激光与光电子学进展 先进成像

基于密集残差移位图卷积的骨架行为识别

杨涛[1,2*]，韩军[1,2]，姜海燕[1,2]

[1]上海大学通信与信息工程学院，上海 200444；

[2]上海先进通信与数据科学研究院，上海 200444

摘要 针对人体骨架行为识别中因时空特征提取不充分、网络计算量大和计算效率低导致相似行为识别结果不理想的问题，提出一种基于密集残差移位图卷积网络的骨架行为识别算法。使用姿态估计算法提取人体骨架信息，经坐标向量计算得到骨架的关节、骨骼以及各自的运动信息，并分别输入网络中。在移位图卷积模块间引入密集残差结构，提高网络性能和提取时空特征的效率。所提算法可应用于日常行为场景，例如：行走、坐下、站起、脱衣服、穿衣服、扔以及摔倒等。其在自制数据集上的识别准确率达到81.7%，在NTU60 RGB+D数据集两种评估标准下的准确率也分别达88.1%和95.3%，验证了算法具有优秀的识别精度。

关键词 图像处理；时空特征；图卷积；密集残差；日常行为

中图分类号 TP391 **文献标志码** A **DOI**：10.3788/LOP220428

Skeleton Action Recognition Based on Dense Residual Shift Graph Convolutional Network

Yang Tao[1,2*], Han Jun[1,2], Jiang Haiyan[1,2]

[1]*College of Communication and Information Engineering, Shanghai University, Shanghai* 200444, *China*;

[2]*Shanghai Institute of Advanced Communication and Data Science, Shanghai* 200444, *China*

Abstract In order to solve the problem of recognition results of similar behaviors not being ideal owing to insufficient extraction of spatio-temporal features, a large amount of network computing, and low computing efficiency in human skeleton behavior recognition, a skeleton behavior recognition algorithm based on dense residual shift bitmap convolution network is proposed. The pose estimation algorithm is used to extract human skeleton information, and the joint, skeleton, and motion information of the skeleton are calculated by coordinate vector, and input into the network respectively. The dense residual structure is introduced between the shift graph convolution modules to improve the network performance and efficiency of extracting spatio-temporal features. The proposed algorithm can be applied to daily behavior, such as walking, sitting, standing up, undressing, dressing, throwing, and falling. The recognition accuracy on the self-made dataset is 81.7%, and under the two evaluation criteria of NTU60 RGB+D dataset, the accuracy is 88.1% and 95.3%, respectively, thus validating that the algorithm has excellent recognition accuracy.

Key words image processing; spatio-temporal feature; graph convolution; dense residual; daily behavior

1 引 言

行为识别技术具有巨大的应用前景和经济价值，在视频监控[1]、人机交互[2]和虚拟现实[3]等领域被广泛应用。早期基于深度学习的方法将骨架序列排列为向量序列和伪图像，并输入卷积神经网络(CNN)[4-5]或循环神经网络(RNN)[6-7]中预测。目前，基于人体骨架序列的主流算法是图卷积网络(GCN)[8]。Yan等[9]提出时空图卷积网络(ST-GCN)，将GCN应用到骨架序列的时空建模中，设计邻接矩阵来表示人体的物理结构，使得网络具有更好的表达能力和更高的性能。文献[10]提出一种自监督的动作连接与结构连接的图卷积网络(AS-GCN)，分别挖掘潜在的关节联系和高阶邻域信息。文献[11]，基于局部图卷积网络(PB-GCN)

收稿日期：2022-01-04；**修回日期**：2022-02-15；**录用日期**：2022-02-25；**网络首发日期**：2022-03-05

基金项目：国家自然科学基金(62071287)

通信作者：*983785320@qq.com

把几何和运动特征相结合代替关节位置坐标。Shi等[12]提出一种双流自适应图卷积网络(2S-AGCN),结合关节点和骨骼长度的信息,并且通过计算不同关节的相关性来实现邻接矩阵的自适应性。文献[13]提出一种移位图卷积网络(Shift-GCN),将移位卷积与GCN结合应用于行为识别任务中。移位图卷积操作由移位图运算和轻量的逐点卷积组成,它通过轻量级的移位操作为空间图与时间图提供了灵活的接收域,大大降低了计算复杂度,但它对时空特征提取不充分,导致一些日常行为的识别准确率不太理想。

针对目前Shift-GCN在日常行为识别准确率不足的问题,本文提出一种基于密集残差移位图卷积网络(DRS-GCN)的人体骨架行为识别算法,主要工作如下:

1)改进Shift-GCN。在时空移位图卷积模块中引入密集残差连接来捕获时域和空域范围上的依赖关系,充分利用骨架数据的时空特征,加强深层特征和浅层特征的联系,提高行为识别的准确率。在自制数据集上的实验结果表明,所提算法能有效鉴别行走、坐下、站起、脱衣服、穿衣服、扔和摔倒等7种日常行为。

2)多流骨架数据。提取原始骨架数据的关节点信息和骨骼长度信息,通过连续两帧间骨架序列的差异计算出人体行为的运动信息;然后将运动信息分别与骨架的关节点和骨骼长度结合形成多流数据,分别输入改进后的模型中进行训练;最后将多流网络的结果使用加权平均的方式融合,构建最终的行为识别系统。

2　基本原理

首先使用姿态估计算法提取每一帧动作视频的人体骨架信息,接着通过关节坐标向量计算得到4种骨架数据:关节(J)信息、骨骼(B)信息、关节-运动(J-M)信息以及骨骼-运动(B-M)信息。将4种数据信息流分别输入DRS-GCN中进行训练,最后输出4个流的Softmax分数,使用加权求和的方式融合,获得最终得分并预测动作。系统框架如图1所示。

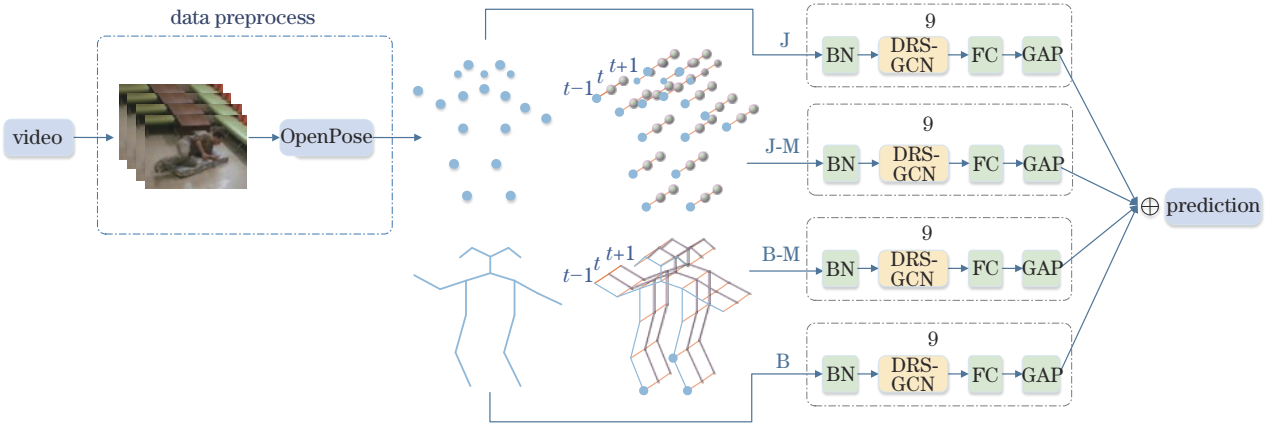

图1　系统框图

Fig. 1　Diagram of system block

2.1　密集残差移位图卷积

利用图卷积神经网络[14]将人体的骨架序列建模为时空图:1)以关节点为顶点,骨骼长度为边,将每一帧人体的物理连接建模为空间图;2)为相邻两帧骨架图之间对应的关节添加时间边,建模为时间图。对于T帧的人体行为视频,每一帧动作均对应骨架序列中的每一幅骨架图。在具有N个关节点的骨架图中,每个关节点可表示为位置坐标向量$\boldsymbol{x}=(x,y,z)$,则N个关节的特征向量集合$\boldsymbol{V}=\{\boldsymbol{x}_i|i=1,2,\cdots,N\}$,节点间的连接关系使用邻接矩阵表示为$\boldsymbol{M}\in\{0,1\}^{N\times N}$。图卷积将空间图表示为$\boldsymbol{G}=(\boldsymbol{V},\boldsymbol{M})$,由于视频中每一帧骨架图邻接矩阵均相同,因此$T$帧的人体行为骨架序列可表示为$\boldsymbol{S}=\{\boldsymbol{G}_t|t=1,2,\cdots,T\}$。

移位图卷积使用轻量级移位卷积操作替代常规图卷积,结合空间移位操作与点卷积,将相邻节点的信息转移到当前节点,同时融合了空间维度和通道维度的信息。空间上采用全局移位卷积操作:设单帧骨架序列特征为$\boldsymbol{F}\in\mathbf{R}^{N\times C}$,其中,$N$为关节点数量、$C$为通道大小,移位之后的特征表示为$\tilde{\boldsymbol{F}}\in\mathbf{R}^{N\times C}$,如图2所示,对$\boldsymbol{F}$的每个节点进行移位操作。以图2中3个节点为

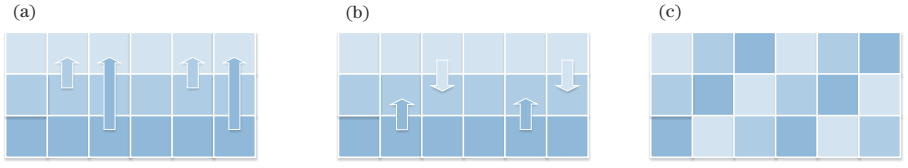

图2　移位卷积操作。(a)节点1移位;(b)节点2移位;(c)移位后的特征图

Fig. 2　Shift convolution operation. (a) Node 1 shift; (b) node 2 shift; (c) feature map after shift

例，图 2(a)是节点 1 的移位操作，图 2(b)是节点 2 的移位操作，3 个节点循环进行移位操作后，交换通道信息得到如图 2(c)所示的特征图。由于人体骨架中每个关节的重要性不同，引入注意力机制 $\tilde{F}_M=\tilde{F}\cdot[\tanh(M)+1]$学习节点之间的关系。时间上采用自适应移位卷积操作：T帧的骨架图形成的时空图特征为 $\tilde{\boldsymbol{F}}\in\mathbf{R}^{N\times C\times T}$，其中，$\boldsymbol{F}=\{\boldsymbol{F}_1,\boldsymbol{F}_2,\boldsymbol{F}_3,\cdots,\boldsymbol{F}_T\}$。每个通道都有一个可学习的参数 $S=\{S_i|i=1,2,\cdots,C\}$，将时移参数从整数约束放宽到实数，非整数位移 $\tilde{F}_{(v,t,i)}$ 使用线性插值法计算得到：

$$\tilde{F}_{(v,t,i)}=(1-\lambda)\cdot F_{(v,\lfloor t+S_i\rfloor,i)}+\lambda\cdot F_{(v,\lfloor t+S_i\rfloor+1,i)},\quad(1)$$

式中：$\lambda=S_i-\lfloor S_i\rfloor$，使用插值的手段弥补整数实数化之后产生的余量。

结合残差网络[15]易于优化和密集连接网络[16]特征复用的特点，提出 DRS-GCN。将 DRS-GCN 中的每一个空间移位卷积模块(S-block)和时间移位卷积模块(T-block)使用密集残差连接方式连接，网络结构如图 3 所示。输入的数据先经过归一化(BN)层，然后通过 9 个 DRS-GCN 模块(下面的 3 个数字分别表示每一层的输入通道数、输出通道数和步幅大小)，接着再输入全局平均池化层(GAP)层，最后通过全连接层(FC)分类。

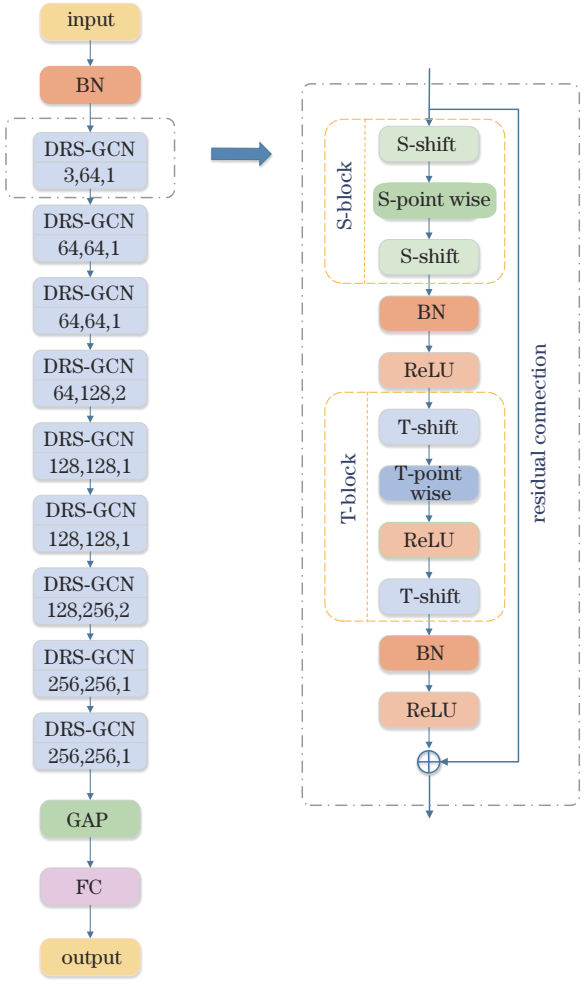

图 3　DRS-GCN 框图

Fig. 3　Diagram of DRS-GCN block

1)残差连接：Shift-GCN 算法采用残差网络中的残差连接方式，残差网络由多个残差块组合而成，每一个残差块可以表示为

$$\boldsymbol{h}_{l+1}=\boldsymbol{h}_l+F(\boldsymbol{h}_l,W),\quad(2)$$

式中：$\boldsymbol{h}_l$为第 l 个残差块的输入；$\boldsymbol{h}_{l+1}$为第 l 个残差块的输出(即 $l+1$ 个残差块的输入)；F 是非线性变换函数；W 是第 l 个残差块的参数。图 4(a)为原始残差块。随着网络层数的加深，残差连接可以有效解决梯度消失等问题，从而保证信息流从浅层顺利传输到较深层，提高网络性能。

2)密集连接：残差连接将输入和输出直接相加，而密集网络采用的密集连接方式是第 l 层将前面的所有层作为输入，在通道维度上进行连接。图 4(b)为密集块，复合函数 H_l 的输出 $\boldsymbol{h}_l$ 表示为

$$\boldsymbol{h}_l=H_l[\boldsymbol{h}_0,\boldsymbol{h}_1,\cdots,\boldsymbol{h}_l],\quad(3)$$

式中：$[\boldsymbol{h}_0,\boldsymbol{h}_1,\cdots,\boldsymbol{h}_l]$表示前面所有特征图的连接；$H_l$是归一化-ReLU-卷积操作的复合函数。由于密集网络有效传递特征和梯度，更充分地利用网络参数，其性能明显优于其他模型。

3)密集残差连接：结合残差块易于优化和密集连接高效利用特征的优点，提出一种新的小型密集残差结构(Dense-Res block)，如图 4(c)所示。新型密集残差结构可以加速网络训练、提高模型的性能，该结构可以表示为

$$\boldsymbol{h}_{l+1}=\boldsymbol{h}_l+F\{H_d[\boldsymbol{h}_0,\cdots,\boldsymbol{h}_d],[W_0,\cdots,W_d]\},\quad(4)$$

式中：$\boldsymbol{h}_l$和$\boldsymbol{h}_{l+1}$分别为网络中第 l 个密集连接残差块的输入和输出；$[\boldsymbol{h}_0,\cdots,\boldsymbol{h}_d]$是第 l 块中从第 0 个到第 d 个卷积层特征映射的连接；F 是非线性变换函数；$[W_0,\cdots,W_d]$是第 l 块中所有的参数。

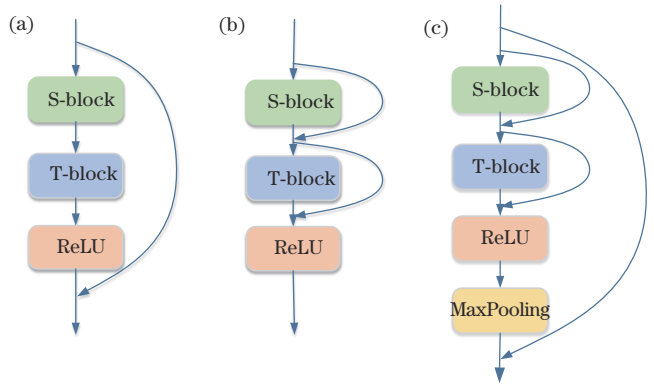

图 4　密集残差网络结构图。(a)残差连接；(b)密集连接；(c)密集残差连接

Fig. 4　Diagrams of dense residual network structure. (a) Residual connection; (b) dense connection; (c) dense residual connection

密集残差连接将复杂的浅层特征输入求和层来实现浅层特征的复用，同时也能够避免因一些层的输入信息和梯度信息消失而造成信息堵塞的问题。在时域图卷积模块里加入最大池化层，池化后帧数变少，相邻几帧的数据通过池化层聚合到同一帧，下一层网络时间上感受野增大，计算量也相对减小。

2.2　多流数据

首先将骨架序列的4种骨架数据流（关节、骨骼、关节-运动和骨骼-运动）输入DRS-GCN中训练；其次通过Softmax分类器分类；最后融合4个流的分数来预测动作标签。

1）关节数据流：通过姿态估计算法提取关节点的位置坐标信息。

2）骨骼（骨长）信息流：定义靠近人体重心的点为源关节点，其坐标定义为 $V_{i,t}=(x_{i,t},y_{i,t},z_{i,t})$；远离重心的点为目标关节点，其坐标为 $V_{j,t}=(x_{j,t},y_{j,t},z_{j,t})$：则骨长信息流 $L_{i,j,t}$ 可以表示为源关节点与目标关节点的差值。

$$L_{i,j,t}=V_{j,t}-V_{i,t}=(x_{j,t}-x_{i,t},y_{j,t}-y_{i,t},z_{j,t}-z_{i,t})。\tag{5}$$

在骨架特征图中，每个骨骼数据没有循环，因此为每个骨骼指定唯一的目标关节。由于根关节点没有分配骨骼，为了匹配关节数据与骨骼数据，为根关节分配一个值为0的空骨骼。

3）关节-运动信息流：关节-运动信息是连续两帧中相同关节的差异，如图5所示，定义在 t 帧上的关节点 i 的坐标为 $V_{i,t}=(x_{i,t},y_{i,t},z_{i,t})$，在 $t+1$ 帧上的关节点 i 则定义为 $V_{i,t+1}=(x_{i,t+1},y_{i,t+1},z_{i,t+1})$，运动信息流 $M_{i,t,t+1}$ 可表示为

$$M_{i,t,t+1}=V_{i,t+1}-V_{i,t}=(x_{i,t+1}-x_{i,t},\ y_{i,t+1}-y_{i,t},z_{i,t+1}-z_{i,t})\quad。\tag{6}$$

4）骨骼-运动信息流：骨骼-运动信息是连续两帧中相同骨骼的差异。定义在 t 帧上的骨骼长度为 k，其坐标为 $L_{k,t}=(x_{k,t},y_{k,t},z_{k,t})$，在 $t+1$ 帧上的骨骼 t 则定义为 $L_{k,t+1}=(x_{k,t+1},y_{k,t+1},z_{k,t+1})$，运动信息流 $M_{k,t,t+1}$ 可表示为

$$M_{k,t,t+1}=L_{k,t+1}-L_{k,t}=(x_{k,t+1}-x_{k,t},\ y_{k,t+1}-y_{k,t},z_{k,t+1}-z_{k,t})\quad。\tag{7}$$

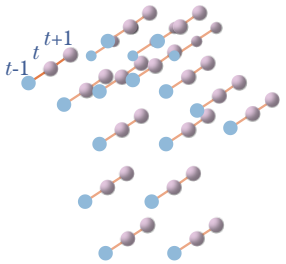

图5　关节-运动图

Fig. 5　Diagram of joint-motion

4种骨架数据分别输入网络中训练，利用Softmax作为激活函数分类后得到预测的分数 P_{J}、P_{B}、P_{JM}、P_{BM}，4个流的分数采用加权规则 $P=W_{\mathrm{J}}P_{\mathrm{J}}+W_{\mathrm{B}}P_{\mathrm{B}}+W_{\mathrm{JM}}P_{\mathrm{JM}}+W_{\mathrm{BM}}P_{\mathrm{BM}}$ 得到最终的预测结果，其中，W_{J}、W_{B}、W_{JM} 和 W_{BM} 分别是关节信息预测结果的加权系数、骨骼信息预测结果的加权系数、关节-运动信息预测结果的加权系数和骨骼-运动信息预测结果的加权系数。

3　实验过程

3.1　数据集预处理

人体日常行为数据集（DAILY）包括7种行为，其中，行走、坐下、站起、脱衣服、穿衣服以及扔取自公开行为识别数据集Northwest UCLA，每一个动作由10名受试者执行1～6次。摔倒行为数据取自公开数据集Le2i，包含home、office、coffee room、lecture room等

图6　DAILY数据集示例。(a)行走；(b)扔；(c)摔倒

Fig. 6　Examples of DAILY dataset. (a) Walking; (b) throwing; (c) falling

4种场景视频。DAILY数据集共有1029个视频序列，根据深度学习的划分策略，选取70% 为训练集，30% 为测试集。视频帧率为30 frame/s，分辨率为320×240。图6为日常行为数据集的样本案例，所示图像分别是行走、扔和摔倒这3类行为视频的分解帧。

使用OpenPose工具箱[17]提取数据集每一帧（每个视频最多取300帧）骨架关节点（18个）的位置特征信息（2D坐标）。图7是OpenPose的骨骼标注，标有序号的圆形节点表示骨骼关节（如序号4表示左手）。OpenPose提取的骨架数据如图8所示，图中分别为图6中3类行为数据对应的骨架图。

NTU60 RGB+D数据集[18]出自新加坡南洋理工大学，由3个Microsoft Kinect v2相机同时捕获。该数据集采集的关节点数为25，相机摆放位置组合有17个，共由56880个动作片段组成，包含有40名演员执行的60个动作分类。采用该数据集的两种评判标准来进行算法评价：1）cross subject（CS），CS表示训练集和验证集中的行为来自不同的演员；2）cross view（CV），CV表示训练集和验证集中的行为来自不同的摄像机。

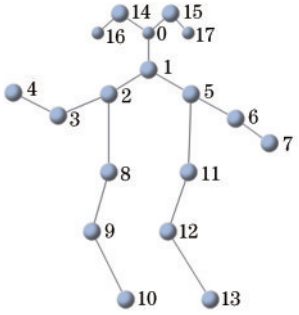

图7　OpenPose中的骨骼标注

Fig. 7　Node labeling of OpenPose

图8　骨架图示例。(a)行走；(b)扔；(c)摔倒

Fig. 8　Examples of skeleton diagram. (a) Walking; (b) throwing; (c) falling

3.2　实验结果与分析

在以上两种数据集上基于PyTorch深度学习框架进行实验，网络模型均采用随机梯度下降（SGD）的优化策略。将交叉熵作为梯度反向传播的损失函数，权重衰减系数设为1×10^{-4}，训练批次大小（batch size）设置为32。

为了验证密集残差图卷积网络的有效性，在日常行为数据集上采用3个指标进行评估，分别为召回率（recall）、精确率（precision）、准确率（accuracy），定义如下：

$$R_{\text{recall}}=\frac{N_{\text{TP}}}{N_{\text{TP}}+N_{\text{FN}}}\times100\%,\tag{8}$$

$$R_{\text{precision}}=\frac{N_{\text{TP}}}{N_{\text{TP}}+N_{\text{FP}}}\times100\%,\tag{9}$$

$$R_{\text{accuracy}}=\frac{N_{\text{TP}}+N_{\text{TN}}}{N_{\text{TP}}+N_{\text{TN}}+N_{\text{FN}}+N_{\text{FP}}}\times100\%,\tag{10}$$

式中：N_{TP}表示正例预测为正例的数量；N_{FP}表示反例预测为正例的数量；N_{TN}表示反例预测为反例的数量；N_{FN}表示正例预测为反例的数量。

在DAILY数据集上进行实验对比仅反映出DRS-GCN算法在单一数据集上的性能表现。为了验证算法的泛化性能，在大型公共数据集NTU60 RGB+D上进行了多流融合和引入密集残差模块的对比实验。

将单流网络和多流（关节、骨骼、关节-运动、骨骼-运动数据）融合网络（4s-DRS-GCN）进行评估比较。表1是在NTU60 RGB+D数据集上多流融合网络的对比结果，可以看出，4s-DRS-GCN在CS和CV上的准确率分别达90.8% 和96.3%，相较于单流网络中仅使用关节数据的DRS-GCN(J)分别提升2.7个百分点和1个百分点。由此可以看出，骨架的高阶信息在行为识别任务中的重要性。

为了验证引入密集残差模块的有效性，将DRS-GCN算法与Shift-GCN算法在DAILY数据集上进行

表 1　NTU60 RGB+D 数据集上多流网络的性能研究

Table 1　Research on performance of multi-stream network on NTU60 RGB+D dataset

Method	CS /%	CV /%
DRS-GCN(J)	88.1	95.3
DRS-GCN(B)	88.9	94.8
DRS-GCN(J-M)	86.8	93.6
DRS-GCN(B-M)	87.0	93.7
4s-DRS-GCN	90.8	96.3

召回率、精确率和准确率这 3 个指标的对比分析。图 9 是识别 7 种日常行为的混淆矩阵，横坐标是预测的标签值，纵坐标是真实标签值。由图 9 可见，整体上网络模型可较好区分 7 个动作，其中，行走与扔最容易混淆。表 2 是 DRS-GCN 与原网络在 DAILY 数据集上的实验数据对比，可以看出，DRS-GCN 在走路、坐下、站起、脱衣服、穿衣服、扔和摔倒这 7 类运动行为下的精确率分别达到了 73.1%、84.4%、86.7%、90.7%、81.4%、61.4%、97.4%，相较于原算法分别提高了

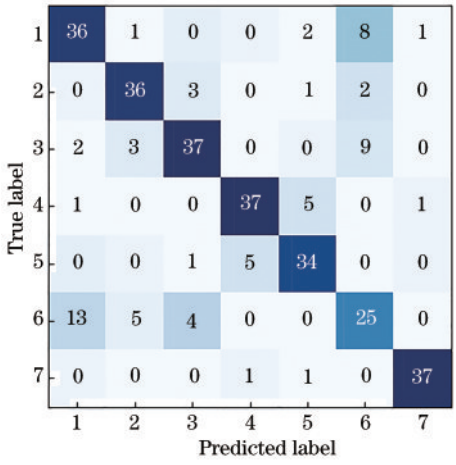

图 9　多类别混淆矩阵

Fig. 9　Confusion matrix of multi-class

表 2　DAILY 数据集上 7 种行为的实验数据对比

Table 2　Experimental data comparison of 7 behaviors on DAILY dataset

Class	Recall /%		Precision /%	
	DRS-GCN	Shift-GCN	DRS-GCN	Shift-GCN
walking	73.1	75.0	73.1	69.2
sitting	90.5	85.7	84.4	80.0
standing	84.8	72.6	86.7	82.2
donning	88.6	84.1	90.7	86.1
doffing	83.3	85.0	81.4	79.1
throwing	60.0	53.2	61.4	56.8
falling	95.0	94.9	97.4	94.9

3.9 个百分点、4.4 个百分点、4.5 个百分点、4.6 个百分点、2.3 个百分点、4.6 个百分点、2.5 个百分点。行走和扔这 2 类动作识别率较低，是由于视频数据集中扔的动作包含行走(即边走边扔)，很容易产生误判。

图 10 展示了两种模型的训练和测试过程，准确率的变化曲线如图 10(a)所示，损失值的变化曲线如图 10(b)所示。根据训练过程中迭代次数(epoch)和损失值(loss)的变化关系设置学习率：实验共 100 个 epoch，初始学习率设为 0.1，在迭代次数为 40、60 和 80 时分别除以 10。从图 10 的实验结果可以看出：网络在迭代 60 个 epoch 后均表现出较好的收敛性，DRS-GCN 中所设计的密集残差结构能够提升网络的时空特征提取能力，网络参数更易优化。为了更直观地比较两种网络的性能，表 3 给出了 DRS-GCN 和原网络在 DAILY 数据集上关于召回率、精确率和准确率这 3 个性能指标的实验对比，DRS-GCN 的准确率提高了 3.9 个百分点，召回率和精确率也分别提高了 3.6 个百分点和 3.9 个百分点。因此 DRS-GCN 在各方面都有明显提升。

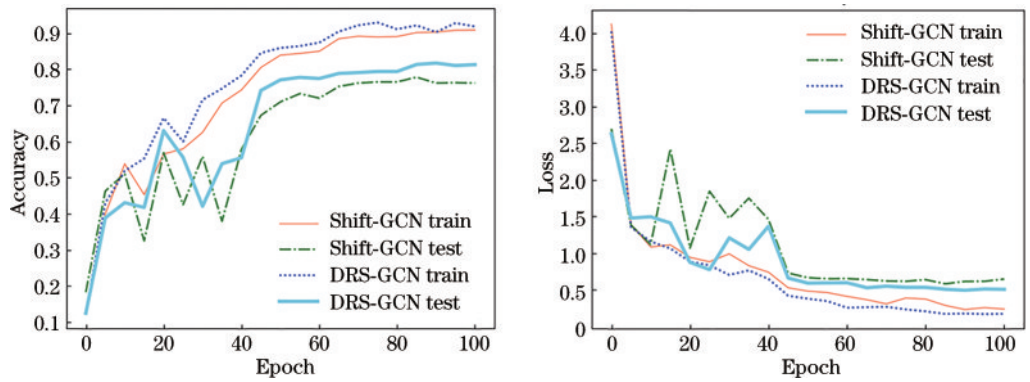

图 10　DAILY 数据集的训练准确率和损失值变化曲线。(a)准确率；(b)损失值

Fig. 10　Curves of training accuracy and loss value on DAILY dataset. (a) Accuracy; (b) loss value

DRS-GCN 与原网络在 NTU60 RGB+D 数据集(CS/CV)上随迭代次数变化的准确率变化图和损失值变化图分别如图 11 和图 12 所示，根据训练过程中迭代次数和损失值的变化关系设置学习率：实验共 120 个 epoch，初始学习率设为 0.1，在迭代次数为 60、80 和 100 时除以 10。从图 11、12 可以看出，DRS-GCN

表3　DAILY数据集上3个指标对比分析

Table 3　Comparative analysis of three indicators on DAILY dataset

Method	Recall /%	Precision /%	Accuracy /%
Shift-GCN[11]	78.6	78.3	77.8
DRS-GCN	82.2	82.2	81.7

将视频中获取的浅层特征和深层特征进行充分融合，提高了网络参数的利用率，收敛过程也更加稳定。

将DRS-GCN算法与行为识别领域中当前先进的算法进行比较，表4是这几种算法在NTU60 RGB+D数据集上的实验对比分析。从表4可知：在CS和CV下DRS-GCN的准确率分别为88.1%和95.3%，浮点运算数(FLOPs)为3.7；与基于循环神经网络的方法VA-LSTM[19]和基于卷积神经网络的方法HCN[20]相比，DRS-GCN在CS(CV)下识别准确率分别提高了8.9%(7.6%)和1.6%(4.2%)；与其他几种基于图卷积神经网络的方法相比，DRS-GCN在识别精度和浮点运算数上的综合表现也更为优越。实验结果表明，所提算法可以在兼顾计算复杂度的同时提高实验精度。

所提算法在自制数据集和公共数据集上的表现，均证明了模型的有效性和泛化性。

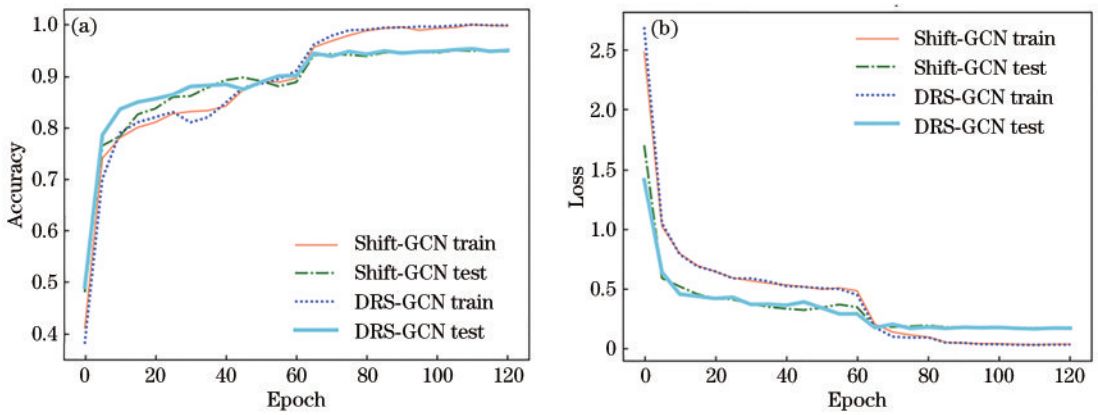

图11　NTU60 RGB+D数据集(CS)的训练准确率和损失值变化曲线。(a) CS准确率；(b) CS损失值

Fig. 11　Curves of training accuracy and loss value on NTU60 RGB+D dataset (CS). (a) Accuracy of CS; (b) loss value of CS

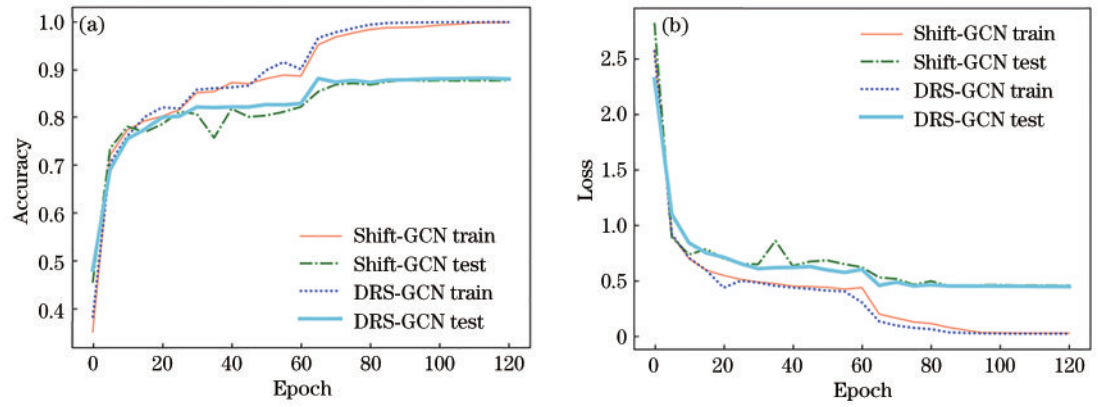

图12　NTU60 RGB+D数据集(CV)的训练准确率和损失值变化曲线。(a) CV准确率；(b) CV损失值

Fig. 12　Curves of training accuracy and loss value on NTU60 RGB+D dataset. (CV). (a) Accuracy of CV; (b) loss value of CV

表4　NTU60 RGB+D数据集上DRS-GCN与先进算法的实验数据对比

Table 4　Comparison of experimental data between DRS-GCN and advanced algorithms on NTU60 RGB+D dataset

Method	Accuracy /%		FLOPs /10^9
	CS	CV	
VA-LSTM[19]	79.2	87.7	
ST-GCN[9]	81.5	88.3	
HCN[20]	86.5	91.1	
AS-GCN[10]	86.8	94.2	27.0
2s-AGCN[12]	88.5	95.1	35.8
Shift-GCN[13]	87.8	95.1	2.5
DRS-GCN	88.1	95.3	3.7
4s-DRS-GCN	90.8	96.3	14.8

4 结 论

提出一种将密集残差连接引入移位图卷积网络中的方案。首先，使用OpenPose姿态估计算法从视频中提取人体骨骼数据；其次，将最新的移位图卷积行为识别算法作为基础网络，在图卷积网络中加入密集残差块来有效提取时空特征；最后，将多种骨架信息流输入改进后的行为识别算法系统中进行融合，在兼顾网络计算量的情况下进一步提高性能。实验结果表明，该算法相较于其他算法有更为出色的表现。但是，在提高精度的同时仍然会对计算量有少量的影响，在未来的工作中将进一步研究如何以更小的计算量实现更高识别精度。

参 考 文 献

[1] 刘宗达，董立泉，赵跃进，等. 视频中快速运动目标的自适应模型跟踪算法[J]. 光学学报，2021，41(18)：1815001.
Liu Z D，Dong L Q，Zhao Y J，et al. Adaptive model tracking algorithm for fast-moving targets in video[J]. Acta Optica Sinica，2021，41(18)：1815001.

[2] 邹梓吟，盖绍彦，达飞鹏，等. 基于注意力机制的遮挡行人检测算法[J]. 光学学报，2021，41(15)：1515001.
Zou Z Y，Gai S Y，Da F P，et al. Occluded pedestrian detection algorithm based on attention mechanism[J]. Acta Optica Sinica，2021，41(15)：1515001.

[3] Sawada J，Kusumoto K，Munakata T，et al. A mobile robot for inspection of power transmission lines[J]. IEEE Power Engineering Review，1991，11(1)：57.

[4] 张文强，王增强，张良. 结合时序动态图和双流卷积网络的人体行为识别[J]. 激光与光电子学进展，2021，58(2)：0210007.
Zhang W Q，Wang Z Q，Zhang L. Human action recognition combining sequential dynamic images and two-stream convolutional network[J]. Laser & Optoelectronics Progress，2021，58(2)：0210007.

[5] Liu H，Tu J H，Liu M Y. Two-stream 3D convolutional neural network for skeleton-based action recognition[EB/OL]. (2017-03-23)[2022-02-01]. https://arxiv. org/abs/1705.08106.

[6] 朱铭康，卢先领. 基于Bi-LSTM-Attention模型的人体行为识别算法[J]. 激光与光电子学进展，2019，56(15)：151503.
Zhu M K，Lu X L. Human action recognition algorithm based on Bi-LSTM-Attention model[J]. Laser & Optoelectronics Progress，2019，56(15)：151503.

[7] Liu J，Shahroudy A，Xu D，et al. Spatio-temporal LSTM with trust gates for 3D human action recognition[M]//Leibe B，Matas J，Sebe N，et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham：Springer，2016，9907：816-833.

[8] Dong X W，Thanou D，Rabbat M，et al. Learning graphs from data：a signal representation perspective[J]. IEEE Signal Processing Magazine，2019，36(3)：44-63.

[9] Yan S J，Xiong Y J，Lin D H. Spatial temporal graph convolutional networks for skeleton-based action recognition[EB/OL]. (2018-01-23)[2022-02-03]. https://arxiv.org/abs/1801.07455.

[10] Li M S，Chen S H，Chen X，et al. Actional-structural graph convolutional networks for skeleton-based action recognition[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)，June 15-20，2019，Long Beach，CA，USA. New York：IEEE Press，2019：3590-3598.

[11] Thakkar K C，Narayanan P J. Part-based graph convolutional network for action recognition[EB/OL]. (2018-09-13)[2022-02-03]. https://arxiv.org/abs/1809.04983.

[12] Shi L，Zhang Y F，Cheng J，et al. Two-stream adaptive graph convolutional networks for skeleton-based action recognition[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)，June 15-20，2019，Long Beach，CA，USA. New York：IEEE Press，2019：12018-12027.

[13] Cheng K，Zhang Y F，He X Y，et al. Skeleton-based action recognition with shift graph convolutional network[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)，June 13-19，2020，Seattle，WA，USA. New York：IEEE Press，2020：180-189.

[14] Henaff M，Bruna J，LeCun Y. Deep convolutional networks on graph structured data[EB/OL]. (2015-06-16)[2022-02-01]. https://arxiv.org/abs/1506.05163.

[15] Hara K，Kataoka H，Satoh Y. Learning spatio-temporal features with 3D residual networks for action recognition[C]//2017 IEEE International Conference on Computer Vision Workshops，October 22-29，2017，Venice，Italy. New York：IEEE Press，2017：3154-3160.

[16] Huang G，Liu Z，van der Maaten L，et al. Densely connected convolutional networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition，July 21-26，2017，Honolulu，HI，USA. New York：IEEE Press，2017：2261-2269.

[17] Cao Z，Simon T，Wei S H，et al. Realtime multi-person 2D pose estimation using part affinity fields[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition，July 21-26，2017，Honolulu，HI，USA. New York：IEEE Press，2017：1302-1310.

[18] Shahroudy A，Liu J，Ng T T，et al. NTU RGB D：a large scale dataset for 3D human activity analysis[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition，June 27-30，2016，Las Vegas，NV，USA. New York：IEEE Press，2016：1010-1019.

[19] Zhang P F，Lan C L，Xing J L，et al. View adaptive recurrent neural networks for high performance human action recognition from skeleton data[C]//2017 IEEE International Conference on Computer Vision，October 22-29，2017，Venice，Italy. New York：IEEE Press，2017：2136-2145.

[20] Li C，Zhong Q Y，Xie D，et al. Co-occurrence feature learning from skeleton data for action recognition and detection with hierarchical aggregation[C]//IJCAI′18：Proceedings of the 27th International Joint Conference on Artificial Intelligence，July 13-19，2018，Stockholm，Sweden. New York：ACM Press，2018：786-792.