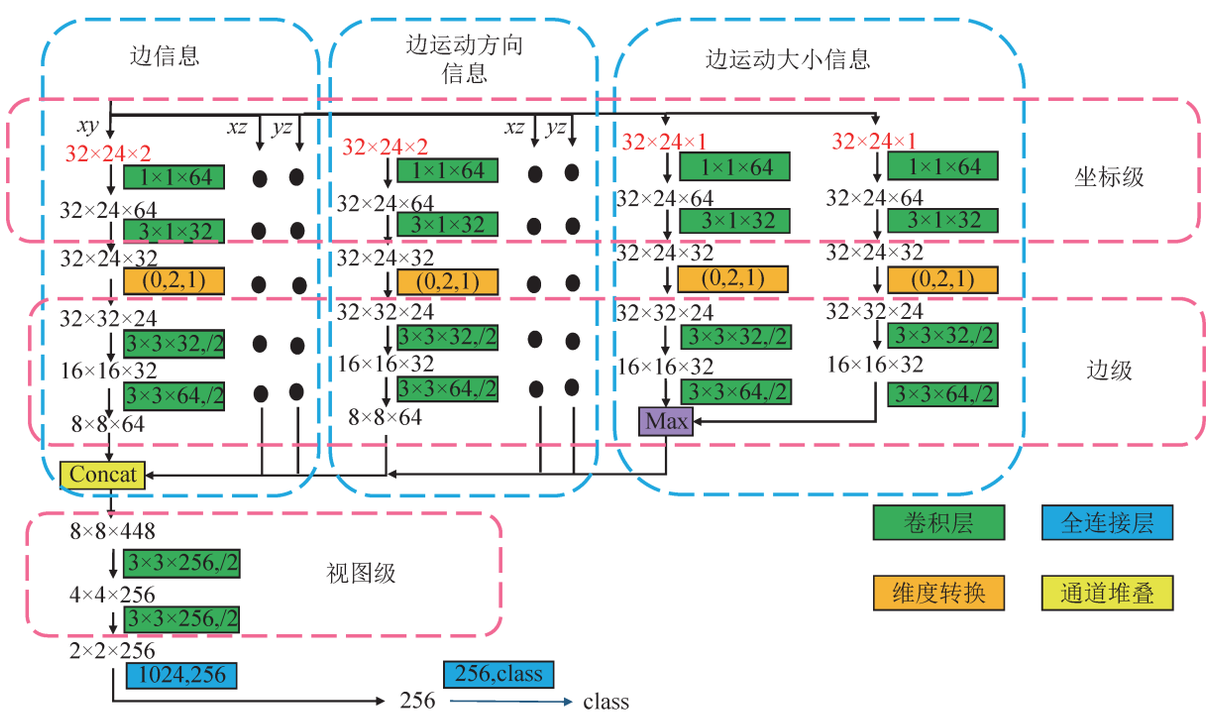

图1 SE-CNN框架图
Fig. 1 SE-CNN framework

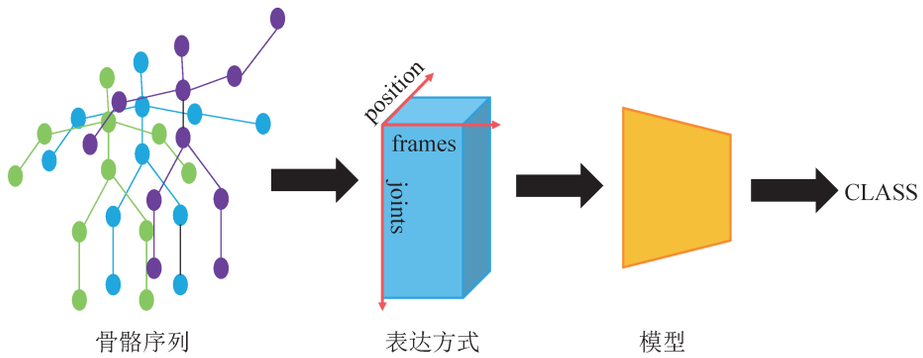

图2 基于CNN的骨骼动作识别流程
Fig. 2 CNN-based skeleton action recognition flow

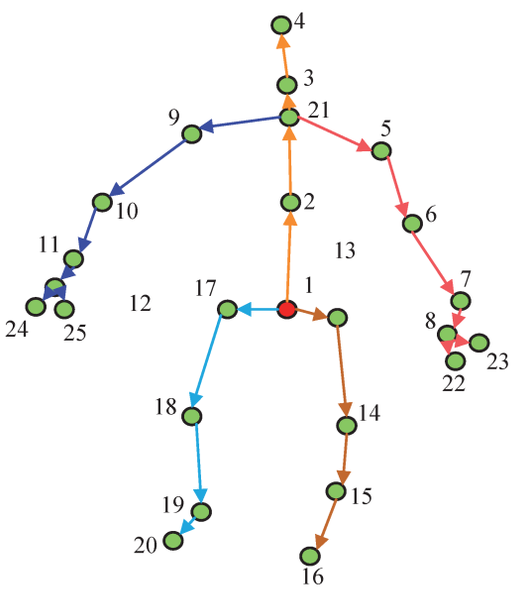

图3 骨骼边的构建
Fig. 3 Construction of skeleton edges

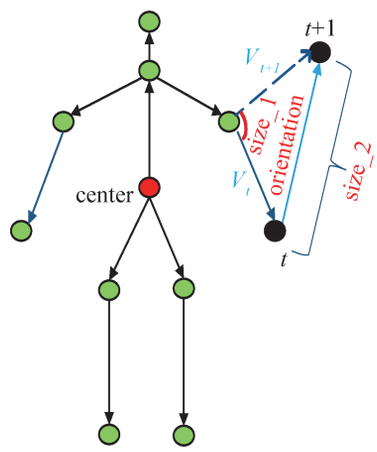

图4 骨骼边的方向和大小信息构建示意图
Fig. 4 Direction and size information of constructed skeleton edges

假设$J_{m,t}$为第m个关节点第t帧的骨骼点坐标，(m,n)为m和n两个相邻关节；S为相邻关节点的集合；T为最终帧数，则骨骼边序列E可表示为

$$E=[J_{m,t}-J_{n,t}],\ (m,n)\in S,\ t=1,2,\cdots,T \tag{1}$$

1.2.2 骨骼边运动的方向和大小

在1.2.1节中，本文采用构建骨骼边的方式表示动作的空间特性，为了更好的探索骨骼边在时间上的变化，本文提出了骨骼边运动的大小与方向信息，如图4所示。图4中，以手肘和手腕节点为例，t和t+1表示时刻，V_t表示t时刻的骨骼边，V_{t+1}表示t+1时刻的骨骼边，orientation向量为骨骼边的方向，size代表骨骼边大小。

边运动的方向：根据骨骼边向量的固有特性，将边的后一帧和前一帧相减形成的边运动矢量，称为边运动的方向信息。通过边的方向信息可以表示骨骼边的时间变化，称为Eo。

边运动的大小：

(1) 根据先验知识，对于每条构建的骨骼边向量，运动更多涉及到边向量终点表示的关节点的变化。因此，本文将每个骨骼边向量对应的关节点(向量终点)在单位时间的变化作为骨骼边在时间上变化的刻画之一。以手肘和手腕连接形成的骨骼边向量为例，手腕是“终点”，所以用手腕关节点在单位时间内变化的距离D作为该骨骼边的大小，所有D的集合称为Es_1。

(2) 为了在多尺度下表示边运动的大小，类似于(1)的方式，本文又提出了骨骼边在单位时间内

变化产生的角度特征，即关节角A来表征边在时间上变化。所有的A集合称为Es_2。

具体表示为

$$Eo = J_{t+1} - J_t,\ t = 1, 2, \cdots, T \tag{2}$$

$$Es\_1 = \{D_1, D_2, \cdots, D_n\},\ n = 1, 2, \cdots, 24 \tag{3}$$

$$D_n = \sqrt{(x_{n,t+1} - x_{n,t})^2 + (y_{n,t+1} - y_{n,t})^2 + (z_{n,t+1} - z_{n,t})^2} \tag{4}$$

$$Es\_2 = \{A_1, A_2, \cdots, A_n\} \tag{5}$$

$$A_n = \cos^{-1} \frac{e_{n,t+1} e_{n,t}}{\|e_{n,t+1}\| \|e_{n,t}\|} \tag{6}$$

式中：e_n为第n组相邻骨骼点组成的骨骼边。

1.3 基于骨骼边的子空间投影网络

本节将具体介绍SE-CNN的具体网络结构，主要包括坐标级、边级和视图级聚合。

1.3.1 3D投影子空间构建

将原始的3D骨骼边序列沿着*XOY*、*XOZ*和*YOZ*3个平面进行投影得到3个2D空间的边子序列。

具体地，如图1所示，将构建的24条边及其对应的边方向分别向3个坐标平面投影，得到3个子空间视角下的6组特征信息，在CNN网络中通过维度拼接的方式进行不同视角下的特征融合。

1.3.2 坐标级特征聚合

在基于CNN的方法中，跟大多数研究者选择对关节点和时间帧建模的方式不同的是，本文选择对骨骼边和时间帧进行建模。

包含2个卷积核大小分别为1×1和3×1的卷积层。其中，时间维度的卷积核大小分别为1和3，可以获取不同时间尺度的关节坐标特征；骨骼边维度的卷积核大小都为1，是为了保证在卷积前后骨骼边维度的信息保持相对稳定(以行和列分别表示伪图像的关节和时间帧维度为例，卷积前后伪图像的行数是不变的，即每一行代表每条骨骼边信息综合坐标的高级特征)。

1.3.3 边级特征聚合

人体动作往往只涉及到特定部位集的变化。例如“穿鞋”这个动作，手部骨骼边和脚部骨骼边组成的骨骼边集合就可以很好的反映出该动作。由于手部和脚部距离较远，受限于规整卷积核的局限性，难以有效的提取特征。

在经过2个坐标级的卷积层后，本文增加了一个维度转换，将骨骼边维度转换到坐标维度。通过这种方式，可以全局聚合骨骼边级特征。在这个模块中包含2个卷积核大小均为3×3的卷积层，通道维度(表示骨骼边特征)分别为32和64。

1.3.4 视图级特征聚合

值得注意的是，本文首先在3D边信息及其运动的方向信息的2D投影子空间中进行坐标级和边级特征聚合，然后结合边运动的方向信息，形成7路包含多视图信息的特征信息后进行堆叠。

本模块中，同样包含2个卷积核大小均为3×3的卷积层，通道中包含不同投影子空间下的视图信息，从而可以全局聚合骨骼的不同视图信息。

2 实验与结果

2.1 数据集及实验细节

NTU-RGB+D 60数据集[8]是目前大规模的室内人体动作识别数据集，由南洋理工大学Rose实验室创建，包含RGB视频、深度图序列、3D骨骼数据和红外视频数据。它由3个方向分别为−45°、0°和+45°的Kinect 2.0摄像头组成，共40个实验者执行60个动作，总样本量为56 880，包括跨对象(cross-subject, cs)和跨视角(cross-view, cv)评估标准。具体来说，在cs评估中，实验对象的一半样本用作训练集，另一半样本用作测试集；在cv评估中，摄像机2和摄像机3采集的样本作为训练集，摄像机1采集的样本作为测试集。

NTU-RGB+D 120数据集[9]是目前最大的人类动作识别骨架数据集，是NTURGB+D 60数据集的扩展。它包含114 480个序列，由106个主体表

演 120 个动作类别组成。骨骼信息包括 25 个/帧不同关节的三维位置。2 种标准评价分别进行跨对象评估和跨设置(corss-set, cset)评价。对于 cs 评估，将 106 名受试者分为训练集和测试集。每组由 53 名受试者组成。对于 cset 评估，选择数据集设置 id 为偶数的样本进行训练，数据集设置 id 为奇数的样本进行测试。

在数据处理中，本文采用文献[10]的方法，从整个序列中进行随机裁剪，对于不同长度的序列，在时间帧的维度上进行双线性插值，最后归一化到一个固定的长度 32。在训练中，epoch 轮数设置为 600，并选择 Adam 优化器。所有实验都在一个 RTX 6000 GPU 上进行，并使用 Pytorch 深度学习框架。

2.2 骨骼边的有效性

为探索骨骼边在动作识别中的有效性，本文用骨骼点和骨骼边进行比较，在 NTU-RGB+D 60 和 NTU-RGB+D 120 数据集下的实验结果见表 1。

表 1 骨骼点和骨骼边在 NTU-RGB+D 60 和 NTU-RGB+D 120 数据集下的比较

Table 1 Comparison of skeleton edges and skeleton points on NTU-RGB+D 60 and NTU-RGB+D 120 datasets %

实验方法	NTU-RGB+D 60		NTU-RGB+D 120	
	cs	cv	cs	cset
骨骼点	80.6	86.2	72.3	73.2
骨骼边	81.6	87.3	73.1	75.0

由表 1 可知，在 NTU-RGB+D 60 数据集的 2 个指标下骨骼边的识别率分别为 81.6% 和 87.3%，同比骨骼点增长 1.0% 和 1.1%。在 NTU-RGB+D 120 数据下，骨骼边在 2 个指标下的识别率分别为 73.1% 和 75.0%，同比骨骼点增长 0.8% 和 1.8%。可以观察到同一基准下骨骼边优于骨骼点，充分说明结合人体自身连接构建的骨骼边的优越性。

2.3 骨骼边方向、大小和投影子空间的消融实验及分析

本文对提出的基于骨骼边的方向、大小和投影子空间信息分别进行递增消融实验，在 NTU-RGB+D 60 和 NTU-RGB+D 120 数据集下的实验结果见表 2。

表 2 骨骼边方向、大小和投影子空间的有效性验证

Table 2 Validation of skeleton edge direction, size, and projection subspace

实验方法	NTU-RGB+D 60/%		NTU-RGB+D 120/%		NTU-RGB+D 60 (for cs metric)
	cs	cv	cs	cset	one epoch training time for models/s
骨骼边	81.6	87.3	73.1	75.0	13
骨骼边方向	84.2	90.5	77.5	79.8	17
骨骼边大小	86.1	91.2	78.1	81.0	23
投影子空间	87.7	91.5	79.6	82.4	49

由表 2 可知，在 NTU-RGB+D 60 和 NTU-RGB+D 120 数据集下，本文提出的骨骼边方向、大小和投影子空间方法在不同的指标中均具有优越性。

具体地，对于 NTU-RGB+D 60 数据集，在 cs 指标下，原始的骨骼边的识别率为 81.6%，分别增加了骨骼边的方向、大小和投影子空间后，达到了 84.2%、86.1% 和 87.7%，同比增长 2.6%、4.5% 和 6.1%；在 cv 指标下，原始的骨骼边的识别率为 87.3%，分别增加了骨骼边的方向、大小和投影子空间后，达到了 90.5%、91.2% 和 91.5%，同比增长 3.2%、3.9% 和 4.2%。对于 NTU-RGB+D 120 数据集，在 cs 指标下，原始的骨骼边的识别率为 73.1%，分别增加了骨骼边的方向、大小和投影子空间后，达到了 77.5%、78.1% 和 79.6%，同比增长 4.4%、5.0% 和 6.5%；在 cset 指标下，原始的骨骼边的识别率为 75.0%，分别增加了骨骼边的方向、大小和投影子空间后，达到了 79.8%、81.0% 和 82.4%，同比增长 4.8%、6.0% 和 7.4%。

DOI: 10.16182/j.issn1004731x.joss.22-1234

特别地，为了探索模型添加在算法效率上的变化，在NTU-RGB+D 60 的cs指标下，本文通过对SE-CNN的相关模块添加后一轮模型训练时长在分析算法损失，由实验可知本文添加的边大小、边方向和子空间投影在提升识别效果的同时也会带来一定的算法效率损失。综上所述，可以说明本文在骨骼边中提出的骨骼边方向、大小和投影子空间具有有效性。

2.4 与其他先进方法的比较

本文在2个数据集下对比了提出的SE-CNN方法和最近方法的性能，具体见表3~4。

表3 本文方法与其他方法在NTU-RGB+D 60数据集上的实验结果对比

Table 3 Comparison of experimental results between proposed method and other methods on NTU-RGB+D 60 dataset

方法	年份	cs/%	cv/%
Deep LSTM[8]	2016	60.7	67.3
VA-LSTM[11]	2017	79.2	87.7
TCN[12]	2018	74.3	83.1
ElAtt-GRU[13]	2018	80.7	88.4
ARRN-GRU[14]	2018	80.7	88.8
ST-GCN[15]	2018	81.5	88.3
AS-GCN[16]	2019	86.8	94.2
PA-GCN[17]	2020	80.4	82.7
LSTM+GCN[18]	2020	84.8	90.2
CNN+LSTM[19]	2021	79.2	85.6
DIF-CNN[20]	2021	81.0	85.8
AFE-CNN[21]	2022	86.2	**92.2**
HCN[10] (base)	2018	84.5	89.1
本文方法		**87.7**	91.5

由表3可知，在基于RNN的方法[8,11-14]中，“ElAttt-GRU[13]”是具有代表性的方法，本文方法与其他相比，在cs和cv指标下的识别率分别有7.0%和3.1%的提升；在基于GCN的方法[15-18]中，本文方法也比经典方法“ST-GCN[15]”在2个指标下高6.2%和3.2%；在基于CNN的方法中，本文方法较基准方法[10]在2个指标下分别高3.2%和2.4%，较最新方法“AFE-CNN[21]”在cross-subject指标下高1.5%，在cv指标下略有下降，说明本文方法对不同对象之间的识别更有优势。也有些方法混合多种框架[18-19]，结果表明本文提出的纯CNN的架构同样具有竞争力。

表4 本文方法与其他方法在NTU-RGB+D 120数据集上的实验结果对比

Table 4 Comparison of experimental results between proposed method and other methods on NTU-RGB+D 120 dataset

方法	年份	cs/%	cset/%
ST-LSTM[22]	2016	56.5	54.1
GCA-LSTM[23]	2017	58.3	59.2
Two-stream network[9]	2019	62.2	61.8
ST-GCN[24]	2018	70.7	73.2
AS-GCN[25]	2019	77.7	78.9
STF-GCN[26]	2021	76.7	79.0
Synthesized CNN[27]	2017	60.3	63.2
Clips+CNN+MTCN[28]	2018	62.2	61.8
SGN[29]	2020	79.2	81.5
AFE-CNN[21]	2022	**80.4**	81.6
HCN[10] (base)	2018	76.5	76.6
本文方法		79.6	**82.4**

由表4可知，基于RNN的方法[9,22-23]近年来研究较少，Two-stream network[9]是该数据集中提供的算法，在2个指标下分别达到了62.2%和61.8%的精度；对于GCN架构[24-26]，本文方法较经典算法ST-GCN[24]在2个指标下分别高8.9%和9.2%；对于基于CNN的方法[10,21,27-29]，本文方法较基准方法在2个指标下分别提升3.1%和5.8%。特别地，本文方法较最新方法“AFE-CNN[21]”在cross-subject指标下略低，cross-set指标略高，说明本文方法对相机高度、远近等更具鲁棒性。

3 结论

本文提出了一种投影子空间下基于骨骼边信息的动作识别方法。首先利用人体物理结构构建的骨骼边信息，在此基础上引出了骨骼边的方向和大小信息，最后在3D动作的二维投影空间下进行动作识别。实验结果表明，本文方法在一定程度上对动作的局部特性具有更好的表征，对视角的多样性具有补充。同时本文的方法也有不足之

处，本文将边级、坐标级和视图级视为同等的重要性，而在一个动作中，不同的边、坐标和视图下可能发挥不同的作用，对动作的表征不够完善。未来工作中，将考虑引入边、坐标和视图的注意力机制，用来探索动作中不同模块发挥作用的重要性。

参考文献：

[1] Jacek Trelinski, Bogdan Kwolek. CNN-based and DTW Features for Human Activity Recognition on Depth Maps [J]. Neural Computing and Applications, 2021, 33(21): 14551-14563.

[2] 刘云, 薛盼盼, 李辉, 等. 基于深度学习的关节点行为识别综述[J]. 电子与信息学报, 2021, 43(6): 1789-1802.
Liu Yun, Xue Panpan, Li Hui, et al. A Review of Action Recognition Using Joints Based on Deep Learning[J]. Journal of Electronics & Information Technology, 2021, 43(6): 1789-1802.

[3] 苏本跃, 孙满贞, 马庆, 等. 单视角下基于投影子空间视图的动作识别方法[J]. 系统仿真学报, 2023, 35(5): 1098-1108.
Su Benyue, Sun Manzhen, Ma Qing, et al. Action Recognition Method Based on Projection Subspace Views under Single Viewing Angle[J]. Journal of System Simulation, 2023, 35(5): 1098-1108.

[4] Du Yong, Fu Yun, Wang Liang. Skeleton Based Action Recognition with Convolutional Neural Network[C]// 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR). Piscataway, NJ, USA: IEEE, 2015: 579-583.

[5] 赵瑛, 陆耀, 张健, 等. 基于深度神经网络的多视角人体动作识别[J]. 系统仿真学报, 2021, 33(5): 1019-1030.
Zhao Ying, Lu Yao, Zhang Jian, et al. Multi-view Human Action Recognition Based on Deep Neural Network[J]. Journal of System Simulation, 2021, 33(5): 1019-1030.

[6] Xu Kailin, Ye Fanfan, Zhong Qiaoyong, et al. Topology-aware Convolutional Neural Network for Efficient Skeleton-based Action Recognition[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto, CA, USA: AAAI Press, 2022: 2866-2874.

[7] Daniel Weinland, Mustafa Özuysal, Pascal Fua. Making Action Recognition Robust to Occlusions and Viewpoint Changes[C]//Computer Vision-ECCV 2010. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010: 635-648.

[8] Amir Shahroudy, Liu Jun, Tian Tsong Ng, et al. NTU RGB+D: A Large Scale Dataset for 3D Human Activity Analysis[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2016: 1010-1019.

[9] Liu Jun, Amir Shahroudy, Mauricio Perez, et al. NTU RGB+D 120: A Large-scale Benchmark for 3D Human Activity Understanding[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(10): 2684-2701.

[10] Li Chao, Zhong Qiaoyong, Xie Di, et al. Co-occurrence Feature Learning from Skeleton Data for Action Recognition and Detection with Hierarchical Aggregation [C]//Proceedings of the 27th International Joint Conference on Artificial Intelligence. Palo Alto, CA, USA: AAAI Press, 2018: 786-792.

[11] Zhang Pengfei, Lan Cuiling, Xing Junliang, et al. View Adaptive Recurrent Neural Networks for High Performance Human Action Recognition from Skeleton Data[C]//2017 IEEE International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2017: 2136-2145.

[12] Kim T S, Reiter A. Interpretable 3D Human Action Analysis with Temporal Convolutional Networks[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway, NJ, USA: IEEE, 2017: 1623-1631.

[13] Zhang Pengfei, Xue Jianru, Lan Cuiling, et al. Adding Attentiveness to the Neurons in Recurrent Neural Networks[C]//Computer Vision-ECCV 2018. Cham: Springer International Publishing, 2018: 136-152.

[14] Li Lin, Zheng Wu, Zhang Zhaoxiang, et al. Relational Network for Skeleton-based Action Recognition[EB/OL]. (2019-04-11) [2022-08-30]. https://arxiv. org/abs/1805.02556.

[15] Yan Sijie, Xiong Yuanjun, Lin Dahua. Spatial Temporal Graph Convolutional Networks for Skeleton-based Action Recognition[C]//Proceedings of the Thirty-second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence. Palo Alto, CA, USA: AAAI Press, 2018: 912.

[16] Li Maosen, Chen Siheng, Chen Xu, et al. Actional-structural Graph Convolutional Networks for Skeleton-based Action Recognition[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2019: 3590-3598.

[17] Qin Yang, Mo Lingfei, Li Chenyang, et al. Skeleton-based Action Recognition by Part-aware Graph Convolutional Networks[J]. The Visual Computer, 2020, 36(3): 621-631.

[18] Si Chenyang, Jing Ya, Wang Wei, et al. Skeleton-based

DOI: 10.16182/j.issn1004731x.joss.22-1234

Action Recognition with Hierarchical Spatial Reasoning and Temporal Stack Learning Network[J]. Pattern Recognition, 2020, 107: 107511.

[19] Zhuang Tianming, Zhao Pengbiao, Xiao Peng, et al. Multi-stream CNN-LSTM Network with Partition Strategy for Human Action Recognition[C]//Proceedings of the 2021 International Conference on Bioinformatics and Intelligent Computing. New York, NY, USA: Association for Computing Machinery, 2021: 431-435.

[20] Chen Han, Jiang Yifan, Hanseok Ko. Action Recognition with Domain Invariant Features of Skeleton Image[C]// 2021 17th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). Piscataway, NJ, USA: IEEE, 2021: 1-7.

[21] Guan Shannan, Lu Haiyan, Zhu Linchao, et al. AFE-CNN: 3D Skeleton-based Action Recognition with Action Feature Enhancement[J]. Neurocomputing, 2022, 514: 256-267.

[22] Liu Jun, Amir Shahroudy, Xu Dong, et al. Spatio-temporal LSTM with Trust Gates for 3D Human Action Recognition[C]//Computer Vision-ECCV 2016. Cham: Springer International Publishing, 2016: 816-833.

[23] Liu Jun, Wang Gang, Duan Lingyu, et al. Skeleton-based Human Action Recognition with Global Context-Aware Attention LSTM Networks[J]. IEEE Transactions on Image Processing, 2018, 27(4): 1586-1599.

[24] Yan Sijie, Xiong Yuanjun, Lin Dahua. Spatial Temporal Graph Convolutional Networks for Skeleton-based Action Recognition[C]//Proceedings of the Thirty-second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence. Palo Alto, CA, USA: AAAI Press, 2018: 912.

[25] Li Maosen, Chen Siheng, Chen Xu, et al. Actional-structural Graph Convolutional Networks for Skeleton-based Action Recognition[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2019: 3590-3598.

[26] Liu Xing, Li Yanshan, Xia Rongjie. Adaptive Multi-view Graph Convolutional Networks for Skeleton-based Action Recognition[J]. Neurocomputing, 2021, 444: 288-300.

[27] Liu Mengyuan, Liu Hong, Chen Chen. Enhanced Skeleton Visualization for View Invariant Human Action Recognition[J]. Pattern Recognition, 2017, 68: 346-362.

[28] Ke Qiuhong, Mohammed Bennamoun, An Senjian, et al. Learning Clip Representations for Skeleton-based 3D Action Recognition[J]. IEEE Transactions on Image Processing, 2018, 27(6): 2842-2855.

[29] Zhang Pengfei, Lan Cuiling, Zeng Wenjun, et al. Semantics-guided Neural Networks for Efficient Skeleton-based Human Action Recognition[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2020: 1109-1118.

