计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.220100061

面向智慧教育行为分析的图卷积骨架动作识别方法

苗启广　辛文天　刘如意　谢　琨　王　泉　杨宗凯
西安电子科技大学计算机科学与技术学院　西安 710071
(qgmiao@xidian.edu.cn)

摘　要　智慧教育即教育信息化,是利用现代信息技术的新一代教育模式,智慧行为分析是智慧教育系统的核心组成。在面对复杂的教室应用场景时,针对传统的行为识别分类算法的精确性与时效性都存在严重不足的问题,提出了一种基于分离与注意力机制的图卷积(Depthwise Separable Attention Graph Convolutional Network,DSA-GCN)骨架动作识别算法。首先,为解决传统算法在通道域信息聚合天生不充分的难题,通过逐点卷积进行多维通道映射,将时空图卷积对输入骨骼序列的原始时空信息的保护能力与深度可分离卷积在空间和通道特征学习上的分离能力相结合,以增强模型特征学习与抽象表达性。其次,采用多维度融合的注意力机制,在空间卷积域利用自注意力与通道注意力机制来提升模型的动态敏感性,在时间卷积域利用时间与通道注意力融合法来增强对关键帧的判别力。实验结果表明,在 NTU RGB+D 和 N-UCLA 两个大型数据集上,DSA-GCN 都获得了优异的性能和效能表现,证明了模型对通道域信息聚合能力的提升。

关键词:行为识别;智慧行为分析;骨架动作分类;图卷积神经网络;深度可分离卷积;注意力机制
中图法分类号　TP391

Graph Convolutional Skeleton-based Action Recognition Method for Intelligent Behavior Analysis

MIAO Qi-guang,XIN Wen-tian,LIU Ru-yi,XIE Kun,WANG Quan and YANG Zong-kai
School of Computer Science and Technology,Xidian University,Xi'an 710071,China

Abstract　Smart education is a new education model using modern information technology,and smart behavior analysis is the core component. In the complex classroom scenarios,traditional action recognition algorithms are seriously deficient in accuracy and timeliness. A graph convolutional method based on separation and attention mechanism (DSA-GCN) is proposed to solve the above problems. First,in order to solve the challenge that traditional algorithms are inherently inadequate in aggregating information in the channel domain,multidimensional channel mapping is performed by point-wise convolution,combining the ability of ST-GC to preserve the original spatio-temporal information with the separation ability of depth-separable convolution in spatial and channel feature learning to enhance model feature learning and abstract expressivity. Second,a multi-dimensional fused attention mechanism is used to enhance the model dynamic sensitivity in the spatial convolution domain using self-attention and channel attention mechanisms,and to enhance the key frame discrimination in the temporal convolution domain using temporal and channel attention fusion method. Experiment results show that DSA-GCN achieves better accuracy and effectiveness performance on NTU RGB+D and N-UCLA datasets,and prove the improvement of the ability to aggregate channel information.

到稿日期:2022-01-06　返修日期:2022-01-09
基金项目:国家新工科研究与实践项目(E-GCJYZL20200818);全国高等院校计算机基础教育研究会计算机基础教育教学研究项目(2021-AFCEC-459);中国成人教育协会"十四五"成人继续教育科研规划重点课题(2021-414ZA);陕西高等教育教学改革研究重点攻关/重点项目(21JG001,21BZ014);广西可信软件重点实验室研究课题(KX202061,KX202041);西安电子科技大学教育教学改革研究重点攻关项目(A21003);新实验开发与新实验设备研制重点项目(SY21022I);西安电子科技大学重庆集成电路创新研究院资助产学研项目(CQIRI-CXYHT-2021-06)
This work was supported by the National New Engineering Research and Practice Project(E-GCJYZL20200818),Computer Basic Education Teaching Research Project of the National Institute of Computer Basic Education Research Association(2021-AFCEC-459),China Adult Education Association's "14th Five-Year Plan" Adult Continuing Education Research Plan Key Project(2021-414ZA),Key Research/Key Projects of Shaanxi Higher Education Teaching Reform Research(21JG001,21BZ014),Guangxi Key Laboratory of Trusted Software(KX202061,KX202041),Xidian University Education and Teaching Reform Research Key Research Project(A21003),New Experimental Development and New Experimental Equipment Development Key Projects(SY21022I) and Academy of Integrated Circuit Innovation of Xidian University in Chongqing IUR Project(CQIRI-CXYHT-2021-06).
通信作者:刘如意(ruyiliu@xidian.edu.cn)

Keywords Action recognition，Smart behavior analysis，Skeleton-based action classification，Graph convolutional neural network，Depth-wise separable convolution，Attention mechanism

1 引言

当前，以大数据分析、人工智能等信息技术为支撑的智慧教育模式已成为教育信息化发展的趋势。通过动作识别等深度学习方法自主解析学生和教师在课堂上的行为表现，进而结合考核指标完成智能分析，是智慧教室的核心组成部分。

为了从多数据维度分析动作识别问题，现有方法利用视觉外观、深度信息、光流，甚至声音等进行融合、辅助识别。与视觉外观数据相比，骨骼数据提供了高度抽象的信息，因为其只包含人体关键关节的二维或三维位置数据，避免了环境噪声（如背景杂乱、光照条件、衣服）的干扰，使得动作识别算法可以专注于动作本身[1-3]。尤其是使用骨架时空图来建模动作关节运动的方法，在 NTU-RGB+D 等大规模人体行为识别数据集上表现优异。

在基于骨架的动作识别领域，时空图卷积网络（Spatial Temporal Graph Convolutional Network）[4]最先使用图卷积（Graph Convolutional Network）对骨架数据进行建模。文献[4]以人体关节的自然连接构建了时空图，提出了一种以图卷积层为基本模块的时空网络模型。2s-AGCN[5]将骨骼的二阶信息（人体骨骼的长度和方向）与一阶信息（关节的坐标）相结合，提出了一种双流自适应 GCN。MS-G3D[6]集成了一种离散的多尺度聚合方案和一种名为 G3D 的时空图卷积算子，组成了强有力的特征提取结构。Dynamic-GCN[7]为增强上下文特征相关性、自动学习骨架拓扑，引入了上下文编码网络。REGINA[8]首先将空间图卷积扩展到时间图卷积，实现了时空统一图架构。在数据构建改进中，Ang-Net[9]融合了三阶特征来有效捕获关节与身体部位之间的关系。MST-GCN[10]引入了一种新颖的渐进多尺度卷积，用于捕获空间和时间域的长短期相关性。CTR-GCN[11]采用多尺度时间卷积并利用原始数据的相关性更好地建模了信道拓扑。但上述采用了时空图卷积架构的方法会产生 4 类共性问题：1）图的拓扑结构在所有层上都是固定的，对不同层中包含的多层次语义进行建模缺乏灵活性；2）图结构具有数据依赖性，对不同动作样本难以保证最优性；3）现有算法的计算复杂度过高，无法满足实际应用的实时性需求；4）设计图卷积核的策略未考虑卷积核的数量与图结构的复杂性之间的关系[12-14]，因此网络的拓扑结构会因过度光滑[15]在学习过程中不能充分捕捉远距离关节之间的依赖联系。

深度可分离卷积由独立输入通道空间卷积（Depth-wise Convolution，DC）和 1×1 映射空间卷积（Point-wise Convolution，PC）组成，该框架减少了卷积运算的参数和计算量，提高了特征表达能力。如果将可分离的思想迁移至时空图卷积的架构中，以逐点卷积进行多维通道映射，可以在极大地增强模型抽象表达能力的同时减少参数量并降低计算复杂度。另一方面，在骨架数据的每个维度中，各个特征对动作识别任务的贡献是不同的，有些甚至会产生噪声，通过注意力机制强化重要的特征，可以提升模型的性能和稳定性。

基于以上研究，我们将深度可分离卷积思想与多维度注意力机制引入时空图卷积中，提升了骨架行为识别任务的最终性能。本文的主要工作如下：1）提出了一种基于可分离卷积策略的通道膨胀方法与自注意力策略相融合的方案，实现了多层次的特征抽取与融合方法，解决了传统 ST-GCN 算法在通道域信息聚合天生不充分的难题；2）提出了维度注意力融合方案，通过扩展注意力机制的思想，在 GCN 与 TCN 层后分别添加空间-通道注意力算法与时间-通道注意力算法，加强了关键节点的特征判别力，集中了骨架数据维度中的重要信息；3）在 NTU RGB+D 和 N-UCLA 两个大型数据集上，本文方案都获得了较好的性能和效能表现，并通过消融实验和可视化方法进行了有效验证。

2 研究背景

2.1 图卷积神经网络

通过计算图拉普拉斯算子的特征分解，在傅里叶域中定义卷积运算。图上的谱卷积可以定义为信号 $\boldsymbol{x}\in\mathbb{R}^N$ 与滤波器 $g_\theta=\mathrm{diag}(\theta)(\theta\in\mathbb{R}^N)$ 在傅里叶域的乘积[16]，如式(1)所示：

$$g_\theta\cdot\boldsymbol{x}=\sigma(\boldsymbol{U}g_\theta(\Lambda)\boldsymbol{U}^{\mathrm{T}}\boldsymbol{x}) \tag{1}$$

其中，$\boldsymbol{U}$ 为归一化图拉普拉斯方程 L 的特征向量矩阵，对角线上元素依次为图中各节点的度，A 为图的邻接矩阵，Λ 为特征值对角矩阵，$U^{\mathrm{T}}\boldsymbol{x}$ 为 $\boldsymbol{x}$ 的图傅氏变换，$g_\theta(\cdot)$ 为 L 的特征向量函数，$\sigma(\cdot)$ 为激活函数，$\Theta=(\theta_1,\theta_2,\cdots,\theta_n)$ 为可学习参数，$\boldsymbol{x}$ 为图上对应每个顶点的特征向量。该方法每一次前向传播均需计算 $\boldsymbol{U}$，$\mathrm{diag}(\theta)$ 及 $\boldsymbol{U}^{\mathrm{T}}$ 三者的矩阵乘积，在大规模图计算中，计算复杂度为 $O(n^2)$。通过采用切比雪夫多项式 $T_k(\boldsymbol{x})$ 计算第 k 阶展开来获得 $g_\theta(\cdot)$ 的近似[17]：

$$g_\theta(\boldsymbol{\Lambda})=\sum_0^{k-1}\beta_k T_k(\bar{\boldsymbol{\Lambda}}) \tag{2}$$

其中，$T_k(\cdot)$ 是 k 阶的切比雪夫多项式，β_k 为训练中迭代更新的参数，$\bar{\boldsymbol{\Lambda}}$ 为特征值对角矩阵，其中元素介于[−1，1]之间。至此，谱图卷积不再依赖整个图，只是依赖距离中心节点 K 步之内的邻居节点。在 GCN 的优化模型中[15]，进一步优化了切比雪夫多项式的二阶特例，通过式(3)计算得到谱图卷积中第 i 个节点的更新特征。

$$X^{l+1}=\sigma(\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{A}\boldsymbol{\Lambda}^{-\frac{1}{2}}X^lW^l) \tag{3}$$

其中，$\boldsymbol{A}$ 表示图的邻接矩阵，$\boldsymbol{\Lambda}$ 为 $\boldsymbol{A}$ 的度矩阵，W 为可训练参数。

2.2 时空图卷积网络（ST-GCN）

ST-GCN[4]模型首次把人体骨架序列模型化为以关节点为顶点、骨骼为边的空间图，然后将相邻两帧中的相同关节连接起来进一步构成时空图。在 ST-GCN 中，每个节点及其邻居根据分区策略被归为 3 类。具体来说，每个节点和它的相邻节点形成一个邻居集，其中中心节点被称为根节点，分割策略将邻居集分为 3 个子集，即根节点本身、向心组和离心组，图的邻接矩阵也被分为 3 部分，即 $\boldsymbol{A}_1$，$\boldsymbol{A}_2$ 和 $\boldsymbol{A}_3$，通过将 3 个矩阵作为邻接矩阵融合的方法求得动作的最终空间特征。

$$X^{l+1}=\sigma\left(\sum_{j=1}^{3}\boldsymbol{\Lambda}_j^{-\frac{1}{2}}\boldsymbol{A}_j\boldsymbol{\Lambda}_j^{-\frac{1}{2}}X^l W_j^l\right) \tag{4}$$

3 算法设计

3.1 参数定义

人体骨架数据的表现形式为一系列帧，每一帧都有一组或多组联合坐标，给定 2D 或 3D 坐标下的身体关节序列，根据人体骨架关节点之间天然的连接关系，构成了天然的骨架图结构。人体的骨架图被定义为 $G=(V,E)$，其中 V 代表 N 个关节的节点集合，E 表示边的骨骼集合，在对称邻接矩阵 $\boldsymbol{A}\in\mathbb{R}^{N\times N}$ 中(假定 G 为无向图)，如果 v_i 和 v_j 关节有骨架直接连结，则 $\boldsymbol{A}_{i,j}=1$，否则 $\boldsymbol{A}_{i,j}=0$。图序列特征表示为 $X=\{x_{t,n}\in\mathbb{R}^C\mid t,n\in\mathbb{Z},1\leqslant t\leqslant T,1\leqslant n\leqslant N\}$，特征张量可以表示为 $\boldsymbol{X}\in\mathbb{R}^{T\times N\times C}$，其中 $\boldsymbol{x}_{t,n}$ 表示在总计 T 帧中，节点 v_n 在 t 时刻的 C 维特征向量。输入动作可以由 A 和 X 来描述，X_t 是 t 时刻的节点特征，$\boldsymbol{\Theta}^{(l)}$ 表示网络在第 l 层的可学习矩阵。

3.2 “分离-注意”图卷积网络

本文提出的基于分离与注意机制的图卷积网络结构如图 1 所示，其包含 9 组图卷积层。其中，动作特征可分为空间特征和时间特征，图卷积部分的功能是抽取空间特征，在每个图卷积后增加一个时间卷积来抽取时间特征，进而每个图卷积和时间卷积构成网络的一个基本模块。将 9 个基本模块堆叠在一起，进一步形成最终的网络，其中基本模块主要由实现空间卷积的可分离-自注意力图卷积模块、时间卷积和注意力模块构成。B_1-B_9 分别表示 9 个基本模块，GAP 表示全局平均池化层。需要特别说明的是，SC-Attention 包含通道注意力机制和自注意力邻接矩阵方法，TC-Attention 包含时间通道融合注意力模块，两者代表的算法存在细节上的差异。本文提出的相关算法的流程如图 2、图 3 所示，模型整体分为两大模块，即训练模块和测试模块。训练模块由两组分支构成，即关节序列流和骨骼序列流。关节序列流由 PC(Point-wise 卷积模块)、DC(depth-wise 卷积模块)和维度注意力模块构成，其中 PC 和 DC 共同构成深度可分离空域卷积，维度注意力模块(特指时间、通道)构成时域卷积。骨骼序列采用的算法流程与关节序列相同，但两者分别训练，即具有不同的网络参数。训练模块的反向传播损失采用 Cross-Entropy-Loss，为完全监督训练。测试模块的双流数据将分别采用最优的第 G 代骨骼模型参数和第 K 代关节模型参数，在网络的最后一层将输出结果求和，输出最高值对应的标签为最终预测结果。

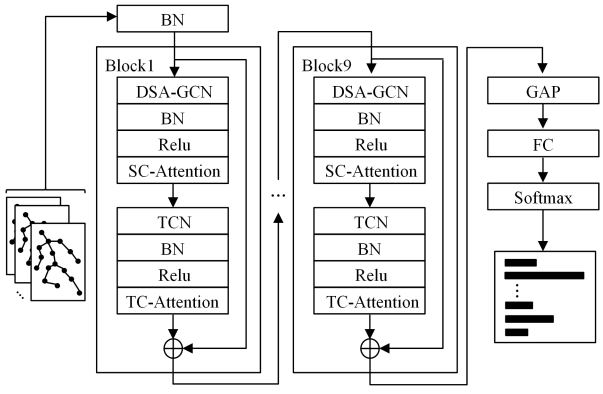

图 1 DSA-GCNs 网络结构

Fig. 1 Network structure of DSA-GCNs

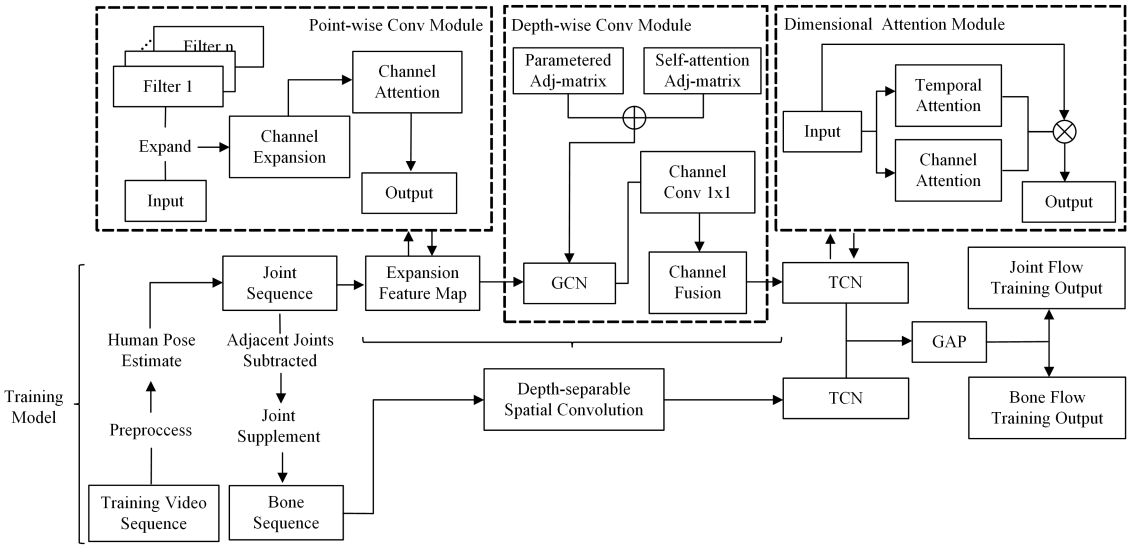

图 2 DSA-GCN 训练流程

Fig. 2 Training process of DSA-GCN

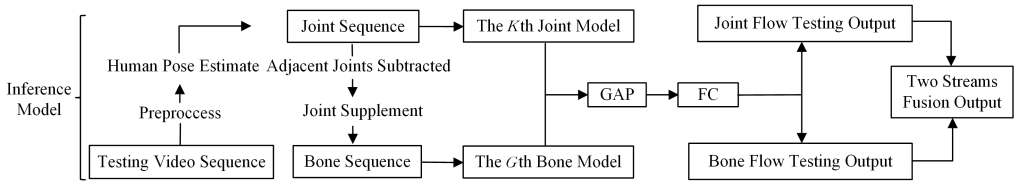

图 3 DSA-GCN 测试流程

Fig. 3 Test process of DSA-GCN

3.2.1 通道膨胀与通道注意力

首先考虑空间域部分，与深度可分离卷积（Depthwise Separable Convolution，DSC）类似，空间图的可分离策略由逐点卷积（PC）和逐通道卷积（DC）组成。在逐点卷积中，引入数倍于输入通道本身的 1×1 滤波器进行卷积运算，将低维信息映射到高维空间。因为原始输入信息的冗杂性极易对模型的稳定性和最终的分类准确性造成干扰，我们进一步采用通道注意力模块来强化与识别任务密切相关的通道，抑制与动作识别任务无关的通道，以提升本文模型的抗噪声性能和鲁棒性。

通道注意力模块的实现过程包括两个阶段：维度压缩与局部跨通道交互。在维度压缩阶段，通过全局平均池化将多维信息压缩至通道表示维度。以单人单动作识别方法为例，输入特征图 f_{in} 为 $C\times T\times V$，C 代表特征通道数，T 代表视频帧数，V 代表骨架图顶点数，通道维度压缩过程如下：

$$\boldsymbol{h}^C=\frac{1}{T\times V}\sum_{i=1}^{T}\sum_{j=1}^{V}f_{\text{in}}^C(i,j) \tag{5}$$

其中，各通道压缩后构成向量矩阵为$\{\boldsymbol{h}^1,\boldsymbol{h}^2,\cdots,\boldsymbol{h}^C\}$。

局部跨通道交互由大小为 k 的 1D 卷积完成，使用 Sigmoid 激活，如式(6)所示：

$$Z^C=\sigma(\delta(W_{1\times1}\boldsymbol{h}^C)) \tag{6}$$

其中，$\sigma(\cdot)$表示 ReLU 函数，$\delta(\cdot)$表示 Sigmoid 激活函数，$W_{1\times1}$表示快速 1D 卷积的权重。1D 卷积实现了局部跨通道交互，其中滤波器内核 k 的大小代表了局部跨通道信息交互的覆盖范围大小，即参与权重预测的相邻通道数量。Z 是包含各个通道权重的集合，最终通过式(7)对输入数据进行重新矫正：

$$\widetilde{f}_{\text{in}}^C=Z^C\otimes f_{\text{in}}^C \tag{7}$$

其中，$\otimes$表示逐通道相乘。

3.2.2 自注意力邻接矩阵

空间图卷积可以概括为通过聚合每个节点及其邻居的特征来学习节点之间的潜在特征和关系，是同时考虑当前图的拓扑结构和全部节点特征来进行相应计算的卷积方法。因此，可以把当前图的空间结构和所有节点的特征作为先验知识来学习潜在的节点特征和节点之间的关系。在 ST-GCN 中，骨架图的关节连通性与多次的图卷积运算操作积累，成为了产生过度光滑的必要条件，对模型有效判别特征的抽象聚合造成了严重的干扰。本文设计了自注意力邻接矩阵模块，用于对现有的多维参数化邻接矩阵进行有效的指导训练，使模型自动学习每个节点与其余所有节点直接的相互依赖关系，然后建立基于这些依赖关系的节点间的连接，这些连接充分反映了节点间的连接强度或连接权重。根据自注意力机制的设计，最终的更新公式如下：

$$\begin{aligned}\boldsymbol{A}^V&=Squ_{C,T}(Trans_{V,T,C}(f_{\text{in}}))^{\mathrm{T}}\otimes Squ_{C,T}\\&\quad(Trans_{V,T,C}(f_{\text{in}}))\\\boldsymbol{A}'&=\boldsymbol{A}\oplus\text{softmax}(\boldsymbol{A}^V)\\f_{\text{in}}^{l+1}&=conv_{1\times1}(\sigma(\sum_{i=1}^{3}\boldsymbol{\Lambda}_i^{-\frac{1}{2}}\boldsymbol{A}_i'\boldsymbol{\Lambda}_i^{-\frac{1}{2}}f_{\text{in}}^l\boldsymbol{W}_i^l))+conv_{1\times1}(f_{\text{in}}^l)\end{aligned} \tag{8}$$

其中，$Squ_i(\cdot)$表示将矩阵的第 i 维的向量数量压缩成 1，$Trans(\cdot)$表示转换，$\oplus$表示元素求和，$\otimes$表示矩阵乘法。

3.2.3 多维度注意力机制

在时域维度采用卷积操作是融合帧间关节信息的唯一方式。针对现有时域工作结构单一、固化的问题，本文设计了时间与通道融合的多维度注意力机制，以增强模型对数据信息的动态敏感性，提升模型的关键特征判别力，排除冗余信息的干扰，降低相似动作误分类的风险。最终的更新方程如下（$Expan$ 表示维度膨胀）：

$$\begin{aligned}T_e&=Squ_{C,V}(Trans_{T,C,V}(f_{\text{in}}))^{\mathrm{T}}\otimes Squ_{C,V}(Trans_{T,C,V}(f_{\text{in}}))\\C_e&=Squ_{T,V}(f_{\text{in}})^{T}\otimes Squ_{T,V}(f_{\text{in}})^{\mathrm{T}}\\f_{\text{out}}^l&=\text{Softmax}(f_{\text{in}}^l\otimes Expan_{V,T,C}(T_e)+f_{\text{in}}^l\otimes Expan_{V,T,C}(C_e))\end{aligned} \tag{9}$$

4 实验及结果

4.1 参数细节

本文在 NTU RGB+D[18] 和 N-UCLA[19] 两个数据集上验证本文方法的有效性。ST-GCN 作为骨干网络，将空间卷积层替换为基于自注意力的深度可分离图卷积，其中通道膨胀为原始输入的 3 倍，对于每一张图采用 6 组参数化邻接矩阵与自注意力相加的方式来获得最终网络邻接矩阵的输入。对于 NTU RGB+D，使用 SGD 算法训练模型，训练轮数为 65，初始学习率为 0.1，在第 35 和 55 轮分别降低为当前学习率的 1/10。*Batch-size* 和 *weight-decay* 分别设置为 64 和 0.0004。对于 N-UCLA，使用 SGD 算法训练模型的 65 个 epoch。初始学习率为 0.1，在第 50 轮降低为当前学习率的 1/10。*Batch-size* 和 *weight-decay* 分别设置为 64 和 0.0001。

4.2 实验分析

4.2.1 算法性能对比

在 NTU RGB+D 和 N-UCLA 数据集上，比较了本文模型与最先进的基于骨架动作识别的模型的性能，如表 1、表 2 所列。将用于比较的方法分为 3 类，自上而下，分别表示基于传统机器学习和基于 RNN 的算法、基于 CNN 的算法和基于 GCN 的算法。为了实现对比的公平性，GCN 相关算法均采用了骨架流与关节流的双流方案[5]。通过对比可知，所提出的 DSA-GCN 算法的性能最优。

表 1 NTU RGB+D 数据集与最优算法的对比

Table 1 Compared with SOTA methods on NTU RGB+D

Algorithm	C-sub/%	C-view/%	From
Lie Group	50.1	52.8	CVPR2014
ST-LSTM+TS	69.2	77.7	ECCV2016
VA-LSTM	79.2	87.7	CVPR2017
DPRL	83.5	89.8	CVPR2018
HCN	86.5	91.1	IJCAI2018
ST-GCN	82.6	89.6	AAAI2018
2s-AGCN	88.5	95.1	CVPR2019
2s Shift-GCN	89.7	96.0	CVPR2020
MS-G3D	91.5	96.2	CVPR2020
Res-GCN	90.9	96.0	ACM2020
STIGCN	90.1	96.1	ACM2020
EfficientGCN	90.9	95.5	TPAMI2021
DSA-GCN(ours)	**90.9**	**96.6**	—

表 2 N-UCLA 数据集与最优算法的对比

Table 2 Compared with SOTA methods on N-UCLA

Algorithm	Top-1/%	From
Lie Group	74.2	CVPR2014
HBRNN-L	79.2	ICCV2015
Ensemble TS-LSTM	89.2	ICCV2017
AGC-LSTM	93.3	CVPR2019
Shift-GCN-4s	94.6	CVPR2020
DC-GCN+ADG	95.3	ECCV2020
Shift-GCN++-4s	95.0	TIP2021
DSA-GCN(ours)	**95.6**	—

4.2.2 与智慧教室相关的动作识别准确率分析

表 3 列出了采用本文模型时，教室中师生常出现的 10 类动作的分类准确率，可以看到，大部分动作的分类准确率均超过 90%，其中拾取动作和举手动作因为整身的动作幅度较大，更容易被准确识别，分别达到了 98.1%和 98.4%的识别准确率；对于违规动作(如玩手机)，也达到了 95.6%的高识别率；但是对于写字这类静态动作的识别，虽然没有达到其他动作的识别精度，仍有 83.1%的识别准确率。

表 3 教室内行为的识别准确率

Table 3 Classroom action recognition accuracy

(单位：%)

Action	*Accuracy*
Drink Water	94.0
Eat Food	91.1
Pick Up	98.1
Stand Up	97.5
Sit Down	97.8
Raise hands	98.4
Play Phone	95.6
Check Time	97.8
Write	83.1
Read	94.3

4.2.3 能效比实验

为了满足智慧教育的需要，模型需要做到高精度、低计算代价。表 4 列出了本文模型相比其他算法的优势。表 4 中，*Param.*(Parameters)代表参数量，*FLOPs* 代表每秒浮点运算次数，*Ra.*(Ratio)表示倍率，其值越低效果就越好。相比本文模型，传统强势算法最多高出 2.2 倍的参数量和 2.0 倍的计算复杂度。同时，本文模型保证准确率不降反增，在 90%+量级的基础上做到最多 3%的准确率提升。

表 4 DSA-GCN 与其他方法的能效比较

Table 4 Effectiveness and efficiency comparison of DSA-GCN with others

Method	*Acc*/%	*Param.*	*Ra.*	*FLOPs*	*Ra.*
ST-GCN	92.7	2.0×10^6	1.3	2.4×10^9	1.2
2S-AGCN	93.7	3.5×10^6	2.2	4.0×10^9	2.0
DC-GCN	95.4	2.5×10^6	1.6	2.7×10^9	1.4
DSA-GCN(ours)	**95.7**	$\mathbf{1.6\times10^6}$	**1.0**	$\mathbf{2.0\times10^9}$	**1.0**

4.2.4 可视化实验

从图 4 中的邻接矩阵可视化可以得到更直观的验证，因为在 ST-GCN 中，依赖于手工设计的重力中心，卷积核的数量固定为 3，虽然 ST-GCN 试图用分层 GCN 聚合更广泛的特征，但在长时间的扩散过程中，节点特征的丰富性会被削减，而本文采用在扩增通道数量的同时扩增邻接矩阵的方式来捕获更丰富特征，同时采用了时间、通道注意力机制来强化这一过程，在后续帧中才能有效捕捉到多种有效细节。相比 ST-GCN 只能捕捉关节之间的局部物理依赖性，DSA-GCN 采用了多组邻接矩阵对关节关系进行持续学习，通过可视化图可以明显观察到不同邻接矩阵中的颜色差异。

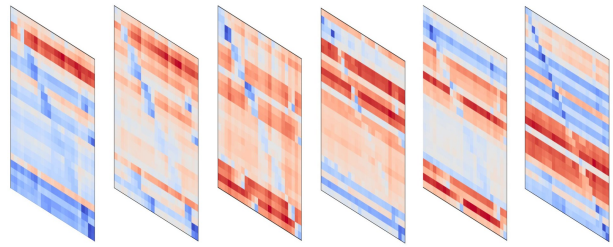

图 4 邻接矩阵可视化

Fig. 4 Visualization of adjacency matrixes

结束语 针对智慧教育中智能行为分析领域的核心难题，即对行为识别算法的准确率和计算复杂度的苛刻要求，本文提出了一种时空图卷积优化方案，通过深度可分离空域图卷积方案，以逐点卷积进行多维通道映射，增加了参数化邻接矩阵的输入数量，在极大地增强模型抽象表达能力的同时，使得模型破除了骨架连通图由连通性预定义的桎梏；通过多维度融合的注意力机制，来提升模型的时域、空域、通道域的动态敏感性，在增强关键特征判别力的同时排除冗余信息的干扰。消融性实验与可视化实验均证明了本文算法的有效性。后续工作可以在提升个别动作识别准确率和继续提出更轻量级的模型这两个方面做出更多的改进。

参 考 文 献

[1] VEMULAPALLI R, ARRATE F, CHELLAPPA R. Human action recognition by representing 3d skeletons as points in a lie group[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014: 588-595.

[2] 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. 2015: 5378-5387.

[3] DU Y, WANG W, WANG L. Hierarchical recurrent neural network for skeleton based action recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 1110-1118.

[4] YAN S, XIONG Y, LIN D. Spatial temporal graph convolutional networks for skeleton-based action recognition[C]// Thirty-second AAAI Conference on Artificial Intelligence. 2018.

[5] 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. 2019: 12026-12035.

[6] LIU Z, ZHANG H, CHEN Z, et al. Disentangling and unifying graph convolutions for skeleton-based action recognition[C]// Proceedings of the IEEE/CVF Conference on Computer Vision

and Pattern Recognition. 2020:143-152.

[7] YE F,PU S,ZHONG Q,et al. Dynamic GCN:Context-enriched topology learning for skeleton-based action recognition[C]// Proceedings of the 28th ACM International Conference on Multimedia. 2020:55-63.

[8] DEGARDIN B,LOPES V,PROENÇA H. REGINA-Reasoning Graph Convolutional Networks in Human Action Recognition [J]. arXiv:2105.06711,2021.

[9] QIN Z,LIU Y,JI P,et al. Leveraging Third-Order Features in Skeleton-Based Action Recognition[J]. arXiv:2105.01563,2021.

[10] CHEN Z,LI S,YANG B,et al. Multi-Scale Spatial Temporal Graph Convolutional Network for Skeleton-Based Action Recognition[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2021:1113-1122.

[11] CHEN Y,ZHANG Z,YUAN C,et al. Channel-wise topology refinement graph convolution for skeleton-based action recognition[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021:13359-13368.

[12] SI C,JING Y,WANG W,et al. Skeleton-based action recognition with spatial reasoning and temporal stack learning[C]// Proceedings of the European Conference on Computer Vision (ECCV). 2018:103-118.

[13] TANG Y,TIAN Y,LU J,et al. Deep progressive reinforcement learning for skeleton-based action recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018:5323-5332.

[14] ZHANG X,XU C,TIAN X,et al. Graph edge convolutional neural networks for skeleton-based action recognition[J]. IEEE Transactions on Neural Networks and Learning Systems,2019,31(8):3047-3060.

[15] KIPF T N,WELLING M. Semi-supervised classification with graph convolutional networks[J]. arXiv:1609.02907,2016.

[16] BRUNA J,ZAREMBA W,SZLAM A,et al. Spectral networks and locally connected networks on graphs[J]. arXiv:1312.6203,2013.

[17] HAMMOND D K,VANDERGHEYNST P,GRIBONVAL R. Wavelets on graphs via spectral graph theory[J]. Applied and Computational Harmonic Analysis,2011,30(2):129-150.

[18] 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. 2016:1010-1019.

[19] WANG J,NIE X,XIA Y,et al. Cross-view action modeling, learning and recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014:2649-2656.

MIAO Qi-guang, born in 1972, Ph. D, professor, Ph. D supervisor, is a senior member of China Computer Federation and AC of CCF YOCSEF. His main research interests include CV and ML.

LIU Ru-yi, born in 1989, Ph., lecturer, is a member of China Computer Federation. Her main research interests include computer vision, big data analysis and object detection in remote sensing.

(责任编辑:李亚辉)