

6-28-2024

Fusing Rotation Angle Coding in Spherical Space for Human Action Recognition

Benyue Su

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

Banguo Zhu

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

Mengjuan Guo

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

Min Sheng

School of Mathematics and Physics, Anqing Normal University, Anqing 246133, China

Follow this and additional works at: <https://dc-china-simulation.researchcommons.org/journal>



Part of the [Artificial Intelligence and Robotics Commons](#), [Computer Engineering Commons](#), [Numerical Analysis and Scientific Computing Commons](#), [Operations Research, Systems Engineering and Industrial Engineering Commons](#), and the [Systems Science Commons](#)

This Paper is brought to you for free and open access by Journal of System Simulation. It has been accepted for inclusion in Journal of System Simulation by an authorized editor of Journal of System Simulation. For more information, please contact xtfzxb@126.com.

Fusing Rotation Angle Coding in Spherical Space for Human Action Recognition

Abstract

Abstract: The existing human action recognition methods focus more on the translation information such as the coordinates and displacements of skeleton structure, and pay less attention to the motion trend of skeleton structure and the rotation information representing the motion direction of joints and bones. A spatio-temporal convolutional neural network method combining the rotation angle coding in spherical space is introduced. The angle information with scale invariance is obtained by mapping the human action in three-dimensional spherical space, and the dynamic angular velocity information is extracted as the angle code to represent the rotation information of joints and bones in the action trajectory. A spatiotemporal feature extraction and co-occurrence module(STCN) is constructed to better capture the spatiotemporal features of data. A suitable fusion strategy is utilized to fuse the translation features and rotation features. The experimental results show that the rotation angle coding benefits the accuracy improvement of motion representation and the effectiveness of the spatio-temporal feature extraction and co-occurrence module.

Keywords

human action recognition, skeleton data, rotation angle encoding, 3D spherical space, spatial-temporal feature

Recommended Citation

Su Benyue, Zhu Bangguo, Guo Mengjuan, et al. Fusing Rotation Angle Coding in Spherical Space for Human Action Recognition[J]. Journal of System Simulation, 2024, 36(6): 1433-1441.

融合球空间下旋转角度编码的人体动作识别

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

摘要: 针对现有的人体动作识别方法较多考虑骨架结构的坐标和位移等平移信息, 较少关注代表骨架结构的运动趋势以及代表关节、骨骼运动方向的旋转信息, 提出一种融合球空间下旋转角度编码的时空卷积神经网络方法。通过人体动作在三维球空间中的映射, 获取具有尺度不变性的角度信息, 提取其动态角速度信息作为角度编码, 表征动作轨迹中关节点和骨骼边的旋转信息; 构建了时空特征提取与共现模块来更好地捕获数据的时空特征; 用合适的融合策略对平移特征和旋转特征进行运动特征融合。实验结果证明了旋转角度编码有利于提升运动表征的准确性, 以及时空特征提取与共现模块的有效性。

关键词: 人体动作识别; 骨架数据; 旋转角度编码; 3D球空间; 时空特征

中图分类号: TP391.9 文献标志码: A 文章编号: 1004-731X(2024)06-1433-09

DOI: 10.16182/j.issn1004731x.joss.23-1263

引用格式: 苏本跃, 朱邦国, 郭梦娟, 等. 融合球空间下旋转角度编码的人体动作识别[J]. 系统仿真学报, 2024, 36(6): 1433-1441.

Reference format: Su Benyue, Zhu Bangguo, Guo Mengjuan, et al. Fusing Rotation Angle Coding in Spherical Space for Human Action Recognition[J]. Journal of System Simulation, 2024, 36(6): 1433-1441.

Fusing Rotation Angle Coding in Spherical Space for Human Action Recognition

Su Benyue^{1,2}, Zhu Bangguo^{1,2}, Guo Mengjuan^{1,2}, Sheng Min³(1. School of Mathematics and Computer, Tongling University, Tongling 244061, China;
2. School of Computer and Information, Anqing Normal University, Anqing 246133, China;
3. School of Mathematics and Physics, Anqing Normal University, Anqing 246133, China)

Abstract: The existing human action recognition methods focus more on the translation information such as the coordinates and displacements of skeleton structure, and pay less attention to the motion trend of skeleton structure and the rotation information representing the motion direction of joints and bones. A spatio-temporal convolutional neural network method combining the rotation angle coding in spherical space is introduced. The angle information with scale invariance is obtained by mapping the human action in three-dimensional spherical space, and the dynamic angular velocity information is extracted as the angle code to represent the rotation information of joints and bones in the action trajectory. A spatio-temporal feature extraction and co-occurrence module (STCN) is constructed to better capture the spatio-temporal features of data. A suitable fusion strategy is utilized to fuse the translation features and rotation features. The experimental results show that the rotation angle coding benefits the accuracy improvement of motion representation and the effectiveness of the spatio-temporal feature extraction and co-occurrence module.

收稿日期: 2023-10-18 修回日期: 2024-01-04

基金项目: 安徽省领军人才团队(皖教秘人[2019]16号); 安徽省高校优秀科研创新团队(2023AH010056); 铜陵学院联合培养研究生科研创新基金(22tlaqsfly1)

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

Keywords: human action recognition; skeleton data; rotation angle encoding; 3D spherical space; spatial-temporal feature

0 引言

人体动作识别被广泛应用于运动分析、机器人技术、医疗保健和医疗康复等领域^[1]。由人体关节坐标序列构成的骨架数据, 表征人体动作的时空运动时具有直接性, 对复杂背景信息的鲁棒性、存储需求低和易处理等优势, 使基于骨架数据的人体动作识别具有广泛的应用前景^[2]。

基于骨架数据的动作识别深度学习方法大致可分为 3 类: 考虑时序关系的循环神经网络(RNN)、基于拓扑图结构的图卷积网络(GCN)和基于伪图像的卷积神经网络(CNN)^[3-4]。其中, RNN 以长向量的形式在时间维度对骨架数据进行建模^[5]。考虑到人体的空间结构信息, GCN 引入人体拓扑图结构^[6], 通过图结构引导的信息聚合来提取动作空间特征。但由于图结构的复杂性, 造成 GCN 中卷积核的尺寸受图结构约束, 使模型缺乏灵活性^[7-8]。CNN 相较 RNN 和 GCN, 以伪图像^[9-10]的形式对骨架数据的时空信息进行编码, 其模型具有轻量、灵活性高、且易于构建^[11-12]等优点。

基于 CNN 的方法提取骨架数据中的时空信息进行动作识别, 其关键在于准确地刻画骨架序列中的时空运动特征^[13]。传统方法多仅对骨架序列的关节坐标, 或人体沿物理连接方向上关节之间的向量差(骨骼边)进行建模^[14-15]。其中, 文献[4]提出了基于节点时间特征学习与共现网络框架(HCN), 对动作的平移信息进行建模。除平移运动外, 人体结构在运动过程中还伴随着关节和骨骼边的转动, 如“弯腰”可视为人体绕尾椎位置转动过程, “拿手机”可抽象为手指指节相对人体结构的连续转动, 表明人体运动中还蕴含着丰富的旋转信息。文献[16]利用两个相邻骨骼边的角度作为依据选择运动信息最大的关节进行动作识别。文献[17]利用全局坐标系下的三维关节角度作为特

征向量进行识别。文献[18]提出一种具有旋转不变性的识别框架, 针对人体关节向量与以人为中心的直角坐标轴间的夹角进行建模。然而, 这些方法多仅在欧氏空间中对旋转角度进行建模。考虑到肢体的运动通常具有高自由度的特点, 在同样具有高自由度的球空间, 可由具有尺度不变性的角度矢量表征肢体结构旋转运动的大小和方向。

在实际场景中, 人体动作可抽象为铰接的肢体构件^[19]的连续运动, 其两个关键要素是各肢体构件随时间变化在空间中的旋转和平移。为准确地刻画动作的旋转信息, 本文提出了一种融合球空间下旋转角度编码的人体动作识别方法(rotational angle encoding spatial-temporal CNN, R-STCNN)。引入人体动作在球空间中动态与静态的旋转角度编码, 通过将人体运动映射到以人为中心的三维超球体, 采用具有尺度不变性的空间角度作为单位矢量, 以旋转角度编码的方式表征骨架数据的关节点和骨骼边的旋转信息。考虑到动作的平移和旋转两种模态在时间维度上的变化, 以静态特征为基础, 进一步刻画了动作的平移速度特征和角速度特征。考虑到人体动作是旋转运动与平移运动的复合过程, 在 R-STCNN 框架中对平移信息和旋转信息进行运动级特征融合, 实现对动作进行识别分类。

1 方法

R-STCNN 体架构设计旨在通过综合考虑人体骨架数据中的平移和旋转信息, 以提高人体动作识别的准确性, 核心在于如何有效地从骨架数据中提取和融合这两种信息, 进而实现对复杂动作的精确识别。总体架构如图 1 所示, 输入数据包括代表关节点和骨骼边平移特征的 joints、bones 和代表旋转信息的 joints_A、bones_A。将处理后的特征输入时空特征提取与共现模块(STCN), 在

坐标级别, STCN 通过卷积操作捕捉单个关节或骨骼边的时间特征和相同时间维度节点空间特征; 在具身级别, 通过对特征图换维后的多次卷积和步长为 2 的最大值池化操作聚合全身关节和骨骼边的全局时空信息, 得到动作数据的静态和动态时空特征。

将旋转信息时空特征(特征 3、特征 4)取 max

融合后与平移信息时空特征(特征 1、特征 2)沿通道维度进行堆叠融合, 融合后的特征图会经过运动级特征提取, 进一步提炼和抽象化特征, 以便于捕捉动作的运动级信息。最后经过全连接层的处理, 通过计算输入特征与动作类别之间的关系, 确定最终的动作类别。

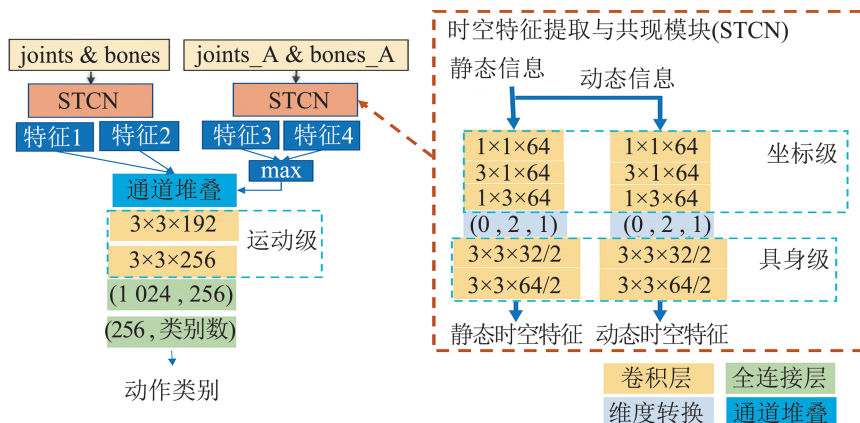


图 1 R-STCNN 框架图
Fig. 1 R-STCNN framework diagram

1.1 基于 CNN 的动作识别

基于 CNN 的动作识别流程图如图 2 所示, 通过将骨架数据序列建模成“伪图像表示”, 构建数据特征进行动作的分类识别^[9-10]。以关节序列 P 的建模为例, 序列的帧数和点数构建为伪图像的列和行, 关节的三维坐标作为伪图像的通道。输入为伪图像数据 $G^{C \times T \times V \times N}$ 时, C 、 T 、 V 、 N 分别表示输入数据的通道数、时间帧、关节数、人数。帧数和关节数为 64 和 25, 单人动作 N 为 1, 双人动作 N 为 2。

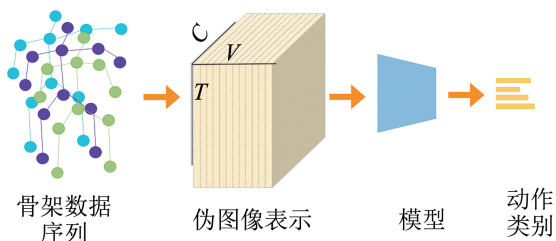


图 2 基于 CNN 方法的动作识别流程
Fig. 2 Action recognition process based on CNN method

1.2 三维球空间中的旋转角度特征

人体的动作是肢体结构旋转与平移的复合运动过程, 为了刻画动作的动态旋转信息, 在前期工作中, 将人体投影到以人为中心的三维球空间通过算法 1 获取静态旋转角度特征, 如图 3 所示。再将旋转信息推广到动态域, 以获取动作的角速度编码, 以表征运动的动态旋转信息。

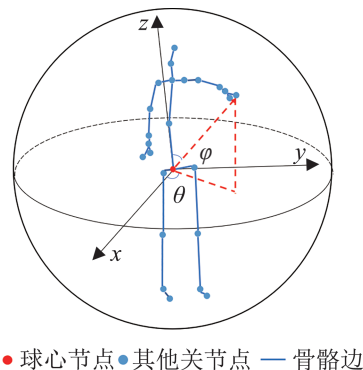


图 3 三维球形空间示意图
Fig. 3 Schematic representation of 3D sphere space

算法1 角度信息获取算法

step 1: 输入骨架数据序列中关节点序列, $P_{i,t}=(x_{i,t}, y_{i,t}, z_{i,t})$, $t \in \{1, 2, \dots, T\}$, $i \in \{1, 2, \dots, V\}$, t 代表时间帧数, i 代表关点序号。同时构造骨骼边: $B_{i,j,t}=(x_{i,t}-x_{j,t}, y_{i,t}-y_{j,t}, z_{i,t}-z_{j,t})$, $t \in \{1, 2, \dots, T\}$, $i, j \in \{1, 2, \dots, V\}$, j 表示与 i 号关节点存在物理连接的关节点序号。

step 2: 定义3个表示方向的单位向量, $\alpha_x=[a_{1x}, a_{2x}, a_{3x}]^T$, $\alpha_y=[a_{1y}, a_{2y}, a_{3y}]^T$, $\alpha_z=[a_{1z}, a_{2z}, a_{3z}]^T$, 通过式(1)构造旋转矩阵 M , 再通过式(2)将原始坐标 P 映射到相对坐标系得到人体坐标系坐标 \hat{P} 。

step 3: 将转换后的三维坐标 $\hat{P}(x, y, z)$ 利用式(3)映射到三维球空间, 如图3所示。用肢体结构运动时产生的方位角 θ 和仰角 φ 作为动作的静态角度信息 $F=[\theta, \varphi]$ 。

$$M=[\alpha_x, \alpha_y, \alpha_z] \quad (1)$$

$$\hat{P}=M^{-1}(P-L) \quad (2)$$

式中: L 为原始坐标系中的原点坐标。

$$\begin{cases} \theta = \arccos\left(\frac{z}{\sqrt{x^2+y^2+z^2}}\right) \\ \varphi = \begin{cases} \pi/2, & x=0, y < 0 \\ -\pi/2, & x=0, y > 0 \\ \arctan(y/x), & x \neq 0, y \in R \end{cases} \end{cases} \quad (3)$$

人体运动是随时间变换的动态过程, 将静态的角度特征推广到动态的角速度特征, 利用运动信息的帧间变化量, 通过式(4)构建关节点和骨骼边的速度信息 v 与角速度信息 ω , 以表征动作的动态特征信息(图1中的动态信息输入), 从而获得关节点和骨骼边的旋转角度编码 $A=\{F, \omega\}$ (图1中的 joints_A 和 bones_A)。

$$\begin{cases} v = P_{t+1} - P_t \\ \omega = F_{t+1} - F_t \end{cases} \quad (4)$$

1.3 时空卷积神经网络

R-STCNN的网络结构如图1所示, 主要包括时空特征提取与共现模块、运动级信息聚合。

1.3.1 时空特征提取与共现模块(STCN)

为获取肢体动作中丰富的时空信息, 本模块针对肢体结构的平移信息和旋转信息在静态、动态两种模态下的特征进行建模。针对坐标级信息和全局结构信息构建2种特征聚合方式: ①坐标级特征聚合, 包含卷积核分别为 1×1 、 3×1 、 1×3 的3个卷积层, 在后2个卷积层中对时间维度的卷积尺寸分别为3和1, 从而获取到不同时间尺度下的坐标级信息。为了进一步捕捉相同时间节点下节点序列的空间信息, 相较于HCN, 本文方法运用第3层的 1×3 卷积, 聚焦单个时间点下的关节点之间的结构关系, 以获取节点序列中坐标的空间关联特征。②具身级特征聚合, 针对时空中全身肢体节点间的关联特征。通过维度转换将关节维度转换到通道维度, 使卷积聚焦于肢体结构的所有关节点和骨骼边信息中的关键运动部位。然后, 运用2层卷积核尺寸为 3×3 的卷积层结合步长为2的最大池化层, 融合不同时间维度中的全局关键时空共现特征。针对不同运动模态的数据序列, 分别获取其静态和动态时空共现特征。

1.3.2 运动级特征聚合模块

将关节点和骨骼边的旋转特征 joints_A、bones_A 和平移特征 joints、bones 输入时空特征提取与共现模块, 以获取两者的静态和动态时空特征; 将融合后的旋转特征与平移特征等3路特征信息沿通道维度进行堆叠; 采用2个卷积核为 3×3 的卷积层对融合后的特征图进行运动级全局特征聚合, 提取更完整的运动特征。最后, 经全连接层输出分类结果。

1.3.3 R-STCNN算法流程

对骨架数据的三维坐标序列进行特征建模, 通过将人体动作映射到三维球空间, 刻画动作的旋转运动信息进行动作识别, 提出融合球空间下旋转角度编码的人体动作识别方法(R-STCNN)。

算法2 R-STCNN算法流程输入: 坐标序列 S 输出: 动作类别 C

step 1: 对输入坐标序列进行数据预处理, 得到坐标序列 P 、骨骼边序列 B ;

step 2: 用式(1)构建旋转矩阵 M , 结合式(2)将坐标序列 P 、骨骼边序列 B 转换于人体坐标系;

step 3: 通过式(3)进行运动空间的转换, 构建人体动作的球空间投影, 并获得动作的静态旋转角度编码 F ;

step 4: 通过式(4)获取角速度信息 ω , 可得旋转角度编码 $A=\{F, \omega\}$, 获得关节点和骨骼边的速度信息 v_P, v_B ;

step 5: 将平移信息 $\text{joints}=\{P, v_P\}$, $\text{bones}=\{B, v_B\}$ 和旋转信息 $\text{joints_A}=\{F_P, \omega_P\}$, $\text{bones_A}=\{F_B, \omega_B\}$ 以静态和动态信息为依据沿通道维度堆叠 $\Psi=\{[P, B], [v_P, v_B]\}$ 、 $\hat{A}=\{[F_P, F_B], [\omega_P, \omega_B]\}$;

step 6: 将 Ψ 、 \hat{A} 输入STCN模块得到平移特征(特征1、特征2)和旋转特征(特征3、特征4);

step 7: 将特征3、特征4经max融合后与特征1、特征2进行通道堆叠融合, 通过运动级特征聚合模块进行特征聚合, 获得综合特征图;

step 8: 将获得的特征图经过2层全连接层处理后得到类别标签从而统计出类别 C 。

2 实验与分析

2.1 数据集与实验细节

NTU RGB+D 60^[20]是目前应用最广泛的基于人体骨架数据的数据集之一, 共60个动作类别, 56880个动作样本, 使用45°、0°和-45°视角的Kinect V2相机, 采集了40个不同受试者的日常动作数据。数据集提供CS和CV评估方式, CS采用20个受试者的数据用作训练集, 另20个受试者的数据作为测试集; CV采用0°和45°相机位置采集的数据作为训练集, -45°位置相机采集的数据作为测试集。

NTU RGB+D 120^[21]是NTU RGB+D 60的扩展

版, 增加了受试者的数量和复杂动作, 以及不同人机距离和相机高度的32种数据采集场景设置, 增加了数据的视图多样性, 大幅度提升了动作识别的难度。NTU RGB+D 120提供CS和Cset评估方式, CS采用53名受试者的数据作为用于训练集, 另53名受试者的数据作为测试集; Cset选取设置编号为偶数的16类样本作为训练集, 编号为奇数的16类样本作为测试集。

本文方法在2个NVIDIA Quadro RTX 6000 GPU上训练深度学习模型, 并使用Pytorch作为深度学习基本框架, 使用Adam优化器对模型进行600个epoch的训练。

2.2 消融实验

2.2.1 旋转角度编码的有效性

为验证旋转角度编码的有效性, 在相同特征提取网络架构(STCN)的前提下, 在2个数据集上的测试结果如表1所示, 融合关节点的动静态旋转角度(joints_A)能够提升识别精度, 这表明旋转信息对于动作识别是有效的。并且, 同时融合骨骼边信息和关节点的旋转信息(\hat{A})仅利用两者的平移信息获得更好的识别效果, 表明旋转信息和平移信息之间具有互补性。即使在更复杂的NTU RGB+D120数据集上, 融合旋转角度编码也能够提升识别效果, 表明球空间下的旋转角度编码能够有效凸显动作的类间差异性, 这进一步证实了融合旋转角度编码的有效性。在时间消耗方面, 从表1单次训练时间列可知, 随着动作信息的增多, 模型所需的推理时间也随之增加。

2.2.2 STCN的有效性

为捕捉动作的时空特征, STCN模块对坐标级的卷积操作中, 采取时空卷积层对单个节点的时间特征和相同时刻节点之间的空间特征进行提取。为验证STCN模块的有效性, 在控制原始输入均为关节点信息 P 时, 将R-STCNN与基线方法HCN进行识别效果的对比, 由表2可知, 具有STCN模块的R-STCNN方法能有效提取动作数据中的时空特征, 提升了对人体动作的识别效果。

表 1 旋转角度编码有效性验证
Table 1 Validation of rotation angle coding

特征	NTU RGB+D60/%				NTU RGB+D120/%				单次训练时间/s
	CV	Add	CS	Add	Cset	Add	CS	Add	NTU-RGB+D 60 (CS)
P	91.9		86.8		76.9		76.5		21
+B	92.3	0.4	87.3	0.5	78.0	1.1	80.0	3.5	34
+joints_A	92.3	0.4	87.1	0.3	77.0	0.1	80.0	3.5	34
+B+ \hat{A}	92.5	0.6	87.8	1.0	78.9	2.0	80.4	3.9	38

表 2 STCN 的有效性验证
Table 2 Effectiveness validation of STCN %

方法	NTU RGB+D60				NTU RGB+D120			
	CV	Add	CS	Add	Cset	Add	CS	Add
HCN(base)	90.3		85.8		76.6		76.5	
R-STCNN	91.9	1.6	86.8	1.0	76.9	0.3	79.5	3.0

2.3 与其他方法的比较

为验证本文算法的识别性能，在 2 个数据集下将 R-STCNN 方法与其他方法进行对比，结果如表 3~4 所示。

表 3 证明了 R-STCNN 方法融合动作旋转信息的合理性，以及在球空间中构建的人体动作的旋转角度编码的有效性。

由表 4 可知，在 NTU RGB+D120 数据集上的 Cest 评估方式下，R-STCNN 略低于 SGN 和 DG-2sCNN。这是由于数据集在采集时人体与相机的距离和相机所处高度的变化引起动作视角的改变，导致在对相同动作在不同距离和视角时的类内相似性减小，造成识别混淆。但 CS 评估方式，R-STCNN 识别率最优，说明 R-STCNN 在针对不同对象的动作识别中更具有鲁棒性。

为研究模型对动作的识别效果，对模型在 NTU RGB+D 60 数据集 CS 方式下的混淆矩阵见图 4。模型的错误分类主要表现在“阅读(11)”“写作(12)”和“穿鞋(16)”“脱鞋(17)”这类与手部关节相关，且具有极大相似性的动作类别之间。这是由于 NTU-RGB+D 数据集中的手部的动作仅采集拇指和中指指尖两个关节的数据，使细微手部动作的类间差异性难以捕捉，造成识别效果受

限。模型训练和测试的损失曲线见图 5，经过多次的测试，模型的损失曲线能够很好地收敛，证明模型对于动作识别具有稳定性。

表 3 NTU RGB+D 60 数据集上的识别率
Table 3 Recognition rates on NTU RGB+D 60 dataset %

方法	CV	Add	CS	Add
STA-LSTM ^[22]	81.2	11.3	73.4	14.4
TCN ^[8]	83.1	9.4	74.3	13.5
VA-LSTM ^[23]	87.7	4.8	79.2	8.6
ElAtt-GRU ^[24]	88.4	4.1	80.7	7.1
PA-GCN ^[25]	82.7	9.8	80.4	7.4
ST-GCN ^[6]	88.3	4.2	81.5	6.3
LSTM+GCN ^[26]	90.2	2.3	84.8	3.0
DIF-CNN ^[27]	85.8	6.7	81.0	6.8
DG-2sCNN ^[28]	91.2	1.3	87.1	0.7
AFE-CNN ^[7]	92.2	0.3	86.2	1.6
HCN ^[4] (base)	89.1	3.4	84.5	3.3
R-STCNN	92.5		87.8	

表 4 NTU RGB+D 120 数据集上的识别率
Table 4 Recognition rates on NTU RGB+D 120 dataset %

方法	CS	Add	Cset	Add
ST-LSTM+Trust Gate ^[29]	56.5	33.9	54.1	24.8
GCA-LSTM ^[30]	58.3	22.1	59.2	19.7
Two-stream network ^[21]	62.2	18.2	61.8	17.1
ST-GCN ^[6]	70.7	9.7	73.2	5.7
AS-GCN ^[31]	77.7	2.7	78.9	0
STF-GCN ^[32]	76.7	3.7	79.0	-0.1
Synthesized CNN ^[33]	60.3	20.1	63.2	15.7
SGN ^[34]	79.2	79.2	81.5	-2.6
DG-2sCNN ^[28]	78.0	2.4	81.0	-2.1
HCN(base)	76.5	3.9	76.6	2.3
R-STCNN	80.4		78.9	

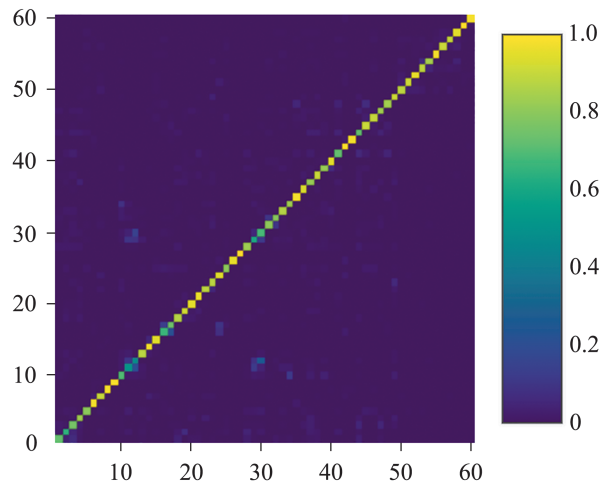


图4 NTU RGB+D 60数据集CS方式下的混淆矩阵
Fig. 4 Confusion matrix under CS criteria of NTU RGB + D 60 dataset

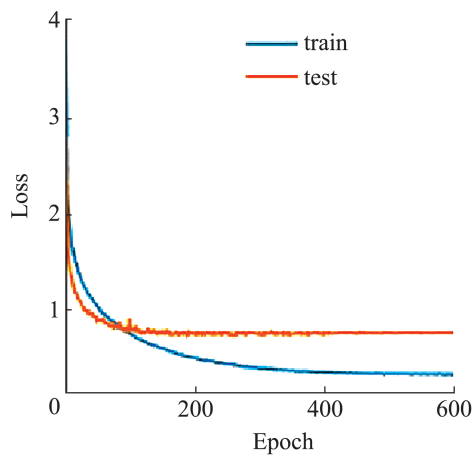


图5 NTU RGB+D 60数据集CS方式下的损失曲线
Fig. 5 Loss curves under CS criteria of NTU RGB+D 60 dataset

3 结论

相较于传统的基于骨架数据的动作识别方法, 本文提出的融合球空间下旋转角度编码的人体动作识别方法(R-STCNN)。通过将人体在传统欧氏空间的动作信息映射到三维球空间, 以获取动作的动态和静态的旋转角度信息, 再结合自身的平移信息, 从旋转加平移的角度刻画人体动作的运动特征。并针对坐标级别的动作时空特征, 提出时空特征提取与共现模块进行特征提取。在两个

大型数据集上验证了旋转角度编码和时空特征提取与共现模块的有效性, 证明本文提出的R-STCNN方法相较先进的基于骨架数据的动作识别方法具有较强的竞争力。

然而, 本文方法在面对相机高度和人机距离变化较大的视图场景和关节数据较少的手部细微动作时, 对动作的识别性能受限, 后续将考虑通过解算相机位置、增加手部数据样本以及进行数据增强等改进措施, 进一步提高模型在面对复杂视图场景时的动作识别性能。

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] Shi Lei, Zhang Yifan, Cheng Jian, et al. Skeleton-based Action Recognition with Directed Graph Neural Networks[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2019: 7904-7913.
- [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] 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.
- [5] Talha S A W, Hammouche M, Ghorbel E, et al. Features and Classification Schemes for View-invariant and Real-time Human Action Recognition[J]. *IEEE Transactions on Cognitive and Developmental Systems*, 2018, 10(4): 894-902.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] Si Chenyang, Jing Ya, Wang Wei, et al. Skeleton-based Action Recognition with Spatial Reasoning and Temporal Stack Learning[C]//Computer Vision - ECCV 2018. Cham: Springer International Publishing, 2018: 106-121.
- [10] Du Y, Fu Y, Wang L. Skeleton Based Action Recognition with Convolutional Neural Network[C]//2015 3rd IAPR Asian Conference on Pattern Recognition. 2015: 579-583.
- [11] Liu Shuhua, Bai Xiaoying, Fang Ming, et al. Mixed Graph Convolution and Residual Transformation Network for Skeleton-based Action Recognition[J]. *Applied Intelligence*, 2022, 52(2): 1544-1555.
- [12] 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.
- [13] Ke Qihong, Bennamoun M, An Senjian, et al. A New Representation of Skeleton Sequences for 3D Action Recognition[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2017: 4570-4579.
- [14] 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.
- [15] Huang Hong'en, Su Hang, Chang Zhigang, et al. Convolutional Neural Network with Adaptive Inferential Framework for Skeleton-based Action Recognition[J]. *Journal of Visual Communication and Image Representation*, 2020, 73: 102925.
- [16] Ofli F, Chaudhry R, Kurillo G, et al. Sequence of the Most Informative Joints (SMIJ): A New Representation for Human Skeletal Action Recognition[J]. *Journal of Visual Communication and Image Representation*, 2014, 25(1): 24-38.
- [17] Ömer Faruk İnce, Ibrahim Furkan Ince, Mustafa Eren Yıldıırım, et al. Human Activity Recognition with Analysis of Angles Between Skeletal Joints Using a RGB-depth Sensor[J]. *ETRI Journal*, 2020, 42(1): 78-89.
- [18] Ibrahim Furkan Ince. Performance Boosting of Scale and Rotation Invariant Human Activity Recognition (HAR) with LSTM Networks Using Low Dimensional 3D Posture Data in Egocentric Coordinates[J]. *Applied Sciences*, 2020, 10(23): 8474.
- [19] Zatsiorsky V M. Kinetics of Human Motion[M]. Champaign, IL, USA: Human Kinetics, 2002.
- [20] Shahroudy A, Liu Jun, Ng T T, 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.
- [21] Liu Jun, Amir Shahroudy, Perez M, 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.
- [22] Song S, Lan C, Xing J, et al. An end-to-end spatio-temporal Attention Model for Human Action Recognition from Skeleton Data[C]//Proceedings of the AAAI conference on artificial intelligence. 2017, 31(1).
- [23] 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.
- [24] 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.
- [25] 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.
- [26] Si Chenyang, Jing Ya, Wang Wei, et al. Skeleton-based Action Recognition with Hierarchical Spatial Reasoning and Temporal Stack Learning Network[J]. *Pattern Recognition*, 2020, 107: 107511.
- [27] 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.
- [28] Su Benyue, Zhang Peng, Sun Manzhen, et al. Direction-guided Two-stream Convolutional Neural Networks for Skeleton-based Action Recognition[J]. *Soft Computing*,

- 2023, 27(16): 11833-11842.
- [29] Liu Jun, Shahroudy A, Xu D, et al. Skeleton-based Action Recognition Using Spatio-temporal LSTM Network with Trust Gates[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(12): 3007-3021.
- [30] 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.
- [31] 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.
- [32] Liu Xing, Li Yanshan, Xia Rongjie. Adaptive Multi-view Graph Convolutional Networks for Skeleton-based Action Recognition[J]. Neurocomputing, 2021, 444: 288-300.
- [33] Liu Mengyuan, Liu Hong, Chen Chen. Enhanced Skeleton Visualization for View Invariant Human Action Recognition[J]. Pattern Recognition, 2017, 68: 346-362.
- [34] 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.