计算机科学
COMPUTER SCIENCE

基于超图卷积和多角度拓扑细化的骨骼行为识别方法

黄倩, 苏新凯, 李畅, 巫义锐

引用本文

黄倩, 苏新凯, 李畅, 巫义锐. 基于超图卷积和多角度拓扑细化的骨骼行为识别方法[J]. 计算机科学, 2025, 52(5): 220-226.

HUANG Qian, SU Xinkai, LI Chang, WU Yirui. Hypergraph Convolutional Network with Multi-perspective Topology Refinement forSkeleton-based Action Recognition [J]. Computer Science, 2025, 52(5): 220-226.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.240600125

基于超图卷积和多角度拓扑细化的骨骼行为识别方法

黄 倩 苏新凯 李 畅 巫义锐
河海大学计算机与软件学院 南京 211106
(huangqian@hhu.edu.cn)

摘 要 由于人体骨架是一个天然存在的拓扑结构,因此图卷积网络(GCNs)被广泛地应用于基于骨骼的人体行为识别。然而,目前的基于GCN的方法只关注关节点对之间的低阶关系,而忽略了潜在的关节点在关节点群中的高阶关系。同时,现有的方法忽略了空间拓扑随时间的动态变化。这些不足影响了模型的表现。为此,利用K-NN计算出相关性高的关节点构成超边,提出了超图构建方法和超边图卷积来动态地学习关节点间的高阶关系。此外,设计了一个从时间和通道角度细化的拓扑图来学习帧级的和通道级的关节点对之间的相关性。最后,开发了一个多角度拓扑细化的超图卷积网络(HyperMTR-GCN)用于骨骼行为识别,其在NTU RGB+D和NTU RGB+D 120数据集上具有显著优势。具体地,所提方法在NTU RGB+D的X-sub基准上比2s-AGCN提高了3.7%,在NTU RGB+D 120的X-sub基准上比2s-AGCN提高了5.7%。

关键词:行为识别;图卷积网络;超图神经网络;骨架建模;拓扑细化
中图分类号 TP391.41

Hypergraph Convolutional Network with Multi-perspective Topology Refinement for Skeleton-based Action Recognition

HUANG Qian,SU Xinkai,LI Chang and WU Yirui
College of Computer Science and Software Engineering,Hohai University,Nanjing 211106,China

Abstract Since the human skeleton is a natural topological structure,graph convolutional networks(GCNs) are widely used for skeleton-based human action recognition. In recent research,skeleton sequences are represented as spatio-temporal graphs and topology graphs are used to model the correlation between human joints. However,current GCN-based methods only focus on pairwise joint relationships and ignore potential high-order relationships beyond pairwise relationships,leading to underutilization of the graph structure of skeleton data. To solve this problem,this paper proposes the concept of hypergraph to represent potential high-order relationships of joints. Since the high-order relationships of joints within each frame in the skeleton sequence may vary,the model dynamically learns the high-order correlations within each frame with the K-NN method and initialize the hypergraph structure using the high-level representation of joints. This hypergraph structure can better learn the high-order relationships between joints as the hyperedges dynamically adjust with the evolution of joint features. In current hypergraph neural networks,hypergraph convolution transforms the hypergraph into a simple graph using the Laplace's transformation and then performs graph convolution. This method does not fully utilize the characteristics of the hypergraph. The proposed hypergraph convolution method better utilizes the relationship between hyperedges and hypernodes in the hypergraph,performing hyperedge graph convolution on each hyperedge to learn the high-order relationships between joints. The second problem with current GCN-based human action recognition methods is that the topology built by GCNs to represent pairwise joint relationships is not dynamic enough,such as using the same topology for all frames in a sample. To fully explore the dynamic correlation between pairwise joints,the frame-wise topology modeling method is proposed to capture correlation between pairwise joints under different frames and channel-level topology modeling method is proposed to capture correlation between different feature types. Finally,a hypergraph convolution network with multi-perspective topology refinement(HyperMTR-GCN) is developed for skeleton-based action recognition,which has a significant advantage on the NTU RGB+D and NTU RGB+D 120 datasets. Specifically,it improves by 3.7% on the X-sub benchmark of NTU RGB+D and by 5.7% on the X-sub benchmark of NTU RGB+D 120 compared to 2s-AGCN.

Keywords Action recognition,Graph convolutional network,Hypergraph neural network,Skeleton modeling,Topology refinement

到稿日期:2024-06-21 返修日期:2024-11-08
通信作者:苏新凯(1784072485@qq.com)

1　引言

人体动作识别是计算机视觉领域的重要方向，被广泛应用于视频监控[1]、人机交互[2]等领域。近年来，基于骨骼的人体行为识别由于计算效率高且对环境变化和相机视角鲁棒而受到广泛的关注。

最近的研究中，骨骼序列以时空图的结构表示。Yan等[3]根据人体关节点之间的自然连接构建空间图，并在连续帧中对同一种关节点添加时序边。他们用拓扑图来建模人体关节之间的相关性。然而，ST-GCN[3]中的拓扑图是预定义的，难以捕获非物理连接信息，这限制了模型的表达能力。为了学习关节点间的非物理连接，Shi等[4]利用注意力机制自适应地学习拓扑图。Ye等[5]利用所有关节点的上下文特征学习任意关节点对之间的相关性。

然而，这些优化的拓扑图有两个缺点。

1)它们只能学习关节点对之间的关系，而忽略了关节群在动作中的作用，导致骨骼数据的图结构没有被充分利用。对于人类的动作而言，每种类型的关节点都有其独特的物理功能，同种类别的关节点组成关节群参与到特定动作中。比如，跳远这个动作，手肘、膝盖和脚跟之间具有相同的物理功能，它们被认为是同一类型的关节点，这些关节点组成一个群体，相互配合，协调地完成跳远这个动作。挖掘同一类型的关节点间的复杂关系和关节点群在人体动作中的作用，可以更好地判别人体行为。因此，仅仅学习表示关节点对之间关系的拓扑结构是不够的。

为了弥补这一不足，利用超图的概念来表示身体关节的高阶关系。超图是一种特殊的图结构，包含顶点和超边。与简单图中的一条边只能连接两个顶点不同，超图中的一条超边可以连接任意多个顶点。基于此，提出了一个新颖的超图卷积模块用于基于骨骼的行为识别。超图卷积模块的主要内容是超图构建和超图卷积。现有的工作[6]用基于稀疏表达的方法构建超图，然而，选取超点和计算重构系数引入了较大的计算量。对此，我们提出了基于K-NN的超图构建方法，用关节点特征间的欧氏距离作为关节点间的相关性度量，连接相关性高的关节点，并用这些关节点的特征构建超边来初始化超图结构。相比于稀疏表达的方法，基于K-NN的方法简单，计算量小。对于超图卷积，现有的方法[6]通过拉普拉斯变换将超图转化为简单图后再进行图卷积操作。然而，这种方法无法很好地动态调整关节点间的高阶关系。对此，我们提出了一种新颖的超图卷积方法，充分利用超图内超边和超点的关系，对每一条超边进行超边图卷积来学习关节点之间的高阶关系。

2)现有的基于GCN的方法中，代表关节点对之间关系的拓扑图包含的动态信息不足，比如忽略了空间拓扑随时间的动态变化，即一个样本中所有的帧共用一个拓扑。从直观上讲，不同帧内的关节点之间的相关性是不同的。为了解决这个问题，提出了多角度拓扑细化模块，从时间和通道的角度出发，分别捕获不同帧下的关节点对之间的相关性和不同特征类型下的相关性。

结合超图模块和多角度拓扑细化模块，开发了基于多角度拓扑细化的超图卷积网络(HyperMTR-GCN)用于基于骨骼的行为识别，如图1所示。图中，HyperMTR-GC基本块包含超图卷积模块(HyperGC)、多角度拓扑细化模块(MTR-GC)和ST-GCN[3]中的时间图卷积模块(TCN)；BN是批量归一化；FC是全连接层。模型将9个HyperMTR-GC基本块堆叠，用于时空建模。接着，利用一个全局平均池化和一个分类器来预测动作。9个基本块的通道数是64，64，64，128，128，128，256，256，256。HyperMTR-GC中的空间图卷积模块和时间图卷积模块(TCN)都添加一个残差连接，以稳定训练。

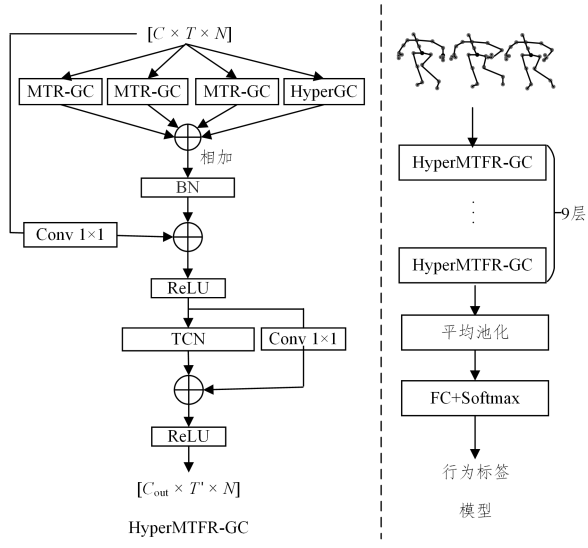


图1　HyperMTR-GCN模型

Fig. 1　HyperMTR-GCN model

本文的贡献如下：

1)提出了超图卷积模块(HyperGC)。采用基于K-NN的方法动态地捕获超边与关节点之间的连接存在性；利用关节点的高层表示来初始化超图结构；对每条超边进行超边图卷积来捕获关节点之间的高阶关系。

2)提出了多角度拓扑细化模块(MTR-GC)。从时间和通道角度学习关节点对之间的相关性，动态构建帧级的拓扑和通道级的拓扑来细化拓扑。

3)设计了一个新颖的神经网络框架。通过合理地结合HyperGC和MTR-GC来捕获低阶和高阶的关节点关系。

4)在NTU RGN+D 60和NTU RGB+D 120这两个公开数据集上的大量实验和分析表明，所提模型具有显著优势。

2　相关工作

2.1　基于骨骼的行为识别

基于骨骼的行为识别中应用最广泛的模型是RNNs，CNNs和GCNs。基于RNN的方法通常将骨骼数据建模为一个坐标向量序列，每个坐标向量代表一个人体关节[7]。基于CNN的方法根据人工设计的转换规则[8]将骨骼数据建模为伪图像。由于骨骼数据天然地以图的形式嵌入，而不是矢量序列或者二维网格，RNNs和CNNs都不能完全表示骨骼数据结构。因此，以图结构为输入的图卷积网络(GCNs)更适用于骨骼行为识别。

2.2 基于 GCN 的骨骼行为识别

基于 GCN 的骨骼行为识别方法可以分为基于频域的方法[9]和基于空间域的方法两大类。基于频域的方法是借助于图的拉普拉斯矩阵的特征值和特征向量来提取图的性质。基于空间域的方法是在表示连接关系的拓扑图上进行图卷积操作来提取特征。相比于基于频域的方法，基于空间域的方法具有更好的灵活性和泛化能力。因此在研究基于 GCN 的骨骼行为识别研究中，基于空间域的方法成为了主流。本文采用基于空间域的方法。

2.2.1 基于图的动作表示

人体骨架被表示成一个图，其中关节为节点，骨骼为边。邻接矩阵描述关节点对之间的连接强度。图卷积中邻接矩阵反映的节点拓扑结构至关重要，它决定着图卷积层中的感受野。许多动作表示的方法侧重于拓扑建模。Yan 等[3]首先构建了具有关节点间的物理连接和帧间连接的时空图来建模骨骼数据，并根据人体的物理连接构建拓扑图。由于这种预定义的拓扑是固定的，因此模型对新样本缺乏通用性，并且不能有效地捕获关节点间的非物理连接关系。Tang 等[10]定义了物理上联通和不连通的边来增强预定义的拓扑结构。Shi 等[4]提出了双流结构的 2s-AGCN，该模型利用自注意力机制来计算关节点之间的相关性，并添加一个完全根据训练数据学习的邻接矩阵来优化拓扑结构。Shi 等[11]设计了有向无环 GCN，提出了一个新颖的定向图神经网络，用于提取关节、骨骼及其关系的信息。Li 等[12]提出的 AS-GCN 试图通过邻接矩阵的高阶多项式来捕捉结构上距离较远的关节点对之间的相关性。Cheng 等[13]提出的 shift-GCN 能够自适应地学习节点之间的关系，其全局位移机制模块将每个节点的感受野覆盖到整个拓扑图。Chen 等[14]放松了其他图卷积的约束，并利用简单有效的相关性建模函数。然而，这些方法没有考虑空间拓扑随时间的动态变化，这限制了模型的性能。同时，Thakkar 等[15]重点研究了身体部位的语义，将人体骨架划分为 4 个子图，并提出了 PB-GCN 来嵌入身体部位的语义。Huang 等[16]使用图的池化和上池化操作学习了身体部位之间的高层关系，并突出了重要部位。这些方法通过嵌入身体部位的语义来捕获多个关节点之间的复杂关系，然而身体部位的划分是固定的，它们忽略了关节点的组合会随样本变化，这限制了模型的性能。

2.2.2 基于超图的动作表示

超图由顶点和超边组成。不同于简单图中两个顶点由一条边连接，超图中的每个超边可以连接任意数量的顶点。在骨骼行为识别中，多个关节点之间的关系往往要比两两关节点对之间的关系复杂得多。超图进一步考虑了数据之间的高阶相关性。Liu 等[17]通过半动态神经网络将人体表示为超图，比 GCN 捕获了更加丰富的信息。Hyper-GNN[6]同时捕获时空信息和高阶依赖，用于基于骨骼的动作识别。

现有的大部分超图模型侧重于超图的构建和超图的学习。Hyper-GNN 选取部分关节点作为超点，用稀疏表达的方法构建超图，然而选取超点和计算重构系数引入了较大的计算量。Song 等[18]采用图卷积的方法来实现超图的学习，他们针对超图定义一个拉普拉斯矩阵，将超图转为简单图表示后再进行图卷积操作。Zhou 等[19]利用注意力机制进行超图的学习，提出了 Hyperformer 用于基于骨骼的动作识别，通过自注意力机制动态调整超边内关节与关节及超边与超边之间的连接强度。然而，超边与关节点之间的连接存在性是预定义的，构建的超图无法很好地表示关节点和超边之间的动态信息。

3 方法

3.1 超图卷积模块

正如前文所说，人体动作中存在关节点之间的高阶关系。为了捕获这种复杂的高阶关系，我们提出了超图卷积模块，其主要内容是超图构建和超图卷积。

3.1.1 基于 K-NN 的超图构建

超图构建过程如图 2 所示，图中使用动作类“跳远”中的一帧来可视化超图的构建过程。图 2(b)中，灰色线表示关节点间的连接强度，灰色实线连接两个特征相似度高的关节点，灰色虚线连接两个特征相似度低的关节点。超图的构建分为 4 个步骤：1)输入原始骨架图。2)选择超点。我们采用 K-NN 方法来选择超点。具体地，给定一个关节点(即质心)，计算出与它相关性高的邻近关节点。选择的邻近关节点的个数是一个预定义的参数 K。3)构建超边。质心关节点和与它相关性高的 $K-1$ 个关节点连接，构成超边。用一个关联矩阵 $H\in\mathbb{R}^{N\times M}$ 表示超边和关节点之间的连接存在性，$H_{ve}=1$ 表示关节点 v 包含在超边 e 中。4)构建超图。根据关联矩阵，将相关性高的关节点的高层特征 $F\in\mathbb{R}^{N\times C}$ 放置到对应的超边里来初始化超图结构 $X\in\mathbb{R}^{M\times K\times C}$。为了减少存储空间，采用稀疏关联矩阵，其中的元素表示关节点索引。值得注意的是，我们的超图构建方法并不是将基于 K-NN 的方法中学习到的关节连接强度显示地表现在超图结构中，而是将相关性高的关节点的特征聚合起来初始化超边结构，然后通过后续的超边图卷积，从这些关节点的特征中学习超边内的节点之间的连接强度。随着关节点特征的演变，关节点之间的连接强度也会动态调整。相比于固定的连接强度，动态调整的超图结构可以更好地学习关节点之间的高阶关系。

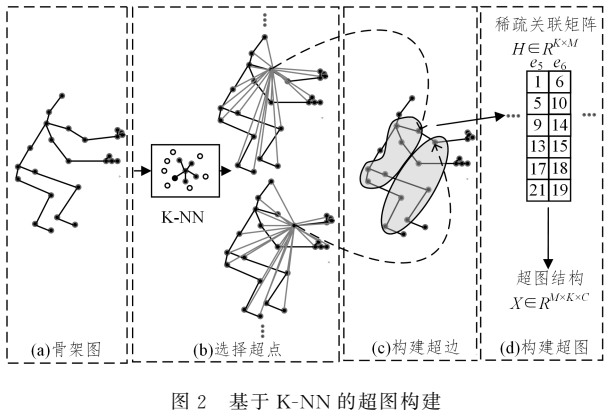


图 2 基于 K-NN 的超图构建

Fig. 2 Hypergraph construction based on K-NN

3.1.2 超图卷积

我们认为超边内的关节点组成了关节点群，捕获关节点的高阶关系即在关节点群中捕获关节点之间的相关性。因此，对超图结构中的每条超边进行超边图卷积(Hyperedge-

GC)，具体过程如图 3 左侧所示。对于每条超边，从关节点特征 $F\in\mathbb{R}^{K\times C}$ 中学习拓扑 $A\in\mathbb{R}^{K\times K}$。经过一维卷积和 softmax 操作后生成的拓扑 A 实现了顶点间和通道间的信息流。然后，将拓扑 A 和超边的特征图 F 相乘进行特征聚合。最后，将聚合后的关节点特征通过一维卷积的方法压缩到质心顶点，来实现关节点之间的高阶关系信息到质心关节点信息的转化。

3.1.3　模块结构

通过结合超图构建和超图卷积，构造了超图卷积模块。如图 3 右侧所示。首先，将输入特征通过 1×1 卷积映射到高层表示。用减少率 r 来降低计算复杂度。然后，将时间池化后的高层表示送到超图构建模块来生成超图。值得注意的是，超图内的每条超边作为一个子图独立地执行超边图卷积，这促进了高相关性的关节点之间的信息传播，同时削弱了相关性不强的关节点之间的信息交流。所有的超边都进行一次超边图卷积之后，再将这些聚合了超边信息的质心关节点特征级联并通过一维卷积进行特征升维。我们的超图卷积模块可以很好地学习到关节点之间的高阶关系，并且可以方便地嵌入到 GCNs 的空间图卷积中。

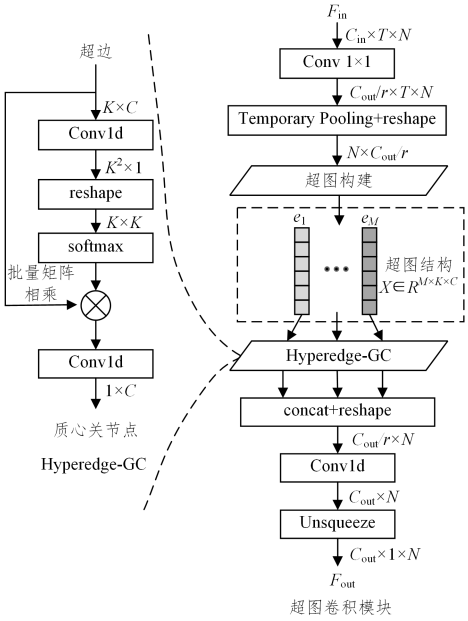


图 3　超图卷积模块结构

Fig. 3　Hypergraph convolution module structure

3.2　多角度拓扑细化模块

现有方法 2s-AGCN[4] 对特定的样本构建特定的拓扑图。然而，它捕获到的关节点间的动态信息不足；在同一样本下，模型对不同帧和不同特征类型的骨架数据共用同一个拓扑图，这限制了拓扑图表达相关性的能力。为了解决这个问题，从时间和通道角度分别捕获不同帧下的相关性和不同特征类型下的关节点相关性；然后，将得到的帧级拓扑和通道级拓扑结合来细化代表人体物理连接的拓扑图。多角度拓扑细化模块的结构如图 4 所示，其中 MTR-GC 包含了帧级拓扑构建(FTC)和通道级拓扑构建(CTC)。FTC 和 CTC 的特征映射部分使用减少率 r 来降低时间复杂度。

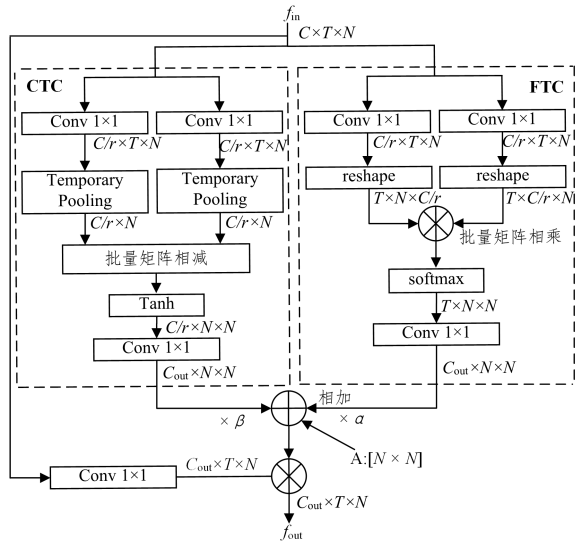


图 4　多角度拓扑细化模块(MTR-GC)

Fig. 4　Multi-perspective topology refinement module

3.2.1　帧级拓扑构建

通过动态地推断帧级的拓扑，来捕获不同帧下关节点之间的相关性。具体来说，帧级拓扑图构建包含两部分：特征映射和相关性建模。采用线性变换将输入特征转化为高层表示。受到 2s-AGCN 中通过归一化的嵌入高斯函数来计算相关性的启发，采用相同的方法计算关节点之间的相似度。与 2s-AGCN 对任意一个关节点将不同帧下的特征平铺后再计算特征相似度不同，文中所提模型对动作样本逐帧地计算关节点之间的特征相似度。具体地，给定尺寸为 $C\times T\times N$ 的输入特征，首先用两个 1×1 卷积作为高斯函数，将输入特征嵌入到 $C/r\times T\times N$；然后将两个嵌入后的特征的尺寸重塑为 $T\times C/r\times N$ 和 $T\times N\times C/r$；最后进行批量相乘并归一化，得到 $T\times N\times N$ 的帧级拓扑。相比于 2s-AGCN，本文方法的优势在于构建了帧级的拓扑，提高了模型捕获动态信息的能力，同时降低了时间复杂度，提高了模型的推理速度。

3.2.2　通道级拓扑构建

通过动态地推断通道级的拓扑来捕获在不同类型的运动特征下关节点之间的相关性。通道级拓扑图的构建由特征映射和相关性建模组成。借鉴 CTR-GCN[14] 中的方法，计算通道维度上关节点之间的距离，并利用非线性变化作为关节点之间的通道特定的拓扑关系。

3.2.3　多角度拓扑细化

邻接矩阵 $\boldsymbol{A}$ 是一个可学习的拓扑，初始化为表示人体物理连接的拓扑。用帧级的拓扑和通道级的拓扑来细化邻接矩阵 $\boldsymbol{A}$。利用 1×1 卷积对帧级拓扑图和通道级拓扑图进行线性变换来统一它们的个数，然后加权后加到邻接矩阵 $\boldsymbol{A}$ 中。具体地，给定帧级拓扑 $\boldsymbol{FT}\in\mathbb{R}^{T\times N\times N}$、通道级拓扑 $\boldsymbol{CT}\in\mathbb{R}^{C_c\times N\times N}$ 和邻接矩阵 $\boldsymbol{A}\in\mathbb{R}^{N\times N}$，细化后的拓扑 $\boldsymbol{R}\in\mathbb{R}^{C_{out}\times N\times N}$ 如下：

$$\boldsymbol{R}=\alpha\cdot conv(\boldsymbol{FT})+\beta\cdot conv(\boldsymbol{CT})+\boldsymbol{A}$$

其中，α 和 β 是可训练的标量，用于调节细化的强度。加法以广播的形式进行。细化的拓扑随着样本的不同而自适应地变化。最后，将细化后的拓扑和高层特征图相乘，进行特征聚合。

4 实验

4.1 数据集

1)NTU RGB+D 60。NTU RGB+ D 60[20]是一个包含56880个骨架动作序列的大规模人体动作识别数据集。行动样本由40名志愿者完成,分为60类。每个样本包含一个动作,并保证最多有2名被试,由3个来自不同视角的Microsoft Kinect v2相机同时拍摄。有两个流行的基准:(1)跨被试(X-sub),即训练数据来自一半的被试,测试数据来自另一半的被试;(2)跨视角(X-view),即训练集来自摄像机ID 2和3,测试集来自摄像机ID 1。

2)NTU RGB+D 120。NTU RGB+ D 120[21]是NTU RGB+ D 60的扩展,它有120个动作类,114480个样本,由106名志愿者完成,由3个摄像头拍摄。样本被收集在不同的位置和背景中,记为32个设置。该数据集的作者推荐了两个基准:(1)跨被试(X-sub),即训练数据来自53名被试,测试数据来自另外53名被试;(2)交叉设置(X-setup),即训练数据来自设置ID为偶数的样本,测试数据来自设置ID为奇数的样本。

4.2 实现细节

所有实验均在PyTorch深度学习框架下的RTX 3090 GPU上进行。本文模型通过动量为0.9的随机梯度下降(Stochastic Gradient Descent,SGD)训练65个历元。为了训练的稳定性,在前5个阶段采用了热身策略,重量衰减为0.0004。对于NTU RGB+ D和NTU RGB+ D 120,批处理大小为64,在历元35和55,学习率设置为0.01。在这些数据集上采用了文献[14]中的数据预处理方法。

4.3 消融实验分析

本节研究了超图卷积模块和多角度拓扑细化模块的影响。为了公平和简洁,所有的消融实验都是在NTU RGB+D的X-sub基准上进行的,仅使用关节点模态作为输入。

4.3.1 模块有效性

采用ST-GCN作为基线,实验结果如表1所列。与基线相比,加入HyperGC模块或者MTR-GC模块后,模型精度均得到较大提升,说明了模块的有效性。而组合了HyperGC和MTR-GC的HyperMTR-GCN的精度达到90%,说明HyperGC捕获的高阶相关性和MTR-GC捕获的关节点对之间的低阶相关性是互补的。

表1 各模块对模型精度的影响

Table 1 Effect of each module on the accuracy of model

Methods	*Accuracy*/%
ST-GCN	81.5
ST-GCN w HyperGC	89.2
ST-GCN w MTR-GC	89.8
HyperMTR-GCN	90.0

4.3.2 超图参数设置

本节讨论了基于K-NN的超图构建的参数设置,如表2所列。因为基于K-NN的超图构建对最近邻个数K和相似性度量比较敏感,所以对不同的邻居个数和相似性度量进行比较。可以观察到,欧氏距离更加适合作为相似性度量。增加邻居个数有助于加强识别性能,但是过多的邻居个数可能会带来噪声。当$K=6$时,效果最好。

表2 在超图构建的不同设置下模型的准确率

Table 2 Accuracy of model under different settings of hypergraph construction

K	Similarity metrics	*Accuracy*/%
4	Cosine	88.4
6	Cosine	89.0
8	Cosine	88.8
4	Cosine	89.6
6	Euclid	90.0
8	Euclid	89.8

4.3.3 多角度拓扑细化

本节讨论了帧级拓扑图和通道级拓扑图对精度的影响,如表3所列。FTC是帧级拓扑构建,CTC是通道级拓扑构建。保持其他模块不变化,用ST-GCN中的ST-GC替代模型中的MTR-GC作为基线。在ST-GC中加入FTC或者CTC后,精度均有所提高,表明了捕获的帧级拓扑和通道级拓扑有利于模型更好地判别人体行为。此外,MTR-GC表现出了更好的性能,说明帧级拓扑图和通道级拓扑图是互补的,MTR-GC可以更全面地学习关节之间的相关性。

表3 帧级拓扑图和通道级拓扑图对精度的影响

Table 3 Effect of frame-wise and channel-wise topology maps on the accuracy

Methods	*Accuracy*/%
ST-GC	89.5
ST-GC w FTC	89.8
ST-GC w CTC	89.7
MTR-GC	90.0

4.4 与现有方法的比较

许多先进的方法采用多流融合框架,我们的框架则采用四流融合方式[5,22-23],分别考虑了关节位置流、骨骼结构流、关节运动流和骨骼运动流。这4个流通过加权融合的方式进行融合,更好地捕获了人体动作中的时空信息。在NTU RGB+D 60和NTU RGB+D 120上,HyperMTR-GCN与现有方法的Top-1精度比较结果如表4所列。

表4 在NTU RGB+ D 60 & 120数据集上与现有方法的比较

Table 4 Comparison with state-of-the-art methods on NTU RGB+D 60&120 datasets

(%)

Methods	NTU-RGB+D		NTU-RGB+D 120		Publication
	X-Sub	X-View	X-Sub	X-Set	
ST-GCN[3]	81.5	88.3	70.7	73.2	AAAI'18
2s-AGCN[4]	88.5	95.1	82.9	84.9	CVPR'19
SGN[24]	89.0	94.5	79.2	81.5	CVPR'20
MS-G3D[25]	91.5	96.2	86.9	88.4	CVPR'20
Dynamic GCN[5]	91.5	96.0	87.3	88.6	ACM MM'20
Hyper-GNN[6]	89.5	95.7	—	—	TIP'21
RA-GCN[26]	87.3	93.6	81.1	82.7	TCSVT'21
MST-GCN[27]	91.5	96.6	87.5	88.8	AAAI'21
Graph2Net[28]	90.1	96.0	86.0	87.6	TCSVT'22
Ta-CNN+[29]	90.7	95.1	85.7	87.3	AAAI'22
SMotif-GCN+TBs[30]	90.5	96.1	87.1	87.7	TPAMI'22
MS&TA-HGCN-FC[31]	90.8	96.4	87.0	88.4	TCSVT'23
ActCLR[32]	88.2	93.6	82.1	84.6	CVPR'23
EfficientGCN[33]	92.1	96.1	88.3	89.1	TPAMI'23
SkeAttnCLR[34]	89.4	94.5	83.4	92.7	IJCAI'23
RVTCLR+[35]	87.5	93.9	82.0	83.4	ICCV'23
HyperMTR-GCN	92.2	96.6	88.6	89.9	

可以观察到,本文模型取得了比主流方法更具有竞争力

的性能。值得注意的是，我们的方法是第一个将基于 K-NN 的超图构建用于行为识别的，并且提出了一个超图卷积方法，它在行为识别中非常有效。

4.5　结果量化分析

为了更好地说明所提方法的有效性，从 NTU RGB+D 数据集里挑选了一些动作类别来进行具体分析。比如在“跳起来”这个动作中，手肘、膝盖和脚跟 6 个关节点组成一个关节点群，且在整个动作过程中的权重最大。这符合人的认知，因为手肘、膝盖和脚跟具有表示身体部位弯曲的物理功能，而且摆手、屈膝和脚跟起跳是辨别“跳起来”这个动作的重要因素。可以将“跳起来”这个动作分为准备起跳和起跳。在起跳阶段，脚跟关节点的权重在关节点群中有所提高，这验证了设计的超边结构和超边图卷积可以使得关节点间的高阶关系随着关节点的特征动态调整。

除了“跳起来”动作，还挑选了一些其他的动作。比如“打电话”和“拍球”的动作主要是上肢动作，手臂和全身其他关节相互作用；“踢足球”动作主要是下肢的动作，下肢作为一个关节点群，在整个运动过程中的权重远远大于其他关节点群的权重。这些关节点群几乎与我们的直观理解相吻合。

此外，对细化后的拓扑图进行具体的分析。在“喝水”这个动作中，同一层内的拓扑图会随着不同帧的骨骼数据而变化，“拿起水杯”时头部关节点和手部关节点的相关性低，“喝水”时两个关节点的相关性高，这符合我们的理解，从而验证了多角度拓扑细化的有效性。

还有一部分的动作类别的正确率低于 80%，如“阅读”和“写作”。我们分析，这两个动作只有手部的变化并且它们之间的区别很小，而我们的模型对多个身体部位相互作用的动作类别表现较好，对于“阅读”这种只有手部细微变化的动作表现相对较差。此外，对时间上相反的一对动作的识别率较低，比如“穿鞋”和“脱鞋”，“戴帽”和“脱帽”。我们分析，由于模型在空间图卷积层并没有考虑时序信息，因此模型对这类动作的识别性能较差。在未来，我们将尝试在帧级拓扑图中嵌入时序信息来提高模型对时间上相反的动作的识别能力。

结束语　本文提出了一个基于多角度拓扑细化的超图卷积网络（HyperMTR-GCN）用于基于骨骼的动作识别。特别地，提出了一个新颖的超图卷积模块来学习关节点间的高阶关系。所提出的多角度拓扑细化模块动态地捕获不同帧下和不同通道内的相关性。在 NTU RGB+D 和 NTU RGB+D 120 上的实验结果表明，与其他基于骨骼的方法相比，HyperMTR-GCN 取得了更好的表现。但是模型对部分动作类型的判别仍然有缺陷和不足，尤其是对局部细微动作和时间上相反的动作的判别。为了解决细微动作识别率差的问题，我们考虑在空间卷积的最后添加注意力机制，对卷积后的高层特征进一步细化，提高局部细微动作的判别能力；利用空间注意力、时间注意力和通道注意力，从多个角度自适应地调整关节之间以及帧间和通道之间的贡献度来更好地识别动作。此外，我们考虑为帧级拓扑嵌入位置编码来提高对时间上相反的动作的识别能力；从超图构建和超图学习两个方面进行优化，尝试不同的超图构建算法，找出更加适合骨骼行为识别应用的超图构建方法；优化超图卷积，为超边图卷积提供更多的理论和合理性解释。

参 考 文 献

[1] JIANG Y G，DAI Q，LIU W，et al. Human action recognition in unconstrained videos by explicit motion modeling [J]. IEEE Transactions on Image Processing，2015，24(11)：3781-3795.

[2] GAUR U，ZHU Y，SONG B，et al. A “string of feature graphs” model for recognition of complex activities in natural videos [C]// Proceedings of the 2011 International Conference on Computer Vision. Barcelona，Spain，2011：2595-2602.

[3] YAN S J，XIONG Y J，LIN D H. Spatial temporal graph convolutional networks for skeleton-based action recognition [C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2018：7444-7452.

[4] SHI L，ZHANG Y F，CHENG J，et al. Two-stream adaptive graph convolutional networks for skeleton-based action recognition [C]// Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach，CA，USA，2019：12018-12027.

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

[6] HAO X K，LI J，GUO Y C，et al. Hypergraph neural network for skeleton-based action recognition [J]. IEEE Transactions on Image Processing，2021，30：2263-2275.

[7] ZHU Y，CHEN W B，GUO G D. Fusing spatiotemporal features and joints for 3D action recognition [C]// Proceedings of the IEEE Confe-rence on Computer Vision and Pattern Recognition Workshops. Portland，OR，USA，2013：486-491.

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

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

[10] TANG Y S，TIAN Y，LU J W，et al. Deep progressive reinforcement learning for skeleton-based action recognition [C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City，UT，USA，2018：5323-5332.

[11] SHI L，ZHANG Y，CHENG J，et al. Skeleton-Based Action Recognition With Directed Graph Neural Networks [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach，CA，USA，2019：7904-7913.

[12] LI M，CHEN S H，CHEN X，et al. Actional-Structural Graph Convolutional Networks for Skeleton-Based Action Recognition [C]// Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach，CA，USA，2019：3590-3598.

[13] CHENG K，ZHANG Y，HE X，et al. Skeleton-Based Action Recognition with Shift Graph Convolutional Network [C]// Pro-

ceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Seattle, WA, USA, 2020: 180-189.

[14] CHEN Y X, ZHANG Z Q, YUAN C F, et al. Channel-wise topology refinement graph convolution for skeleton-based action recognition[C]// Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision(ICCV). Montreal, QC, Canada, 2021: 13339-13348.

[15] THAKKAR K, NARAYANAN P J. Part-based graph convolutional network for action recognition[C]// Proceedings of the Brit. Mach. Vis. Conf. (BMVC). 2018: 270-283.

[16] HUANG L, HUANG Y, OUYANG W, et al. Part-Level Graph Convolutional Network for Skeleton-Based Action Recognition [C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2020: 11045-11052.

[17] LIU S Y, LV P, ZHANG Y Z, et al. Semi-dynamic hypergraph neural network for 3d pose estimation[C]// Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence. Yokohama, Yokohama, Japan, 2021.

[18] BAI S, ZHANG F H, TORR P H S. Hypergraph convolution and hypergraph attention [J]. Pattern Recognition, 2021, 110(1): 1-8.

[19] ZHOU Y X, LI C, CHENG Z Q, et al. Hypergraph Transformer for Skeleton-based Action Recognition [EB/OL]. https://api.semanticscholar.org/CorpusID:253581243.

[20] 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 2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Las Vegas, NV, USA, 2016: 1010-1019.

[21] LIU J, SHAHROUDY A, 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] LI C, MAO Y C, HUANG Q, et al. Scale-Aware Graph Convolutional Network with Part-Level Refinement for Skeleton-Based Human Action Recognition [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(6): 4311-4324.

[23] ZHU X W, HUANG Q, LI C, et al. Skeleton-Based Action Recognition with Combined Part-Wise Topology Graph Convolutional Networks[C]// Pattern Recognition and Computer Vision (PRCV 2023). 2023: 43-59.

[24] ZHANG P F, LAN C L, ZENG W J, et al. Semantics-guided neural networks for efficient skeleton-based human action recognition[C]// Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Seattle, WA, USA, 2020: 1109-1118.

[25] LIU Z Y, ZHANG H W, CHEN Z H, et al. Disentangling and unifying graph convolutions for skeleton-based action recognition[C]// Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA, 2020: 140-149.

[26] SONG Y F, ZHANG Z, SHAN C F, et al. Richly activated graph convolutional network for robust skeleton-based action recognition [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 31(5): 1915-1925.

[27] FENG D, WU Z C, ZHANG J, et al. Multi-scale spatial temporal graph neural network for skeleton-based action recognition [J]. IEEE Access, 2021, 9: 58256-58265.

[28] WU C, WU X J, KITTLER J. Graph2net: Perceptually-enriched graph learning for skeleton-based action recognition [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(4): 2120-2132.

[29] XU K L, YE F F, ZHONG Q Y, et al. Topology-aware convolutional neural network for efficient skeleton-based action recognition [C]// AAAI Conference on Artificial Intelligence. 2021: 2866-2874.

[30] WEN Y H, GAO L, FU H B, et al. Motifgcns with local and non-local temporal blocks for skeleton-based action recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(2): 2009-2023.

[31] HUANG Z X, QIN Y S, LIN X, et al. Motiondriven spatial and temporal adaptive high-resolution graph convolutional networks for skeleton-based action recognition [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2023, 33(4): 1868-1883.

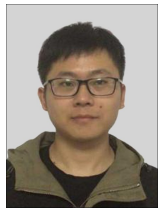
[32] LIN L, ZHANG J, LIU J. Actionlet-Dependent Contrastive Learning for Unsupervised Skeleton-Based Action Recognition [C]// Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Vancouver, BC, Canada, 2023: 2363-2372.

[33] SONG Y F, ZHANG Z, SHAN C F, et al. Constructing stronger and faster baselines for skeleton-based action recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(2): 1474-1488.

[34] HUA Y, WU W, ZHENG C, et al. Part Aware Contrastive Learning for Self-Supervised Action Recognition[C]// Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence. Macao, China, 2023: 855-863.

[35] ZHU Y S, HAN H, YU Z T, et al. Modeling the relative visual tempo for self-supervised skeleton-based action recognition [C]// 2023 IEEE/CVF International Conference on Computer Vision(ICCV). Paris, France, 2023: 13867-13876.

HUANG Qian, born in 1981, Ph.D, is a senior member of CCF (No. 08758S). His main research interests include industry-specific multi-media computing and so on.



SU Xinkai, born in 1998, postgraduate, is a member of CCF(No. T0865G). His main research interests include computer vision and so on.

(责任编辑:柯颖)