**Artificial Intelligence and Robotics Research** 人工智能与机器人研究, 2026, 15(6), 1376-1390
Published Online November 2026 in Hans. https://www.hanspub.org/journal/airr
https://doi.org/10.12677/airr.2026.156125

Hans汉斯

# 基于骨架的渐进式模式重构与阶段式协同对比人体行为识别研究

**马黎明，喻扬涛**

云南民族大学数学与计算机科学学院，云南 昆明

收稿日期：2026年9月20日；录用日期：2026年10月27日；发布日期：2026年11月5日

## 摘　要

**骨架动作识别旨在从人体关节序列中学习具有判别性的时空表示。针对现有方法拓扑先验易弱化、相似动作细粒度差异建模不足以及时间维区间协同约束缺乏等问题，本文提出一种面向骨架动作识别的渐进式模式重构与阶段式协同对比方法。针对拓扑先验弱化，引入样本自适应骨骼先验融合策略，将样本级几何先验注入拓扑学习，以缓解原始骨骼语义偏离；针对细粒度差异建模不足，设计渐进式模式重构模块，两阶段串联捕获主导模式并补充局部判别信息；针对区间协同约束缺乏，构建阶段式协同对比损失，在时间轴划分阶段区间并施加监督对比，以增强类内紧凑性与类间可分性。实验结果表明，所提方法在NTU RGB + D60、NTU RGB + D120和NW-UCLA等主流数据集上取得了具有竞争力的识别性能，消融实验进一步验证了各组成模块的有效性。结果表明，所提方法能够有效增强骨架拓扑建模和细粒度动作表示能力。**

## 关键词

**骨架动作识别，图卷积网络，自适应骨骼拓扑学习，渐进式模式重构，阶段式协同对比学习**

# A Study on Skeleton-Based Human Action Recognition via Progressive Pattern Reconstruction and Stage-Wise Collaborative Contrastive Learning

**Liming Ma, Yangtao Yu**

School of Mathematics and Computer Science, Yunnan Minzu University, Kunming Yunnan

Received: September 20, 2026; accepted: October 27, 2026; published: November 5, 2026

文章引用: 马黎明, 喻扬涛. 基于骨架的渐进式模式重构与阶段式协同对比人体行为识别研究[J]. 人工智能与机器人研究, 2026, 15(6): 1376-1390. DOI: 10.12677/airr.2026.156125

## Abstract

**Skeleton-based action recognition aims to learn discriminative spatiotemporal representations from human joint sequences. Existing methods still face three challenges: weakened topological priors, insufficient modeling of fine-grained differences between similar actions, and limited interval-level collaborative constraints along the temporal dimension. To address these issues, this paper proposes a progressive pattern reconstruction and stage-wise collaborative contrastive learning method for skeleton-based action recognition. First, a sample-adaptive skeletal prior fusion strategy injects sample-level geometric priors into topology learning to alleviate deviations from the original skeletal semantics. Second, a progressive pattern reconstruction module employs a two-stage reconstruction process to capture dominant patterns and complementary local discriminative information. Third, a stage-wise collaborative contrastive loss divides the temporal axis into multiple stage intervals and applies supervised contrastive constraints to improve intra-class compactness and inter-class separability. Experimental results on NTU RGB + D60, NTU RGB + D120, and NW-UCLA demonstrate that the proposed method achieves competitive recognition performance, and ablation studies further verify the effectiveness of each component. The results show that the proposed method can effectively enhance skeletal topology modeling and fine-grained action representation.**

## Keywords

**Skeleton-Based Action Recognition, Graph Convolutional Network, Adaptive Skeletal Topology Learning, Progressive Pattern Reconstruction, Stage-Wise Collaborative Contrastive Learning**

Copyright © 2026 by author(s) and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

## 1. 引言

人体动作识别在智能监控、人机交互、行为分析等场景中具有重要应用价值。与基于 RGB 视频的方法相比，骨架动作识别以人体关键点序列作为输入，具有表示紧凑、对光照与背景变化不敏感、易于建模人体结构关系等优点，因此近年来受到广泛关注。早期方法多依赖人工设计的时空特征，难以充分描述复杂动作中的局部变化与整体结构；随后，基于循环神经网络或卷积神经网络的方法逐渐出现，但这类方法通常将骨架序列视为向量序列或伪图像处理，未能直接利用人体骨架天然的图结构信息。随着图卷积网络的发展，基于 GCN 的方法将关节视为节点、骨骼连接视为边，使信息沿骨架图在相邻关节间传递聚合，从而在保留人体结构关系的同时联合建模空间依赖与时序动态，已成为该领域的重要研究方向[1]-[3]。

尽管现有 GCN 方法已取得显著进展，但仍存在两方面关键不足。其一，自适应拓扑学习在提升表达能力的同时，容易使模型在再学习过程中偏离原始骨骼连接提供的结构先验。原始骨骼拓扑不仅是初始化图，还体现了稳定的人体结构语义；若其在拓扑重建中被持续削弱，模型虽获得更大自由度，却可能学到不符合人体结构的关系模式。其二，现有方法在相似动作的细粒度区分上仍显不足。许多类别在整体运动趋势上高度相似，真正起区分作用的往往是局部模式差异、关节协同方式以及动作过程中的细微演化。一方面，现有方法难以在整体表示中同时兼顾主导模式与局部判别信息，容易使关键细节被整体相似性掩盖；另一方面，现有方法多依赖整段动作的全局表示进行监督或对比，容易将不同时间阶段的信息压缩为一个整体，使局部阶段内的关节协同差异被平均化。因此，相似类别在细节上的判别仍不够清晰。

针对上述问题，本文提出一种面向骨架动作识别的渐进式模式重构与阶段式协同对比方法。针对拓扑再学习易偏离原始骨骼结构先验的问题，引入样本自适应骨骼先验融合策略，在样本级骨骼几何结构信息引导下，将骨骼结构先验持续注入拓扑学习，以缓解原始骨架语义的削弱。针对相似动作细粒度区分不足的问题，一方面设计渐进式模式重构模块，通过两阶段串联重组深层表示，先捕获主导模式，再补充局部判别信息，以增强相似动作的区分能力；另一方面构建阶段式协同对比损失(SCCL)，将中高层时空特征的时间轴划分为若干阶段区间，提取区间内关节协同表示并施加监督对比约束，以提升同类动作在对应区间上的表示一致性与判别性。实验表明，所提方法在多个主流数据集上取得了具有竞争力的性能，消融实验验证了各模块的有效性。

本文的主要贡献可概括如下：

(1) 提出一种样本自适应骨骼先验融合策略，在拓扑再学习过程中持续引入样本级骨骼几何先验，以缓解数据驱动拓扑优化对原始骨骼语义的偏离。

(2) 提出一种渐进式模式重构模块，通过两阶段串联建模逐步提取主导模式并补充局部判别信息，从而增强动作表示能力。

(3) 提出一种阶段式协同对比损失(SCCL)，在中高层时空特征上划分阶段区间并对区间协同表示施加监督对比约束，以提升阶段内表示的一致性与判别性。

## 2. 相关工作

### 2.1. 基于 GCN 的骨架动作识别

近年来，基于图卷积网络(GCN)的骨架动作识别方法取得了显著进展，已成为该领域最具代表性的技术路线之一。围绕 GCN 的骨架动作识别研究，现有工作主要集中于两个方向：一类关注拓扑建模能力，通过可学习邻接矩阵、通道级拓扑细化或样本相关动态关系建模增强对复杂关节依赖的表达能力；另一类关注骨架表示本身的判别性，通过额外约束、拓扑增强或特征重组提取更具区分性的动作特征。总体来看，现有 GCN 方法已从依赖固定骨骼拓扑，逐步发展到兼顾拓扑学习与表示增强的综合框架。

尽管如此，现有方法仍存在两个突出不足。首先，引入可学习拓扑虽提升了对复杂关节关系的适应能力，但也导致骨骼原始结构先验逐渐弱化。现有主流 GCN 通常将邻接矩阵与模型参数联合优化，使基于骨骼物理连接初始化的拓扑知识在训练中逐渐衰减，模型对原始人体拓扑变得不敏感。其次，现有方法在相似动作的细粒度区分上仍显不足。许多方法更易学习粗粒度运动模式，难以稳定捕捉关键部位内部更细微、更具区分性的运动细节，从而导致分类混淆。

基于上述分析，本文关注三个相互关联的问题：如何在拓扑再学习过程中保留骨骼结构先验，如何在整体模式相近时揭示关键局部差异，以及如何在时间维表示上增强区间级协同表示的一致性与判别性。围绕这些问题，本文分别从骨骼先验保持、渐进式模式重构和阶段式协同对比三方面，对现有 GCN 骨架动作识别框架进行补充。

### 2.2. 自适应拓扑学习与骨骼先验建模

在骨架动作识别中，拓扑建模始终是 GCN 方法的核心问题之一。早期方法通常采用由骨骼物理连接定义的固定邻接矩阵，以刻画人体关节之间的自然连接关系[1]。这类方式能够提供清晰而稳定的人体结构先验，但表达能力有限，难以覆盖复杂动作中大量潜在的非局部关节依赖。随着研究深入，越来越多方法开始引入可学习拓扑或样本相关动态图结构，以增强对远程关节关系和复杂交互模式的建模能力[2]-[5]，自适应拓扑学习已成为提升骨架动作识别性能的重要方向。

然而，拓扑学习能力的增强并不意味着骨架结构信息能够被自然保留。许多现有 GCN 在训练中将邻

接矩阵与模型参数联合优化，虽提高了拓扑表达自由度，但会导致基于骨骼物理连接初始化的拓扑知识逐渐衰减，使模型对原始骨骼拓扑变得不敏感。骨骼连接提供的人体结构归纳偏置并不会在训练中被自动保持，反而可能随参数更新被更自由但不一定更合理的关系模式替代。

现有自适应拓扑方法主要通过数据驱动方式增强图结构的表达能力。例如，CTR-GCN [3]通过共享拓扑与通道特定相关性相结合，实现通道级动态拓扑细化，其重点在于增强不同通道上的关节关系建模能力；InfoGCN [4]通过上下文相关的内在拓扑与关节相对位置表示增强骨架特征；DeGCN [5]则通过学习空间图和时间图上的可变形采样位置，自适应选择更具判别性的关节及感受野。上述方法均提升了骨架拓扑或空间关系的自适应建模能力，但其核心仍侧重于由网络学习动态关系、相对位置表示或采样位置，并未显式将由当前样本骨架几何结构直接构建的先验持续注入各层拓扑学习过程中。相比之下，本文提出的 SAPF 首先根据每个输入样本的关节几何关系构建样本级相似度先验，再将该先验与不同图卷积层自身的可学习拓扑进行融合，使各层在保留独立拓扑建模能力的同时共享与当前样本相关的几何参考。因此，SAPF 的重点并非替代已有自适应拓扑，而是在其基础上增加显式、样本相关且跨层持续作用的骨骼几何约束，以缓解拓扑再学习过程中原始骨骼结构语义被逐渐弱化的问题。

尽管现有方法已增强了骨架拓扑的自适应建模能力，但对当前样本几何关系的显式利用仍较有限。不同动作、姿态和个体差异会导致关节间几何关系发生变化，因此仅依赖全局共享或网络隐式学习的拓扑，仍难以同时兼顾结构稳定性与样本适应性。基于此，本文采用样本自适应骨骼先验融合策略，将由当前样本骨架几何关系构建的先验持续注入各层可学习拓扑，以补充动态拓扑学习并保持骨骼结构约束。

### 2.3. 模式重构与对比学习

除了拓扑建模之外，如何进一步提升骨架表示的判别能力，也是骨架动作识别中的重要研究方向。已有工作开始通过记忆单元[6]、代表性模式组合或原型机制[7]增强动作表示，其中，原型类方法通常利用一组可学习的代表性模式对输入特征进行匹配或重构，以增强特征的结构化表达能力。然而，现有原型或记忆库方法大多侧重于从整体特征中检索代表性模式，通常未显式区分主导动作模式与剩余细粒度关系之间的层次作用。对于高度相似的动作类别而言，决定类别差异的往往并非整体运动模式，而是少量关节关系在局部过程中的细微变化。与上述方法不同，本文提出的 PPR 并非直接对整体动作表示进行一次性原型匹配，而是首先将高层节点特征转换为关节对关系 token，并采用两阶段渐进式重构：第一阶段从完整关系表示中提取跨样本较稳定的主导模式，第二阶段进一步针对第一阶段未充分解释的残差信息进行模式寻址与细节补偿。由此，PPR 将模式重构从单阶段整体表示扩展为“主导模式–残余细节”的递进式关系建模，使模式库不仅用于压缩或重构已有表示，还能够显式强化相似动作之间具有判别性的局部关系信息。

另一方面，对比学习近年来也被广泛用于表示学习，并逐渐应用于骨架动作识别任务[8]。现有方法通常通过构造正负样本关系，约束全局动作表示、类别中心或局部时序特征，以提升类内紧凑性和类间可分性。其中，全局对比方法更关注整段动作语义的一致性，类中心约束则强调样本表示与类别原型之间的距离关系，而部分时序对比方法开始利用局部时间片段进行表示学习[9]。然而，这类方法通常以单个样本特征或时间片段本身作为对比对象，对同一动作阶段内多个关节之间的协同结构关注不足。相比之下，本文提出的 SCCL 并非直接对原始帧级特征或整段全局表示施加对比约束，而是首先将中高层时空特征沿时间轴划分为多个阶段区间，并在每个区间内构造关节协同表示，再利用标签监督对对应阶段的协同向量进行对比学习。由此，SCCL 将对比约束从“全局动作级”或“局部时间片段级”进一步扩展到“阶段区间–关节协同结构”层面，使监督信号能够更直接地作用于动作过程中的局部协同关系，从而增强相似动作在细粒度阶段上的判别能力。

## 3. 提出的方法

本节将详细介绍本文提出的整体框架。围绕骨架动作识别中拓扑先验易衰减、细粒度模式差异难以刻画以及时间维表示缺乏区间级协同判别约束等问题，本文从拓扑建模、表示增强和训练约束三个层面进行改进。具体而言，本文主要包含三个核心组成部分：样本自适应骨骼先验融合模块、渐进式模式重构模块以及阶段式协同对比损失(SCCL)。其中，样本自适应骨骼先验融合模块用于在拓扑再学习过程中持续保留骨骼结构先验；渐进式模式重构模块用于由主导模式到局部差异逐步增强高层表示；SCCL 则在中高层特征的阶段区间上对区间协同表示施加监督对比约束，以提升阶段内表示的一致性与判别性。下面按照由整体到局部的顺序，先介绍预备知识与总体框架，再分别对各模块进行详细说明。

### 3.1. 总体框架

本文整体框架如图 1 所示，由骨架图表示学习主干和训练期辅助约束分支构成，主要包括样本自适应骨骼先验融合模块、渐进式模式重构模块和阶段式协同对比损失(SCCL)。

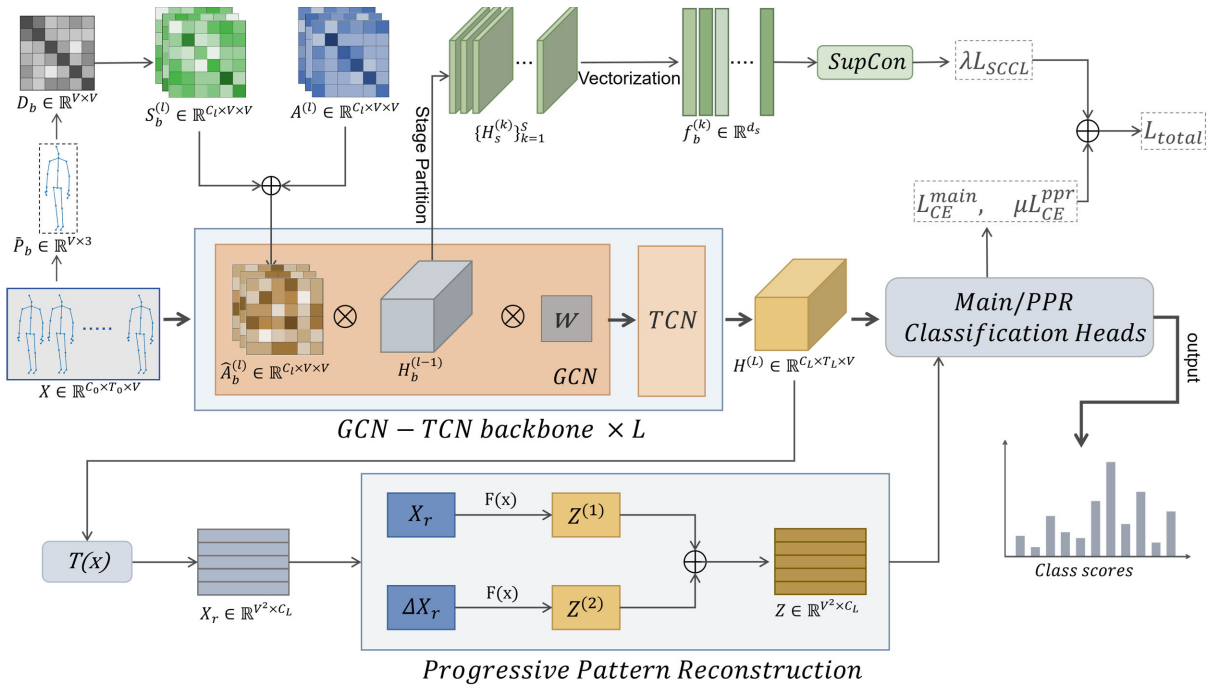



**Figure 1.** Overall framework of the proposed method
**图 1.** 所提方法的整体框架

由于人体骨架具有天然的结构属性，骨架动作序列通常可表示为图

$$\mathcal{G} = (\mathcal{V}, \mathcal{E}) \tag{1}$$

其中，$\mathcal{V}$ 表示关节集合，$\mathcal{E}$ 表示由人体骨骼物理连接所定义的边集合，$V = |\mathcal{V}|$ 表示关节数量。给定一个 batch 的输入骨架序列，记其特征张量为

$$\boldsymbol{X} \in \mathbb{R}^{B\times C_0\times T_0\times V\times M} \tag{2}$$

其中，$B$ 表示 batch size，$C_0$ 表示输入通道数，通常对应三维关节坐标，$T_0$ 表示输入时间长度，$V$ 表示关节数，$M$ 表示人数维度。对于 NTU RGB + D 系列数据集，通常有 $V = 25$，且每个样本最多包含两个人。

为简化后续公式表示，本文将人数维 $M$ 合并到 batch 维中，并使用

$$\boldsymbol{X} \in \mathbb{R}^{B' \times C_0 \times T_0 \times V}, \quad B' = B \times M \tag{3}$$

表示后续输入特征，其中该记号简化不影响模型的实际实现。在骨架动作识别中，GCN 通常通过堆叠多个图卷积层来提取骨架序列的时空表示。对第 $l$ 层图卷积而言，其基本传播过程可抽象表示为

$$\boldsymbol{H}^{(l)} = \sigma\left(\hat{\boldsymbol{A}}^{(l)} \boldsymbol{H}^{(l-1)} \boldsymbol{W}^{(l)}\right) \tag{4}$$

其中，$\boldsymbol{H}^{(l-1)}$ 和 $\boldsymbol{H}^{(l)}$ 分别表示第 $l-1$ 层和第 $l$ 层的隐藏特征，$\hat{\boldsymbol{A}}^{(l)}$ 表示第 $l$ 层实际用于前向传播的拓扑矩阵，$\boldsymbol{W}^{(l)}$ 表示可学习的特征投影参数，$\sigma(\cdot)$ 为非线性激活函数。在具体实现中，$\boldsymbol{W}^{(l)}$ 通常由 $1 \times 1$ 卷积完成通道映射，而 $\hat{\boldsymbol{A}}^{(l)}$ 负责沿关节维进行图聚合。

在骨架图表示学习主干中，本文引入样本自适应骨骼先验融合模块，以在拓扑学习中保留骨骼结构先验。其次，在主干网络的高层时空表示上，PPR 采用两阶段串联重构，先提取主导模式，再对未充分表达的部分进行细节补偿，以增强相似动作的区分能力。最后，SCCL 作为训练期辅助约束，将中高层特征划分为若干时间阶段，并对各阶段的关节协同表示施加监督对比约束，以增强阶段级表示的判别性。

## 3.2. 样本自适应骨骼先验融合模块

骨架动作识别中的图卷积通常依赖邻接结构来定义关节间的信息传播路径与强度。然而，固定的骨骼连接仅提供通用结构先验，难以反映具体样本在当前动作中的真实几何结构与关节空间耦合差异。为此，我们提出样本自适应骨骼先验融合模块(Sample-Adaptive Skeletal Prior Fusion, SAPF)，其核心目标是在每个样本上构建一张由关节几何关系诱导的相似度先验图，并在每个图卷积基础块内部与该层的可学习拓扑参数融合，得到层特定的前向邻接结构，从而增强对细粒度动作差异的建模能力。

### 3.2.1. 关节坐标定义

为构造样本级几何先验，我们从输入中提取关节坐标，并记第 $b$ 个样本中第 $v$ 个关节在第 $t$ 帧的坐标为

$$\boldsymbol{p}_{b,v,t} \in \mathbb{R}^3 \tag{5}$$

### 3.2.2. 样本级几何先验：从距离到相似度

几何先验应尽可能稳定且抗噪，因此我们首先沿时间维对坐标进行平均池化，得到样本级平均骨架构型：

$$\bar{\boldsymbol{p}}_{b,v} = \frac{1}{T_0} \sum_{t=1}^{T_0} \boldsymbol{p}_{b,v,t} \in \mathbb{R}^3 \tag{6}$$

随后对每个样本 $b$ 计算任意关节对 $(i, j)$ 的欧氏距离，形成距离矩阵：

$$D_b(i, j) = \left\| \bar{\boldsymbol{p}}_{b,i} - \bar{\boldsymbol{p}}_{b,j} \right\|_2, \quad D_b \in \mathbb{R}^{V \times V} \tag{7}$$

考虑到不同个体的骨架尺度差异可能导致数值不稳定，我们对距离矩阵进行归一化：

$$\tilde{D}_b = \frac{D_b}{\max(D_b) + \epsilon} \tag{8}$$

图卷积中的邻接权重通常遵循“越相关权重越大”的语义，而距离越小通常意味着更强的局部关联。因此我们将归一化距离转换为相似度(亲和度)先验矩阵：

$$S_b = 1 - \tilde{D}_b, \quad S_b \in \mathbb{R}^{V \times V} \tag{9}$$

由此，$S_b(i,j)$越大表示关节$i$与$j$在当前样本结构下越接近，从而提供样本自适应的几何结构先验。

#### 3.2.3. 层内融合：生成通道级前向拓扑

在本文采用的 DeGCN [5]骨干网络中，第$l$个图卷积基础块具有对应的可学习拓扑表示。记该层生成的通道级拓扑为：

$$\boldsymbol{A}_b^{(l)} \in \mathbb{R}^{C_l \times V \times V} \tag{10}$$

其中，$C_l$表示第$l$层的通道数。另一方面，本文由输入骨架计算得到的样本级几何相似度先验为

$$\boldsymbol{S}_b \in \mathbb{R}^{V \times V} \tag{11}$$

由于$\boldsymbol{S}_b$不包含通道维度，本文将其在通道维上进行广播，使其与当前基础块的通道级拓扑具有一致的形状：

$$\boldsymbol{S}_b^{(l)} = \mathrm{Broadcast}_{C_l}\left(\boldsymbol{S}_b\right), \quad \boldsymbol{S}_b^{(l)} \in \mathbb{R}^{C_l \times V \times V} \tag{12}$$

随后，在第$l$层内部将通道级拓扑与样本级几何先验进行融合：

$$\hat{\boldsymbol{A}}_b^{(l)} = \left(1-\alpha_l\right)\boldsymbol{A}_b^{(l)} + \alpha_l \boldsymbol{S}_b^{(l)}, \quad \alpha_l \in [0,1] \tag{13}$$

其中，$\hat{\boldsymbol{A}}_b^{(l)}$表示融合后的通道级前向拓扑。通过这种方式，样本级几何先验$\boldsymbol{S}_b$被注入到当前基础块的通道级拓扑中，从而在保留通道级拓扑建模能力的同时，引入稳定的样本几何结构约束。

#### 3.2.4. 用于图卷积传播

在图卷积传播时，第$l$层以融合后的拓扑$\hat{\boldsymbol{A}}_b^{(l)}$作为前向拓扑进行消息聚合。将总体框架中的基本传播公式$\boldsymbol{H}^{(l)} = \sigma\left(\hat{\boldsymbol{A}}^{(l)}\boldsymbol{H}^{(l-1)}\boldsymbol{W}^{(l)}\right)$中的$\hat{\boldsymbol{A}}^{(l)}$替换为样本自适应拓扑$\hat{\boldsymbol{A}}_b^{(l)}$，即可得到第$l$层实际用于前向传播的图卷积形式：

$$\boldsymbol{H}_b^{(l)} = \sigma\left(\hat{\boldsymbol{A}}_b^{(l)}\boldsymbol{H}_b^{(l-1)}\boldsymbol{W}^{(l)}\right) \tag{14}$$

其中，$\boldsymbol{H}_b^{(l-1)}$为第$l-1$层特征，$\boldsymbol{W}^{(l)}$与$\sigma(\cdot)$的含义与总体框架一致。对于当前基础块中的通道级拓扑，式(14)可理解为对通道级图聚合过程的抽象表示，即融合后的$\hat{\boldsymbol{A}}_b^{(l)}$与当前层 backbone 的拓扑张量格式保持一致，并参与对应的图聚合操作。

需要强调的是，$\boldsymbol{S}_b$由输入骨架实时计算，并不作为可学习参数。训练时，梯度通过上述融合过程与图卷积传播过程自然回传，用于更新各层的$\boldsymbol{A}^{(l)}$、backbone 参数以及可选的$\alpha_l$，从而在引入样本几何先验的同时保持训练语义清晰与端到端可优化性。

### 3.3. 渐进式模式重构模块

相似动作往往具有高度重叠的整体运动模式，真正决定类别差异的通常是少量关键关节在局部阶段的细微变化。这类细节线索占比小、易受噪声与类内变化影响，直接使用原始表示进行判别时容易被整体相似性掩盖。为此，我们提出渐进式模式重构模块：在“关节对(joint-pair)”粒度上对关系表示进行可学习的检索与组装式重构，以结构化方式强化关键细节并提升相似动作的可分性。

输入组织：由节点特征到关节对 token

为使后续模式重构在关节对(joint-pair)粒度上进行，受 CTR-GCN [3]的启发，本文使用其类似的特征转换方法$T(x)$，先将 backbone 输出的高层节点特征转换为关节对关系 token。给定最后一个基础块输出的高层特征

$$\boldsymbol{H}^{(L)} \in \mathbb{R}^{B' \times C_L \times T_L \times V} \tag{15}$$

其中 $C_L$ 和 $T_L$ 分别表示最后一层特征的通道数和时间长度。随后，本文沿时间维进行聚合，得到关节级高层表示：

$$\bar{\boldsymbol{H}} \in \mathbb{R}^{B' \times V \times C_L} \tag{16}$$

在此基础上，通过可学习的关系生成算子对任意关节对 $(i, j)$ 建模其关系表示，得到关节对关系张量

$$\boldsymbol{R} \in \mathbb{R}^{B' \times V \times V \times C_r} \tag{17}$$

其中 $C_r$ 表示关系 token 的通道维度。进一步地，将 $\boldsymbol{R}$ 重塑为关节对 token 序列：

$$\boldsymbol{X}_r \in \mathbb{R}^{B' \times V^2 \times C_r} \tag{18}$$

其中，每一个 token 对应一个关节对关系表示。与节点级特征相比，$\boldsymbol{X}_r$ 显式刻画任意关节对之间的关系，因此更适合作为后续渐进式模式重构的基本单元。

渐进式两阶段重构，给定关节对 token 序列

$$\boldsymbol{X}_r \in \mathbb{R}^{B' \times V^2 \times C_r} \tag{19}$$

本文采用两阶段“先主后细”的渐进式重构。模块维护两套可学习的模式库(memory bank)与寻址矩阵(query)，分别用于主导模式重构与细节补偿重构。为避免与关系张量 $\boldsymbol{R}$ 混淆，本文使用 $\boldsymbol{P}^{(1)}$ 和 $\boldsymbol{P}^{(2)}$ 表示两阶段的模式寻址权重。

第一阶段直接对完整的关节对 token 序列进行模式寻址，以捕获覆盖面更广、跨样本更稳定的主导关系模式：

$$\boldsymbol{P}^{(1)} = \mathrm{softmax}\left( \frac{\boldsymbol{X}_r \left( \boldsymbol{W}_{\mathrm{query}}^{(1)} \right)^{\top}}{\tau} \right), \quad \boldsymbol{Z}^{(1)} = \boldsymbol{P}^{(1)} \boldsymbol{W}_{\mathrm{mem}}^{(1)} \tag{20}$$

其中，$\boldsymbol{W}_{\mathrm{query}}^{(1)}$ 与 $\boldsymbol{W}_{\mathrm{mem}}^{(1)}$ 分别表示第一阶段的寻址矩阵与模式库，$\tau$ 为温度系数。

为突出第一阶段尚未充分建模的细粒度关系信息，本文在主导模式重构之后构造残差关系表示。第一阶段通过模式库对完整关节对 token 进行重构，主要提取具有较强共性和稳定性的主导关系模式；因此，将原始关系表示与第一阶段重构结果作差，可以削弱已经被主导模式充分解释的成分，使剩余表示更加集中于第一阶段未能完整覆盖的局部关系变化。具体定义为

$$\Delta \boldsymbol{X}_r = \boldsymbol{X}_r - \boldsymbol{Z}^{(1)} \tag{21}$$

其中，$\Delta \boldsymbol{X}_r$ 表示主导模式重构后剩余的关系成分。第二阶段并不直接对原始 token 再次进行独立重构，而是以该残差表示作为寻址输入，并通过第二阶段模式库重新组合得到细节补偿表示。由于最终重构结果仍由模式库中的可学习模式组合形成，而不是直接保留残差中的全部扰动信息，因此第二阶段的作用更侧重于从未被第一阶段充分解释的关系中提取具有判别性的补充模式，而不是简单累积原始噪声。

第二阶段以残差 token $\Delta \boldsymbol{X}_r$ 作为输入，进一步对细节关系模式进行寻址与重构：

$$\boldsymbol{P}^{(2)} = \mathrm{softmax}\left( \frac{\Delta \boldsymbol{X}_r \left( \boldsymbol{W}_{\mathrm{query}}^{(2)} \right)^{\top}}{\tau} \right), \quad \boldsymbol{Z}^{(2)} = \boldsymbol{P}^{(2)} \boldsymbol{W}_{\mathrm{mem}}^{(2)} \tag{22}$$

其中，$\boldsymbol{W}_{\mathrm{query}}^{(2)}$ 与 $\boldsymbol{W}_{\mathrm{mem}}^{(2)}$ 分别表示第二阶段的寻址矩阵与模式库。最终，本文将主导模式重构结果与细节补偿结果相加，得到渐进增强的关节对关系表示：

$$\boldsymbol{Z}=\boldsymbol{Z}^{(1)}+\boldsymbol{Z}^{(2)},\quad \boldsymbol{Z}\in\mathbb{R}^{B'\times V^2\times C_r} \tag{23}$$

为使重构后的关节对关系表示具有直接判别性，本文在 PPR 输出 $\boldsymbol{Z}$ 上引入辅助分类监督。具体而言，对 $\boldsymbol{Z}$ 在 token 维度进行全局池化，得到样本级重构表示

$$\boldsymbol{g}^{\mathrm{ppr}}=\mathrm{Pool}\left(\boldsymbol{Z}\right) \tag{24}$$

随后通过 PPR 分类头得到 logits：

$$\boldsymbol{o}^{\mathrm{ppr}}=f_{\mathrm{cls}}^{\mathrm{ppr}}\left(\boldsymbol{g}^{\mathrm{ppr}}\right) \tag{25}$$

PPR 辅助分类损失定义为

$$\mathcal{L}_{\mathrm{CE}}^{\mathrm{ppr}}=\mathrm{CE}\left(\boldsymbol{o}^{\mathrm{ppr}},y\right) \tag{26}$$

该损失用于约束重构后的关系表示具有类别判别性，并与主干分类损失共同优化模型。

### 3.4. 阶段式协同对比损失

骨架动作识别的中高层时空特征在时间维上具有更强的语义抽象性，但其判别性往往仍受相似动作的全局重叠所限制。我们观察到，同类样本在相同时间位置附近往往应呈现更一致的关节协同结构。为此，我们提出 SCCL：将特征的时间轴划分为若干阶段区间，对每个区间提取关节协同描述，并在 batch 内施加监督对比约束，使同类在对应阶段区间的协同模式更紧凑、异类更可分，从而增强过程级细粒度判别能力。

SCCL 作用于主干网络第 5 个 GCN-TCN 基础块输出的时空特征。该层输出通道数为 128，时间长度为 32 帧，在已经获得较强动作语义表示的同时，仍保留了相对充分的时间分辨率。因此，本文选择第 5 个基础块输出作为 SCCL 的输入，记为

$$\boldsymbol{H}_s\in\mathbb{R}^{B'\times C_s\times T_s\times V} \tag{27}$$

其中，$C_s$ 表示特征通道数，$T_s$ 表示该中高层特征的时间长度。本文将时间维划分为 $S$ 个阶段区间，并根据后续消融实验默认设置 $S=8$。第 $k$ 个阶段区间对应的时间索引集合记为 $\mathcal{T}_k$。

对于第 $b$ 个样本和第 $k$ 个阶段区间，首先在时间索引集合 $\mathcal{T}_k$ 内对特征进行平均池化，得到该区间的特征表示：

$$\bar{\boldsymbol{H}}_b^{(k)}=\frac{1}{|\mathcal{T}_k|}\sum_{t\in\mathcal{T}_k}\boldsymbol{H}_{b,:,t,:},\quad \bar{\boldsymbol{H}}_b^{(k)}\in\mathbb{R}^{C_s\times V} \tag{28}$$

随后，通过分段特征之间的内积构造区间协同矩阵：

$$\boldsymbol{M}_b^{(k)}=\left(\bar{\boldsymbol{H}}_b^{(k)}\right)^{\top}\bar{\boldsymbol{H}}_b^{(k)},\quad \boldsymbol{M}_b^{(k)}\in\mathbb{R}^{V\times V} \tag{29}$$

其中，$\boldsymbol{M}_b^{(k)}$ 描述第 $b$ 个样本在第 $k$ 个阶段区间内不同关节之间的协同关系。最后，取协同矩阵的上三角元素并展平，得到区间协同向量：

$$\boldsymbol{f}_b^{(k)}=\mathrm{vec}\left(\mathrm{upper}\left(\boldsymbol{M}_b^{(k)}\right)\right) \tag{30}$$

对于第 $k$ 个阶段区间，收集 batch 内所有样本的区间协同向量，得到

$$\boldsymbol{F}^{(k)}=\left\{\boldsymbol{f}_b^{(k)}\right\}_{b=1}^{B'} \tag{31}$$

在标签监督下，本文对每个阶段区间分别施加监督对比约束，使同类样本在区间协同空间中更加接

近，而不同类样本相互远离。具体地，SCCL 损失定义为

$$\mathcal{L}_{\mathrm{SCCL}}=\frac{1}{S}\sum_{k=1}^{S}\mathrm{SupCon}\left(\boldsymbol{F}^{(k)},y,\tau_c\right) \tag{32}$$

其中，$\mathrm{SupCon}(\,)$表示监督对比损失[10]，$y$ 表示样本标签，$\tau_c$ 表示对比学习温度系数。

主干网络输出的高层特征经过全局池化和主分类头得到主分支预测，其交叉熵损失记为

$$\mathcal{L}_{\mathrm{CE}}^{\mathrm{main}}=\mathrm{CE}\left(\boldsymbol{o}^{\mathrm{main}},y\right) \tag{33}$$

结合 PPR 辅助分类损失与 SCCL 损失，本文最终训练目标定义为

$$\mathcal{L}_{\mathrm{total}}=\mathcal{L}_{\mathrm{CE}}^{\mathrm{main}}+\mu\mathcal{L}_{\mathrm{CE}}^{\mathrm{ppr}}+\lambda\mathcal{L}_{\mathrm{SCCL}} \tag{34}$$

其中，$\mu$ 和 $\lambda$ 分别表示 PPR 辅助分类损失和 SCCL 损失的权重。推理阶段仅使用主分支分类输出，不额外引入 SCCL 分支。

## 4. 实验结果

为评估所提方法的有效性，本文在多个骨架动作识别基准数据集上开展实验，重点验证 SAPF、PPR 和 SCCL 的有效性及关键超参数的影响。第 4.1 节介绍数据集与评估协议，第 4.2 节给出实现细节，第 4.3 节与代表性方法进行比较，第 4.4 节进行消融分析。

### 4.1. 数据集

NTU RGB + D60 [11]包含 60 类动作和 56,880 个骨架序列，每个序列包含人体 25 个关节的三维坐标。本文按照官方协议在 Cross-Subject (X-Sub)和 Cross-View (X-View)两种设置下进行评估，其中 X-Sub 按受试者划分训练集和测试集，X-View 使用摄像机 2、3 的样本训练，摄像机 1 的样本测试。

NTU RGB + D120 [12]是 NTU RGB + D 的扩展版本，包含 120 类动作和 113,945 个骨架序列，每个序列同样包含 25 个关节的三维坐标。本文按照官方协议报告 Cross-Subject (X-Sub)和 Cross-Setup (X-Set)结果，其中 X-Sub 按 106 名受试者划分训练集和测试集，X-Set 按 setup ID 的奇偶性划分训练集和测试集。

NW-UCLA [13]包含 10 类动作和 1494 个骨架序列，由三个不同视角的 Kinect 摄像机采集。按照文献[13]的标准评估协议，本文使用前两个视角的样本训练，并使用剩余一个视角的样本测试。

### 4.2. 实现细节

所有实验均基于 PyTorch 框架在 NVIDIA RTX 3090ti GPU 上实现。我们采用 DeGCN [5]作为基线骨干网络，并使用 NTU 空间图进行骨骼建模。对于 NTU RGB + D 数据集，每个骨骼序列包含 25 个关节点，最多支持两个人。输入序列被调整为 64 帧。除非另有说明，关节点流和骨骼流分别训练，然后用于性能评估。

我们提出的逐层样本自适应拓扑融合模块被插入到图卷积块中。对于每个输入样本，首先根据关节点坐标计算样本级的几何先验。该先验随后传入每个图卷积块，并与块内特定层的可学习拓扑进行融合。这样一来，不同层可以保留各自的拓扑参数，同时共享相同的样本级几何参考。

训练过程中，采用带动量的随机梯度下降(SGD)进行优化，其中使用 Nesterov 动量。初始学习率设为 0.1，权重衰减设为 0.0004，总训练轮数为 80。在前 5 个 epoch 采用预热策略。学习率在第 35、60 和 75 个 epoch 时衰减。训练批次大小设为 64。采用混合精度训练以提升训练效率。

数据增强方面，训练时使用随机旋转。训练采样区间设为[0.5, 1]，测试采样区间设为[0.95]。对于

SCCL，时间分段数设为 $S=8$，对比学习温度系数设为 $\tau_c=0.1$，SCCL 损失权重设为 $\lambda=0.4$。对于 PPR 辅助分类分支，损失权重设为 $\mu=0.2$。评估时报告 Top-1 和 Top-5 准确率。

### 4.3. 与现有先进方法的比较

本文在三个数据集上与具有代表性的骨架动作识别方法进行比较，包括固定拓扑、自适应拓扑、通道级拓扑细化和可变形图卷积等方法。实验结果均按照各数据集的官方评估协议报告，并在必要时区分不同输入流和集成设置。

表 1 给出了 NTU RGB + D60 上的比较结果。本文方法在 X-Sub 和 X-View 协议下分别取得 93.4%和 97.5%的 Top-1 准确率，表明 SAPF、PPR 与 SCCL 的联合设计能够有效提升骨架动作表示能力。

**Table 1.** Comparisons with state-of-the-art methods on the NTU RGB + D60 dataset in accuracy (%)
**表 1.** NTU RGB + D60 数据集上与现有先进方法的准确率(%)比较

| Model | Venue | X-Sub | X-View |
|---|---|---|---|
| ST-GCN [1] | AAAI18 | 81.5 | 88.3 |
| AS-GCN [14] | CVPR19 | 86.8 | 94.2 |
| 2s-AGCN (2-ensemble) [2] | CVPR19 | 88.5 | 95.1 |
| Shift-GCN (4-ensemble) [15] | CVPR20 | 90.7 | 96.5 |
| MS-G3D (2-ensemble) [16] | CVPR20 | 91.5 | 96.2 |
| CTR-GCN (4-ensemble) [3] | ICCV21 | 92.4 | 96.8 |
| InfoGCN (4-ensemble) [4] | CVPR22 | 92.7 | 96.9 |
| DeGCN (3-ensemble) [5] | TIP24 | 93.1 | 97.1 |
| Ours | - | 93.4 | 97.5 |

表 2 给出了 NTU RGB + D120 上的比较结果。本文方法在 X-Sub 和 X-Set 协议下分别取得 90.5%和 91.9%的 Top-1 准确率，在更大规模的数据集上仍保持了稳定的识别性能。

**Table 2.** Comparisons with state-of-the-art methods on the NTU RGB + D120 dataset in accuracy (%)
**表 2.** NTU RGB + D120 数据集上与现有先进方法的准确率(%)比较

| Model | Venue | X-Sub | X-Set |
|---|---|---|---|
| ST-GCN [1] | AAAI18 | 71.7 | 72.2 |
| AS-GCN [14] | CVPR19 | 78.2 | 77.7 |
| 2s-AGCN (2-ensemble) [2] | CVPR19 | 82.7 | 84.5 |
| Shift-GCN (4-ensemble) [15] | CVPR20 | 85.9 | 87.6 |
| MS-G3D (2-ensemble) [16] | CVPR20 | 86.9 | 88.4 |
| CTR-GCN (4-ensemble) [3] | ICCV21 | 88.9 | 90.6 |
| InfoGCN (4-ensemble) [4] | CVPR22 | 89.4 | 90.7 |
| DeGCN (4-ensemble) [5] | TIP24 | 90.4 | 91.5 |
| Ours | - | 90.5 | 91.9 |

表 3 给出了 NW-UCLA 上的比较结果，本文方法取得 97.0%的 Top-1 准确率，表现出较好的跨视角识别能力。

**Table 3.** Top-1 accuracy comparison with state-of-the-art methods on the Northwestern-UCLA dataset
**表 3.** Northwestern-UCLA 数据集上与现有先进方法的 Top-1 准确率比较

| Method | Venue | Acc. (%) |
|---|---|---|
| Shift-GCN (4-ensemble) [15] | CVPR20 | 94.6 |
| CTR-GCN (4-ensemble) [3] | ICCV21 | 96.5 |
| InfoGCN (4-ensemble) [4] | CVPR22 | 96.6 |
| DeGCN (4-ensemble) [5] | TIP24 | 96.8 |
| Ours | - | 97.0 |

## 4.4. 消融实验

在本小节中，本文对所提出方法的不同组成部分和关键超参数进行消融分析。考虑到在完整 NTU RGB + D60 数据集上反复训练会带来较高计算开销，本文从 NTU RGB + D60 数据集中构建了一个固定的缩减子集用于消融实验。完整 NTU RGB + D60 数据集包含 56,880 条骨架序列，而本文用于消融分析的缩减子集约包含 10,000 条样本，并保持与原数据集一致的动作类别定义和评估形式。需要强调的是，该缩减子集仅用于方法内部的组件分析和参数敏感性验证，不用于与已有方法进行公开基准上的绝对性能比较。所有消融实验均采用相同的数据划分、输入模态、数据预处理方式和训练策略，以保证不同模型变体之间的公平比较。除特别说明外，所有结果均报告 Top-1 准确率。

### 4.4.1. 组件消融实验

为验证各组成模块的有效性，本文在基线模型上逐步加入 SAPF、PPR 和 SCCL，实验结果如表 4 所示。可以看到，基线模型在 NTU RGB + D60 缩减子集上取得 85.4%的 Top-1 准确率。在加入 SAPF 后，模型准确率提升至 86.6%，相比基线提高 1.2 个百分点，说明样本级几何先验的引入能够有效增强拓扑建模能力。进一步加入 PPR 后，准确率提升至 86.9%，表明渐进式模式重构能够在已有拓扑增强的基础上继续改善高层关系表示。最后，当 SAPF、PPR 和 SCCL 同时使用时，模型取得 87.4%的最高准确率，相比基线提升 2.0 个百分点。上述结果表明，三个模块并非孤立地发挥作用，而是在逐步叠加过程中形成互补，共同提升了模型对骨架动作的判别能力。

**Table 4.** Ablation study on the contribution of each component of the proposed method on the reduced NTU RGB + D60 subset
**表 4.** 所提方法各组成模块在 NTU RGB + D60 缩减子集上的消融实验

| Baseline | SAPF | PPR | SCCL | Acc. (%) |
|---|---|---|---|---|
| ✓ | – | – | – | 85.4 |
| ✓ | ✓ | – | – | 86.6 (↑1.2) |
| ✓ | ✓ | ✓ | – | 86.9 (↑1.5) |
| ✓ | ✓ | ✓ | ✓ | 87.4 (↑2.0) |

### 4.4.2. SAPF 融合系数的影响

为分析 SAPF 融合系数 $\alpha$ 的影响，实验结果如表 5 所示。当 $\alpha = 0.2$ 时取得最高准确率 87.4%，说明

适度引入样本级几何先验能够有效补充 backbone 原有拓扑。当 $\alpha$ 继续增大时性能未进一步提升，而可学习系数取得 86.7%，表明当前设置下固定且适度的融合比例更为稳定。

**Table 5.** Ablation study on the SAPF fusion coefficient $\alpha$ on the reduced NTU RGB + D60 subset
**表 5.** SAPF 融合系数 $\alpha$ 在 NTU RGB + D60 缩减子集上的消融实验

| $\alpha$ | Acc. (%) |
|---|---|
| 0.0 | 86.9 |
| 0.1 | 87.2 |
| 0.2 | 87.4 |
| 0.3 | 87.1 |
| 0.5 | 86.9 |
| 0.7 | 87.0 |
| Learnable | 86.7 |

#### 4.4.3. 模式库大小分配的影响

为验证 PPR 两阶段模式库分配策略，本文比较不同 Stage 1/Stage 2 配置，结果如表 6 所示。其中，“100/0”表示仅采用第一阶段模式库。采用“50/50”时准确率最高，为 87.2%，而单阶段“100/0”为 86.8%。结果表明，第二阶段的细节补偿能够在主导模式建模基础上进一步增强细粒度关系表示，较均衡的两阶段分配效果更好。

**Table 6.** Ablation study on the pattern bank size allocation in PPR on the reduced NTU RGB + D60 subset
**表 6.** PPR 中不同模式库大小分配在 NTU RGB + D60 缩减子集上的消融实验

| Pattern Bank Size (Stage 1/Stage 2) | Acc. (%) |
|---|---|
| 50/50 | 87.2 |
| 60/40 | 86.9 |
| 70/30 | 86.6 |
| 80/20 | 86.7 |
| 90/10 | 86.6 |
| 100/0 | 86.8 |

#### 4.4.4. 时间分段数的影响

为分析 SCCL 时间分段数 $S$ 的影响，本文比较 $S=2,4,6,8$ 的结果，如表 7 所示。当 $S=8$ 时取得最高准确率 87.3%。分段过少会使局部协同变化被较长时间区间平均，而分段设置不合适也可能导致区间表示不稳定，因此本文采用 $S=8$。

进一步分析 SCCL 损失权重 $\lambda$，结果如表 8 所示。当 $\lambda=0.4$ 时取得最高准确率 87.4%，表明适当增强区间级监督对比约束有助于提升中高层特征的判别性。由于 SCCL 仅在训练阶段参与优化，因此不会增加推理阶段的额外计算开销。

**Table 7.** Ablation study on the number of temporal segments used in SCCL on the reduced NTU RGB + D60 subset
**表 7.** SCCL 时间分段数在 NTU RGB + D60 缩减子集上的消融实验

| Number of Segments S | Acc. (%) |
|---|---|
| 2 | 86.7 |
| 4 | 86.9 |
| 6 | 86.6 |
| 8 | 87.3 |

**Table 8.** Ablation study on the SCCL loss weight $\lambda$ on the reduced NTU RGB + D60 subset
**表 8.** SCCL 损失权重 $\lambda$ 在 NTU RGB + D60 缩减子集上的消融实验

| SCCL Loss Weight $\lambda$ | Acc. (%) |
|---|---|
| 0.1 | 86.7 |
| 0.2 | 87.0 |
| 0.3 | 86.5 |
| 0.4 | 87.4 |

## 5. 结论

本文针对骨架动作识别中拓扑先验易弱化、相似动作细粒度差异建模不足以及时间维区间协同约束缺乏等问题，提出了一种结合样本自适应骨骼先验融合、渐进式模式重构和阶段式协同对比学习的骨架动作识别方法。首先，SAPF 根据输入骨架的几何关系构建样本级先验，并将其逐层融入图卷积网络的可学习拓扑，以增强拓扑建模的稳定性与样本适应能力。其次，PPR 在关节对关系粒度上采用两阶段渐进式重构，通过主导模式建模与残差信息补偿强化高层表示中的局部判别信息。最后，SCCL 将中高层时空特征划分为多个阶段区间，并通过监督对比约束增强同类样本在对应阶段上的协同一致性与不同类别之间的可分性。

实验结果表明，所提方法在 NTU RGB + D60、NTU RGB + D120 和 NW-UCLA 数据集上均取得了具有竞争力的识别性能。其中，在 NTU RGB + D60 的 X-Sub 和 X-View 协议下分别取得 93.4%和 97.5%的 Top-1 准确率，在 NTU RGB + D120 的 X-Sub 和 X-Set 协议下分别达到 90.5%和 91.9%，在 NW-UCLA 数据集上达到 97.0%。消融实验进一步验证了 SAPF、PPR 和 SCCL 各组成模块的有效性，并表明三个模块在拓扑建模、高层关系表示和时间阶段判别约束方面具有互补作用。综上，本文方法能够从骨骼结构、关系模式和阶段协同三个层面提升骨架动作表示的判别能力。

尽管本文方法取得了较好的识别性能，但仍存在一定局限性。首先，SAPF 主要依据输入骨架的关节几何关系构建样本级先验，其表示仍可能受到视角变化以及不同样本骨架坐标分布差异的影响，未来可进一步研究具有更强视角不变性的骨骼几何先验。其次，PPR 中模式库的规模及两阶段模式分配方式目前仍采用固定设置，其在不同数据集和动作类别上的最优配置可能存在差异，后续可探索自适应模式库更新与动态容量分配机制。最后，SCCL 采用固定数量的时间阶段划分，对于动作持续时间和执行速度差异较大的样本，固定分段可能难以实现更加精确的阶段对齐，未来可进一步研究可学习或动作自适应的时间划分策略。

## 参考文献

[1] Yan, S., Xiong, Y. and Lin, D. (2018) Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action

Recognition. *Proceedings of the AAAI Conference on Artificial Intelligence*, **32**, 7444-7452. https://doi.org/10.1609/aaai.v32i1.12328

[2] Shi, L., Zhang, Y., Cheng, J. and Lu, H. (2019) Two-Stream Adaptive Graph Convolutional Networks for Skeleton-Based Action Recognition. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Los Angeles, 16 June 2019, 12026-12035. https://doi.org/10.1109/cvpr.2019.01230

[3] Chen, Y., Zhang, Z., Yuan, C., Li, B., Deng, Y. and Hu, W. (2021) Channel-Wise Topology Refinement Graph Convolution for Skeleton-Based Action Recognition. 2021 *IEEE/CVF International Conference on Computer Vision* (*ICCV*), Montreal, 10-17 October 2021, 13359-13368. https://doi.org/10.1109/iccv48922.2021.01311

[4] Chi, H., Ha, M.H., Chi, S., Lee, S.W., Huang, Q. and Ramani, K. (2022) InfoGCN: Representation Learning for Human Skeleton-Based Action Recognition. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), New Orleans, 19 June 2022, 20186-20196. https://doi.org/10.1109/cvpr52688.2022.01955

[5] Myung, W., Su, N., Xue, J. and Wang, G. (2024) DeGCN: Deformable Graph Convolutional Networks for Skeleton-Based Action Recognition. *IEEE Transactions on Image Processing*, **33**, 2477-2490. https://doi.org/10.1109/tip.2024.3378886

[6] Liu, D., Xu, H., Wang, J., Lu, Y., Kong, J. and Qi, M. (2021) Adaptive Attention Memory Graph Convolutional Networks for Skeleton-Based Action Recognition. *Sensors*, **21**, Article 6761. https://doi.org/10.3390/s21206761

[7] Snell, J., Swersky, K. and Zemel, R.S. (2017) Prototypical Networks for Few-Shot Learning. *Advances in Neural Information Processing Systems*, **30**, 4077-4087.

[8] Guo, T., Liu, H., Chen, Z., Liu, M., Wang, T. and Ding, R. (2022) Contrastive Learning from Extremely Augmented Skeleton Sequences for Self-Supervised Action Recognition. *Proceedings of the AAAI Conference on Artificial Intelligence*, **36**, 762-770. https://doi.org/10.1609/aaai.v36i1.19957

[9] Gao, X., Yang, Y., Zhang, Y., Li, M., Yu, J. and Du, S. (2023) Efficient Spatio-Temporal Contrastive Learning for Skeleton-Based 3-D Action Recognition. *IEEE Transactions on Multimedia*, **25**, 405-417. https://doi.org/10.1109/tmm.2021.3127040

[10] Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., Isola, P., Maschinot, A., Liu, C. and Krishnan, D. (2020) Supervised Contrastive Learning. *Advances in Neural Information Processing Systems*, **33**, 18661-18673.

[11] Shahroudy, A., Liu, J., Ng, T. and Wang, G. (2016) NTU RGB+D: A Large Scale Dataset for 3D Human Activity Analysis. 2016 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Las Vegas, 27-30 June 2016, 1010-1019. https://doi.org/10.1109/cvpr.2016.115

[12] Liu, J., Shahroudy, A., Perez, M., Wang, G., Duan, L. and Kot, A.C. (2020) NTU RGB+D 120: A Large-Scale Benchmark for 3D Human Activity Understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **42**, 2684-2701. https://doi.org/10.1109/tpami.2019.2916873

[13] Wang, J., Nie, X., Xia, Y., Wu, Y. and Zhu, S. (2014) Cross-View Action Modeling, Learning, and Recognition. 2014 *IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, 24-27 June 2014, 2649-2656. https://doi.org/10.1109/cvpr.2014.339

[14] Li, M., Chen, S., Chen, X., Zhang, Y., Wang, Y. and Tian, Q. (2019) Actional-Structural Graph Convolutional Networks for Skeleton-Based Action Recognition. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Los Angeles, 16 June 2019, 3595-3603. https://doi.org/10.1109/cvpr.2019.00371

[15] Cheng, K., Zhang, Y., He, X., Chen, W., Cheng, J. and Lu, H. (2020) Skeleton-Based Action Recognition with Shift Graph Convolutional Network. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), 14-19 June 2020, 183-192. https://doi.org/10.1109/cvpr42600.2020.00026

[16] Liu, Z., Zhang, H., Chen, Z., Wang, Z. and Ouyang, W (2020). Disentangling and Unifying Graph Convolutions for Skeleton-Based Action Recognition. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), 14-19 June 2020, 143-152. https://doi.org/10.1109/cvpr42600.2020.00022