

## 基于3D骨架相似性的自适应移位图卷积神经网络人体行为识别算法

闫文杰, 尹艺颖

引用本文

闫文杰, 尹艺颖. 基于3D骨架相似性的自适应移位图卷积神经网络人体行为识别算法[J]. 计算机科学, 2024, 51(4): 236-242.

YAN Wenjie, YIN Yiyang. [Human Action Recognition Algorithm Based on Adaptive Shifted Graph Convolutional Neural Network with 3D Skeleton Similarity](#)[J]. Computer Science, 2024, 51(4): 236-242.

---

## 相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

**Similar articles recommended (Please use Firefox or IE to view the article)**

[基于双路先验自适应图神经常微分方程的交通流预测](#)

Traffic Flow Prediction Model Based on Dual Prior-adaptive Graph Neural ODE Network  
计算机科学, 2024, 51(4): 151-157. <https://doi.org/10.11896/jsjcx.230100066>

[基于图卷积神经网络的节点分类方法研究综述](#)

Review of Node Classification Methods Based on Graph Convolutional Neural Networks  
计算机科学, 2024, 51(4): 95-105. <https://doi.org/10.11896/jsjcx.230600071>

[融合关系传递信息的双图文档级关系抽取方法](#)

Method of Document Level Relation Extraction Based on Fusion of Relational Transfer Information Using Double Graph  
计算机科学, 2023, 50(12): 229-235. <https://doi.org/10.11896/jsjcx.230500010>

[基于动态负采样的图卷积协同过滤推荐模型](#)

Dynamic Negative Sampling for Graph Convolution Network Based Collaborative Filtering Recommendation Model  
计算机科学, 2023, 50(11A): 230200149-7. <https://doi.org/10.11896/jsjcx.230200149>

[多流融合的轻量级图卷积行为识别算法](#)

Lightweight Graph Convolution Action Recognition Algorithm Based on Multi-stream Fusion  
计算机科学, 2023, 50(11A): 220800147-6. <https://doi.org/10.11896/jsjcx.220800147>

# 基于 3D 骨架相似性的自适应移位图卷积神经网络人体行为识别算法

闫文杰 尹艺颖

河北工业大学人工智能与数据科学学院 天津 300401

**摘要** 图卷积神经网络(Graph Convolutional Neural network,GCN)在基于 3D 骨架的人体行为识别领域取得了良好效果。然而,现有的大多数 GCN 方法对行为动作图的构建都是基于人体物理结构的手动设置,训练阶段各个图节点只能根据手动设置建立联系,无法感知动作作为过程中骨骼节点之间产生的新联系,导致图拓扑结构不合理和不灵活。移位图卷积网络通过改变图网络结构使得感受野更加灵活,并且在全局移位角度取得了良好效果。因此,提出了一种基于自适应移位图卷积神经网络(Adaptive Shift Graph Convolutional Neural network,AS-GCN)的人体行为识别算法来弥补前述 GCN 方法的不足。AS-GCN 借鉴了移位图卷积网络的思想,提出用每个人体动作的本身特点来指导图神经网络进行移位操作,以尽可能准确地选定需要扩大感受野的节点。在基于骨架的通用动作识别数据集 NTU-RGBD 上,所提算法在骨骼有无物理关系约束的前提下条件下均进行了实验验证。与现有的先进算法相比,AS-GCN 算法的动作识别准确率在有骨骼物理约束的条件下的 CV 和 CS 角度上平均提高了 12% 和 4.84%;在无骨骼物理约束的条件下的 CV 和 CS 角度上平均提高了 20% 和 14.49%。

**关键词:** 骨架动作分类;图卷积神经网络;行为识别;自适应移位

**中图分类号** TP391

## Human Action Recognition Algorithm Based on Adaptive Shifted Graph Convolutional Neural Network with 3D Skeleton Similarity

YAN Wenjie and YIN Yiyang

School of Artificial Intelligence, Hebei University of Technology, Tianjin 300401, China

**Abstract** Graph convolutional neural network(GCN) has achieved good results in the field of human action recognition based on 3D skeleton. However, in most of the existing GCN methods, the construction of the behavior diagram is based on the manual setting of the physical structure of the human body. In the training stage, each graph node can only establish the connection according to the manual setting, which cannot perceive new connections between bone nodes during action, resulting in the unreasonable and inflexible topology of the graph. The shifted graph convolutional neural network(Shift-GCN) makes the receptive field more flexible by changing its structure, and achieves satisfied results in the global shift angle. In order to tackle the above problems of graph structure, an adaptive shift graph convolutional neural network(AS-GCN) is proposed to make up for the above shortcomings. AS-GCN draws on the idea of shifted graph convolutional neural network, and proposes to use the characteristics of each human action to guide the graph network to perform shift operation, so as to select the nodes that need to expand the receptive field as accurately as possible. On the general skeleton-based action recognition dataset NTU-RGBD, the AS-GCN is verified by extensive experiments under the premise of whether the skeleton has physical relationship constraints or not. Compared with the existing advanced algorithms, the accuracy of action recognition of AS-GCN is improved by 12% and 4.84% respectively in CV and CS angles on average with skeleton physical constraints. While under the condition of no skeleton physical constraint, the average improvement is 20% and 14.49% in CV and CS angles, respectively.

**Keywords** Skeleton-based action classification, Graph convolutional neural network, Action recognition, Adaptive shift

## 1 引言

基于视觉的人体行为识别技术在日常生活中有广阔的应用前景和潜在的经济价值,它能够广泛应用于公共安全、人机交互以及自动驾驶等领域<sup>[1-2]</sup>。基于视觉的人体行为识别的研究,从数据上看有 RGB 数据和深度数据这两类。深度数据中基于骨架数据<sup>[3-4]</sup>的动作识别方法因对动态环境和复杂背景的强适应性而受到广泛关注。针对骨架数据的人体行为识别研究主要利用骨架数据中视频各个帧之间的节点变化来描述行为。

的输出作为当前时刻的输入来形成其结构内部的递归连接。Zhang 等<sup>[5]</sup>提出了一种简单而有效的元素注意门,它可以很方便地添加到 RNN 的任意神经元中,使神经元具有注意能力,能够自适应地调节模型的输入,完成对骨架序列的行为识别。

卷积神经网络(CNN)凭借优秀的高级信息提取能力,能够高效、轻松地学习高级语义线索。Li 等<sup>[6]</sup>提出了一种基于 CNN 的融合模型,通过灰度值编码将每个三维骨架序列的时空信息输入 CNN 融合模型中进行骨架行为识别。

图卷积神经网络(GCN)是一种能对图数据进行深度学习的方法。人体 3D 骨架数据是自然的拓扑图,顶点表示关节,边表示连接关节的肢节,可以用图卷积网络来发掘骨架

循环神经网络(RNN)能够处理骨架序列数据,将上一时刻

到稿日期:2022-12-20 返修日期:2023-03-30

基金项目:国家自然科学基金(61702157)

This work was supported by the National Natural Science Foundation of China(61702157).

通信作者:闫文杰(wenjieyanhit@163.com)

之间的空间联系,将图卷积操作拓展到时域上,就能同时发掘空间和时间特征。Li 等<sup>[7]</sup>提出了一种基于注意力机制的时空图卷积网络模型,在时空卷积模块当中增加了空间注意力机制和时间注意力机制,能同时捕捉空间特征与时间特征,还能够关注到运动过程中非运动区域的信息,进行骨架序列的人体行为识别。

文献[8]提出采用时空图卷积神经网络(Spatial Temporal Graph Convolutional Networks,ST-GCN)的方法进行行为识别,作者从物理学角度出发,将身体各部分的运动分为同心运动和偏心运动,并且将节点范围划分为三部分(源点、近心点和远心点)。大量的实验结果表明,该方法构建的时空结构能够有效地进行人体动作识别。Cheng 等<sup>[9]</sup>利用移位图卷积神经网络(Shift Graph Convolutional Neural network,Shift-GCN)来进一步增强空间图的表达能力,同时提高了空间图和时间图感受野的灵活性。本文受移位图卷积神经网络算法的启发,认为移位思想能够解决在特征提取时提取的人体信息具有局部性的问题,将节点的特征进行分区交换以扩大感受野,能够在对点卷积时考虑到动作完成需要其他部分的肢体配合,因此需要相关关节或骨骼的信息。Shift-GCN 利用局部移位和全局移位这两种方式进行实验验证,局部移位通过实验可以看出定义好的相关关节或骨骼是不准确的;全局移位即人体所有关节或骨骼进行信息交换又存在信息的冗余问题。

Shi 等<sup>[10-11]</sup>提出自适应图卷积网络,通过端到端的方式进行自适应的学习,增加了每个骨架样本的图拓扑结构。本文受上述方法的启发,通过数据驱动的方式,在有骨骼物理约束的前提下,利用人体物理结构和不同关节之间的相似度来指导 Shift-GCN 选择交换的节点和特征,从而缓解 Shift-GCN 处理局部信息不准确以及全局信息冗余问题;在无物理

骨骼约束的前提下,直接利用全部人体关节点进行图的拓扑学习,计算关节之间的相似度,通过相似度选择在 Shift-GCN 网络中需要进行特征交换的节点。经实验验证,无物理骨骼约束的方法能够显著提高人体行为识别关节模态的准确率。

本文的主要贡献为:

(1) 以数据驱动方式表示节点关系。利用不同动作的自身特点,在骨骼有无物理约束的前提下,通过节点特征学习,计算节点相似度,结合人体骨架的自然连接结构,共同表示构成图节点的拓扑结构。该方法能够成功解决因手动设置节点相关性而带来的节点间关系不准确的问题,使动作行为过程中产生相关性的节点关系也能够被充分感知。

(2) 利用骨骼数据的邻接矩阵指导节点进行特征交换。通过学习得到的图节点的拓扑结构,根据各个节点之间的相关性来判断需要进行移位交换的节点,进而能够通过邻接矩阵指导移位卷积神经网络进行图节点间的交换。该方法能够有效解决采用图卷积神经网络进行行为识别时图卷积网络的感受野不灵活的问题。

## 2 本文算法

本文提出的算法主要基于图卷积神经网络及残差网络的思想。在骨骼有物理约束的前提下,首先,将人体骨架数据进行预处理,得到人体骨骼数据和关节数据。其次,利用数据中关节的坐标以及人体物理结构自然连接模型构造空间结构特征和时序特征,并将二者相结合构成人体骨架的时空特征。再次,分别将关节数据和骨骼数据输入时空卷积神经网络中进行特征提取,特征提取过程中需要进行节点空间特征交换和时间特征交换。最后,将双流数据的训练结果进行融合得到最终的分类结果。算法框架图如图 1 所示。

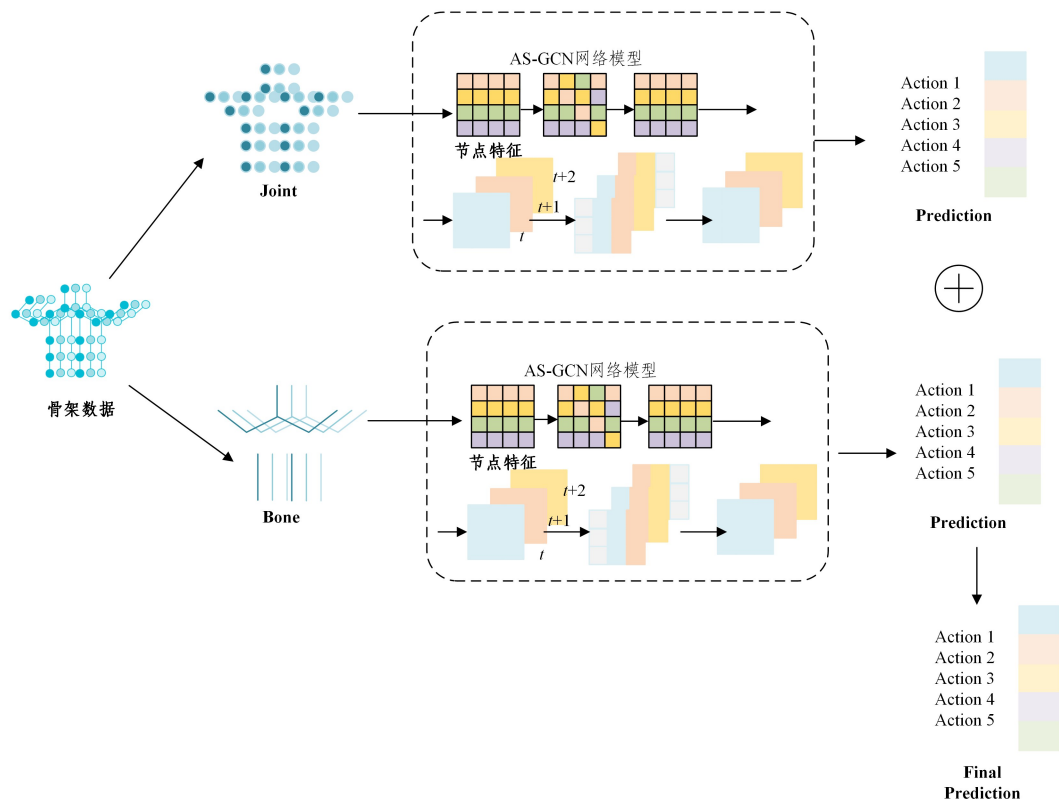


图 1 有骨骼物理约束条件的行为识别算法框图

Fig. 1 Block diagram of behavior recognition algorithm with skeletal physical constraints

在骨骼无物理约束的前提下,首先,将骨架数据进行预处理,得到人体关节数据。其次,利用关节坐标进行时空特征构建,得到人体时空骨架特征。再次,将关节数据输入

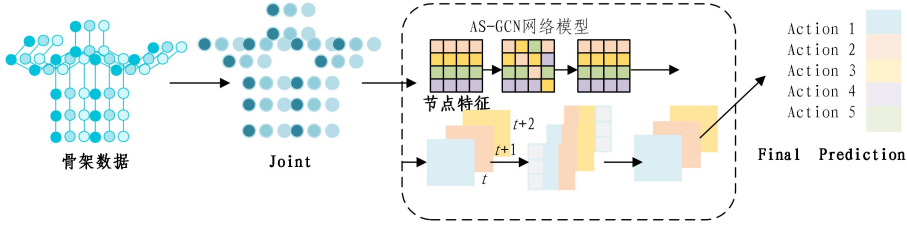


图2 无骨骼物理约束条件的行为识别算法框图

Fig. 2 Block diagram of behavior recognition algorithm without skeletal physical constraints

## 2.1 图构造

本文采用 NTU-RGBD 数据集<sup>[12]</sup>,该数据集定义了 25 个人体关节点,如图 3 所示。

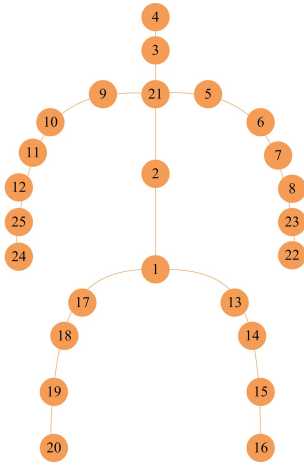


图3 NTU-RGBD 数据集人体骨骼关节点图示

Fig. 3 Human skeleton junction in NTU-RGBD dataset

视频数据一般由很多视频帧构成。骨架数据作为矢量序列,每个视频帧中的每个矢量表示每个采样时刻对应的人体关节坐标。根据人体骨骼的分布(如图 3 所示),将关节点进行标号处理得到骨骼关节点数据即 Joint,将在人体结构上存在连接的关节点进行连接,得到人体骨骼的生成数据即 Bone。处理后的人体关节点数据和骨骼数据如图 4 所示,其中  $t$  代表帧次。

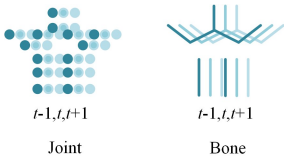


图4 关节数据和骨骼数据

Fig. 4 Joint data and bone data

## 2.2 图卷积神经网络构建

### 2.2.1 邻接矩阵构建

#### (1)有骨骼物理约束条件的邻接矩阵构建

本文构造的邻接矩阵  $\mathbf{A}$  中存在的物理连接关系如图 3 所示。但图 3 构造的邻接矩阵  $\mathbf{A}$  只是人体物理结构的直接连接。演示者进行动作演示时,不同部位的关节会产生新的物理连接关系。例如人在进行穿鞋的动作时,人体的手和脚会产生交互。此时人体的手脚因穿鞋的动作而建立了一个新的

时空卷积神经网络进行特征提取,特征提取过程中需要进行节点空间特征交换和时间特征交换。最后,进行分类预测,得到最终的分类结果。算法框图如图 2 所示。

连接,那么仅使用邻接矩阵  $\mathbf{A}$  进行人体关节点间的物理连接表示是不合理的。本文在此阶段采用人体物理结构形成的节点连接和动作形成的节点连接构成的复合节点图表示当前演示者动作的邻接矩阵。

$$\mathbf{U} = \mathbf{A}_k + \mathbf{S}_k \quad (1)$$

$\mathbf{U}$  是用于指导节点特征交换的复合矩阵, $\mathbf{A}_k$  为人体按照物理结构自然连接骨架图的矩阵表示,而  $\mathbf{S}_k$  则是人体动作数据的相关图。为了确定两个节点之间是否存在连接以及连接的强度,本文采用归一化嵌入高斯函数来计算两个节点之间的相似性:

$$f(v_i, v_j) = \frac{e^{\theta(v_i)^T \varphi(v_j)}}{\sum_{j=1}^N e^{\theta(v_i)^T \varphi(v_j)}} \quad (2)$$

其中,  $N$  是人体物理结构图总的节点数;  $\theta$  和  $\varphi$  是功能函数;  $f(v_i, v_j)$  表示两个节点之间的相似度。

由于归一化高斯函数有 softmax 的功能,因此可以根据式(3)进行计算得到  $\mathbf{S}_k$ :

$$\mathbf{S}_k = \text{softmax}(f_{in}^T \mathbf{W}_{\theta}^T \mathbf{W}_{\varphi} \mathbf{f}_{in}) \quad (3)$$

其中,  $\mathbf{W}_{\theta}$  和  $\mathbf{W}_{\varphi}$  分别是嵌入函数  $\theta$  和  $\varphi$  的参数。矩阵  $\mathbf{S}_k^{ij}$  中的每个元素表示为节点  $v_i$  和  $v_j$  之间的相似度,节点间的相似度由式(2)计算得到,并且其取值范围为  $0 \sim 1$ 。图 5 展示了  $\mathbf{S}_k$  在实验代码中的计算流程,首先,将给定的输入特征  $\mathbf{f}_{in} \in \mathbb{R}^{N \times C \times T \times V}$  嵌入到功能函数  $\theta$  和  $\varphi$  中,利用  $1 \times 1$  的卷积完成对嵌入特征图的重新排列和重塑,经过排列和重塑后特征的维度分别为  $N \times V \times CT$  和  $N \times CT \times V$ ,二者相乘即可得到维度为  $N \times N$  的相似矩阵  $\mathbf{S}_k$ 。

通过图 3 所示的人体物理结构图,可以将其关节点进行连接得到骨骼数据。同时可以通过式(2)和式(3)的计算方法得出关节点和骨骼数据之间的矩阵关系,其中每个元素的数值代表两个关节点之间是否存在物理连接或是动作连接关系。

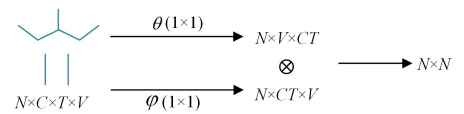


图5 人体骨骼节点的相似度计算

Fig. 5 Similarity calculation of bone nodes

#### (2)无骨骼物理约束条件的邻接矩阵构建

前面论述了在有骨骼物理约束条件下的节点间关系的表达方法。在无骨骼物理约束条件下,本文能够通过原始的关节数据节点,学习动作演示者在进行动作展示时各个骨骼关节点

之间的动态关系。人体骨骼的物理约束当中人的左右胳膊之间并没有直接的物理关联,但在图 6 所展示的射箭动作当中,人的左右手及其胳膊需要一起发力,此时就能够根据射箭动作的动态关节数据,建立起相应关节之间的物理关系图。

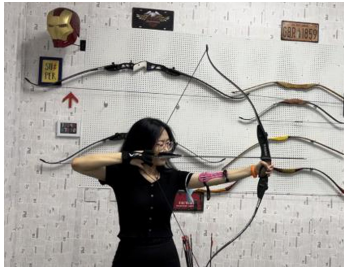


图 6 射箭动作

Fig. 6 Archery action

因此,本文通过直接计算节点间相似度的方式来得到关节之间的连接矩阵关系  $S$ ,计算方法如式(2)和式(3)所示。

### 2.2.2 特征交换

首先,本文在上一小节中根据人体骨骼节点图中各个关节的标号,且用无向图进行拓扑表示,从而得到了邻接矩阵  $A$ 。其次,根据式(1)~式(3)得到了人体骨骼节点的复合拓扑结构图,并且采用矩阵  $U$  表示图各个节点间的关系。最后,通过矩阵  $U$  指导骨骼节点间的特征交换。

受 Shi 等<sup>[10-11]</sup>的工作的启发,本文的节点交换从卷积操作的感受野角度出发,利用移位图卷积神经网络的移位思想对节点的特征进行交换。以图 7 中人体骨架结构的 10 个节点为例,选择其中 4 个节点 1,4,6,8 构建遵循人体物理结构连接的邻接矩阵,如矩阵  $A$  所示。上述节点形成的特征如图 7(c)所示。根据矩阵  $A$  中节点之间是否有物理或是动作联系来确定其特征是否需要交换。如需交换,则按照图 7(d)所示的方法进行,交换之后得到的新节点特征图如图 7(e)所示,在此基础上进行逐点卷积操作,如式(4)所示:

$$\mathbf{F}'_{v_i} = \mathbf{F}_{v_j} \parallel \mathbf{F}_{v_i} \quad (4)$$

其中,  $\mathbf{F}'_{v_i}$  是经过节点特征交换之后的节点  $v_i$  的特征,  $\mathbf{F}_{v_j}$  为节点  $v_j$  的原始特征,而  $\mathbf{F}_{v_i}$  是邻接矩阵  $A$  中与交换节点有关系的其他节点。

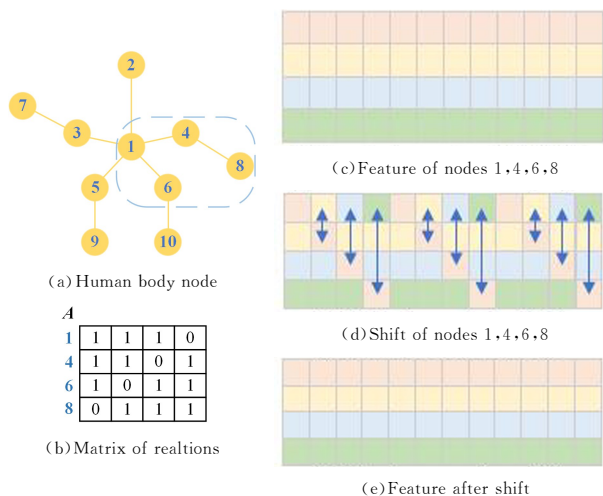


图 7 移位节点的特征交换

Fig. 7 Feature exchange of the shifted nodes

### 2.2.3 移位图卷积神经网络

本文的骨架数据经预处理后表示为具有 25 个关节点和 300 幅数据帧的时空关系图,用  $G=(V,E)$  表示。其模型结构分为空间图卷积和时间图卷积两个部分。

对于常规空间图卷积而言,骨架数据的邻接矩阵  $A \in \{0,1\}^{N \times N}$  通常被划分为 3 个分区:

(1)源节点本身。

(2)远心点:空间位置上比源节点更远离整个骨架重心的节点集合。

(3)近心点:空间位置上比源节点更靠近整个骨架重心的节点集合。

空间图的卷积运算如下:

$$\mathbf{F}_{out} = \sum_{p \in P} \mathbf{A}_p \mathbf{F}_{in} \mathbf{W}_p \quad (5)$$

其中,  $p$  表示节点本身、远心点或近心点,  $P$  表示人体结构中的关节点集合;  $\mathbf{A}_p$  为对应分区的邻接矩阵,  $\mathbf{W}_p$  为其权重矩阵;  $\mathbf{F}_{in}$  为输入特征矩阵;  $\mathbf{F}_{out}$  为输出特征矩阵。

如前所述,文中的邻接矩阵  $A$  通过数据驱动的方式求得。本文算法利用骨架数据本身的特点,使其自适应地进行节点间关系的学习,而后指导不同节点进行特征交换。本文的图卷积运算步骤如式(6)和式(7)所示:

$$\mathbf{F}_{out} = \mathbf{F}_{in} \mathbf{W} \quad (6)$$

$$\mathbf{F}_{in} = \mathbf{F}'_{v_i}, i=1,2,\dots,N \quad (7)$$

其中,  $\mathbf{F}_{out}$  为输出特征矩阵;  $\mathbf{W}$  为进行  $1 \times 1$  逐点卷积的权重矩阵;  $\mathbf{F}_{in}$  为输入特征矩阵,并且其特征是通过式(1)所得的矩阵  $U$  进行指导交换后得到的。

对于时间图卷积而言,其由人体行为数据视频帧中骨骼点数据的时间维度构造而成。朴素的时态移位图卷积将时态移位与时态逐点卷积相结合,将通道平均分成  $2t+1$  个区,之后每个通道分别在时间维度上位移  $-t, -t+1, \dots, 0, \dots, t-1, t$  的距离,移出的通道被截断,空通道用 0 来进行填充,如图 8 所示。朴素的时间图卷积通过这种交换不同数据帧上特征的方式扩大了感受野。通常,基于 GCN 的动作识别方法中常规时间卷积核的大小为  $9^{[10-11,13]}$ 。与常规时间卷积相比,朴素时间移位图卷积的计算成本降低了 90%。

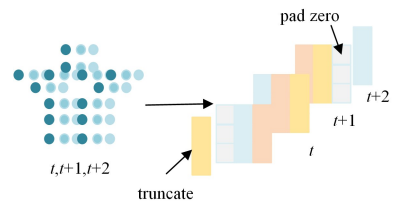


图 8 朴素时间移位卷积

Fig. 8 Naive time-shifted convolution

朴素时间移位图卷积需要手动进行参数设置。由于在进行动作分类时,不同的人体行为数据集需要的感受野是不同的,而且在不同的移位图卷积网络层中感受野通常也是不相同的,因此,手动进行参数设置并不科学,这会限制朴素时间移位图卷积的泛化能力。故本文中时间图卷积的构造部分采用自适应时间移位图卷积,通过自适应的方式来解决朴素时间移位图卷积中存在的问题。

在给定的骨架数据中,其序列特征为  $\mathbf{F}^{N \times T \times C}$ ,让每个通道都有一个可学习的参数  $S$ 。自适应方式使得参数  $S$  的值从

整数约束放宽到了实数。当  $S$  是非整数时, 计算方式如下:

$$\mathbf{F}'_{(v,t,i)} = (\alpha) \cdot \mathbf{F}_{(v,\lfloor t+S \rfloor,i)} + (1-\alpha)\mathbf{F}_{(v,\lfloor t+S \rfloor+1,i)} \quad (8)$$

其中,  $\mathbf{F}$  表示输入的特征矩阵,  $v$  表示节点,  $t$  表示某一时刻的数据帧,  $i=1,2,\dots,C$  表示通道。在  $t$  帧的特征通过  $\lfloor t+S \rfloor$  帧及相邻帧  $\lfloor t+S \rfloor+1$  来表示, 并且用参数  $\alpha$  和  $(1-\alpha)$  进行调节, 这样既避免了手动进行参数设置, 也能够扩大时间卷积时  $t$  帧的感受野。通过将此操作与逐点卷积相结合, 就可得到自适应时间移位卷积。

### 2.3 基于 3D 骨骼相似性的自适应移位卷积神经网络模型

从整体上看, 本模型的网络结构共有 9 个基本块 (B1-B9)。每个块的 3 个数字分别代表输入通道数、输出通道数和步幅。GAP 表示全局平均池化层。如图 9 所示, 网络的整体架构由这些基本块的堆栈构成。在网络的输入部分添加数据 BN 层, 使输入数据规范化。每个块的输出通道数分别为 64, 64, 64, 128, 128, 128, 256, 256 和 256。每个基本块均由空间卷积和时间卷积两部分构成, 在进行空间特征提取之前, 采用 2.2.1 节构建的矩阵进行 2.2.2 节所述方式进行特征交换, 然后根据 2.2.3 节所述空间卷积部分进行空间特征提取, 接着根据 2.2.3 节所述时间卷积部分进行时间卷积感受野的扩大与时间特征提取, 如图 10 所示。经过 9 个这样的基本块的操作之后, 在末端执行全局平均池化, 以将不同样本的特征映射为相同大小。最终的输出经过 softmax 层后获得预测结果。

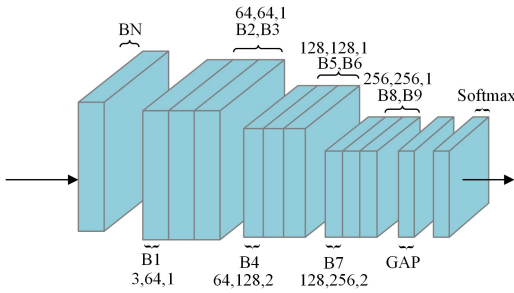


图 9 网络模型

Fig. 9 Network model

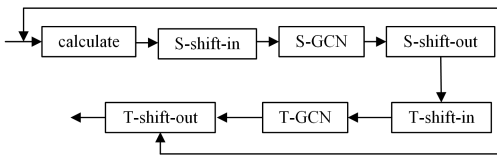


图 10 基本块中的具体操作

Fig. 10 Specific operations in basic block

## 3 实验与结果分析

### 3.1 实验设置

#### 3.1.1 实验数据集

NTU-RGBD 是目前较为流行的公开的人体行为识别数据集, 本文主要利用其中的骨架数据。该数据集在进行训练时, 通常划分为两类数据: 一类是从相机视角即 Xview 维度进行划分, 通过 3 个配置相同的相机分别从水平角度的  $-45^\circ, 0^\circ, 45^\circ$  进行演示者的动作采集, 每位演示者做两遍相同的动作, 一次对着左边的相机, 一次对着右边的相机; 而另一类是从人员角度即 Xsubject 维度进行划分, 对每个演示者唯一的 id 进行标识。

#### 3.1.2 评估方案

有骨骼物理条件约束下的 AS-GCN-A 算法和无骨骼物理条件约束下的 AS-GCN-S 算法的目的都是识别出每个动作演示者所做的动作, 因此本文将动作识别的准确率作为评价指标, 即 Top-1。由于 NTU-RGBD 数据集的划分方式有两种, 因此本文在进行对比时将 Xview 维度称为 CV, 将 Xsubject 维度称为 CS。

#### 3.1.3 实验实现细节

本文按照 8:2 的比例将数据分为训练集和测试集。在所有的实验中, 利用训练动态调整学习率, 以确保达到最佳性能。具体来讲, 本文的权值衰减为 0.0001, 每次训练 16 个样本数据, 初始学习率为 0.1, 在第 40, 60, 80 个训练轮次进行学习率的缩减。除此之外, 无骨骼物理条件约束下的 AS-GCN-S 算法还需要设置  $S_t$  的阈值为 0.3。

### 3.2 对比实验

有骨骼物理条件约束下的 AS-GCN-A 算法采用了两类数据流: 第一类使用原始骨架坐标作为输入, 称为“关节流”, 即“joint”; 第二类利用骨骼的二阶信息即骨骼信息作为输入, 即“bone”。将这两类数据流的 softmax 分数相加以获得最终的识别结果。无骨骼物理条件约束下的 AS-GCN-S 算法是单数据流模式, 其主要利用关节流进行节点关系矩阵构造, 在没有骨骼约束的条件下, 根据演示者动作进行节点关系自适应学习, 而后指导节点特征交换, 从而进行动作识别。

相比利用图结构进行动作识别的算法 2s-AGCN, MS-AAGCN, SGN<sup>[14]</sup>, AS-GCN-A 算法在 CV 维度上准确率分别高出 21%, 17%, 0.17%; 在 CS 维度上准确率分别高出 6.89%, 6.89%, 1.45%。

相比利用不同的网络模块结构进行动作识别的算法 GCN-NAS<sup>[15]</sup> 和 Shift-GCN, AS-GCN-A 算法在 CV 维度上比 GCN-NAS 高出 13%; 在 CS 维度上比 GCN-NAS 高出 5.17%, 比 Shift-GCN 高出 3.45%。

相比提升参数数量的动作识别算法 Decouple GCN<sup>[16]</sup>, AS-GCN-A 算法在 CV 维度上准确率高出 21%; 在 CS 维度上准确率高出 5.17%。

表 1 有骨骼物理条件约束的 Top-1 精度对比

Table 1 Top-1 accuracy comparison with skeletal physical conditions

Algorithms	Top-1 (%)	
	CV	CS
Shift-GCN	96.00	81.03
2S-AGCN	75.00	77.59
MS-AAGCN	79.00	77.59
SGN	95.83	83.03
GCN-NAS	83.00	79.31
Decouple GCN	75.00	79.31
AS-GCN-A(Ours)	<b>96.00</b>	<b>84.48</b>

注: 表中所用算法数据流为 joint 和 bone。

相比利用图结构进行动作识别的算法 2s-AGCN 和 MS-AAGCN, AS-GCN-S 算法在 CV 维度上准确率分别高出 21%, 34%; 在 CS 维度上准确率分别高出 12.08%, 12.07%。

相比利用不同的网络模块结构进行动作识别的算法 GCN-NAS 和 Shift-GCN, AS-GCN-S 算法在 CV 维度上比

GCN-NAS 高出 24%;在 CS 维度上比 GCN-NAS 高出 29.31%,比 Shift-GCN 高出 8.63%。

相比提升参数数量的动作识别算法 Decouple GCN, AS-GCN-S 算法在 CV 维度上准确率高出 21%;在 CS 维度上准确率高出 10.35%。

从表 1 和表 2 中的实验数据可知,与目前最先进的算法相比,本文提出的算法 AS-GCN-A 和 AS-GCN-S 在 Top-1 评价指标的 CV 角度的识别精度与 Shift-GCN 相当,比其他基线算法的识别精度都要高。从 CS 的角度看,本文提出的算法 AS-GCN-A 的识别精度比其他先进算法都要高。此外,在无骨骼物理条件约束的条件下的 AS-GCN-S 算法的识别精度要远远高于目前的其他先进算法和本文提出的 AS-GCN-A 算法。该实验结果也有力地验证了本文算法章节的理论推导。

表 2 无骨骼物理条件约束下的 Top-1 精度对比

Table 2 Top-1 accuracy comparison without skeletal physical conditions

Algorithms	Top-1 (%)	
	CV	CS
	Shift-GCN	96.00
2s-AGCN	75.00	77.58
MS-AAGCN	62.00	77.59
GCN-NAS	72.00	60.35
Decouple GCN	75.00	79.31
AS-GCN-S(Ours)	<b>96.00</b>	<b>89.66</b>

注:表中采用算法的数据流为 joint。

### 3.3 消融实验

在本节中,使用 NTU-RGBD 数据集上的 Xsubject 基准测试模型中各个组件的有效性。将与本文算法最为相近的 Shift-GCN 算法作为实验中的基线,在 NTU-RGBD 数据集上的原始性能为 81.03%,下述消融实验的所有结果如表 3 所列。

#### 3.3.1 移位交换中节点相似性消融实验

AS-GCN-A(A)模型是由人体骨架数据的物理连接构成的邻接矩阵来引导 AS-GCN 扩大节点的感受野,并且根据固定的邻接矩阵结构来引导关节的位置交换。表 3 中的实验结果表明 AS-GCN-A(A)模型比 Baseline 算法的识别准确率高出 0.56%。

AS-GCN-A(S)模型则是通过直接计算节点间的相似性来得到骨骼数据的邻接矩阵,更多地关注做动作时每个节点之间产生的关系,从而引导移位图卷积神经网络进行节点位置的交换。

Baseline 算法在增大每个节点感受野时需要遍历每个节点,NTU-RGBD 数据集一共有 25 个关节,将关节的特征进行分区,每一层网络在进行训练时,换进和换出一共需要交换  $2 \times 25 \times$  分区数。AS-GCN-A 虽然遍历了 25 个节点,但由于本文采用的邻接矩阵为无向图,网络模型在训练时只需要遍历上三角矩阵,因此可以将交换次数缩减为原来的一半。此外,实验结果表明 AS-GCN-A 的准确率比 Baseline 高出 3.45%,从而验证了本文提出的有骨骼物理约束条件下以数据驱动方式进行邻接矩阵学习并且指导移位卷积思想的有效

性。无骨骼物理约束条件下的 AS-GCN-S 算法不但减少了骨骼节点特征交换的次数,而且动作识别的准确率比 Baseline 算法高出 8.63%,这同样验证了本文提出的无骨骼物理约束条件下以数据驱动方式进行节点相似度邻接矩阵学习并且指导其进行骨骼节点移位方法的有效性。

#### 3.3.2 特征移位消融实验

如表 3 所列,S-GCN 模型是利用节点特征矩阵表明各节点之间的相关性,而后利用节点相关性表达的图结构进行训练,最后得到行为识别结果,其准确率为 77.58%。经过对比分析发现,本文提出的 AS-GCN-A 模型和 AS-GCN-S 模型增加了特征移位操作,使动作识别的准确率比 S-GCN 模型高出 6.9% 和 12.0%,验证了移位思想中特征交换方式的有效性。

表 3 AS-GCN 网络模型中不同部分的有效性比较

Table 3 Validity comparison of different parts of AS-GCN

network module (%)	
Model	CS
Baseline	81.03
AS-GCN-A(A)	81.59
AS-GCN-A(S)	81.03
S-GCN	77.58
ST-GCN	22.41
AS-GCN-A	<b>84.48</b>
AS-GCN-S	<b>89.66</b>

#### 3.3.3 自适应时空卷积模块消融实验

ST-GCN 模型是一种时空图卷积神经网络的模型,根据节点位置信息确定节点之间的相关性,而后利用节点相关性表达的图结构进行训练,最后得到行为识别结果。ST-GCN 模型在利用图结构进行训练的过程中没有利用本文提出的节点相似性计算进行特征交换,并且在时间卷积中没有根据节点相关性灵活地调整感受野。其虽然利用时空卷积来提取特征,但并不能根据不同动作进行自适应,所以识别效果较差。

相比 ST-GCN 模型,AS-GCN-A 模型和 AS-GCN-S 模型增加了时空图卷积的自适应特点,能够根据视频中动作节点信息来计算节点相似性,利用节点相似性指导不同节点进行特征交换,从而灵活地调节空间图的感受野,并且,它们还利用时间特征交换方式来自适应调整时间图的感受野。表 3 中的结果验证了本文提出的模型自适应时空卷积这一部分的有效性,AS-GCN-A 和 AS-GCN-S 模型动作识别的准确率分别高出 ST-GCN 模型 62.07% 和 67.25%。

**结束语** 本文提出了以数据驱动的方式在有骨骼数据物理约束的条件下进行骨骼数据邻接矩阵的学习,进而指导自适应移位图卷积神经网络进行骨骼节点位置的交换。AS-GCN-A 和 AS-GCN-S 方法不仅提高了算法的灵活性和通用性,而且有效提升了在 NTU-RGBD 数据集上的动作识别的准确度。本文通过理论分析以及实验验证,表明从骨骼数据本身以及动作本身出发,能够更加精准地挖掘出骨骼节点在人体行为识别中的动态关系,为人工智能的识别算法提供更加实时和准确的邻接矩阵关系,从而为人体行为识别领域的算法研究提供了理论和实践基础。在未来的研究中,课题组

拟进一步探究骨骼节点数据图的拓扑结构,并且从动作执行时的人体物理结构的角度出发,研究有向图表示对行为识别算法的影响。

## 参 考 文 献

- [1] KONG W, LIU Y, LI H, et al. A survey of action recognition methods based on graph convolutional network [J]. *Control and Decision*, 2021, 36(7): 1537-1546.
- [2] LIANG X, LI W X, ZHANG H N. Review of research on human action recognition methods [J]. *Application Research of Computers*, 2022, 39(3): 651-660.
- [3] ZHAO X H, YE S, LI X. Multi-algorithm Fusion Behavior Classification Method for Body Bone Information Reconstruction [J]. *Computer Science*, 2022, 49(6): 269-275.
- [4] MIAO G Q, XIN W T, LIU R Y, et al. Graph Convolutional Skeleton-based Action Recognition Method for Intelligent Behavior Analysis [J]. *Computer Science*, 2022, 49(2): 156-161.
- [5] ZHANG P, XUE J, LAN C, et al. EleAtt-RNN: Adding attentiveness to neurons in recurrent neural networks [J]. *IEEE Transactions on Image Processing*, 2019, 29: 1061-1073.
- [6] LI M, SUN Q. 3D Skeletal Human Action Recognition Using a CNN Fusion Model [J]. *Mathematical Problems in Engineering*, 2021(18): 6650632, 1-6650632, 11.
- [7] LI Y Z, YUAN J Z, LIU H Z. Human skeleton-based action recognition algorithm based on spatiotemporal attention graph convolutional network model [J]. *Journal of Computer Applications*, 2021, 44(7): 1915-1921.
- [8] YAN S, XIONG Y, LIN D. Spatial temporal graph convolutional networks for skeleton-based action recognition[C]// *Thirty-second AAAI Conference on Artificial Intelligence*. Palo Alto, California USA; AAAI Press, 2018: 7444-7452.
- [9] CHENG K, ZHANG Y, HE X, et al. Skeleton-based action recognition with shift graph convolutional network[C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. San Francisco, CA, USA; IEEE, 2020: 183-192.
- [10] SHI L, ZHANG Y, CHENG J, et al. Two-stream adaptive graph convolutional networks for skeleton-based action recognition [C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. San Francisco, CA, USA; IEEE, 2019: 12026-12035.
- [11] SHI L, ZHANG Y, CHENG J, et al. Skeleton-based action recognition with multi-stream adaptive graph convolutional networks [J]. *IEEE Transactions on Image Processing*, 2020, 29: 9532-9545.
- [12] SHAHROUDY A, LIU J, NG T T, et al. Ntu rgb+ d: A large scale dataset for 3d human activity analysis[C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. San Francisco, CA, USA; IEEE, 2016: 1010-1019.
- [13] LI B, LI X, ZHANG Z, et al. Spatio-temporal graph routing for skeleton-based action recognition[C]// *Proceedings of the AAAI Conference on Artificial Intelligence*. Palo Alto, California USA; AAAI Press, 2019, 33(01): 8561-8568.
- [14] ZHANG P, LAN C, ZENG W, et al. Semantics-guided neural networks for efficient skeleton-based human action recognition [C]// *proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. San Francisco, CA, USA; IEEE, 2020: 1112-1121.
- [15] PENG W, HONG X, CHEN H, et al. Learning graph convolutional network for skeleton-based human action recognition by neural searching[C]// *Proceedings of the AAAI Conference on Artificial Intelligence*. Palo Alto, California USA; AAAI Press, 2020, 34(3): 2669-2676.
- [16] CHENG K, ZHANG Y, CAO C, et al. Decoupling gcn with dropgraph module for skeleton-based action recognition [C] // *European Conference on Computer Vision*. Cham: Springer, 2020: 536-553.



**YAN Wenjie**, born in 1983, Ph.D, associate professor. His main research interests include deep learning and computer vision.

(责任编辑:何杨)