DOI:10.19651/j.cnki.emt.2415570

一种基于特征位移的手势识别方法*

刘 翔[1,2] 刘新妹[1,2] 李传坤[1,2] 张晋钊[1,2]
(1.中北大学信息探测与处理山西省重点实验室 太原 030051;
2.中北大学省部共建动态测试技术国家重点实验室 太原 030051)

摘 要:针对卷积操作受到遍历规则的限制,只能提取单个骨骼节点的特征信息,不能对相邻节点之间的有效特征信息进行融合,导致表达能力有限的问题,提出了一种基于特征位移模块的手势识别神经网络。该网络采用常规时空图卷积神经网络的架构,并将常规时空卷积模块替换为特征位移模块,实现相邻节点特征信息之间的融合。利用特征位移模块对位移信道进行重新排序,实现提取骨骼节点的全局化特征信息,进一步完成对手势信息的高效准确分类。并在公开数据集 DHG-14/28 和 FPHA 上验证该特征位移模块,在 14 类、28 类和 FPHA 手势数据集的分类准确度分别达到了 95.11%、93.01%和 92.67%。实验结果表明,该网络模型能够更好更有效的挖掘全局特征信息,在常见的手势识别数据集上达到了优秀的性能。

关键词:手势识别;卷积神经网络;特征位移;图卷积神经网络;深度学习
中图分类号:TP391;TN99 **文献标识码**:A **国家标准学科分类代码**:510.4040

A gesture recognition method based on feature displacement

Liu Xiang[1,2] Liu Xinmei[1,2] Li Chuankun[1,2] Zhang Jinzhao[1,2]
(1. School of Information and Communication Engineering, North University of China, Taiyuan 030051, China;
2. State Key Laboratory of Dynamic Testing Technology, North University of China, Taiyuan 030051, China)

Abstract: The convolutional operation is constrained by traversal rules, limiting the extraction of feature information from individual skeletal nodes and preventing effective fusion of feature information between adjacent nodes, resulting in limited expressive power. In response to this issue, a gesture recognition neural network based on a Feature Displacement Module is proposed. This network adopts the architecture of conventional spatiotemporal graph convolutional neural networks and replaces the conventional spatiotemporal convolution module with the Feature Displacement Module to achieve fusion of feature information between adjacent nodes. By reordering the displacement channels through the Feature Displacement Module, global feature information of skeletal nodes is extracted, further enabling efficient and accurate classification of gesture information. The Feature Displacement Module is validated on the public dataset DHG-14/28 and FPHA, achieving classification accuracies of 95.11%, 93.01% and 92.67% for 14-class, 28-class and FPHA gesture datasets. The experimental results demonstrate that this network model can better and more effectively mine global feature information, achieving excellent performance on common gesture recognition datasets.

Keywords: gesture recognition; convolutional neural network; feature shift; graph convolutional networks; deep learning

0 引 言

非接触的手势识别是人工智能和计算机视觉领域的研究热点,在人机交互中有着大量的应用[1-3]。利用计算机分析和识别手势信息,使人机交互近似于人与人之间的沟通交流,让使用者有更加自然舒适的实际体验。如今,基于身体和手部骨骼的深度传感器[4-5](如英特尔 RealSense 和微软 Kinect)已经变得易于获取。因此,利用姿态或骨骼的手势识别越来越受到人们的关注[6-10]。手势识别可以分为基于彩色(RGB)图像[11-13]和基于手部骨骼[7-8]的两大类方法。前者通过提取图像特征进行分类识别,易受到环境噪声(如背景颜色、光线亮度等)的干扰[14];后者采用手骨骼数据作

收稿日期:2024-03-03
*基金项目:山西省回国留学人员科研项目(2017-090)、山西省重点研发项目(201903D121058)资助

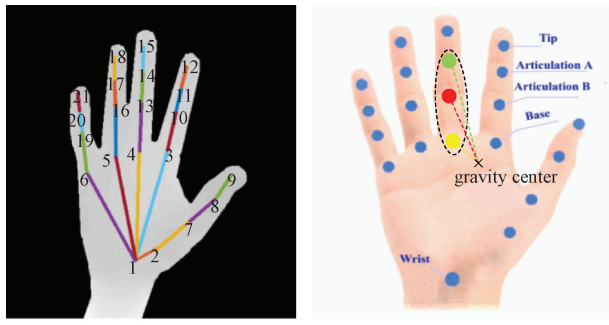
为输入，提取骨骼特征作为分类依据，能够有效避免噪声干扰，具有鲁棒性并且计算成本更低。

随着图像识别算法更新迭代，利用卷积神经网络强大的特征提取能力提取图像中铰链特征，从时空域特征推断手势的识别准确率有了很大的提升。首先是基于传统卷积神经网络（convolutional neural network，CNN）网络的方法率先提出。Devineau 等[9]采用并行 CNN 骨骼关节的三维（3 dimensions，3D）坐标序列进行时间卷积处理，并将每个关节的时域特征融合成一个全连接层和 Softmax 函数进行手势分类。Narayana 等[10]利用空间注意力焦点构建了 12 流网络，每个网络从不同的模态中提取局部特征。然后设计了一个稀疏网络架构来融合 12 个信道。Molchanov 等[15]使用循环三维传统卷积神经网络（3D convolutional neural network，3DCNN）进行手势识别。在循环 3DCNN 中，使用时间分类对视频进行同步分割，并从多模态数据中对动态手势进行分类。然而，这些基于 CNN 的方法参数量大、训练时间长，而且不能有效地表达关节之间的依赖关系。因为手部关节的运动轨迹并不规则，是分布在非欧几里得域中的。为了获得非欧几里得图的结构性质，图卷积网络（graph convolutional network，GCN）被提了出来，用图来表达关节之间的依赖关系。Li 等[16]采用基于手势的图卷积网络（hand gesture based graph convolutional network，HGCN）来捕获不同关节的连接和运动信息，该网络设置了 4 种类型的边来捕获非相邻节点之间的关系，但是缺乏不同关节间的重要性体现。Nguyen 等[17]利用神经网络利用正定矩阵从手部关节中挖掘手势表示。Yan 等[18]提出了时空图卷积网络（spatial-temporal graph convolutional network，ST-GCN）来建模人体关节间的时空相关性，在 GCN 中预先定义了一个基于人体结构的静态拓扑，并在连续时间帧中添加关节之间联系。然而，这种静态拓扑在推理过程中是固定的，严重限制了 GCN 的表达能力。许多研究提出通过设计动态拓扑图来改进 ST-GCN，例如，Li 等[19]设计了一种编码器-解码器结构来捕获关节的相关性，Liu 等[20]使用自注意力机制动态生成拓扑关系。此外，Chen 等[21]通过动态学习不同的拓扑图并聚合不同通道中的联合特征来进行识别，其中 GCN 模型被强制聚合具有不同动态拓扑的不同通道中的骨骼特征。上述方法网络参数多，计算量大，重点关注的是局部关节的联系，忽略了关节之间的相关性。另一部分研究则通过结合注意力机制，使用多类型数据等手段提高模型的性能。例如，Shi 等[22]采用非局部网络计算不同关节的依赖关系生成自适应邻接矩阵，并构建双流网络实现不同模态信息的互补。Liang 等[23]利用面部遮罩生成网络来指导手势识别网络的特征提取和聚合过程。Li 等[24]利用自注意机制增强空间特征的学习能力与残差双向独立递归神经网络（residual-connection enhanced bidirectional independently recurrent neural network，RBi-IndRNN）的长时间特征学习能力，对给定相应特征的两个节点之间的相关性进行建模，以捕获不同的手势的关节特征。胡宗承等[25]对深度信息和三维骨骼信息分别提取空间和时间特征判断手势类别，一定程度上提高了识别准确率。上述网络需要对双流网络进行联合评估，模型复杂，并且增加了计算时间，而且空间图的感受域不够灵活，感受野受到一定的限制，不能充分挖掘手势的空域结构特征。例如，在“写”手势中，拇指尖和食指尖之间的联系可能很强，但在“戳”和“敲”手势中就不是这样了。关节间的联系对于精准的手势识别尤为重要。

针对上述问题，本文将位移卷积引入空间卷积，提出一种空间特征移位卷积模块，将当前节点的特征与相邻节点的特征相融合，实现了特征信息在空间维度和通道维度上的融合，使每个节点的接受域覆盖整个手部骨骼，为空间卷积提供了灵活的接受域。

1　理论与方法

手进行动作时，将手指看成以关节为核心的运动群体，5 个手指运动群体联合运动形成手势动作。因此，将关节作为节点，以手部关节的物理连接随时间运动轨迹为边，构建时空图。如图 1(a)所示，点表示关节点，线表示手骨骼的物理连接。将手骨关节的时间序列以 3D 的形式表示后，通过分析其运动模式来识别手势。



(a) FPHA节点　　(b) 空间分类策略

图 1　手势节点示意图

1.1　图卷积网络

如图 1(a)所示，对于一个给定的 T 帧视频，从每帧中提取出 N 个关节的手骨骼数据，将其形式化的表示为无向时空图 $\boldsymbol{G}=(\boldsymbol{V},\boldsymbol{E})$。设 $\boldsymbol{V}=\{v_{t,i}\mid t=1,\cdots,T;\ i=1,\cdots,N\}$ 表示骨骼数据中的节点集合，将节点特征向量 $\boldsymbol{F}(v_{t,i})$ 作为网络的输入。其中，$v_{t,i}$ 表示第 t 帧的第 i 个关节，特征向量由 $v_{t,i}$ 的坐标向量和预测置信度组成。用 $\boldsymbol{E}$ 表示时空图中的边，将 $\boldsymbol{E}$ 分为两个边集：1)内部边：第 τ 帧内的手部骨骼节点 i 和节点 j 的物理连接，记为 $\boldsymbol{E}_S(\tau)=\{v_{t,i}v_{t,j}\mid t=\tau,(i,j)\in H\}$，其中，$H$ 是手势自然连接的集合。2)帧间边：同一骨骼节点 i 在时间帧 τ 和 $\tau+1$ 之间的时序连接，记为 $\boldsymbol{E}_F=\{v_{t,i}v_{(t+1),i}\}$。通过内部边和帧间边将骨骼序列图中的所有节点连接起来，对于任意一个关节点 i，$\boldsymbol{E}_F$ 表示了它随时间运动的轨迹。接着，将骨骼节点

的相邻节点集合定义为邻接矩阵 $\boldsymbol{A}\in\{0,1\}^{N\times N}$。如果关节 i 和关节 j 连通,表示为 $\boldsymbol{A}_{i,j}=1$,不连通则表示为 $\boldsymbol{A}_{i,j}=0$。由于每个节点的邻接节点数量不同,如图 1(a)所示,节点 1 有 5 个邻接点,节点 2 有 2 个邻接点,节点 9 有 1 个邻接点,所以本文采用类似文献[18]的关节空间配置策略来表征骨骼数据的局部结构。如图 1(b)所示,首先通过计算每帧中所有手部关节的平均坐标来生成一个重心,其次将重心的位置作为参考点,将每帧的手部骨骼在空间上划分为 3 个子集:1) 根节点:执行卷积操作的节点;2) 向心集:比根节点更靠近手部关节重心的相邻节点;3)离心集:比根节点距离手部骨骼重心更远的相邻节点。借助以上邻接矩阵和分区策略,将手势动作的骨骼坐标表示为 $\boldsymbol{X}\in R^{N\times T\times d}$,其中,$N$ 表示关节点数,T 表示时间帧,d 表示关节坐标的维度。将每个节点上的卷积操作表示为:

$$\boldsymbol{F}'=\sum_{p\in \mathrm{P}}\bar{\boldsymbol{A}}_p\boldsymbol{F}\boldsymbol{W}_p \tag{1}$$

其中,$F'\in\boldsymbol{R}^{N\times C'}$ 表示输出特征矩阵,$\boldsymbol{F}\in\boldsymbol{R}^{N\times C}$ 表示输入特征矩阵,C 和 C'分别表示输入和输出的通道维数,$p\in$ {根节点,离心点,向心点},表示空间分区,$\bar{\boldsymbol{A}}_p$ 表示归一化邻接矩阵,$\boldsymbol{W}_p\in\boldsymbol{R}^{1\times1\times C\times C'}$ 表示每一个分区的 1×1 卷积核的权重参数。$\bar{\boldsymbol{A}}_p$ 可具体表示为:

$$\bar{\boldsymbol{A}}_p=\boldsymbol{\Lambda}_p^{-\frac{1}{2}}\boldsymbol{A}_p\boldsymbol{\Lambda}_p^{-\frac{1}{2}}\in R^{N\times N} \tag{2}$$

其中,$\boldsymbol{A}_p$ 为邻接矩阵,$\boldsymbol{R}^{N\times N}$ 为邻接节点集。

$$\boldsymbol{\Lambda}_p^{ii}=\sum_j(\boldsymbol{A}_p^{ij})+\alpha \tag{3}$$

其中,$\wedge_p^{ii}$ 表示度矩阵,$\boldsymbol{A}_p^{ij}$ 表示节点 i 和节点 j 的邻接矩阵,α 设置为 0.001,以避免 $\boldsymbol{A}_p$ 中的空行。

因为时间连接是通过连接相邻帧的骨骼节点来建立的,所以大多数基于图卷积网络的模型[17-20]使用一维卷积作为网络的时间卷积,卷积核表示为 k_t,通常设置为 9。

图卷积网络如图 2 所示,正则卷积核是 3 个点卷积核的融合,每个卷积核对指定空间分区进行卷积,空间分区由 3 个不同的邻接矩阵表示,分别表示“向心”、“根节点”、“离心”。

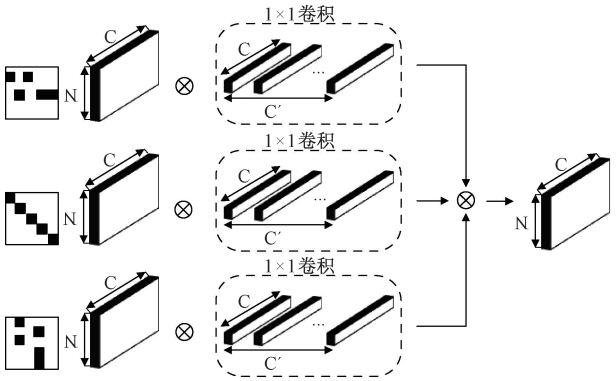
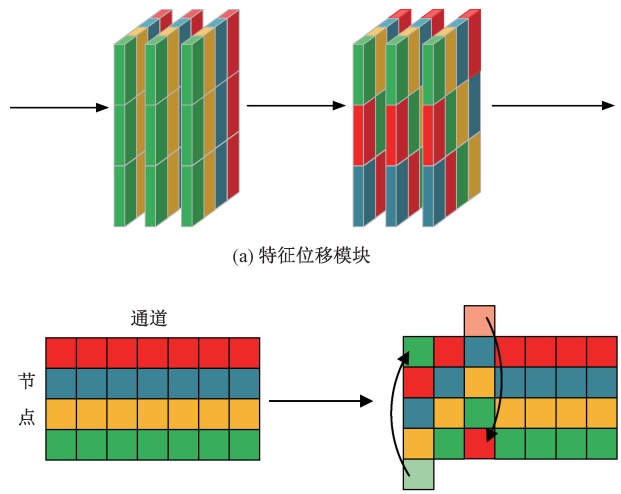


图 2　GCN 卷积

1.2　特征位移模块

本模块的主要作用对位移信道进行重新排序,将相邻节点的特征信息转移到当前卷积节点,实现相邻节点的特征融合,达到提取骨骼节点的全局化特征信息,如图 3(a)所示。位移卷积在 CNN 网络[26]中有效的代替了常规卷积,但是图像特征的排列是有序的,人体关节之间的连接时无序的,不同的骨骼节点有不同数量的邻居节点。因此,本文构建特征位移模块具体操作如下:



(a) 特征位移模块

(b) 位移示意图

图 3　特征位移示意图

位移卷积算子的数学形式可表达为:

$$\tilde{G}_{k,l,m}=\sum_{i,j}\tilde{F}_{i,j,m}F_{k+\wedge_{i},l+\wedge_{j},m} \tag{4}$$

其中,$\tilde{F}_{k,j,m}$ 是位移卷积核,(k,l)和(i,j)沿空间维度的索引 m 进入通道。每一个位移卷积核都是一个单独的算子,卷积核中只有元素(i_m,j_m)处为 1,其他全部为 0,可表示为:

$$\tilde{F}_{i,j,m}\begin{cases}1, & i=i_m,j=j_m\\ 0, & \text{其他}\end{cases} \tag{5}$$

其中,i_m, j_m 是通道的相关索引。对于输入的 M 个通道的张量,分别对应了 M 个位移卷积核。

位移卷积由两个操作组成。首先,设置不同方向的位移卷积核,可以将不同的通道向不同的方向位移,接着配合点卷积实现跨通道交换信息,即可实现空间域和通道域的信息提取。其次,位移卷积的感受野更加灵活。移位卷积可以通过增加移位距离来扩大其感受野,避免了因增大卷积核而增加的计算成本。将每个通道的移位值表示为 S_i,$i=1,2\cdots,C$,其中,$S_i=(x_i,\ y_i)$表示二维移位向量,所以位移卷积的感受野可表示为各相反方向位移向量的并集:$R=\{-S_1\}\cup\{-S_2\}\cup\cdots\cup\{-S_C\}$。

以图 3(b)所示为例,说明特征位移模块的位移操作。不同的特征在每一行使用不同的颜色表示。假设一个只有 4 个关节,7 个通道的手部骨骼,首先将所有节点视为一个全连通图,并分割 7 个通道,从给定的节点开始,其他节点的特征顺序位移到该节点的通道上,以此类推,直到最后一

个通道。位移结束后，每个节点都融合了所有节点的特征信息。从自身节点在给定的骨骼特征映射 $\boldsymbol{F}\in\boldsymbol{R}^{N\times C}$ 的情况下，第 i 个通道的移位距离为 i mod N，移位后的通道用于填充相应的空白空间。为了能够更好的融合特征，本文将特征位移分为向上移动和向下移动。这意味着不同节点之间的连接强度是相同的。但是对于某个手势，各个节点的重要性是不同的，所以在节点进行特征位移融合后，所以在节点进行特征位移融合后，加入 *Mask* 掩码，将相邻节点与本节点的连接强度进行选择性计算。最终的卷积结果计算式表达为：

$$\tilde{\boldsymbol{F}}_M=\tilde{\boldsymbol{F}}\cdot Mask=\tilde{\boldsymbol{F}}\cdot(\tanh(M)+1) \tag{6}$$

其中，$\tilde{\boldsymbol{F}}_M$ 为卷积输出结果，$\tilde{\boldsymbol{F}}$ 是位移卷积核，tanh 为激活函数，M 为输入张量的通道数。

1.3　网络架构

网络模型的输入张量大小为 (N,C,T,V)，N 表示 batch size；C 表示通道数；$T=20$ 表示视频帧数；V 表示手部骨骼节点数。如图 4 所示，该网络由 9 个特征位移模块和一个全局池化层(global pooling)组成，特征位移模块的输出通道数分别为：64、64、64、128、128、128、256、256、256。每个特征位移模块包含 2 个批量正则化(BatchNorm2d，BN)层、1 个空间特征位移层、1 个时间卷积层、1 个二维卷积操作(Conv2D)、1 个注意力机制和激活函数(ReLU)。由于关节在不同时间帧的位置变化较大，首先将骨骼数据经过 BN 层归一化处理后，分为两路，第一路提取骨骼数据的时空间特征；第二路数据经过卷积步长设置为 2 的二维卷积层和 BN 层获得残差(residual，Res)。将两路的输出矩阵按位相加后输入到空间注意力机制中，对其进行加权分析得到本层特征位移模块处理后的特征矩阵。骨骼数据经过 9 层特征位移模块交替实现空间和时间维度的特征信息聚合之后，对得到的结果进行全局池化，获得每个骨骼节点的 256 维特征向量，最后通过全局池化层输出手势分类的概率值。为了避免过拟合，在每个特征位移模块之后设置丢弃层，以 0.5 的概率随机放弃特征；将第 5 层和第 8 层特征位移层作为池化层，时间卷积步长设置为 2，每经过一次，特征图大小就会缩减至原先的一半。

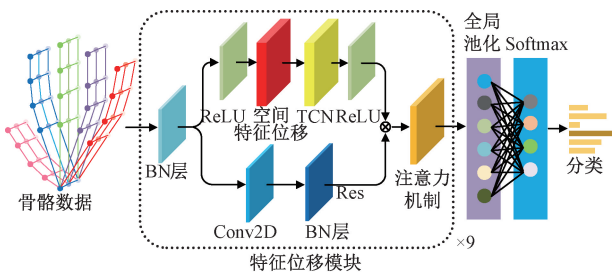


图 4　特征位移图卷积网络示意图

2　实验与比较

为了验证特征位移模块对手势识别准确率的提升，在公开的动态手势数据集(dynamic hand gesture dataset-14/28，DHG-14/28)[8]和第一人称手部动作数据集(First-Person Hand Action Benchmark，FPHA)[27]两个数据集上进行验证。首先，将训练集骨骼数据文件输入到网络模型中进行训练，获得最优的权重参数，然后将验证集骨骼数据文件输入到具有最优权重的网络模型中，根据验证集运行结果和真实结果的对比，获得相应的分类准确度。在 2.3 节中提出消融实验，进一步验证特征位移模块的有效性。具体细节描述如下。

2.1　数据集

本实验使用 3 个具有代表性的数据集进行验证。

DHG-14/28 数据集由英特尔 RealSense 深度相机捕获，其中包含手部骨骼数据(三维坐标(x，y，z))和深度图像序列。该数据集由 20 名参与者的 2 800 个手势视频序列组成，每个手势动作时长 20～50 帧。共包含 14 个手势，并使用手指的数量，分为 14 个或 28 个标签。手势类别标签为："滑动 X (SX)"、"向下滑动(SD)"、"逆时针旋转(RCC)"、"点击(T)"、"顺时针旋转(RC)"、"向右滑动(SR)"、"捏(P)"、"向上滑动(SU)"、"摇晃(SH)"、"抓取(G)"、"向左滑动(SL)"、"滑动 V (SV)"、"张开(E)"、"滑动+(S+)"。

第一人称动态手势数据集，包含 45 种不同手势类别的 1 175 个动作样本，具有高视点、速度、主体内变异性和主体间风格、视点和尺度变异性。该数据集的手势动作由 6 个演员在 3 种不同的场景(厨房、办公室和社交)中操控 26 种不同物体所形成的，被准确标注了标签与类别的彩色和深度(RGB-D)视频序列共计 10 万帧以上。每个物体至少有 1 个关联动作(例如，钢笔-"写")，最多有 4 个关联动作(例如，海绵-"洗"，"抓"，"挤"和"翻")。与 DHG-14/28 数据集相比，FPHA 数据集有 21 个手部关节，手掌关节缺失。

2.2　实验环境

本次实验的数据集训练验证在 Quardro RTX GPU 6000 上完成，软件开发环境为 Python3.7，PyTorch1.2.0，CUDA10.0。DHG-14/28 和 FPHA 数据集的批大小分别设置为 64 和 32，优化函数为随梯度下降法(stochastic gradient descent，SGD)，初始学习率设置为 0.01。在训练过程中，最优权重模型由验证集的准确率决定，且每迭代 30 个 epoch 学习率减小为原来的 0.1，直到学习率降到 0.000 01 时，停止训练。

2.3　数据集实验结果

FPHA 训练集和验证集的精确度如图 5 所示。精度曲线在 20 次迭代之前迅速上升，50 次迭代之后精确度的震荡趋于平稳，说明特征位移模块可以将手部骨骼节点的特征信息进行快速的融合，加快模型收敛速度，减小网络计算量的目标。如图 5 所示，本网络模型有效地避免了过拟合的问题。为了数据的严谨性，数据集的识别率由 5 折交叉验证法得出，如表 1 所示，14 类、28 类手势和 FPHA 在各

折的准确率,最后 5 折的均值作为数据集的识别率。

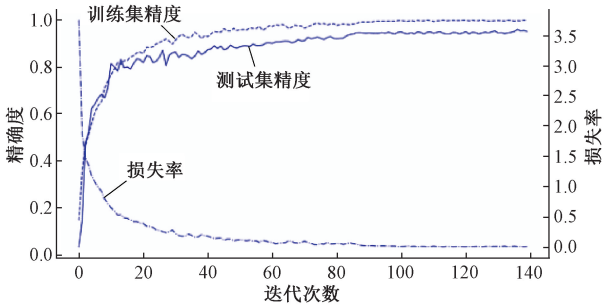


图 5　FPHA 数据集的精度曲线

表 1　5 折交叉验证结果

交叉验证	14 类手势/%	28 类手势/%	FPHA/%
1 折	96.25	94.15	93.15
2 折	94.60	93.50	92.12
3 折	96.00	91.55	90.92
4 折	94.95	93.22	93.61
5 折	93.75	92.63	93.55
平均	**95.11**	**93.01**	**92.67**

同时,为了验证特征位移模块对于手势识别准确率的提高,本文在保持基础图卷积网络结构和最优权重参数一致的条件下,构建一个去除特征位移模块的网络模型,将其训练准确度与本文模型进行对比验证。结果如表 2 所示,于原始 GCN 网络相比,与本文网络在 DHG-14/28 和 FPHA 的准确度分别提高 2.28%、2.47%和 2.89%。实验表明:局部特征与全局特征的结合对于手势识别的精确度有显著的提升。

表 2　位移卷积模块有效性验证

模型	GCN/%	本文的网络/%
14 类手势	92.83	95.11
28 类手势	90.54	93.01
FPHA	89.78	92.67

2.4　与现有方法对比

在公开数据集 DHG-14/28 上,现有算法准确度如表 3 所示,基于 GCN 的方法[27-29]可以更好的捕捉手部关节的空间关系,性能也更好。本文在保留图卷积强大的空间特征提取的同时,融合全局特征与局部特征,进一步优化手势识别能力,提高了所有手势的识别性能。如图 6 所示,除了“抓取 (G)”和“捏 (P)”手势外,大多数手势都可以被有效识别。这主要是因为一些只使用拇指和食指的抓取手势与“捏”动作相似。因此,探索手势细微差异的网络,是下一步的主要研究方向。

表 3　在 DHG-14/28 数据集上方法对比

模型	14 类手势/%	28 类手势/%	平均准确率/%
ST-GCN[18]	91.2	87.1	89.15
2s-AGCN[22]	92.56	89.53	91.04
HPEV+HMM+FRPV[28]	92.58	88.86	90.7
CTR-GCN[21]	92.82	89.55	91.18
RBI-indrnn[24]	94.05	91.90	92.97
TD-GCN[29]	93.9	91.4	92.65
TF-MG[28]	93.29	92.25	92.77
Ours	95.11	93.01	94.06

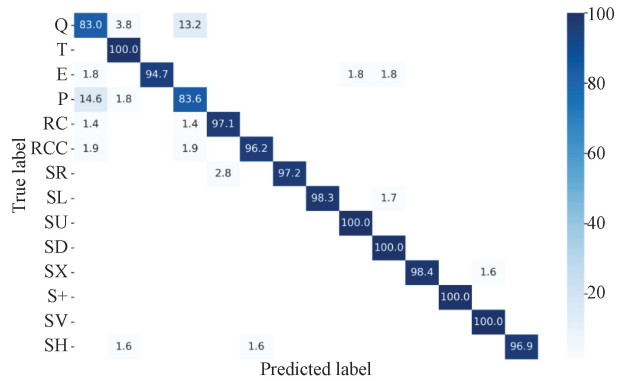


图 6　数据集 DHG-14 的混淆矩阵

在公开数据集 FPHA 上,现有算法准确度如表 4 所示,本文网络获得了最佳的识别率。如图 7 所示,大多数手势都能被准确识别。但对于“打开钱包(open wallet)”、“打开眼镜(unfold glasses)”、“从信封里拿信(take letter from envelope)”等手势识别较差。因为这些手势涉及到手与物体的交互,仅通过骨骼数据实现准确识别较为困难。

表 4　在 FPHA 数据集上方法对比

模型	准确率/%
ST-GCN[18]	84.39
2s-AGCN[22]	89.09
HPEV+HMM+FRPV[28]	90.96
Two-stream NN[24]	90.26
MRMML[30]	83.33
U-SPDNet[31]	87.83
MDCNN-HAHG[32]	90.49
Ours	92.67

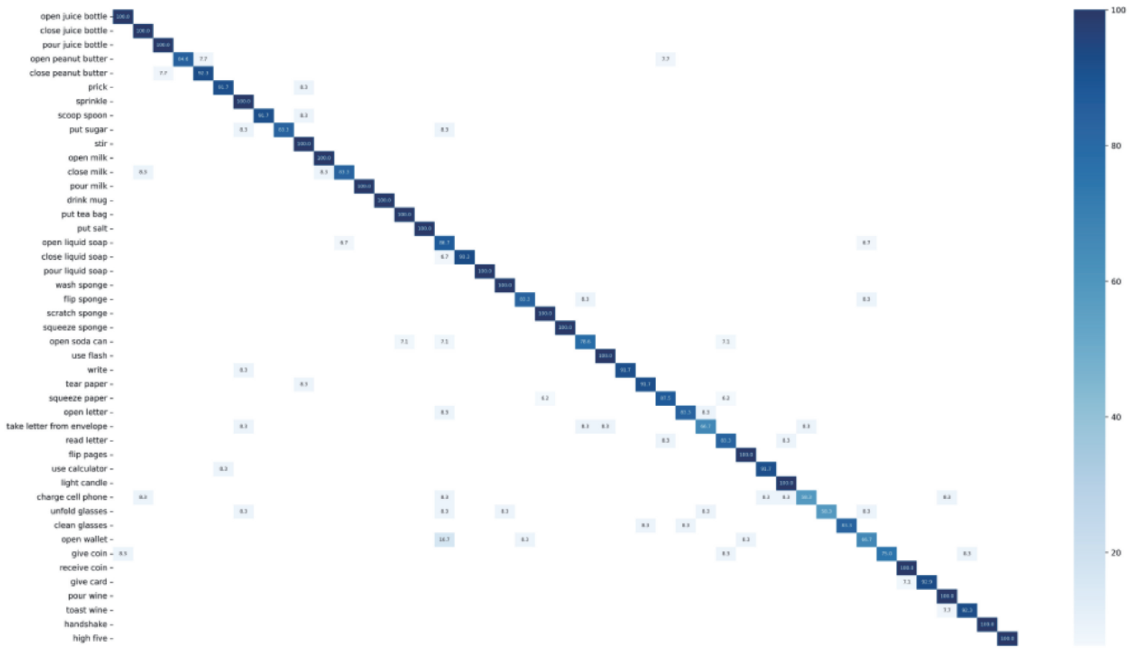


图 7　数据集 FPHA 的混淆矩阵

3　结　　论

本文提出一种基于特征位移的图卷积神经网络手势识别方法，用于基于手部骨骼的动态手势识别。动态手势骨骼数据以图拓扑结构输入模型中进行拟合，在特征位移卷积模块中对位移信道进行重新排序，将相邻节点的特征信息转移到当前卷积节点，使每个节点的感受野扩展到整个手部骨骼，达到融合骨骼节点全局化特征的目的，在保留图卷积强大的空间特征提取能力的同时，结合局部特征与全局特征，进一步提高对手势的识别能力。在数据集 DHG-14/28 和 FPHA 的验证结果表明，本文的方法在手势识别任务上取得了优秀的性能。

参考文献

[1] CHENG L, LIU Y, HOU Z G, et al. A rapid spiking neural network approach with an application on hand gesture recognition [J]. IEEE Transactions on Cognitive and Developmental Systems, 2019, 13(1): 151-161.

[2] XUE Y, JU Z, XIANG K, et al. Multimodal human hand motion sensing and analysis-a review[J]. IEEE Transactions on Cognitive and Developmental Systems, 2018, 11(2): 162-175.

[3] YANG Y, DUAN F, REN J, et al. Performance comparison of gestures recognition system based on different classifiers [J]. IEEE Transactions on Cognitive and Developmental Systems, 2021, 13(1): 141-150.

[4] CHEN C. Deep manifold structure transfer for action recognition [J]. IEEE Transactions on Image Processing, 2019, 28(9): 4646-4658.

[5] CAO C, LAN C, ZHANG Y, et al. Skeleton-based action recognition with gated convolutional neural networks [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2018, 29 (11): 3247-3257.

[6] DE SMEDT Q, WANNOUS H, VANDEBORRE J P. Skeleton-based dynamic hand gesture recognition [C]. IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2016.

[7] 孟杰，杨鹏程，杨朝，等. 基于 Mediapipe 的幻影成像装置自然手势交互系统设计[J]. 国外电子测量技术，2023, 42(3): 116-122.

[8] 程换新，成凯，程力，等. 基于残差融合双流图卷积网络的手势识别方法[J]. 电子测量技术，2022, 45(9): 20-24.

[9] DEVINEAU G, MOUTARDE F, XI W, et al. Deep learning for hand gesture recognition on skeletal data[C]. 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition, 2018.

[10] NARAYANA P, BEVERIDGE J R, DRAPER B A. Gesture recognition: focus on the hands [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018.

[11] 王剑波，朱欣娟，吴晓军. 融合静态手势特征和手部运动轨迹特征的手势交互方法[J]. 国外电子测量技术，2021, 40(7): 14-18.

[12] 卞雨玮,华立涛,周媛. 基于对比学习的信息缺失手势识别新方法[J]. 电子测量技术,2023,46(7):180-186.

[13] 郭鹏,肖秦琨,赵一丹. 基于深度图像的手势识别研究[J]. 国外电子测量技术,2019,38(10):6-12.

[14] CHAKRABORTY B K, SARMA D, BHUYAN M K, et al. Review of constraints on vision-based gesture recognition for human-computer interaction[J]. IET Computer Vision, 2018, 12(1):3-15.

[15] MOLCHANOV P, YANG X, GUPTA S, et al. Online detection and classification of dynamic hand gestures with recurrent 3D convolutional neural networks[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016.

[16] LI Y, HE Z, YE X, et al. Spatial temporal graph convolutional networks for skeleton-based dynamic hand gesture recognition[J]. EURASIP Journal on Image and Video Processing, 2019,2019(1):1-7.

[17] NGUYEN X S, BRUN L, LEZORAY, et al. A neural network based on SPD manifold learning for skeleton-based hand gesture recognition[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2019.

[18] YAN S, XIONG Y, LIN D. Spatial temporal graph convolutional networks for skeleton-based action recognition[J]. Proceedings of the AAAI conference on artificial intelligence, 2018,32(1):7444-7452.

[19] LI M, CHEN S, CHEN X, et al. Actional-structural graph convolutional networks for skeleton-based action recognition[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2019.

[20] LIU Z, ZHANG H, CHEN Z, et al. Disentangling and unifying graph convolutions for skeleton-based action recognition[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2020.

[21] CHEN Y, ZHANG Z, YUAN C, et al. Channel-wise topology refinement graph convolution for skeleton-based action recognition[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2021.

[22] SHI L, ZHANG Y, CHENG J, et al. Two-stream adaptive graph convolutional networks for skeleton-based action recognition[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018.

[23] LIANG H, FEI L, ZHAO S, et al. Mask-guided multiscale feature aggregation network for hand gesture recognition[J]. Pattern Recognition, 2024, 145: 109901.

[24] LI C, LI S, GAO Y, et al. A two-stream neural network for pose-based hand gesture recognition[J]. IEEE Transactions on Cognitive and Developmental Systems, 2021,14(4):1594-1603.

[25] 胡宗承,段晓威,周亚同,等. 基于多模态融合的动态手势识别研究[J]. 计算机工程与科学,2023,45(4):665-673.

[26] WU B, WAN A, YUE X, et al. Shift: A zero flop, zero parameter alternative to spatial convolutions[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018.

[27] GARCIA-HERNANDO G, YUAN S, BAEK S, et al. First-person hand action benchmark with RGB-D videos and 3D hand pose annotations[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018.

[28] LIU J, LIU Y, WANG Y, et al. Decoupled representation learning for skeleton-based gesture recognition[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2020.

[29] LIU J, WANG X, WANG C, et al. Temporal decoupling graph convolutional network for skeleton-based gesture recognition[J]. IEEE Transactions on Multimedia,2023,26:811-823.

[30] WANG R, WU X J, CHEN K X, et al. Multiple riemannian manifold-valued descriptors based image set classification with multi-kernel metric learning[J]. IEEE Transactions on Big Data, 2020,8(3):753-769.

[31] WANG R, WU X J, XU T, et al. U-SPDNet: An SPD manifold learning-based neural network for visual classification[J]. Neural Networks, 2023, 161:382-396.

[32] SUBHASHINI S, REVATHI S N, An accurate estimation of hand gestures using optimal modified convolutional neural network[J]. Expert Systems with Applications, 2024,249(B):123351.

作者简介

刘翔,硕士研究生,主要研究方向为深度学习、手势识别等。

E-mail:1756350656@qq.com

刘新妹,教授,硕导,博士,主要研究方向为电子信息工程、智能检测、测试计量技术及仪器、传感器技术等。

李传坤,副教授,博士,主要研究方向为深度学习、模式识别等。

张晋钊,硕士研究生,主要研究方向为目标检测。