

中文引用格式:余洋,蒋琳,胡启军,等. 基于多元拓扑图卷积网络的建筑工人不安全行为识别[J]. 中国安全科学学报,2026,36(4): 19-27.

英文引用格式:Yu Yang, Jiang Lin, Hu Qijun, et al. Recognition of construction workers' unsafe behaviors based on a multi-component topology graph convolutional network [J]. China Safety Science Journal, 2026, 36(4): 19-27.

基于多元拓扑图卷积网络的建筑工人 不安全行为识别^{*}

余洋¹讲师, 蒋琳¹, 胡启军^{**2}教授, 何乐平³副教授, 蔡其杰³讲师, 白羽³讲师
(1 西南石油大学 计算机与软件学院, 四川 成都 610500; 2 西南交通大学 土木工程学院,
四川 成都 610031; 3 西南石油大学 土木工程与测绘学院, 四川 成都 610500)

中图分类号:X910 文献标志码:A DOI: 10.16265/j.cnki.issn1003-3033.2026.04.1106

基金项目:国家自然科学基金资助(U23A2046, 62441610);四川省“天府万人计划”天府科技菁英项目(川万人第658号);四川省科技计划项目(2025YFHZ0015)。

【摘要】 为有效识别高空作业环境下建筑工人的不安全行为,提出一种基于多元拓扑图卷积网络(MCT-GCN)的工人不安全行为识别模型。首先,设计数据预处理模块,将视频监控数据转换为三维骨架数据;其次,构建三元动态邻接图卷积,融合可学习结构先验拓扑、通道相关性自适应拓扑和相对位置编码拓扑,动态构建出适应不同动作的邻接矩阵;然后,提出多尺度可分离时间卷积,将标准卷积分解为深度时序卷积与逐点卷积,分别对施工人员的动作时序特征和空间分布特征独立建模;最后,在公共数据集和自建的工人不安全行为数据集上进行试验验证和对比分析。结果表明:该模型在识别准确率和跨场景泛化能力上均优于现有方法;在自建的数据集上,该模型达到95.8%的最高识别准确率,能够显著提升复杂动态施工环境下工人不安全行为的识别准确度。

【关键词】 多元拓扑图卷积网络(MCT-GCN); 建筑工人; 不安全行为识别; 多尺度; 骨架

Recognition of construction workers' unsafe behaviors based on a multi-component topology graph convolutional network

Yu Yang¹, Jiang Lin¹, Hu Qijun², He Leping³, Cai Qijie³, Bai Yu³

(1 School of Computer Science and Software Engineering, Southwest Petroleum University, Chengdu
Sichuan 610500, China; 2 School of Civil Engineering, Southwest Jiaotong University, Chengdu
Sichuan 610031, China; 3 School of Civil Engineering and Geomatics, Southwest Petroleum
University, Chengdu Sichuan 610500, China)

Abstract: In order to effectively identify the unsafe behaviors of construction workers in high-altitude environments, a recognition model based on MCT-GCN was proposed. Firstly, a data preprocessing module was designed to convert video surveillance data into three-dimensional skeleton data. Secondly, a tri-component dynamic adjacency graph convolution was constructed. It integrated learnable structural prior topology, channel correlation adaptive topology, and relative position encoding topology to dynamically

* 文章编号:1003-3033(2026)04-0019-09; 收稿日期:2025-11-14; 修稿日期:2026-01-22

** 通信作者:胡启军(1977—),男,湖南衡阳人,博士,教授,博士生导师,主要从事工程安全与深度学习研究。E-mail: huqijunswpu@163.com。

build adjacency matrices adapted to different actions. Furthermore, a multi-scale separable temporal convolution was proposed to decompose standard convolutions into deep temporal convolution and point-by-point convolution, thereby independently modeling the temporal characteristics and spatial distribution characteristics of construction workers' actions. Finally, experimental verification and comparative analysis were conducted on both public datasets and a self-built dataset of workers' unsafe behaviors. The results demonstrate that the proposed model outperforms existing methods in terms of recognition accuracy and cross-scene generalization. On the self-built dataset, the model achieved a peak recognition accuracy of 95.8%, significantly enhancing the detection of workers' unsafe behaviors in complex and dynamic construction environments, and making a significant contribution to the development of intelligent monitoring and control in the construction industry.

Keywords: multi-component topology graph convolutional network (MCT-GCN); construction workers; unsafe behavior recognition; multi-scale; skeleton

0 引言

在建筑领域中,工人不安全行为是导致安全事故发生的主要原因之一^[1]。据中华人民共和国住房和城乡建设部统计,2013—2020年间建筑行业共发生5 014起生产安全事故,造成5 956人死亡,超过75%的安全事故由高处坠落、物体打击、起重机械伤害等造成^[2]。美国劳工局发布的数据显示,2023年美国建筑行业有1 075名工人丧生,是所有行业中死亡人数最多的行业,其中,超过1/3的死亡是由坠落、滑倒或绊倒造成^[3]。然而,施工现场现有的监测方法主要依赖专业人员主观判断和自身经验,成本相对较高,且难以全面覆盖和实时监控施工现场,导致施工现场管理效果不佳^[4]。因此,为切实降低施工领域事故发生率,加强对工人不安全行为的监管显得尤为重要。

近年来,基于骨架数据的识别方法逐渐展现出独特优势。工人不安全行为通常由多个相关的骨架变化构成,而这些变化能够更清晰地反映动作细节。与传统图像识别方法相比,骨架数据具有低维度和抽象运动特征,为行为建模提供了新思路^[5]。刘耀等^[6]提出一种以骨骼关键点为依据的时空图卷积神经网络,结合时间和空间信息来检测工人爬梯子时的异常操作。甘文霞等^[7]优化了PoseConv3D模型,设计了一种专门用于检测施工人员临边危险操作的新型识别方法,为施工安全管理提供了有效的技术保障。张泽辉等^[8]提出一种工人高处攀爬不安全行为识别模型,该模型集成人体姿态估计算法与不安全行为识别算法,通过检测人体关键点坐标来判断工人攀爬高处时的危险行为,有效提升了安全管理水平。虽然基于骨架的动作识别方法已在工

人不安全行为检测中得到广泛应用,但在动态长序列场景下,静态邻接矩阵无法动态调整节点时序关系,且固定长度滑动窗口难以跨窗口捕获长程依赖,导致模型感受野受限、信息表达不足,难以维持稳定性能。

鉴于此,笔者拟通过提取人体关节序列构建不安全行为数据集,提出一种用于建筑工人不安全行为识别的多元拓扑图卷积网络(Multi-Component Topology Graph Convolutional Network, MCT-GCN),利用动态拓扑表征与多尺度时空特征建模机制,解决现有模型对复杂动作特征表达不足的难题,以期提升工人不安全行为识别的准确率。

1 建筑工人不安全行为识别模型

在高空作业环境下,建筑工人不安全行为识别模型结构如图1所示。在数据预处理模块,将视频监控数据提取为骨架数据,并将其划分为关节流、骨架流、关节运动流和骨架运动流4类数据,分别输入MCT-GCN识别模块识别不安全行为。在MCT-GCN识别模块,采用双分支结构,每个分支由一个初始块和 L 个时空卷积堆叠而成(取 $L=9$)。初始块负责提取输入数据的基础空间与时间特征;时空卷积由三元动态邻接图卷积(Tri-Component Dynamic Adjacency Graph Convolution, TriDA-GC)和多尺度可分离时间卷积(Multi-Scale Separable Temporal Convolution, MSS-TC)构成,负责捕获更深层次的时空依赖关系。各分支网络提取的特征向量经全局平均池化处理后输入全连接层进行初步分类预测,最终通过多流特征融合机制实现动作类别的准确识别。

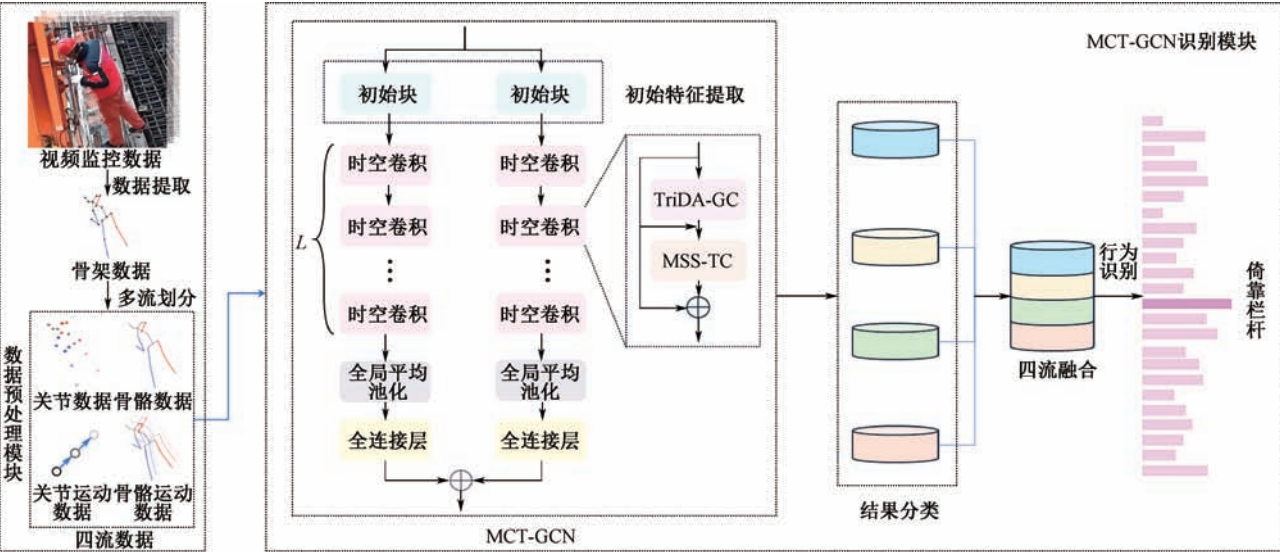


图 1 建筑工人不安全行为识别模型

Fig. 1 Construction workers' unsafe behavior recognition model

1.1 数据预处理模块

预处理原始视频监控数据,将视频帧转换为结构化的三维骨架序列数据。整体预处理流程如图 2 所示。

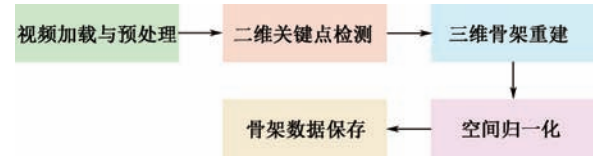


图 2 数据预处理流程

Fig. 2 Data preprocessing process

首先,对所有视频监控数据进行统一解码与格式转换,将其统一为 MP4 并调整至 25 帧/s,以保证后续分析的时间步长一致。其次,在视频流上逐帧调用姿态估计算法,提取视频监控数据中每一帧工人的关键点二维坐标及其置信度;对于置信度低于阈值的关键点,采用前后相邻帧的线性插值方法补全,以保证骨骼序列在时间维度上的连续性。随后,采用三维骨架重建算法将深度信息与二维关节点融合,生成包含 17 个关节点的三维骨架,骨架提取结果如图 3 所示。三维骨架的关节点依次为臀部、右臀部、右膝盖、右脚踝、左臀部、左膝盖、左脚踝、脊柱、胸部、颈部、头部、左肩、左手肘、左手腕、右肩、右手肘、右手腕。并以脊柱中心作为参考原点,三维平移所有关节坐标,并将依据检测到的高度和肩宽进行等比例尺度归一化,以消除不同视角和拍摄距离引入的空间变异。最后,将所得三维骨架数据保存为 npy 格式,供后续滑动窗口划分与模型训练使用。

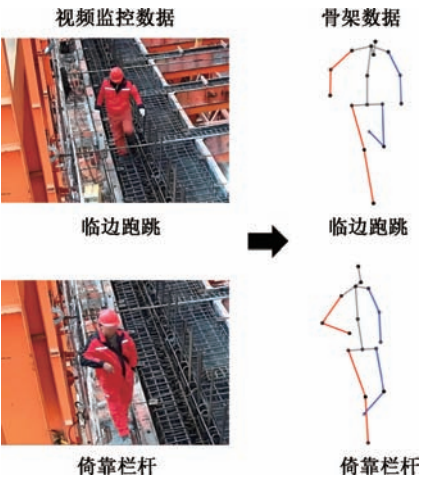


图 3 骨架提取结果

Fig. 3 Skeleton extraction results

1.2 MCT-GCN 识别模块

1.2.1 三元动态邻接图卷积

由于传统静态图卷积仅依赖固定邻接矩阵,难以自适应地捕捉时空图数据中的动态语义与结构信息,限制模型对复杂动作的识别能力。针对这一局限,提出 TriDA-GC。对于单个样本骨骼特征 $X \in \mathbb{R}^{C \times T \times V}$, 其中, C 、 T 和 V 分别为通道数、时间帧和骨架关节点数, TriDA-GC 从不同建模视角构建 3 类可学习拓扑关系,包括可学习结构先验拓扑 $P \in \mathbb{R}^{V \times V}$ 、通道相关性自适应拓扑 $Z \in \mathbb{R}^{V \times V}$ 和相对位置编码拓扑 $E \in \mathbb{R}^{V \times V}$ 。通过融合上述 3 种拓扑信息,在全局连通模式与局部结构感知层面动态生成任务感知的三元动态邻接矩阵 $M \in \mathbb{R}^{V \times V}$, 矩阵结构如图 4 所示。

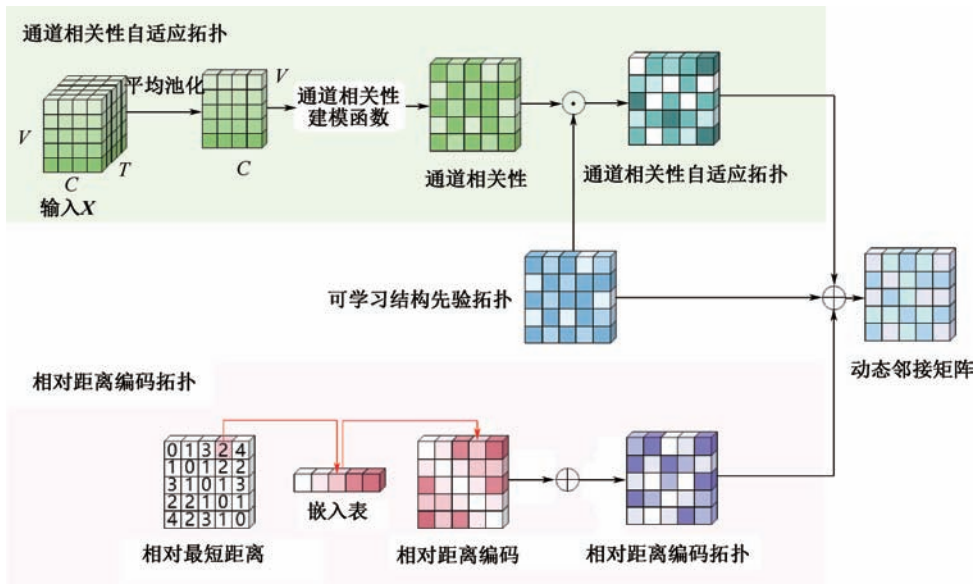


图4 三元动态邻接矩阵结构

Fig. 4 Ternary dynamic adjacency matrix structure

1) 可学习结构先验拓扑。在 TriDA-GC 中, 设计一个可学习结构先验拓扑 P 。为保证训练初期仅保留自环连接, 将其初始化为单位矩阵, 并注册为可训练参数, 在反向传播中根据下游任务损失动态更新。当 $P_{i,j}$ 增大时, 增强关节 v_i 和 v_j 之间的信息传递; 当 $P_{i,j}$ 减小时, 弱化其连接。这种方式既保留了骨架原有的局部结构, 也能捕获潜在的长程依赖和关键关节关系, 从而能够自适应地捕捉最有利的图拓扑模式。

2) 通道相关性自适应拓扑。为使邻接矩阵能够根据输入样本中的动作特征动态调整边权, 基于 Alowonou 等^[9]的工作, 引入通道相关性自适应拓扑 Z 。

首先, 将输入特征 X 沿时间维度进行平均池化, 得到表示空间结构的特征矩阵 $S \in \mathbb{R}^{V \times V}$, 计算如下:

$$S = \frac{1}{T} \sum_{t=1}^T X_{c,t,v} \quad (1)$$

式中 $X_{c,t,v}$ 为张量 X 在第 c 个通道、第 t 帧、第 v 个关节处的特征值。

对 S 中的任意一对关节点 (v_i, v_j) , 引入建模函数计算其通道维度上的皮尔逊相关系数 $B_{i,j}$, 计算公式为:

$$B_{i,j} = \frac{\sum_{c=1}^C (S_{c,i} - \mu_i)(S_{c,j} - \mu_j)}{\sqrt{\sum_{c=1}^C (S_{c,i} - \mu_i)^2} \sqrt{\sum_{c=1}^C (S_{c,j} - \mu_j)^2}} \quad (2)$$

式中: $S_{c,i}$ 和 $S_{c,j}$ 分别为顶点 v_i 与 v_j 在通道 c 上的特

征标量; μ_i 和 μ_j 为顶点 v_i 与 v_j 在所有通道上的平均特征值。

通过遍历图中所有的关节对, 构建通道相关性矩阵 $B \in \mathbb{R}^{V \times V}$, 其中, 第 i 行第 j 列的元素即为 $B_{i,j}$ 。随后, 将通道相关性矩阵 B 通过激活函数 $\sigma(\cdot)$ 进行非线性映射, 并和结构先验拓扑 P 结合, 得到通道相关性自适应拓扑 Z 的计算公式:

$$Z = \sigma(B) \odot P \quad (3)$$

式中 $\odot$ 为元素乘。

3) 相对距离编码拓扑。在动态邻接矩阵中融入节点间的图结构距离信息, 引入 BlockGCN^[10] 中的相对距离编码拓扑 E 。通过初始骨结构计算任意 2 关节点 v_i 与 v_j 之间的最短路径跳数, 构建相对距离拓扑 $H \in \mathbb{R}^{V \times V}$, 其中, $H_{i,j}$ 为关节点 v_i 与 v_j 间的最短路径跳数。对于每个可能的跳数 d , 分配一个可学习的标量编码 w_d 。通过索引映射, 将相对距离拓扑 H 转化为相对距离编码拓扑 E , 计算如下:

$$E = \phi(H) \quad (4)$$

式中 $\phi(\cdot)$ 为索引映射函数。该矩阵能够捕捉骨架结构中不同距离节点对特征聚合的影响权重。

最终的动态邻接矩阵 M 计算如下:

$$M = P + Z + E \quad (5)$$

动态邻接矩阵 M 既保留了骨架的基础自环与局部连接, 又能基于样本内部特征与节点间距离, 自适应地调整各边的权重。随后, 将 M 作为图卷积中的邻接矩阵, 对节点特征执行加权聚合, 以提取骨架动作的多尺度空间表示。

1.2.2 多尺度可分离时间卷积

提出 MSS-TC 模块以实现时空特征的解耦建模,其结构如图 5 所示。设输入的特征为 $\mathbf{X}' \in \mathbb{R}^{C' \times T \times V}$, 其中, C' 为输入通道数;该模块利用深度可分离卷积将标准卷积分解为沿时间维度的深度时序卷积和跨通道的逐点卷积,并通过四分支并行架构精确提取短期依赖。其中,深度时序卷积的卷积核长度为 k , 动态特征提取分支通过设置分组数为 $C'/4$ 以显著降低计算量。各分支特征最终在通道维度进行拼接,生成输出特征 $\mathbf{X}'' \in \mathbb{R}^{C'' \times T \times V}$, 其中, C'' 为输出通道数;在保留了帧间关联信息的同时,显著增强了模型对复杂长序列动作的表征能力。

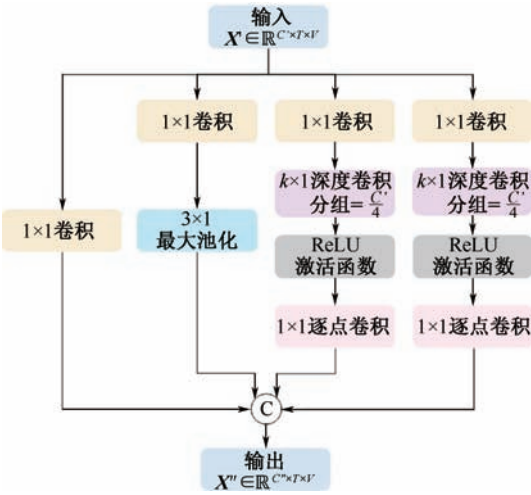


图 5 多尺度可分离时间卷积结构

Fig. 5 Multi-scale separable temporal convolution structure

1.2.2.1 深度时序卷积

对于每个连续采样位置 $\tilde{t}^{(r)}_i$, 在其邻域范围内构建时间感受野,并在每个通道上独立执行卷积操作。设卷积核长度为 k , 则对于第 r 个采样偏移下的第 i 个下采样位置,其第 c 个通道的计算公式如下:

$$Y_{i,c} = \sigma \left(\psi \left(\sum_{n=0}^{k-1} W_{n,c} \cdot \varpi(F_c, \tilde{t}^{(r)}_i)_{t-n} \right) \right) \quad (6)$$

式中: $\psi(\cdot)$ 为归一化函数; $W_{n,c}$ 为第 c 个通道在时间维第 n 个位置上的可学习卷积核权重; $\varpi(\cdot, \cdot)_{t-n}$ 为线性插值采样函数在时间步 $t-n$ 处从通道特征 F_c 中提取的值。

该操作在每个通道内独立构建长度为 k 的时序感受野,既能保证因果性,又能高效提取局部时序特征。

1.2.2.2 空间逐点卷积

在完成深度时序卷积并获得激活输出后,使用 1×1 点卷积对通道维度进行信息融合。对于输出

的第 c' 个通道,其计算过程如下式:

$$Y_s = \psi \left(\sum_{c=0}^C W_{c,c'} \cdot Y_{i,c} \right) \quad (7)$$

式中: $W_{c,c'}$ 为逐点卷积核在第 c 个输入通道与第 c' 个输入通道之间的映射权重; $Y_{i,c}$ 为深度时序卷积处理后,第 i 个位置、第 c 个通道的特征。该操作有效融合了时序卷积提取的特征信息,并输出具有指定通道数的特征表示。

通过上述 2 步,深度可分离卷积以极低的参数与计算开销分别抽取时序与空间特征,为后续堆叠的图卷积和残差连接提供了一组紧凑而高效的时空表示。

2 不安全行为识别模型评估

2.1 数据集

2.1.1 NTU RGB+D 60

南洋理工大学 RGB + D 60 (Nanyang Technological University Red - Green - Blue + Depth 60, NTU RGB+D 60) 动作识别数据集^[11]是专门用于动作识别的 3D 人体骨骼数据集,涵盖 60 多种不同的动作类型,数据集总共有 56 880 个样本。这些样本由 40 名年龄处于 10~35 岁的志愿者完成采集,采集过程采用 3 个不同角度的摄像头拍摄,分别为 -45°、0 和 45°。每一帧的 3D 关节信息通过 Kinect 深度传感器采集,每个人体骨骼由 25 个关键点组成,这些关节的位置采用相机坐标系的三维坐标的形式表示。在每个视频当中,最多会出现 2 个人。该数据集提供了 2 种较为常用的评测方式,一种是跨主体 (Cross-Subject, X-Sub) 的方法,该方法按照被试 ID 划分数据,以此保证训练集和测试集中的人员不会重复,训练集包含 40 320 个视频,测试集包含 16 560 个视频;另一种是跨视角 (Cross-View, X-View) 的方法,训练集由摄像机 2 和 3 录制的共计 37 920 个视频组成,测试集是摄像机 1 拍摄的 18 960 个视频。这样的划分方式能够让模型适应不同的人 and 不同的视角,充分体现出强大的泛化能力,所以成为骨骼动作识别里一个重要的标准数据集。

2.1.2 NTU RGB+D 120

NTU RGB+D 120 动作识别数据集^[12]是 NTU RGB+D 60 的扩展版本,其动作类别从 60 个扩展到 120 个,参与者数量由 40 人增加到 106 人,视频样本总数从 56 880 个增加到 114 480 个。该数据集是由 106 名志愿者在不同场景和背景下录制拍摄完

成,并且采用3种不同的摄像机角度进行记录。数据集设置了32个不同的场景组合,同时引入2种评测方式,交叉被试(Cross-Subject, X-Sub)是根据106名被试的编号划分,训练集有63 026个样本,测试集有50 922个样本,交叉集(Cross-Set, X-Set)是根据集合编号奇偶性划分,其中训练集有54 471个样本,测试集有59 477个样本。

2.2 试验配置与评价指标体系

2.2.1 模型参数设置

MCT-GCN 识别模型的训练参数设置见表1。该模型在PyTorch 框架下训练,批量大小为64,共设置80个训练轮次。由于原始数据样本的帧数各不相同,为确保训练时输入数据的一致性,将所有样本统一归一化为64帧。为稳定早期优化过程,初始学习率设为0.1,并在训练前5轮中采用学习率预热策略以加速收敛。在模型训练过程中,在第5~59轮之间,采用余弦退火策略来动态调整学习率。为进一步提升训练稳定性,学习率在第60轮时衰减至0.001;并在第75轮再次衰减至0.000 1,然后在剩余的训练周期中保持不变。在模型优化方面,选择随机梯度下降作为优化器,设置动量因子为0.9,以加快模型收敛速度。同时引入L2正则化,将其衰减系数设置为0.000 4,以降低模型过拟合的风险。

表1 模型参数设置

Table 1 Model parameter settings

网络模型参数	参数配置
深度学习框架	PyTorch
训练轮次	80
批次大小	64
学习率	0.1
学习率衰减策略	余弦退火
优化器	随机梯度下降
动量因子	0.9
权重衰减	0.000 4

2.2.2 评价指标

首位准确率 A_1 是衡量施工现场工人不安全行

为识别性能的核心指标。作为人体骨架动作识别研究中的基础评判依据, A_1 相比前五准确率 A_5 、平均精度均值(Mean Average Precision, mAP)等指标, A_1 能够更直观地反映模型对最可能类别的判别能力,其计算见下式:

$$A_1 = \frac{T}{T + F} \times 100\% \tag{8}$$

式中: T 为被正确预测的样本数量; F 为预测出现错误的样本数量。 T 和 F 之和为测试集中的全部样本总数。

在具体实现中,每条测试视频片段经模型推理

后输出一个分数最高的类别标 $\hat{y}_i$,若 $\hat{y}_i$ 等于其真实标签 y_i ,则将该样本计入 T ;否则计入 F 。最终, A_1 用来评估模型对单次预测中概率最高的类别进行正确识别的能力。所有试验结果均以上述 A_1 进行横向对比与性能评估。

2.3 对比试验

为验证评估该模型的动作识别性能,选取现有文献中的多种主流方法进行对比,包括时空GCN(Spatial Temporal GCN, ST-GCN)^[13]、动作结构GCN(Actional Structural GCN, AS-GCN)^[14]、反馈GCN(Feedback GCN, FGCN)^[15]和通道移位GCN(Shift GCN, ShiftGCN)^[16]。MCT-GCN 采用多流融合策略,综合利用关节流、骨骼流、关节运动流和骨骼运动流,其中,关节流将原始骨骼坐标当作输入,骨骼流反映骨骼各点间的空间差异,使用差分空间坐标作为输入,关节运动流与骨骼运动流分别捕捉关节与骨骼在时间维度上的动态变化。在NTU-
RGB+D 60数据集的X-Sub划分上,各流在训练集和测试集的准确率曲线如图6所示。试验结果表明:模型约在60个epoch后趋于收敛,表现出良好的训练稳定性。

融合上述4个特征流得到动作识别结果,见

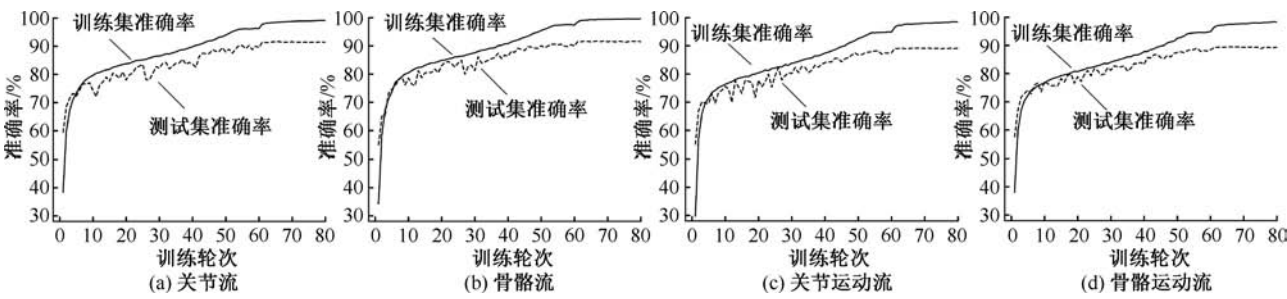


图6 各流在训练集和测试集准确率变化曲线

Fig. 6 Accuracy curves of each stream on the training and test sets

表 2。由表 2 可知:MCT-GCN 在整体性能显著优于现有最先进方法。在 NTU-RGB+D 60 数据集中,MCT-GCN 在 X-Sub 和 X-View 基准测试的准确率分别达到 93.5%和 97.3%,相比于首个基于 GCN 的模型,ST-GCN^[13] 提升幅度为 12.0%和 9.0%,较 ShiftGCN^[16] 分别提升 2.8%和 0.8%。在规模更大

的 NTU-RGB+D 120 数据集上,MCT-GCN 展现出优越的泛化能力,在 X-Sub 和 X-Set 基准上分别取得 90.5%和 91.8%的准确率。试验结果表明:文中方法整体性能比当前主流方法更优,该方法在大规模复杂数据场景中具备一定的稳定性与适应性。

表 2 NTU-RGB + D 数据集与其他方法准确率比较

Table 2 Comparison of accuracy with other existing methods on NTU-RGB+D dataset

方法	NTU-RGB+D 60		NTU-RGB+D 120	
	X-Sub	X-View	X-Sub	X-Set
ST-GCN ^[13]	81.5	88.3	70.7 [*]	73.2 [*]
AS-GCN ^[14]	86.8	94.2	77.9 [*]	78.5 [*]
FGCN ^[15]	90.2	96.3	85.4	87.4
ShiftGCN ^[16]	90.7	96.5	85.9	87.6
MCT-GCN	93.5	97.3	90.5	91.8

注:^{*}表示使用官方发布的代码实现的结果。

3 不安全行为识别模型应用

3.1 数据集构建

数据来源于国家“172”重点工程亭子口灌区一期总干渠施工现场的视频监控数据,工程范围跨越横跨广元市苍溪县、南充市阆中市及仪陇县。核心结构观音寺渡槽与乱石窖渡槽采用造槽机成槽浇筑,设备跨度 42 m,单次最大浇筑质量 1 500 t,属于大型移动式混凝土浇筑模架,渡槽最大架空高度约 65 m,造槽机操作平台位于 60 m 以上高空,属于特级高处作业环境^[17]。

结合事故发生概率、现场潜在风险评估和相关文献中对高风险行为的研究^[7,18-19],选取 4 典型危险动作构建工人不安全行为数据集,其共包含 222 组不安全行为样本,涵盖了倚靠护栏(102 组)、临边跑跳(50 组)、打电话(47 组)及吸烟(23 组)等类别。

3.2 验证结果与分析

对高风险的倚靠栏杆、临边跑跳、打电话和吸烟 4 类典型动作识别结果可视化,以直观展现 MCT-GCN 对复杂动态施工环境中动态骨架特征敏感捕捉能力。图 7 为从连续视频监控序列中选取一帧可视化结果。其中,图 7d 中同时出现了倚靠栏杆和吸烟 2 个不安全动作。由于数据标注阶段对每段骨架序列仅赋予突出的单一类别标签,模型在推理过程中会根据整段序列中骨架关键点的动态变化趋势,通过 MCT-GCN 对空间与时序特征进行联合建模。随后,分类器输出各类别对应的预测概率,并选取概率最高的类别作为最终识别结果。这种基于概率的决策机制既保证了预测结果与标注方式的一致性,也在一定程度上提升了模型在复杂场景下的分类稳定性和可解释性。



图 7 建筑工人不安全行为识别可视化结果

Fig. 7 Visualization results of construction workers' unsafe behavior identification

为评估 MCT-GCN 相较于主流识别模型在施工现场不安全行为识别任务中的性能差异,将其与 ST-GCN^[13]、通道级拓扑细化 GCN (Channel-wise

Topology Refinement GCN, CTR-GCN)^[20] 和 ShiftGCN^[16] 等先进方法对比,结果见表 3。试验显示,MCT-GCN 的识别准确率达到 95.8%,相较于最

早提出的 ST-GCN^[13] 提升 11.2%, 和最新的 ShiftGCN^[16] 相比提高 2.8%, 此结果证明 MCT-GCN 在动态骨架序列建模方面具备卓越性能, 有助于提升建筑安全智能管理的可靠性。

表 3 工人不安全行为数据集识别准确率

Table 3 Recognition accuracy comparison on the worker unsafe behavior dataset

行为识别模型	准确率/%
ST-GCN ^[13]	84.6
CTR-GCN ^[20]	89.2
ShiftGCN ^[16]	93.0
MCT-GCN	95.8

在自建工人不安全行为数据集上绘制 4 种模型的混淆矩阵(图 8), 以此直观反映它们在不同行为类别方面的识别能力差异。结果显示, MCT-GCN 在

所有行为类别上均保持了较高的识别能力。尤其对打电话和吸烟这 2 类相似度较高的动作, 依然保持良好的识别能力。模型能够有效捕捉手腕与头部之间的局部几何关系及其随时间变化的动态特征。在打电话动作中, 手腕相对头部的位置在时间维度上更加稳定, 通常持续位于头部侧方; 而吸烟动作则表现为手腕在头部前方区域的上下微幅摆动, 呈现一定的短周期节律。MCT-GCN 通过三元动态邻接图卷积, 在手腕、肘、肩、头部、颈部和胸部等关键关节之间的空间关系进行自适应建模, 并结合多尺度时序卷积感知不同时间尺度下的时序特征变化, 从而有效区分 2 类在静态姿态上相似、但动态模式存在差异的动作类型。在这 2 类动作中模型的单类准确率均达到 95.7%, 进一步验证了其对于细粒度行为差异的表征能力。

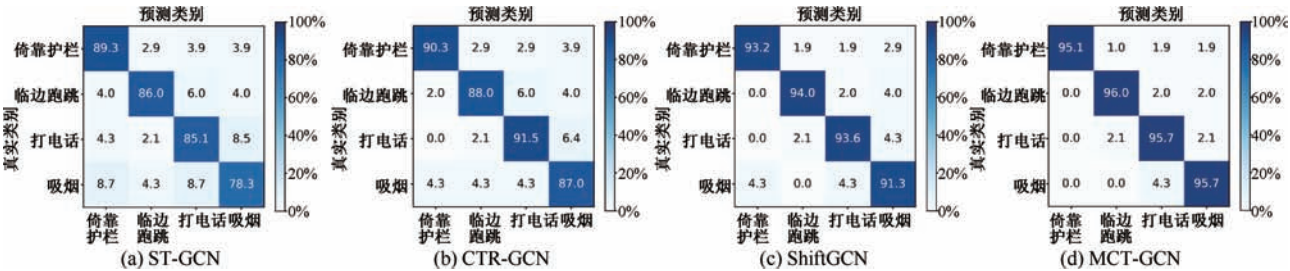


图 8 建筑工人不安全行为识别结果
Fig. 8 Identification results of construction workers' unsafe behaviors

4 结 论

- 1) MCT-GCN 模型通过三元动态邻接图卷积和多尺度可分离时间卷积, 能够增强对复杂环境变化的鲁棒性, 有效提升工人不安全行为识别的准确率。试验结果表明: 该模型在实际应用中表现出优异的识别精度。
- 2) 工人不安全行为数据集涵盖倚靠护栏、临边跑跳、打电话与吸烟 4 类典型行为, 共计 222 个样

本。该数据集通过对视频帧进行骨架提取, 形成三维骨架数据序列, 可为复杂动态环境下不安全行为的特征建模提供可靠的数据基础。

- 3) 基于 MCT-GCN 的识别方法在单视角自建数据集上取得出色的识别准确率, 未来可进一步丰富危险动作类别并持续优化模型, 以提升模型识别准确率与泛化能力, 满足智能化生产管理动态需求, 推动该技术在实际场景中的落地应用。

参 考 文 献

[1] 叶贵, 李学征, 杨丽萍, 等. 建筑工人不安全行为量化分类研究[J]. 安全与环境学报, 2021, 21(6): 2 617-2 627.
Ye Gui, Li Xuezheng, Yang Liping, et al. Study on the quantitative classification of construction workers' unsafe behaviors[J]. Journal of Safety and Environment, 2021, 21(6): 2 617-2 627.

[2] 丛志明, 马敬妍. 建筑业人员安全管理困境分析及改进研究[J]. 管理科学与工程, 2024, 13(3): 568-577.
Cong Zhiming, Ma Jingyan. Analysis and improvement research on the difficulties of personnel safety management in the construction industry[J]. Management Science and Engineering, 2024, 13(3): 568-577.

[3] Bureau of Labor Statistics. Fatal occupational injuries by industry sectors and selected events or exposures, 2023[EB/OL]. (2024-12-19). [2026-01-8]. <https://www.bls.gov/news.release/cfoi.t02.htm>.

[4] Wang Yingchen, Cui Jingyao, Zhang Yikai, et al. Study and action plan on the key factors influencing unsafe behaviors

- by construction workers[J]. Buildings, 2024, 14(7): DOI: 10.3390/buildings14071973.
- [5] 边存灵,吕伟刚,冯伟. 骨架人体行为识别研究回顾、现状及展望[J]. 计算机工程与应用,2024,60(20):11-25.
Bian Cunling, Lyu Weigang, Feng Wei. Skeleton-based human action recognition: history, status and prospects[J]. Journal of Computer Engineering & Applications, 2024, 60(20): 11-25.
- [6] 刘耀,焦双健. ST-GCN 在建筑工人不安全动作识别中的应用[J]. 中国安全科学学报, 2022, 32(4): 30-35.
Liu Yao, Jiao Shuangjian. Application of ST-GCN in unsafe action identification of construction workers[J]. China Safety Science Journal, 2022, 32(4): 30-35.
- [7] 甘文霞,张宇轩,耿晶,等. 改进 PoseConv3D 模型在建筑工人临边不安全行为识别中的应用[J]. 安全与环境学报, 2024, 24(7): 2 712-2 720.
Gan Wenxia, Zhang Yuxuan, Geng Jing, et al. Application of improved PoseConv3D model in recognition of unsafe behaviors of construction workers near the edge[J]. Journal of Safety and Environment, 2024, 24(7): 2 712-2 720.
- [8] 张泽辉,张乾隆,徐晓滨,等. 基于视觉的工人高处攀爬不安全行为识别模型[J]. 中国安全科学学报, 2025, 35(2): 144-151.
Zhang Zehui, Zhang Qianlong, Xu Xiaobin, et al. Unsafe behavior recognition model of high climbing workers based on vision[J]. China Safety Science Journal, 2025, 35(2): 144-151.
- [9] Alowonou K C, Han J H. MSA-GCN: exploiting multi-scale temporal dynamics with adaptive graph convolution for skeleton-based action recognition[J]. IEEE Access, 2024, 12: 54 681-54 694.
- [10] Zhou Yuxuan, Yan Xudong, Cheng Zhiqi, et al. Blockgcnn: redefine topology awareness for skeleton-based action recognition[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 2 049-2 058.
- [11] 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: 1 010-1 019.
- [12] 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, 2019, 42(10): 2 684-2 701.
- [13] Yan Sijie, Xiong Yuanjun, Lin Dahua. Spatial temporal graph convolutional networks for skeleton-based action recognition[C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2018: 7 444-7 452.
- [14] Li Maosen, Chen Siheng, Chen Xu, et al. Actional-structural graph convolutional networks for skeleton-based action recognition[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 3 595-3 603.
- [15] Yang Hao, Yan Dan, Zhang Li, et al. Feedback graph convolutional network for skeleton-based action recognition[J]. IEEE Transactions on Image Processing, 2021, 31: 164-175.
- [16] Cheng Ke, Zhang Yifan, He Xiangyu, et al. Skeleton-based action recognition with shift graph convolutional network[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 183-192.
- [17] GB 3608—2025 高处作业分级[S].
GB 3608—2025 Classification of work at height[S].
- [18] Wu Han, Han Yu, Zhang Meng, et al. Identifying unsafe behavior of construction workers: a dynamic approach combining skeleton information and spatiotemporal features[J]. Journal of Construction Engineering and Management, 2023, 149(11): DOI: 10.1061/JCEMD4.COENG-13616.
- [19] Li Ping, He Youshi, Li Zhengguang. Study on influencing factors of construction workers' unsafe behavior based on text mining[J]. Frontiers in Psychology, 2022, 13: DOI: 10.3389/fpsyg.2022.886390.
- [20] Chen Yuxin, Zhang Ziqi, Yuan Chunfeng, 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: 13 359-13 368.



作者简介: 余洋 (1987—),男,四川泸州人,博士,讲师,硕士生导师,主要从事人因安全、工程安全、工业互联网、工业大数据等方面的研究。E-mail:sccdyuyang@163.com。