

文章编号:1007-2780(2022)04-0530-09

基于轻量级图卷积网络的校园暴力行为识别

李 颀¹, 邓耀辉^{2*}, 王 娇²

(1. 陕西科技大学 电子信息与人工智能学院, 陕西 西安 710021;

2. 陕西科技大学 电气与控制工程学院, 陕西 西安 710021)

摘要:针对卷积神经网络和图卷积网络的两类算法在校园暴力行为识别中识别速度和识别率不高的问题,本文提出一种结合多信息流数据融合和时空注意力机制的轻量级图卷积网络。以人体骨架为研究对象,首先融合关节点和骨架相关的多信息流数据,通过减少网络参数量来提高运算速度;其次构建基于非局部运算的时空注意力模块关注最具动作判别性的关节点,通过减少冗余信息提高识别准确率;接着构建时空特征提取模块获得关注关节点时空关联信息;最终由 Softmax 层实现动作识别。实验结果表明:在校园安防实景中对拳打、脚踢、倒地、推搡、打耳光和跪地 6 种典型动作识别准确率分别为 94.5%, 97.0%, 98.5%, 95.0%, 94.5%, 95.5%, 识别速度最大为 20.6 fps。在 UCF101 数据集上对比两类基准网络,识别速度和准确率均有提升,验证了方法对其他动作的通用性,可以满足对校园典型暴力行为识别的实时性和可靠性要求。

关键词:校园暴力行为识别; 图卷积网络; 数据融合; 时空注意力模块

中图分类号:TP391.4 **文献标识码:**A **doi:**10.37188/CJLCD.2021-0229

Campus violence action recognition based on lightweight graph convolution network

LI Qi¹, DENG Yao-hui^{2*}, WANG Jiao²

(1. School of Electrical Information and Artificial Intelligence, Shaanxi University of Science and Technology, Xi'an 710021, China;

2. School of Electrical and Control Engineering, Shaanxi University of Science and Technology, Xi'an 710021, China)

Abstract: Aiming at the problem of low recognition speed and recognition rate of convolution neural network and graph convolution network in campus violence recognition, this paper proposes a lightweight graph convolution network combined with multi-information flow data fusion and spatio-temporal attention mechanism. The human skeleton is taken as the research object. Firstly, the multi-information flow data related to joint points and skeleton are fused to improve the operation speed by reducing the number of net-

收稿日期:2021-08-31; **修订日期:**2021-10-27.

基金项目:西安市科技计划(No. 201806117YF05NC13(1)); 西安市未央区科技计划(No. 201305); 陕西科技大学博士科研启动基金(No. BJ13-15)

Supported by Xi'an Science and Technology Plan Project(No. 201806117YF05NC13(1)); Xi'an Weiyang District Science and Technology Plan Project(No. 201305); Doctoral Research Start-up Fund of Shaanxi University of Science and Technology(No. BJ13-15)

*通信联系人, E-mail: 173743077@qq.com

work parameters. Secondly, the spatio-temporal attention module based on nonlocal operation is constructed to focus on the most action discriminant nodes, and the recognition accuracy is improved by reducing redundant information. Then, the spatio-temporal feature extraction module is constructed to obtain the spatio-temporal correlation information of the concerned nodes. Finally, action recognition is realized by Softmax layer. The experimental results show that the recognition accuracy of boxing, kicking, falling, pushing, earlighting and kneeling in campus security scene is 94.5%, 97.0%, 98.5%, 95.0%, 94.5% and 95.5%, respectively, and the maximum recognition speed is 20.6 fps. Compared with the two benchmark networks on UCF101 dataset, the recognition speed and accuracy are improved, which verifies the universality of the method for other actions. Therefore, it can meet the real-time and reliability requirements of typical campus violence identification.

Key words: campus violence action recognition; graph convolution network; information flow data fusion; spatio-temporal attention module

1 引言

我国校园安全在依赖人工巡查的基础上,逐步向智能化方向发展,有关人脸检测^[1]与人脸识别^[2]系统应用已经非常广泛,然而缺乏成熟的异常行为识别系统。深度学习中基于卷积神经网络的暴力行为识别方法受图像光照和颜色等因素影响较大,识别速度和准确率有待大幅提高^[3]。人体骨架序列不受光照和颜色影响,可以表征人体关节点和骨架变化与人体行为的关联信息,但基于骨架数据的图卷积网络的方法识别速度和识别率未能满足实际应用,有望通过改进图卷积网络提高实时性和可靠性。

早期人体行为识别通过专家手工设计特征模拟关节之间的相关性实现^[4]。Yang和Tian采用朴素贝叶斯最近邻分类器(Naive-Bayes-Nearest-Neighbor, NBNN)实现了多类动作的识别^[5],但手工提取和调参表征能力有限且工作量大;Li和He等人通过深度卷积神经网络(Convolutional Neural Network, CNN)提取不同时间段的多尺度特征并得到最终识别结果,但映射过程信息丢失、网络参数量庞大^[6];Zhao和Liu等人通过对原始骨架关节坐标进行尺度变换后输入残差独立循环神经网络(Recurrent Neural Network, RNN)得到识别结果,表征时间信息的能力增强,但易丢失原始关节点之间的关联信息^[7];Yan和Xiong等人首次提出用图卷积网络(Graph Convolutional Network, GCN)进行行为识别,避免了手工设计遍历规则带来的缺陷^[8]。

基于人体骨架的行为识别受光照和背景等因素影响非常小,与基于RGB数据的方法相比具有很大优势。人体的关节骨架数据是一种拓扑图,图中每个关节点在相邻关节点数不同的情况下,传统的卷积神经网络不能直接使用同样大小的卷积核进行卷积计算去处理这种非欧式数据^[9]。因此,在基于骨架的行为识别领域,基于图卷积网络的方法更为适合。从研究到应用阶段的转换,需要在保证准确率的同时实现网络的轻量化:(1)需要在多种信息流数据构成的数据集上分别多次训练,融合各训练结果得到最终结果,增加了网络参数量和计算复杂度;(2)输入的骨架序列中,存在冗余的关节点信息,导致识别速度和识别率降低。

2 轻量级图卷积网络搭建

2.1 图卷积网络

以图像为代表的欧式空间中,将图像中每个像素点当作一个结点,则结点规则排布且邻居结点数量固定,边缘上的点可进行Padding填充操作。但在图结构这种非欧空间中,结点排布无序且邻居结点数量不固定,无法通过传统的卷积神经网络固定大小的卷积核实现特征提取,需要一种能够处理变长邻居结点的卷积核^[10]。对图而言,需要输入维度为 $N \times F$ 的特征矩阵 X 和 $N \times N$ 的邻接矩阵 A 提取特征,其中 N 为图中结点数, F 为每个结点输入特征个数。相邻隐藏层的结点特征变换公式为:

$$H^{i+1} = f(H^i, A), \quad (1)$$

其中 i 为层数, 第一层 $H^0 = X$; $f(\bullet)$ 为传播函数, 不同的图卷积网络模型传播函数不同。每层 H^i 对应 $N \times F^i$ 维度特征矩阵, 通过传播函数 $f(\bullet)$ 将聚合后的特征变换为下一层的特征, 使得特征越来越抽象。

2.2 轻量级图卷积网络框架

为了使人体骨架序列中的动作特征被充分利用, 且在识别准确率提高的同时实现动作识别模型的轻量化, 本文提出了一种结合多信息流数据融合和时空注意力机制的轻量级自适应图卷积网络。以输入的人体骨架序列为研究对象, 首先融合关节点信息流、骨长信息流、关节点偏移信息流和骨长变化信息流 4 种数据信息; 接着构建基于非局部运算的可嵌入的时空注意力模块, 关注信息流数据融合后人体骨架序列中最具动作判别性的关节点; 最后通过 Softmax 得到对动作片段的识别结果, 网络主体框架如图 1 所示。

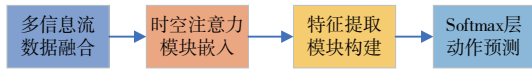


图 1 网络框架

Fig. 1 Network framework

2.3 多信息流数据融合

现阶段基于图卷积的方法^[11]多采用在多种不同数据集下多次训练, 根据训练结果进行决策级融合, 导致网络参数量大。因此, 在训练之前对原始关节点坐标数据进行预处理, 实现关节点信息流、骨长信息流、关节点偏移信息流和骨长变化信息流的数据级融合, 减少网络参量, 从而降低计算要求。

人体骨架序列关节点的定义如公式(2)所示:

$$s = \{V_{i,t} | i = 1, 2, \dots, N; t = 1, 2, \dots, T\}, \quad (2)$$

其中: T 为序列中的总帧数, N 为总关节点数 18, i 为在 t 时刻的关节点。融合多信息流之前, 需要进行骨架序列 s 的多样化预处理。关节点信息流由人体姿态估计算法 OpenPose 获取到的 18 个关节点坐标得到, 相对于动作捕获设备成本大幅降低^[12-13]。其他信息流数据定义如下。

骨长信息流(Bone Length Information Flow): 将靠近人体重心的关节点定义为源关节点, 坐标

表示为 $V_{i,t} = (x_{i,t}, y_{i,t})$; 远离重心点的关节点定位为目标关节点, 坐标表示为 $V_{j,t} = (x_{j,t}, y_{j,t})$ 。通过两关节点作差获取骨长信息流:

$$B_{i,j,t} = V_{j,t} - V_{i,t} = (x_{j,t} - x_{i,t}, y_{j,t} - y_{i,t}). \quad (3)$$

关节点偏移信息流(Joint Difference Information Flow): 定义第 t 帧的关节点 i 的坐标表示为 $V_{i,t} = (x_{i,t}, y_{i,t})$, 第 $t+1$ 帧的关节点 i 的坐标表示为 $V_{i,t+1} = (x_{i,t+1}, y_{i,t+1})$, 关节点偏移信息流可通过对相邻帧同一关节点坐标位置作差获得:

$$JD_{i,t,t+1} = V_{i,t+1} - V_{i,t} = (x_{i,t+1} - x_{i,t}, y_{i,t+1} - y_{i,t}). \quad (4)$$

骨长变化信息流(Change of Bone Length Information Flow): 相邻两帧中, 同一节骨骼由于动作变化导致所表现出的长度不同, 由公式(3)定义第 t 帧的骨长信息流为 $B_{i,j,t}$, 则第 $t+1$ 帧的骨长信息流为 $B_{i,j,t+1}$, 通过对相邻帧同一骨骼长度作差获得骨长变化信息流:

$$CB_{i,j,t,t+1} = B_{i,j,t+1} - B_{i,j,t}. \quad (5)$$

如图 2 所示, 根据对关节点信息流、骨长信息流、关节点偏移信息流和骨长变化信息流的定义, 将多数据流加权融合成单一的特征向量, 骨架序列维度由 $4 \times T \times J \times C_1$ 变为 $1 \times T \times J \times 4C_1$:

$$\text{Fusion} = \{\omega_1 V_{i,t} + \omega_2 B_{i,j,t} + \omega_3 JD_{i,t,t+1} + \omega_4 CB_{i,j,t,t+1}, \\ i = 1, 2, \dots, N; t = 1, 2, \dots, T\}, \quad (6)$$

其中: 权重 $\omega_1 \sim \omega_4$ 由关节点偏移度 σ_1 ($\sigma_1 \in [0^\circ \sim 360^\circ]$) 和骨长变化度 σ_2 ($\sigma_2 \in [0 \sim 100\%]$) 决定, σ_1 为前一帧坐标点 $V_{i,t}$ 与后一帧坐标点 $V_{i,t+1}$ 分别和坐标原点所构成直线的夹角, σ_2 如式(7)定义:

$$\sigma_2 = \frac{|V_{j,t+1} - V_{i,t+1}| - |V_{j,t} - V_{i,t}|}{|V_{j,t} - V_{i,t}|}, \quad (7)$$

式中: 绝对值运算代表骨骼长度, 当 $\sigma_1 \geq 30^\circ$ 且 $\sigma_2 \leq 50\%$ 时, ω_1 和 ω_3 权值为 2, ω_2 和 ω_4 权值为 1;

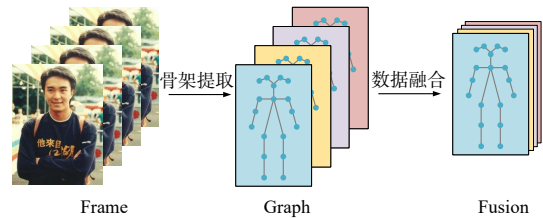


图 2 信息流数据融合

Fig. 2 Data fusion of information flow

当 $\sigma_1 \leq 30^\circ$ 且 $\sigma_2 \geq 50\%$ 时, ω_1 和 ω_3 权值为1, ω_2 和 ω_4 权值为2;当 σ_1 和 σ_2 都小于阈值时,权值均为1;当 σ_1 和 σ_2 都大于阈值时,权值均为2。通过计算关节点偏移程度以及骨长变化程度,为变化程度大的信息流数据赋予了更高的权重,从而增强了信息流对动作的表征。再使用融合后的单一特征向量表示多信息流数据,将训练次数由4次减少为1次,降低了总体参数量,从而提高网络运算速度。

2.4 时空注意力模块构建

在保证网络运算速度提升的基础上,也要保证动作识别的准确性。一段人体骨架序列包含时间域和空间域的所有信息,但是只有对拳打、脚踢和倒地动作具有判别性的关节点关联信息值得关注,注意力机制大多只是去除无关项而关注感兴趣动作区域,但真正的冗余信息来自两个方面:(1)拳打动作发生时,只有肩膀、手肘和手腕3个关节点相互之间相关性强;脚踢动作发生时,只有髋、膝盖、脚踝跟3个关节点相互之间相关性强,这些关键关节点与其他关节点相关性弱或不相关。(2)受到暴力拳打或脚踢而倒地后,各关节点偏移幅度较小,前后帧的各关节点相关性几乎不变,无需对后一帧骨架信息进行提取。

将每个关节点偏移度 $\sigma_i \geq 30^\circ$ 的关节点定义为源关节点,每次选取一个源关节点,其他关节点则为目标关节点,神经网络中的局部运算方法只能对目标关节点遍历后单独计算两两的相关性,使源关节点丢失全局表征能力。为了表征所有目标关节点对源关节点的相关性,如图3所示,将非局部运算(Non-local operations)的思想融入时空注意力模块,并在特征输入后添加尺寸为 2×2 、步长为2的最大池化层(Maxpool layer),以保证压缩数据和参数数量的同时尽可能保留原有特征。

时空注意力模块(Spatio-temporal Attention Module, STA)包含一个空间注意力模块和时间注意力模块,其中空间注意力模块(Spatial Attention Module, SA)捕获帧内关节相关性,时间注意力模块(Temporal Attention Module, TA)捕获帧间关节的相关性,最终二者与输入特征相加融合。时空注意力模块输出特征的维度和输入相同,因此可以嵌入图卷积网络的网络结构之间。

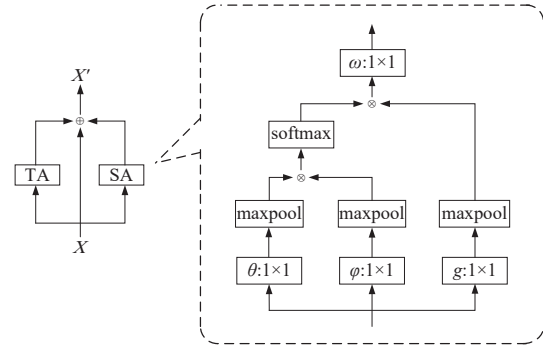


图3 时空注意力模块

Fig. 3 Spatio-temporal attention module

模块功能的实现分为4个步骤:

(1)输入特征 X 的维度为 $T \times N \times C$,其中 T 、 N 和 C 分别对应帧、关节和通道的数目,将空间注意力模块的输入特征表示为 $z = [z_1^s, z_2^s, \dots, z_N^s] \in \mathbb{R}^{T \times N \times C}$ 。

(2)将特征嵌入到高斯函数(θ 和 φ ,卷积内核尺寸 1×1)中计算任意位置两个关节 i 和 j 的相关性,由 j 进行枚举,得到关节点 i 的加权表示:

$$y_i^s = \frac{1}{C(z^s)} \sum_j f(z_i^s, z_j^s) g(z_j^s), \quad (8)$$

其中: z_i^s 和 z_j^s 分别表示关节点 i 和 j 的特征;函数 g 用来计算关节点 j 特征表示, $g(z_j^s) = W_g^s z_j^s$, W_g^s 是待学习的权重矩阵;高斯函数 f 定义为:

$$f(z_i^s, z_j^s) = e^{\theta(z_i^s)^T \varphi(z_j^s)}, \quad (9)$$

其中: $\theta(z_i^s) = W_\theta^s z_i^s$, $\varphi(z_j^s) = W_\varphi^s z_j^s$,设定 $C(z^s) = \sum_j f(z_i^s, z_j^s)$ 为相关性表示的归一化因子。为了降低计算成本及最大程度保留低阶特征,在函数 θ 、 φ 和 g 之后添加了尺寸为 2×2 、步长为2的最大池化层。

(3)通过函数 ω 使加权后得到空间注意力信息 $o_i^s (o_i^s \in \mathbb{R}^{T \times N \times C})$:

$$o_i^s = W_o^s y_i^s. \quad (10)$$

(4)将时间注意力模块的输入特征表示 $z' = [z_1^t, z_2^t, \dots, z_N^t] \in \mathbb{R}^{T \times N \times C}$;重复(2)、(3)得到时间注意力信息 $o_i^t (o_i^t \in \mathbb{R}^{T \times N \times C})$,与空间注意力信息和输入特征相加融合得到时空注意力信息 $o_i (o_i \in \mathbb{R}^{T \times N \times C})$:

$$o_i = o_i^s + o_i^t + z_i. \quad (11)$$

通过基于非局部运算的注意力机制得到具有判别性的关节点时空关联信息,去除了动作区

域无关项和输入的冗余关节信息的干扰,减少了不必要的计算,从而提高了准确率。

2.5 时空特征提取模块构建

为了提取骨架序列在空间和时间维度上的特征,首先利用时空图卷积网络 and 空间划分策略对动态骨架进行建模,原始表达式为:

$$X_{out} = \sum_i^K W_i (X_{in} A_i) \odot M_i, \quad (12)$$

其中, X_{in} 和 X_{out} 分别为图卷积输入和输出特征, K 为空间域卷积核尺寸, W_i 为权重, A_i 为关节点 i 的邻接矩阵, $\odot$ 代表点乘, M_i 为赋予连接权重的关节点映射矩阵。

使用预先定义好的骨架结构数据无法对所有未知动作准确识别,因此需要设计一种具有自适应性的邻接矩阵 A_i ,使得图卷积网络模型具有自适应性。因此,为了在网络学习中改变骨架序列图的拓扑结构,将式(12)中决定拓扑结构的邻接矩阵和映射矩阵分成 A_i 、 H_i 和 L_i ,自适应图卷积模块框图如图4所示,输出特征重新构造为:

$$X_{out} = \sum_i^K W_i X_{in} (A_i + H_i + L_i). \quad (13)$$

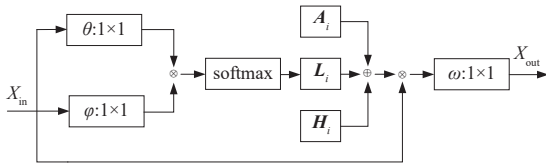


图4 自适应图卷积模块

Fig. 4 Adaptive graph convolutional module

图4中 θ 和 φ 即式(9)中高斯嵌入函数,卷积内核尺寸为 1×1 ;第一部分 A_i 仍为关节点 i 的邻接矩阵;第二部分 H_i 作为对原始邻接矩阵的加法补充,能通过网络训练不断迭代更新;第三部分 L_i 由数据不断驱动更新来学习连接权重,关节点相关性可由式(8)计算得到后与 1×1 卷积相乘得到相似性矩阵 L_i :

$$L_i = \text{softmax}(X_{in}^T W_{\theta i}^T W_{\varphi i} X_{in}). \quad (14)$$

通过以上计算,构建出具有自适应性的图卷积模块,接下来对骨架序列包含的时空信息进行提取。

本文提出的时空特征提取模块如图5所示。在每次完成卷积操作后通过BN(Batch normalization)层将数据归一化,再通过ReLU层提高模型表达能力。可嵌入的时空注意力模块STA已

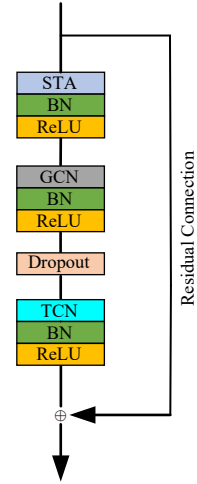


图5 时空特征提取模块

Fig. 5 Spatio-temporal feature extracting module

在2.4一节中搭建完成,将特征输入提取模块后对感兴趣动作关节点进行提取。接着通过具有自适应性的GCN在空间维度上获得骨架数据中同一帧各关节点的相关性,通过时间卷积网络(Temporal Convolutional Network, TCN)在时间维度上获得相邻帧同一关节点的关系。丢弃层(Dropout)减少隐层结点的相互作用避免了图卷积网络的过度拟合,参数设置为0.5,同时为了增加模型稳定性进行了残差连接。

2.6 整体网络结构搭建

如图6所示,将9个时空特征提取模块 $B_1 \sim B_9$ 进行堆叠,从特征输入 X 到行为标签 Label 输出方向上,BN层用于骨架图输入后进行标准化, $B_1 \sim B_3$ 输出特征维度为 $\text{Batch} \times 64 \times T \times N$, $B_4 \sim B_6$ 输出特征维度为 $\text{Batch} \times 128 \times T/2 \times N$, $B_7 \sim B_9$ 输出特征维度为 $\text{Batch} \times 256 \times T/4 \times N$,其中通道数分别为64,64,64,128,128,128,256,256,256。在空间和时间维度上应用全局平均池化操作(Global Average Pooling, GAP)

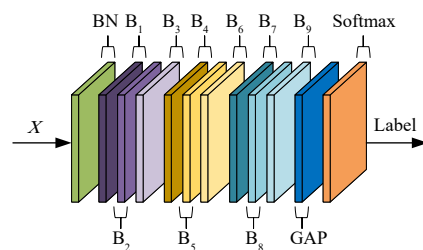


图6 整体网络架构

Fig. 6 Overall network architecture

将样本的特征图大小进行统一,最终使用 softmax 层得到 0~1 的数据进行人体行为的识别。

3 实验结果与分析

3.1 实验配置

实验平台的配置为 8 代 i7 CPU, 64 G 内存, 4 TB 固态硬盘存储, 显卡为 RTX2080Ti。深度学习框架为 PyTorch1.3, Python 版本为 3.6。优化策略采用随机梯度下降(Stochastic gradient descent, SGD), 每批次训练样本数(Batch size)设置为 64, 迭代次数(Epoch)设置为 60, 初始学习率(Learning rate)为 0.1, Epoch 达到 20 以后学习率设置为 0.01。

3.2 行为识别实验

3.2.1 校园安防实景测试

本文面向实际应用,对校园马路、操场和湖边等不同场景制作了 12 000 个视频片段,拳打、脚踢、倒地、推搡、打耳光和跪地 6 种典型动作各 2 000 个,单个时长不大于 5 s。所有人员身高、体重和身体比例等方面有所差异,以增强模型

的泛化能力。根据实验配置进行训练,图 7 为模型的训练损失与综合测试准确率的变化曲线。

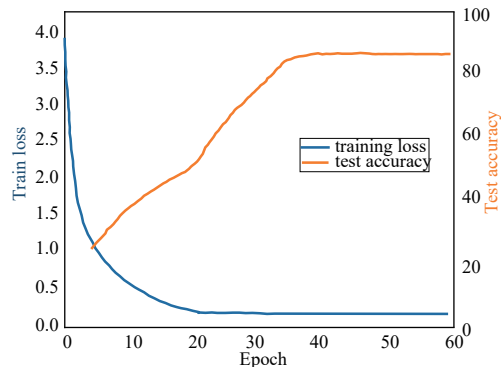


图7 模型训练损失与测试准确率变化图

Fig. 7 Variation diagram of model training loss and test accuracy

可以看出随着迭代次数的增长,模型的训练损失逐渐下降。当 epoch 在 20 左右时,由于学习率的下降,测试准确率开始大幅提高;当 epoch 超过 35 之后,训练损失与测试准确率几乎保持不变。使用训练好的模型分别对 6 类动作对应的测试集进行测试,主要识别过程如图 8 所示。

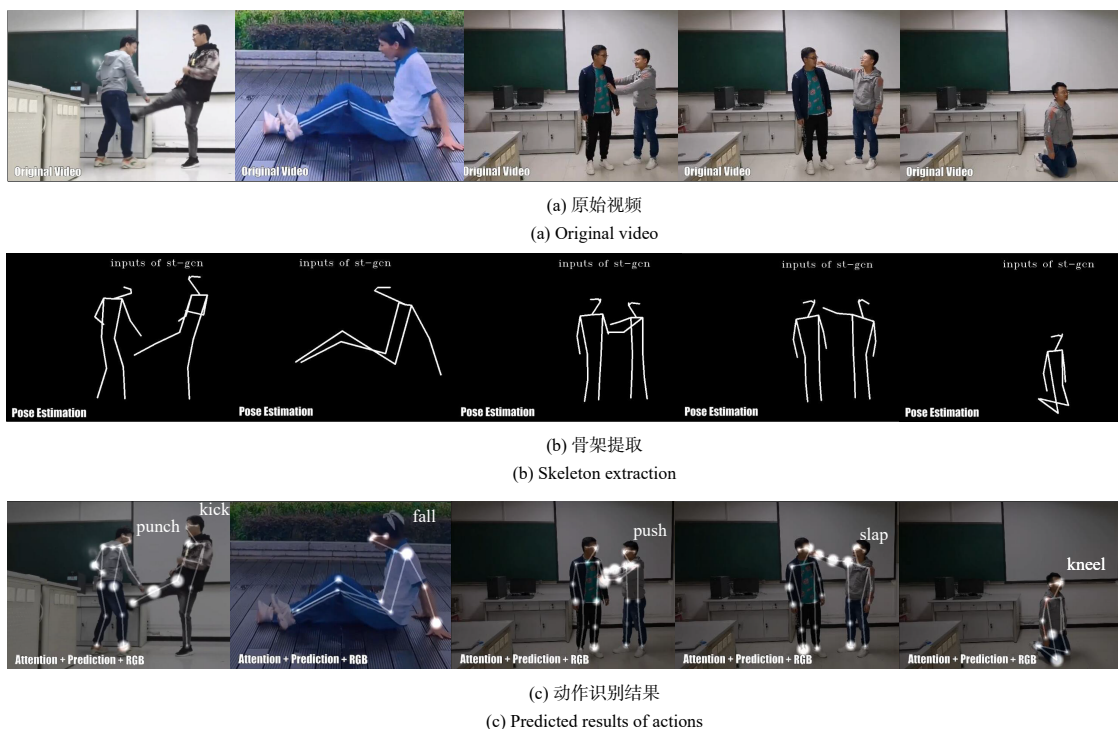


图8 6种典型动作识别过程

Fig. 8 Six typical action recognition processes

图 8 中处理的 5 组动作片段从左至右分别为拳打和脚踢、倒地、推搡、打耳光及跪地,图 8(a)是原视频;图 8(b)是对输入的含有拳打和脚踢动作的视频片段使用 OpenPose 进行人体关节点提取,正确匹配各关节点后得到人体骨架;图 8(c)是将骨架序列输入本文改进的时空图卷积网络得到动作片段的识别结果。改进后模型的处理速度最大可达 20.6 fps,对校园安防实景中拳打、脚踢、倒地、推搡、打耳光和跪地 6 种典型动作识别准确率分别为 94.5%, 97.0%, 98.5%, 95.0%, 94.5%, 95.5%, 测试结果如表 1 所示。

表 1 6 种典型动作识别结果

Tab. 1 Six typical action recognition results

动作类别	样本个数/ 个	识别准确率/ %	最大识别速度/ fps
拳打	200	94.5	20.6
脚踢	200	97.0	20.6
倒地	200	98.5	20.6
推搡	200	95.0	20.6
打耳光	200	94.5	20.6
跪地	200	95.5	20.6

为了验证不同体型(身高、体重和肩宽表示)人员对识别准确率存在影响,选取参与数据集制作的 1~6 号实验人员,每次使用由单一实验人员获取的 6 种典型动作片段作为训练集,将由其他 5 个实验人员获取的 6 种动作片段作为测试集,并记录对所有动作的平均识别准确率,实验参数及结果如表 2 所示。

由表 2 数据可知,使用单一实验人员所拍摄

表 2 不同体型人员动作识别结果

Tab. 2 Action recognition results of personnel with different body types

测试集 人员	身高/cm	体重/kg	肩宽/cm	平均识别 准确率/%
1 号	173	76	45	83.5
2 号	179	67	41	68.7
3 号	154	46	35	72.4
4 号	168	59	37	85.6
5 号	176	80	48	85.0
6 号	163	103	47	64.5

的 6 类动作片段作为数据集进行训练,并分别对其他人员的动作片段测试,测试结果最佳仅为 85.6%,而使用所有实验人员视频片段识别准确率在 94.5% 以上,说明了不同人员体型的差异性可以增强模型的泛化能力,即鲁棒性。

表 2 的 1~6 号实验人员中,2 号的体型为 179 cm/67 kg,身材过瘦;3 号的体型为 155 cm/46 kg,身材矮小,但身高体重比例正常;6 号的体型为 163 cm/103 kg,身材肥胖;1 号、4 号和 5 号体型基本正常。不同体型的人做同一种动作时,姿态检测算法获取的 18 个人体骨骼点坐标有差异,从而骨长也会产生差异,关节点信息流、骨长信息流、关节点偏移信息流和骨长变化信息流 4 种数据信息也有区别。因为 2 号过瘦,各关节点坐标较为集中,而 6 号过胖,各关节点坐标较为分散,导致 2 号和 6 号的平均识别准确率最低,仅为 68.7% 和 64.5%;而 3 号身材比例正常,但身高过于矮小,也导致了关节坐标点分布不均匀,72.4% 的准确率低于其他正常体型。

综上,在数据集的制作过程中所有人员体型差异的多样性可以增强模型的泛化能力,实验结果也表明本文方法可快速有效地识别出校园暴力的典型动作。

3. 2. 2 方法对比实验

为了验证本文方法的有效性,采用具有挑战性的 UCF101 数据集进行行为识别对比实验。该数据集有 101 类动作,13 320 段视频,在人员姿态、外观、摄像机运动状态、和物体大小比例等方面具有多样性。

按照 6:2:2 的比例,参与训练和验证的视频数据 10 656 个,测试视频 2 664 个,使用表 3 中 5 种方法进行对比实验,在当前配置下对视频片段处理速度由 9.2~15.5 fps 最大提高至 19.3 fps,对数据集中 101 类动作平均识别准确率以及参数量变化对比结果如表 3 所示,并在表 4 中给出了数据集中 6 种动作的识别准确率。

表 3 数据表明:本文方法(无注意力模块)相对于两种卷积神经网络的方法,参数量分别减少约 92.6% 和 94.7%,而识别准确率提高 21.4% 和 4.0%;相对于改进前时空图卷积网络的方法,参数量减少约 59.6%,而准确率提高 1.2%。说明本文的多信息流数据融合方法可有效减少网

表 3 不同识别方法的对比结果

Tab. 3 Comparison results of different recognition methods

方法	参数量/ MB	平均识别 准确率/%	最大识别 速度/fps
Slow Fusion CNN ^[14]	17. 03	65. 4	15. 5
VA-CNN ^[15]	24. 03	82. 8	9. 2
ST-GCN ^[7]	3. 10	85. 6	12. 7
Ours(无注意力模块)	1. 25	86. 8	18. 7
Ours(含注意力模块)	0. 78	89. 7	19. 3

表 4 6种数据集动作识别结果

Tab. 4 Six typical action recognition results

动作类别	识别准确率/%
拳打	85. 2
脚踢	90. 7
行走	92. 4
下蹲	88. 2
跳远	87. 6
推	91. 5

络参数量,实现网络轻量化。其中,使用基于非局部运算的时空注意力机制相对于未使用时参数量减少约 37. 6%,准确率提高 2. 9%,说明改进后的时空注意力机制可有效减少冗余关节信息,提高了特征的利用率,从而提高了识别准确率。表 4 数据列出了改进后方法在 UCF101 数据集中 6 种动作的识别准确率。由于该数据集中动作片段来源于不受约束的网络视频,存在相机运

动、部分遮挡和低分辨率等影响导致视频质量差,实验中在 OpenPose 进行人体关节点提取阶段 csv 文件中所存的关节点坐标有部分缺失,因此相较于表 1 中实测数据集识别准确率均偏低。

综上,本文方法在保证准确率提升的同时实现了网络的轻量化,从而提高了可靠性与实时性。

4 结 论

针对校园智能安防识别速度和识别率不高导致可靠性和实时性差的问题,本文提出了一种基于轻量级图卷积的人体骨架数据的行为识别方法,通过多信息流数据融合与自适应图卷积相结合的方式,同时通过嵌入时空注意力模块提高特征的利用率,在校园安防实景中对拳打、脚踢、倒地、推搡、打耳光和跪地 6 种典型动作识别准确率分别为 94. 5%,97. 0%,98. 5%,95. 0%,94. 5%,95. 5%,识别速度最快为 20. 6 fps,且验证了模型的泛化能力。同时在行为识别数据集 UCF101 上验证了方法的通用性,可以扩展至人体其他动作,在参数量比原始时空图卷积网络减少了 74. 8% 的情况下,平均识别准确率由 85. 6% 提高到 89. 7%,识别速度最大提高至 19. 3 fps,能够较好地完成校园实际安防中出现最多的典型暴力行为识别任务。

参 考 文 献:

[1] 林国军,杨明中,陈明举,等. 一种肤色定位的人脸检测算法[J]. 液晶与显示,2019,34(1):70-73.
LIN G J, YANG M Z, CHEN M J, *et al.* Face detection algorithm based on skin color location [J]. *Chinese Journal of Liquid Crystals and Displays*, 2019, 34(1): 70-73. (in Chinese)

[2] 李鹏飞,许金凯,韩文波,等. 基于 S3C2440 的人脸识别平台的设计[J]. 液晶与显示,2014,29(3):417-421.
LI P F, XU J K, HAN W B, *et al.* Design of face identification platform based on S3C2440 [J]. *Chinese Journal of Liquid Crystals and Displays*, 2014, 29(3): 417-421. (in Chinese)

[3] 胡韬. 基于深度学习卷积神经网络的人体行为识别研究[J]. 科技传播,2020,12(6):130-131.
HU T. Research on human behavior recognition based on deep learning convolutional neural network [J]. *Public Communication of Science & Technology*, 2020, 12(6): 130-131. (in Chinese)

[4] 蔡强,邓毅彪,李海生,等. 基于深度学习的人体行为识别方法综述[J]. 计算机科学,2020,47(4):85-93.
CAI Q, DENG Y B, LI H S, *et al.* Survey on human action recognition based on deep learning [J]. *Computer Science*, 2020, 47(4): 85-93. (in Chinese)

[5] YANG X D, TIAN Y L. Effective 3D action recognition using EigenJoints [J]. *Journal of Visual Communication and Image Representation*, 2014, 25(1): 2-11.

- [6] LI B, DAI Y C, CHENG X L, *et al.* Skeleton based action recognition using translation-scale invariant image mapping and multi-scale deep CNN [C]//*Proceedings of 2017 IEEE International Conference on Multimedia & Expo Workshops*. Hong Kong, China: IEEE, 2017: 601-604.
- [7] YAN S J, XIONG Y J, LIN D H. Spatial temporal graph convolutional networks for skeleton-based action recognition [C]//*Proceedings of the 32nd AAAI Conference on Artificial Intelligence*. New Orleans: AAAI Press, 2018: 7444-7452.
- [8] 赵明富, 刘帅, 宋涛, 等. 基于残差独立循环神经网络的空间增强人体骨架行为识别[J]. 激光杂志, 2020, 41(7): 37-43.
- ZHAO M F, LIU S, SONG T, *et al.* Spatial enhancement of human skeleton behavior recognition based on residual independent recurrent neural network [J]. *Laser Journal*, 2020, 41(7): 37-43. (in Chinese)
- [9] 曹毅, 刘晨, 黄子龙, 等. 时空自适应图卷积神经网络的骨架行为识别[J]. 华中科技大学学报(自然科学版), 2020, 48(11): 5-10.
- CAO Y, LIU C, HUANG Z L, *et al.* Skeleton-based action recognition based on spatio-temporal adaptive graph convolutional neural-network [J]. *Journal of Huazhong University of Science and Technology (Nature Science Edition)*, 2020, 48(11): 5-10. (in Chinese)
- [10] 王健宗, 孔令炜, 黄章成, 等. 图神经网络综述[J]. 计算机工程, 2021, 47(4): 1-12.
- WANG J Z, KONG L W, HUANG Z C, *et al.* Survey of graph neural network [J]. *Computer Engineering*, 2021, 47(4): 1-12. (in Chinese)
- [11] 孔伟, 刘云, 李辉, 等. 基于图卷积网络的行为识别方法综述[J]. 控制与决策, 2021, 36(7): 1537-1546.
- 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. (in Chinese)
- [12] 朱洪堃, 殷佳炜, 冯文字, 等. 一种轻量化实时人体姿势检测模型研究与应用[J]. 系统仿真学报, 2020, 32(11): 2155-2165.
- ZHU H K, YIN J W, FENG W Y, *et al.* Research and application of a lightweight real-time human posture detection model [J]. *Journal of System Simulation*, 2020, 32(11): 2155-2165. (in Chinese)
- [13] 朱建宝, 许志龙, 孙玉玮, 等. 基于OpenPose人体姿态识别的变电站危险行为检测[J]. 自动化与仪表, 2020, 35(2): 47-51.
- ZHU J B, XU Z L, SUN Y W, *et al.* Detection of dangerous behaviors in power stations based on OpenPose multi-person attitude recognition [J]. *Automation & Instrumentation*, 2020, 35(2): 47-51. (in Chinese)
- [14] KARPATY A, TODERICI G, SHETTY S, *et al.* Large-scale video classification with convolutional neural networks [C]//*Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition*. Columbus: IEEE, 2014: 1725-1732.
- [15] ZHANG P F, LAN C L, XING J L, *et al.* View adaptive neural networks for high performance skeleton-based human action recognition [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019, 41(8): 1963-1978.

作者简介:



李 颀(1973—), 女, 陕西西安人, 博士, 教授, 2013年于西北工业大学获得博士学位, 主要从事机器视觉、信息融合等方面的研究。E-mail: liqidq@sust.edu.cn



邓耀辉(1996—), 男, 陕西西安人, 硕士研究生, 2019年于陕西科技大学获得学士学位, 主要从事信息融合和智能识别等方面的研究。E-mail: 173743077@qq.com