计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.201200205

融合因果关系和时空图卷积网络的人体动作识别

叶松涛[1]　周扬正[1]　范红杰[2]　陈正雷[3]
1 湘潭大学计算机学院　湖南 湘潭 411105
2 中国政法大学科学技术教学部　北京 102249
3 国家体育总局武术研究院　北京 100029
(yesongtao@xtu.edu.cn)

摘　要　基于人体骨骼的动作识别因具有简洁、鲁棒的特点,近年来受到了广泛的关注。目前大部分基于骨骼的动作识别方法,如时空图卷积网络(ST-GCN),通过提取连续帧的时间特征和帧内骨骼关节的空间特征来区分不同的动作,取得了良好的效果。考虑人体运动中存在的因果性关系,提出了一种融合因果关系和时空图卷积网络的动作识别方法。针对计算关节力矩获取权重复杂的情况,根据关节之间的因果关系为骨骼图分配边权重,并将权重作为辅助信息增强图卷积网络,来提高驱动力较强的关节在神经网络中的权重,降低重要性低的关节的关注度,增强重要性高的关节的关注度。相比 ST-GCN 等方法,在 Kinetics 公开数据集上,所提方法无论是 Top-1 还是 Top-5 都有较大的提升,在构建的真实太极拳数据集上的识别精度达 97.38%(Top-1)和 99.79%(Top-5),证明了该方法可以有效地增强动作特征,提升识别的准确率。

关键词:动作识别;因果关系;权重嵌入;时空图卷积网络;收敛交叉映射
中图法分类号　TP391.4

Joint Learning of Causality and Spatio-Temporal Graph Convolutional Network for Skeleton-based Action Recognition

YE Song-tao[1],ZHOU Yang-zheng[1],FAN Hong-jie[2] and CHEN Zheng-lei[3]
1 School of Computer Science,Xiangtan University,Xiangtan,Hunan 411105,China
2 Department of Science and Technology Teaching,China University of Political Science and Law,Beijing 102249,China
3 Wushu Research Institute,General Administration of Sport of China,Beijing 100029,China

Abstract　In recent years,skeleton based human action recognition has attracted extensive attention due to its simplicity and robustness. Most of the skeleton based human action recognition methods,such as spatio-temporal graph convolutional network (ST-GCN),distinguish different actions by extracting the temporal features of consecutive frames and the spatial features of skeleton joints within frames,achieve good results. In this paper,considering the causality of human action,we propose an action recognition method combining causality and spatio-temporal graph convolutional network. In view of the complexity of obtaining weight,we propose a method to calculate joint weight based on causality. According to the causality,we assign weights to skeleton graph,and use weights as auxiliary information to enhance graph convolutional network to improve the weight of some joints with strong driving force in the neural network,so as to reduce the attention of low importance joints and enhance the attention of high importance joints. Compared with ST-GCN,our method improves the recognition accuracy of both Top-1 and Top-5,and the recognition accuracy reaches 97.38% (Top-1) and 99.79% (Top-5) on the real TaiChi dataset,which strongly prove that our method can effectively learn and enhance the discriminative features.

Keywords　Action recognition,Causality,Weight embedding,Spatio-temporal graph convolutional neural network,Convergent cross mapping

1　引言

人体骨骼中关节的层次结构和不同语义角色为动作识别提供了重要信息。动作识别通过对预先分割好的时域序列进行时空信息建模,学习视频中所包含的语义和运动特征信息,以此来构建视频内容与动作类别之间的映射关系,从而对人体行为进行分类[1]。动作识别在视频理解、智能监控、行人跟踪、人机交互等领域有着广泛的应用[2-4]。动作数据通常以RGB视频或骨骼数据的方式进行呈现。由于骨骼信息删除了背景、光照等因素,只关注人体姿态和位置,因此骨骼信息对于视角、身体比例、运动速度、衣服纹理和背景的变化更具有稳健性和鲁棒性[5-6]。此外,骨骼数据量更小,极大地减小了模型复杂度和计算量。这些优势,使得基于骨骼的动作识别成为了计算机视觉领域的研究热点[7-8]。

基金项目:国家自然科学基金(61802327);湖南省自然科学基金(2018JJ3511)
This work was supported by the National Natural Science Foundation of China(61802327) and Natural Science Foundation of Hunan Province (2018JJ3511).
通信作者:范红杰(hjfan@cupl.edu.cn)

基于骨骼的动作识别主要可分为基于手工特征的方法和基于深度学习的方法。传统的基于手工特征方法通常使用手工特征对人体进行建模。这些方法使用浅层结构，限制了学习能力，无法全面捕捉时空特征，具有一定的局限性[9-10]。循环神经网络[11](Recurrent Neural Networks，RNNs)和卷积神经网络[12](Convolutional Neural Networks，CNNs)相继被应用到动作识别任务中。基于深度学习的动作识别方法以一种端到端的形式，通过网络自主地学习视频中的动作特征来完成分类。例如，利用循环神经网络[13]将骨骼数据处理成坐标矩阵序列或利用卷积神经网络进行动作预测[7]。深度学习导致了基于骨骼建模方法的激增[14-15]。此外，Huang 等[16]总结了目前几种主流的 3D 卷积框架，将其在相应数据集上进行对比和分析，以此得到每种框架的优势及弊端，从而寻找与实际情景相适应的最优框架。

图卷积网络(Graph Convolutional Network，GCN)作为 CNNs 的一种，可以有效提取非欧数据特征[17]。Yan 等[18]首先将图卷积网络应用于基于骨骼的动作识别，提出了时空图卷积网络(Spatial-Temporal Graph Convolutional Networks，ST-GCN)模型。模型根据关节的自然连接特征，在单个视频帧中构建骨骼图，并利用时间边缘连接的帧构造整个时空。图卷积网络成为目前骨骼动作识别的主要方法，后续的许多工作在此基础上进行了深入[19-21]，例如 Shi 等[19]设计了一个基于 ST-GCN 的双流自适应图卷积网络，并引入非局部块自适应的方式来学习节点间的连接。为了获取更丰富的关节相关性依赖关系，Li 等[20]提出了一种动作结构图卷积网络(AS-GCN)。Wen 等[21]提出了基于 motif 的图卷积对层次空间结构进行编码，并利用可变的时间密集块来挖掘人体骨骼序列中面向不同时间范围的信息。

基于深度学习的方法对人体动作识别进行了多角度研究，取得了较好的识别性能。但是这些方法仍然存在一些局限，例如忽略了人体结构中骨骼关节之间的相关性，未考虑不同动作中关节的权重变化等。关节被认为是一个刚体的端点，不同的关节在不同的人体动作中发挥的作用不同，因此具有关键作用的关节在决定动作所属类别时应该占有更大的比重。而将人体关节简化为多刚体模型，并通过解偏微分方程组计算各个关节力矩，再根据关节力矩大小为每个关节分配不同权重，此类方法由于方程数量过多而计算量过大，不适用于本就已经较为复杂的图卷积网络。而局部注意力模型的主要缺点是只利用动作序列的局部变化来获得注意力权重，并且准确的注意力权重不易获得。

针对上述问题，本文提出了一种融合因果关系和时空图卷积网络的人体骨骼动作识别方法。该方法首先根据关节坐标序列计算关节的因果关系系数，构建因果系数矩阵；然后将因果系数矩阵应用于图卷积网络，图卷积网络根据关节在运动过程中的重要性分配不同的权重，关注运动过程中影响力较大的关节，忽视作用较小的关节，从而有效地学习动态特征，增加识别的准确度。

本文主要贡献有 3 个方面：

(1)考虑人体运动中存在的因果性关系，为骨骼数据构建时空图模型，提出了一种融合因果关系和时空图卷积网络的动作识别方法；

(2)针对计算关节力矩获取权重复杂的情况，通过一种基于因果性计算关节权重的方法，根据关节之间的因果关系为骨骼图分配边权重，并将权重作为辅助信息增强图卷积网络，来提高某些驱动力较强的关节在神经网络中的权重，从而使神经网络降低重要性低的关节关注度，增强重要性高的关节关注度；

(3)本文在 Kinetics 公开数据集上，相比 ST-GCN 方法，所提方法无论是 Top-1 还是 Top-5 都有明显的提升，在构建的真实太极拳数据集上的识别精度达 97.38%(Top-1)和 99.79%(Top-5)。

2 相关工作

2.1 基于骨骼的动作识别

传统基于手工特征的骨骼动作识别通过手工设计不同的特征提取方法来捕获关节运动的动态。例如，Hussein 等[23]建立时域分层的协方差矩阵描述子来表示关节的运动轨迹，Wang 等[24]使用关节的相对位置作为特征，Vemulapalli 等[25]使用身体各部位之间的旋转和平移，提取特征后使用传统机器学习算法对特征进行分类，从而将动作分类。由于深层神经网络能够更好地学习特征表示，一些基于骨骼的动作识别研究由手工设计特征转向基于深度学习的方法。

2.2 基于深度学习的动作识别方法

基于深度学习的方法分为两个阶段，初期研究者使用循环神经网络 RNN[6,26]或 Temporal CNN[7,27]以端到端的方式学习动作识别模型。这些方法大多数直接将骨骼坐标序列作为输入特征，或者将骨骼坐标序列转换为灰度图像再输入网络进行分类。然而，RNN 和 CNN 不能完整地表示骨骼结构。根据人体自然结构，图模型比较适合骨骼数据的表示。因此，Yan 等[18]首先将图卷积网络应用于基于骨骼的动作识别，提出了 ST-GCN 网络模型。之后在 ST-GCN 的基础上，Shi 等[19]提出 2s-AGCN，其改进了 GCN 模块使之能自适应地学习图的拓扑结构，除了骨骼数据外，还使用之前从未有人注意到的骨骼信息作为第二信息流来提升识别效果。Bin 等[28]使用 SGR 组件在基于空间的子组聚类上发现关节之间的连通性，测量关节时间轨迹的相关性。Maosen 等[20]扩展了骨骼图结构，以捕获特定于动作的潜在依赖关系。

3 融合因果和时空图卷积的模型设计

如图 1 所示，本文提出了一种融合因果关系和时空图卷积网络的人体动作识别模型。

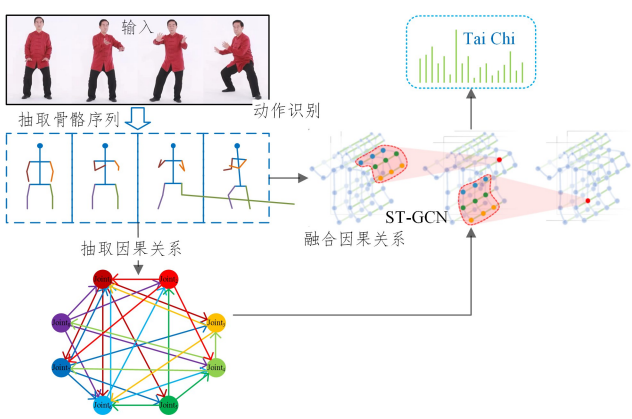


图 1 融合因果关系和时空图卷积网络的人体动作识别模型

Fig. 1 Skeleton-based action recognition model based on causality and spatio-temporal graph convolutional network

该模型首先根据关节坐标序列计算关节的因果关系，构

建因果系数矩阵；然后将矩阵应用于 ST-GCN 图卷积网络，网络对骨骼数据提取深层特征，从而有效地学习动态特征，增加动作识别的准确度。

3.1 时空图模型

设长度为 T 帧的骨骼序列时空图为 $G(V,E)$，点集 V 包含所有时刻的关节，边集 E 由两个子集构成，一个边子集为每帧的骨骼间连接 E_s（图 2 蓝色连线），E_s 反映了空间属性；另一个边子集为帧间对应关节相连 E_F（图 2 橙色连线），E_F 反映了时间属性。$G(V,E)$ 的顶点特征为关节坐标向量 $F(v_{ti})$，顶点 v_{ti} 代表关节 i 在第 t 帧的坐标。

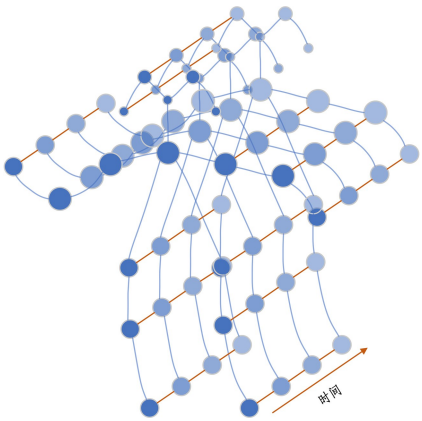


图 2 骨骼时空图示例（电子版为彩色）

Fig. 2 Schematic diagram of skeleton spatial-temporal graph

3.2 图卷积

给定 $G(V,E)$，对 v_{ti} 图卷积：

$$f_{\text{out}}(v_{ti})=\sum_{v_{tj}\in B(v_{ti})}\frac{1}{Z_{ti}(v_{tj})}f_{in}(v_{tj})\cdot w(l_{ti}(v_{tj})) \tag{1}$$

其中，$B(v_{ti})=\{v_{tj}\mid d(v_{tj},v_{ti})\leqslant D\}$ 为 v_{ti} 的卷积采样区域；f_{in} 为 v_{ti} 的输入特征；w 是权重函数，为输入特征提供权重向量。由于传统卷积采样区域大小固定，且权重向量数量和采样区域大小相等，而 B 中的顶点数量是变化的，因此在图卷积时需要将映射的顶点与权重向量对应。映射 $l_{ti}:B(v_{ti})\rightarrow\{0,\cdots,K-1\}$ 将相邻节点映射到子集标签，每个邻居节点根据子集标签找到对应的权重向量。通常 K 设为 3，将划分为 3 个子集：第一个子集 S_1 是顶点本身（图 3 红色节点）；第二个子集 S_2 是向心子集，包含更靠近人体重心的节点（图 3 绿色节点）；第三个子集 S_3 是离心子集，包含离重心更远的节点（图 3 浅蓝色节点）。

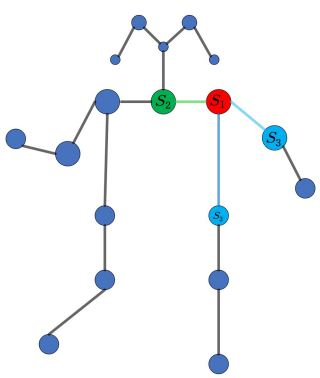


图 3 子集映射策略（电子版为彩色）

Fig. 3 Subset mapping strategy

图卷积网络通常由多层时空图卷积层构成，每层依次进行空间图卷积和时间图卷积来提取图的高层特征，最后再进行池化和 softmax 处理。

$G(V,E)$ 的特征由 (C,T,N) 张量表示，其中 C 为通道数，T 为时间长度，N 为顶点数量。根据式(1)，在多维张量上的图卷积公式如式(2)所示：

$$f_{\text{out}}=\sum_{v}^{K_v}\boldsymbol{W}_k(f_{\text{in}}\boldsymbol{A}_k) \tag{2}$$

其中，$\boldsymbol{A}_k=\boldsymbol{\Lambda}_k^{-\frac{1}{2}}\bar{\boldsymbol{A}}_k\boldsymbol{\Lambda}_k^{-\frac{1}{2}}\in\mathbb{R}^{N\times N}$ 为邻接矩阵，K_v 代表卷积核大小，其元素 $\boldsymbol{A}_k^{ij}$ 表示顶点 v_j 是否在 v_i 的子集中，$\boldsymbol{W}_k\in\mathbb{R}^{C_{\text{out}}\times C_{\text{in}}\times 1\times 1}$ 是权重向量。矩阵 $\boldsymbol{A}_k$ 决定了顶点特征是否会与权重向量协同进行计算。由于 $\boldsymbol{A}_k$ 是根据骨骼结构预先定义，即两个不同的动作视频，只要提取人体骨骼的结构相同，在图卷积时 $\boldsymbol{A}_k$ 是相同的。然而，不同的人体动作关节的重要程度不同，因此需要在邻接矩阵 $\boldsymbol{A}_k$ 中体现关节重要性的差异。为了解决这一问题，本文将因果性关系融合到图卷积网络中，用以增强动作识别效果。

3.3 融合因果关系的动作识别

人体多刚体模型以关节为转轴，使用拉格朗日方法进行建模，并通过偏微分方程组计算关节力矩[17-18]。刚体模型具有联动性，一个刚体运动时会通过转轴带动其他刚体运动。这种联动关系显然具有因果性，力矩大的关节发力时会更多地带动其他关节，发力关节是被带动关节的“因”。由于人体动力系统非常复杂，建模后方程数量多，如果要通过方程组计算关节间作用的强弱程度，计算复杂度过高；而且多刚体模型只能计算单一时刻的关节力矩，计算连续时间段内的关节力矩更加复杂。因此，通过多刚体模型分析关节相互作用强弱程度及因果关系较为困难。

3.3.1 关节因果关系计算

收敛交叉映射(Convergent Cross Mapping，CCM)是一种计算复杂系统中时间序列特殊相关性的方法[29]。该方法用于衡量估计值与 X 的相似性检验变量间的因果关系，能够克服多刚体模型建模复杂和计算复杂度高的困难，可以快速地检验关节因果关系，已有广泛的应用[30-31]。本文基于 CCM 分配边权重，方法步骤如下：

(1)构建阴影流

E 维延迟向量由点的 X_t 历史点构成，描述了一段时间关节的变化，可表示为：

$$\widetilde{X}_t=(X_t,X_{t-\tau},X_{t-2\tau},\cdots,X_{t-(E-1)\tau}) \tag{3}$$

其中，关节坐标序列 X 的阴影流 $\widetilde{X}$ 是 X 中每个点 X_t 延迟向量的集合，τ 为步长。

(2)寻找最邻近点，创建权重

针对 $\widetilde{X}_t$，找到与 t 时刻关节位置变化最相似的其他 $E+1$ 个时间点，计算 $\widetilde{X}_t$ 与其他延迟向量之间的欧氏距离。计算公式如式(4)所示：

$$d_i=D(\widetilde{X}_t,\widetilde{X}_{t_i^\wedge}) \tag{4}$$

选距离最小的 $E+1$ 个延迟向量作为最近邻点，得到时间点集合 $\{\tilde{t}_1,\tilde{t}_2,\cdots,\tilde{t}_{E+1}\}$ 和距离集合 $\{d_1,d_2,\cdots,d_{E+1}\}$。针对距离集合计算 $\widetilde{X}_t$ 的权重 $w_i=\dfrac{u_i}{N}$，$u_i=\mathrm{e}^{-d_i/d_1}$，$N=\sum_{j=1}^{E+1}u_j$。

(3)计算 X 对 Y 的估计值和相关系数

利用每个点的权重，对坐标序列 Y 中的每个点 Y_t 进行加

权求和，从而有 X 对 Y 的估计：

$$Y_t|\widetilde{X}=\sum_{i=1}^{E+1}w_iY_{\hat{t}_i} \tag{5}$$

本文通过计算原始坐标序列 Y 和估计值 $Y|\widetilde{X}$ 的皮尔逊相关系数 $C_{YX}=[\rho(Y,Y|\widetilde{X})]^2$ 来衡量 X 历史信息对 Y 的估计能力。

(4)计算边权重矩阵

一对关节 X 和 Y，可得两个相关系数 C_{XY} 和 C_{YX}，分别为 X 估计 Y 和 X 估计 Y。假设时空图有 N 个关节，计算两两关节因果系数，可得因果系数矩阵 $\boldsymbol{C}=\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}cor(l_j,l_i)$，$i$ 和 j 为关节编号。本文对系数矩阵 $\boldsymbol{C}$ 的每一行用 softmax 进行归一化，归一化后的系数作为 $G(V,E)$ 边权重进行嵌入。

3.3.2 边权重嵌入

受 CNN 和 RNN 的注意力机制启发，本文根据因果系数改变 $G(V,E)$ 边集的边权重。为了给 $G(V,E)$ 中不同的边分配不同的权重，本文将式(2)修改为：

$$f_{\text{out}}=\sum_{v}^{K_v}\boldsymbol{W}_k(f_{\text{in}}(\boldsymbol{A}_k\odot\hat{\boldsymbol{C}})) \tag{6}$$

其中，$\odot$ 代表矩阵逐元素积。在式(2)中，$\boldsymbol{A}_k$ 通过矩阵元素值为 0 或 $1/|S_{ik}|$ 表示两个顶点在子集是否相连。但在运动过程中，不同的关节间作用力强弱程度不同。本文将 $\boldsymbol{A}_k$ 与预计算的边权重矩阵 $\hat{\boldsymbol{C}}$ 逐元素积。由于矩阵 $\boldsymbol{A}_k$ 大小为 $N\times N$，边权重矩阵 $\hat{\boldsymbol{C}}$ 为 $N\times N$，$\hat{\boldsymbol{C}}$ 与 $\boldsymbol{A}_k$ 进行逐元素后矩阵大小也是 $N\times N$，因此用 $(\boldsymbol{A}_k\odot\hat{\boldsymbol{C}})$ 替换式(2)中的 $\boldsymbol{A}_k$ 并不会出现矩阵大小不匹配的问题。

如图 4 所示，上部分的骨骼标注 $\hat{\boldsymbol{C}}$ 中部分边权重对应了下部分的太极拳动作。由于太极拳动作的主要运动部位为手臂，且动作幅度较大，而下肢关节运动幅度较小，因此上肢关节间的边权相应地要大于下肢关节。

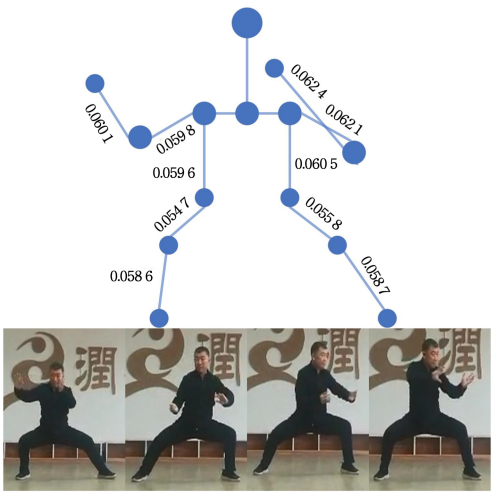


图 4 边权重示意图

Fig. 4 Schematic diagram of edge weight

本文利用 CCM 方法检验时间变量因果关系的特性，计算每个顶点之间的因果关系，将关节因果关系转换为边权重，然后将边权重通过逐元素积嵌入至图邻接矩阵，降低重要性低的关节关注度，增强重要性高的关节关注度，从而提高某些驱动力较强的关节在神经网络中的权重。

3.3.3 网络结构

本文采用 ST-GCN 作为基础网络。为了应用边权重矩阵，本文增加边权重矩阵作为输入，并在卷积运算前将边权重矩阵与邻接矩阵逐元素积。时空图卷积层结构如图 5 所示。

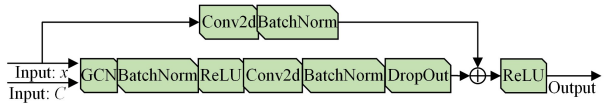


图 5 时空图卷积层结构

Fig. 5 Spatial-temporal graph convolutional block

4 实验与结果分析

4.1 实验步骤与设置

如图 6 所示，本文首先使用 OpenPose[32] 方法从视频中提取骨骼关节坐标，并通过数据清洗对缺失数据、非法值等情况进行处理。为了增强泛化能力，本文对关节坐标采取归一化、模拟相机移动、数据填充等操作进行数据增强。在此基础上，再计算每个视频的边权重矩阵，并将边权重矩阵与骨骼数据一起输入网络进行训练。

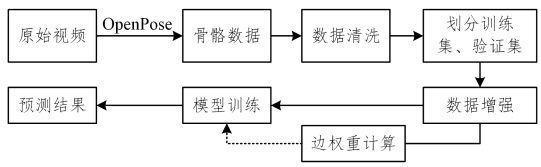


图 6 实验步骤流程

Fig. 6 Workflow of experiment

4.2 基于 Kinetics 数据集的实验评估

为了评估本文方法效果，本文采用 Kinetics 数据集[1)] 将其与 Feature Enc[33]，P-LSTM[24]，Res-TCN[34] 和 ST-GCN[18] 4 种动作识别方法进行了实验对比。Kinetics 数据训练集有 15 440 个视频，测试集 1 241 个视频，骨骼序列长度均填充至 300 帧。每种网络默认训练 65 个 Epoch，Batchsize 设为 32，用 0.1 学习率的随机梯度下降进行学习，每 10 个 epoch 学习率减小0.01。计算边权重矩阵时，延迟向量维度 E 设置为 4，步长 τ 为 1。本文使用 Top-1 和 Top-5 准确率作为实验效果的有效性评估，准确率越高说明识别效果越好。

从表 1 可以发现，由于增加了边权重因果相关性，本文的动作准确率均高于其他方法。本文方法相比 Feature Enc 方法，Top-1 提升了 8%，Top-5 提升了 9.93%；相比 P-LSTM，Top-1 提升了 3.47%，Top-5 提升了 3.31%；相比 Res-TCN，Top-1 提升了 2.81%，Top-5 提升了 1.53%；相比 ST-GCN，Top-1 提升了 2.29%，Top-5 提升了 2.21%。

表 1 Kinetics 数据集实验结果

Table 1 Results on Kinetics

(单位：%)

Kinetics 数据集	Top-1	Top-5
Feature Enc	35.76	68.72
P-LSTM	40.29	75.34
Res-TCN	40.95	77.12
ST-GCN	41.47	76.33
本文方法	**43.76**	**78.65**

1) http://deepmind.com/research/open-source/kinetics

值得一提的是，如表 2 所列，相比 ST-GCN 模型，本文方法在某些动作分类上的 Top-1 准确率具有 2%(baby waking up)到 20%(blasting sand)不等的提升，尽管在某些例如 applauding 和 baking cookies 等动作的识别度稍有降低。分析其原因，这是由于 Kinetics 数据集画质参差不齐，动作规范程度不高，且许多动作类别与身体运动没有强关联。此外，由于图卷积网络无法捕获整体的运动状态，因此与人体运动关联性较大的动作识别准确率较高，而与人体运动关联性较小的识别准确率较低，这一现象在文献[7]中也有提及。

表 2 Kinetics 部分动作的 Top-1 准确率

Table 2 Top-1 accuracy of some kinetics actions

(单位：%)

动作类别	ST-GCN	本文方法
beatboxing	36.7	**42.8(+6.1)**
auctioning	52.0	**64.0(+8.0)**
applauding	18.0	16.0(−2.0)
arm wrestling	75.5	**79.6(+4.1)**
applying cream	36.0	32.0(−4.0)
baby waking up	60.0	**62.0(+2.0)**
baking cookies	16.0	14.3(−1.7)
belly dancing	66.9	**71.4(+4.5)**
blasting sand	22.0	**42.0(+20.0)**
bandaging	26.0	**32.0(+4.0)**
biking through snow	66.0	**74.0(+8.0)**
bench pressing	74.0	70.0(−4.0)
answering questions	36.0	**40.0(+4.0)**

4.3 基于太极拳数据集的实验评估

本文还构建了一个真实的太极拳视频数据集，该数据集采集自 132 位太极拳学员视频，包含 2956 个视频片段，主要内容为陈氏太极老架一路一式到六式及收势，细分为 25 类动作。图 7 为太极拳数据集示例。本文采用 2000 个视频片段作为训练集，956 视频片段作为测试集。

图 7 太极拳数据集示例

Fig. 7 Schematic diagram of TaiChi dataset

如表 3 所列，大部分模型在太极拳数据集中都能达到很好的识别效果。

表 3 在太极数据集上最近技术的比较

Table 3 Comparison of recent technologies on TaiChi dataset

(单位：%)

太极拳数据集	Top-1	Top-5
P-LSTM	94.67	99.16
Res-TCN	95.82	99.37
2s-AGCN[19]	95.71	99.27
ST-GCN	96.95	99.58
本文方法	**97.38**	**99.79**

本文方法能达到表 3 所列的表现，原因如下：

(1)本文采集的太极拳视频均为正面拍摄，且拍摄对象始终保持在画面内。

(2)视频场景几乎没有额外物体遮挡人体和背景干扰，这使得提取出的骨骼数据完整度和准确率较高，完整度对动作识别十分重要。

(3)拍摄对象表演太极拳套路完成度较高，不规范动作少，这使得网络能更好地提取特征。由于增加了因果系数作为边权重，本文能凸显人体运动过程中的主要关节，其效果仍然要优于其他方法。这说明根据关节因果系数的图卷积网络更偏向于某些关节点。该模型已部署于爱太极应用的动作库采集中。

结束语 本文从因果性的角度出发考虑人体运动时关节的权重问题，提出了一种融合因果关系和时空图卷积网络的人体动作识别方法。受 RNN 和 CNN 中注意力机制的启发，本文将因果关系作为辅助信息去增强图卷积网络，从而有效提高某些驱动力较强的关节在神经网络中的权重。本文在 Kinetics 公开数据集和构建的真实太极拳数据集进行实验评估，证明该方法可以有效地学习动态特征，增加动作识别的准确度。

未来的工作中，本文将尝试融合其他模态信息，例如融合 RGB 和骨架数据，并在统一的框架下结合基于骨架的动作识别与姿态估计方法。

参 考 文 献

[1] STEFAN M, CRISTIAN S. Actions in the Eye: Dynamic Gaze Datasets and Learnt Saliency Models for Visual Recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 37(7): 1408-1424.

[2] RONALD P. A Survey on Vision-based Human Action Recognition[J]. Image and Vision Computing, 2010, 28(6): 976-990.

[3] SUDHA M R, SRIRAGHAV K, JACOB S G, et al. Approaches and Applications of Virtual Reality and Gesture Recognition: A review[J]. International Journal of Ambient Computing and Intelligence (IJACI), 2017, 8(4): 1-18.

[4] WANG P, LI W, OGUNBONA P, et al. RGB-D-based Human Motion Recognition with Deep Learning: A Survey[J]. Computer Vision & Image Understanding, 2018, 171: 118-139.

[5] SONG S J, LAN C L, XING J L, et al. An End-to-End Spatio-Temporal Attention Model for Human Action Recognition from Skeleton Data[C]// Proceeding of Thirty-First AAAI Conference on Artificial Intelligence. CA: AAAI, 2017: 4263-4270.

[6] LIU J, SHAHROUDY A, XU D, et al. Spatio-Temporal LSTM with Trust Gates for 3D Human Action Recognition[C]// Proceeding of 14th European Conference on Computer Vision. Cham: Springer, 2016: 816-833.

[7] KE Q, BENNAMOUN M, AN S, et al. A New Representation of Skeleton Sequences for 3D Action Recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2017: 3288-3297.

[8] SI C, JING Y, WANG W, et al. Skeleton-based Action Recognition with Hierarchical Spatial Reasoning and Temporal Stack Learning Network[J]. Pattern Recognition, 2020, 107: 107511.

[9] PRESTI L L, LA CASCIA M. 3D Skeleton-based Human Action

Classification：A survey[J]. Pattern Recognition，2016，53：130-147.

[10] FERNANDO B，GAVVES E，ORAMAS J M，et al. Modeling Video Evolution for Action Recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York：IEEE Press，2015：5378-5387.

[11] YU Y，SI X，HU C，et al. A Review of Recurrent Neural Networks：LSTM Cells and Network Architectures[J]. Neural Computation，2019，31(7)：1235-1270.

[12] KRIZHEVSKY A，SUTSKEVER I，HINTON G E. ImageNet Classification with Deep Convolutional Neural Networks[J]. Communications of the ACM，2017，60(6)：84-90.

[13] ZHANG P，LAN C，XING J，et al. View Adaptive Recurrent Neural Networks for High Performance Human Action Recognition from Skeleton Data[C]// Proceedings of the IEEE International Conference on Computer Vision. New York：IEEE Press，2017：2117-2126.

[14] LUO H L，TONG K，KONG F S. The Progress of Human Action Recognition in Videos Based on Deep Learning：A Review[J]. Acta Electronica Sinica，2019，47(5)：1162-1173.

[15] QIAN H F，YI J P，FU Y H. Review of Human Action Recognition Based on Deep Learning. Journal of Frontiers of Computer Science and Technology，2021，15(3)：438-455.

[16] HUANG H X WANG R P，LIU X Y. Review of Human Action Recognition Technology Based on 3D Convolution[J]. Computer Science，2020，47(S2)：139-144.

[17] NIEPERT M，AHMED M，KUTZKOV K. Learning Convolutional Neural Networks for Graphs[C]// Proceedings of the 33rd International Conference on Machine Learning. New York：PMLR，2016：2014-2023.

[18] YAN S J，XIONG Y J，LIN D H. Spatial Temporal Graph Convolutional Networks for Skeleton-based Action Recognition[C]// Proceeding of Thirty-second AAAI Conference on Artificial Intelligence. CA：AAAI，2018：7444-7452.

[19] SHI L，ZHANG Y F，CHENG J，et al. Two-Stream Adaptive Graph Convolutional Networks for Skeleton-Based Action Recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York：IEEE Press，2019：12026-12035.

[20] LI M S，CHEN S H，CHEN X，et al. Actional-Structural Graph Convolutional Networks for Skeleton-Based Action Recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York：IEEE Press，2019：3595-3603.

[21] WEN Y H，GAO L，FU H B，et al. Graph CNNs with Motif and Variable Temporal Block for Skeleton-Based Action Recognition[C]// Proceeding of Thirty-Third AAAI Conference on Artificial Intelligence. CA：AAAI，2019：8989-8996.

[22] HUSSEIN M E，TORKI M，GOWAYYED M A，et al. Human Action Recognition Using A Temporal Hierarchy of Covariance Descriptors on 3D Joint Locations[C]// Proceeding of the Twenty-Third International Joint Conference on Artificial Intelligence. CA：AAAI，2013：2466-2472.

[23] WANG J，LIU Z，WU Y，et al. Mining Action let Ensemble for Action Recognition with Depth Cameras[C]// Proceeding of IEEE Conference on Computer Vision and Pattern Recognition. NJ：IEEE，2012：1290-1297.

[24] VEMULAPALLI R，ARRATE F. Human Action Recognition by Representing 3d Skeletons As Points in A Lie Group[C]// Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition. NJ：IEEE，2014：588-595.

[25] SHAHROUDY A，LIU J. NTU RGB+D：A Large Scale Dataset for 3D Human Activity Analysis[C]// Proceeding of the IEEE conference on Computer Vision and Pattern Recognition. NJ：IEEE，2016：1010-1019.

[26] ZHANG S，LIU X，XIAO J. On Geometric Features for Skeleton-based Action Recognition using Multilayer LSTM Networks[C]// Proceeding of IEEE Winter Conference on Applications of Computer Vision (WACV). NY：IEEE Computer Society，2017：148-157.

[27] LI C，ZHONG Q，XIE D，et al. Skeleton-based Action Recognition with Convolutional Neural Networks[C]// Proceeding of IEEE International Conference on Multimedia & Expo Workshops (ICMEW). NJ：IEEE，2017：597-600.

[28] LI B，LI X，ZHANG Z，et al. Spatio-temporal Graph Routing for Skeleton-based Action Recognition[C]// Proceeding of the AAAI Conference on Artificial Intelligence，2019(33)：8561-8568.

[29] SUGIHARA G，MAY R，YE H，et al. Detecting causality in complex ecosystems[J]. Science，2012，338(6106)：496-500.

[30] LIU H，LEI M，ZHANG N，et al. The Causal Nexus Between Energy Consumption，Carbon Emissions and Economic Growth：New Evidence from China，India and G7 Countries Using Convergent Cross Mapping[J]. PloS one，2019，14(5)：e0217319.

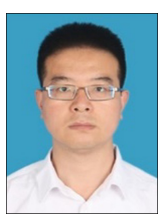
[31] BARRAQUAND F，PICOCHE C，DETTO M，et al. Inferring Species Interactions Using Granger Causality and Convergent Cross Mapping[J]. Theoretical Ecology，2020，14：87-105.

[32] CAO Z，HIDALGO G，SIMON T，et al. OpenPose：Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields[C]// Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. NJ：IEEE，2017：7291-7299.

[33] FERNANDO B，GAVVES E，ORAMAS J M，et al. Modeling Video Evolution for Action Recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. NJ：IEEE，2015：5378-5387.

[34] KIM T S，REITER A. Interpretable 3d Human Action Analysis with Temporal Convolutional Networks[C]// Proceedings of 2017 IEEE conference on Computer Vision and Pattern Recognition Workshops (CVPRW). NJ：IEEE，2017：1623-1631.

YE Song-tao，born in 1983，Ph.D，associate professor. His main research interests include truth discovery，data analysis and mining and action recognition.



FAN Hong-jie，born in 1984，Ph.D，lecturer. His main research interests include knowledge graphs，data exchange and data analysis and mining