基于时空图神经网络的手势识别

袁 冠[1,2],邴 睿[1],刘 肖[1],代 伟[3],张艳梅[1],蔡 卓[1]
(1. 矿山数字化教育部工程研究中心,江苏徐州 221116; 2. 中国矿业大学计算机科学与技术学院,江苏徐州 221116;
3. 中国矿业大学信息与控制工程学院,江苏徐州 221116)

摘 要: 随着感知计算以及传感器集成技术的发展,使用各种传感设备实时捕捉的手势运动数据,为人机交互提供了新的驱动力,并被广泛地应用于智能家居、远程医疗、虚拟现实等领域. 由于手势动作具有时序性与空间连接性,因此在手势识别中需要考虑手势空间连接关系和手势长距离依赖特性. 然而现有的手势识别方法忽略了上述两种特性,导致识别精度不高. 本文提出了基于时空图神经网络的手势识别算法,该方法从传感器空间分布角度出发,基于传感器的空间位置信息,借助图神经网络(Graph Neural Networks,GNN)对手势数据之间的空间关联性进行表征,并引入门控循环单元(Gated Recurrent Unit,GRU)解决手势的时序性和长距离依赖问题,增强手势识别性能. 在多种数据集上的实验结果证明本文方法可行且有效.

关键词: 手势识别; 多传感器融合; 时空图神经网络; 循环神经网络
中图分类号: TP391.4 **文献标识码:** A **文章编号:** 0372-2112(2022)04-0921-11
电子学报 URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20211069

Spatial-Temporal Graph Neural Network based Hand Gesture Recognition

YUAN Guan[1,2], BING Rui[1], LIU Xiao[1], DAI Wei[3], ZHANG Yan-mei[1], CAI Zhuo[1]
(1. *Digitization of Mine, Engineering Research Center of Ministry of Education, Xuzhou, Jiangsu* 221116, *China*;
2. *School of Computer Science and Technology, China University of Mining and Technology, Xuzhou, Jiangsu* 221116, *China*;
3. *School of Information and Control Engineering, China University of Mining and Technology, Xuzhou, Jiangsu* 221116, *China*)

Abstract: With the development of perceptual computing and sensor integration technology, hand gesture motion data collected by various sensor devices provides a new data-driven way for human-computer interaction, and widely used in smart home, telemedicine, virtual reality and other fields. Due to hand gestures have temporality and spatial connectivity, it is necessary to consider spatial connection and long-distance dependence of hand gesture in gesture recognition. However, existing hand gesture recognition models ignore the aforementioned two problems, resulting in low recognition accuracy. Therefore, we propose a spatial-temporal graph neural network based hand gesture recognition model(STGNN-HGR). From the perspective of spatial distribution of sensors, based on the spatial location information of sensors, the model represents spatial correlation of hand gesture data with the help of graph neural networks(GNN), and introduces gated recurrent unit (GRU) to solve temporality and long-distance dependence in dynamic hand gestures, so as to enhance the performance of gesture recognition. The experimental results on a variety of datasets show that our model is feasible and effective.

Key words: hand gesture recognition; multi-sensor fusion; spatial-temporal graph neural network; recurrent neural network

1 引言

近年来,感知计算、传感器集成技术的快速发展,使得全方位感知的人机交互成为了可能. 手势识别具有便利性、自然化和用户友好等特点,给人机交互带来了新机遇[1]. 利用手势进行交互,不仅可以保留用户原有的交互习惯,还可以丰富人机交互的内涵和形式[2,3]. 基于可穿戴多传感器的手势识别主要利用多传感器融合的方法采集手部姿态及运动数据,并进行计

收稿日期:2021-08-11;修回日期:2021-12-12;责任编辑:王天慧
基金项目:国家自然科学基金(No.71774159);中国博士后科学基金(No.2021T140707,No.2018M642358)

算、分析及识别. 将其作为指令输入,实现用户与设备之间的交互,充分体现了交互方式的自然性、人机关系的和谐性、交互途径的隐含性以及感知通道的多样性. 因此,基于可穿戴多传感器的手势识别已经成为现阶段人机交互研究的热点,被广泛地应用于虚拟现实[4]、健康医疗[5,6]、工业控制[7]、智能家居[8]及军事作战[9]等多个领域.

根据模型识别方法的不同,现有的手势识别算法可以分为两大类:基于统计模型的方法和基于深度学习的方法. 常用于手势识别的统计模型有支持向量机(Support Vector Machine, SVM)[10]、隐马尔可夫模型(Hidden Markov Model, HMM)[11]. 例如:Chen等人[12]使用MYO腕带采集手势数据,并对原始手势数据进行预处理,降低噪声,检测肌肉活动区域,然后通过滑动窗口对传感器数据进行特征提取,最后,结合SVM完成对手势的识别;Kumar等人[13]借助耦合隐马尔可夫模型,提出了多传感器融合手势识别方法,该方法克服了HMM中使用观察状态的缺点,在状态空间中提供信息交互,从而提高了手势识别的性能. 统计模型在小样本、小标签的数据集上取得了良好的识别效果[14]. 但是随着手势数据维度的增多、手势数据轨迹复杂化,统计模型识别性能大大降低,主要原因在于手动提取的统计特征无法表征手势数据之间的类内相似性和类间相异性,并受限于专业领域知识差异,人工特征提取对结果影响比较大.

深度学习在手势识别方向应用广泛,简化了统计模型的特征工程过程,避免了人工干扰,实现了跨领域的知识共享. 常用于手势识别的神经网络模型有卷积神经网络[15]、循环神经网络[16]以及图神经网络[17]. 在手势识别过程中经常利用不同模型的优势互补信息,增强手势识别的效果. 比如Nunez等人[18]考虑到使用长短期记忆单元(Long Short Term Memory, LSTM)提取手势数据的时序特征时,需要依赖上一时刻的信息,无法并行执行,导致模型训练速度慢,首先使用卷积神经网络(Convolution Neural Network, CNN)提取手势数据特征,再通过权值共享减少网络模型训练的参数量;Chen等人[19]为避免传统机器学习方法提取手势数据特征时的人工干扰,使用CNN自动提取手势数据的隐性特征,结合SVM完成对手势的识别,提高了手势识别精度;刘肖等人[20]通过卷积神经网络提取手势数据特征,结合多分类器进行决策融合,提高手势识别的准确率;为充分融合不同模型之间的优势,Zhu等人[21]采用混合深度模型对手势数据进行识别,该模型由卷积神经网络和长短期记忆单元组成.

虽然现有的深度学习模型在一定程度上增强了手势识别性能,但是忽略了手势数据的空间关联性[22],即相邻关节点的手势在空间上相互连接、相互影响. 因此本文提出了基于时空图神经网络的手势识别模型(Spatial-Temporal Graph Neural Network based Hand Gesture Recognition, STGNN-HGR),可以提取手势数据的空间关联信息与长距离依赖关系,以实现更优的识别精度. 在中国手语及标准军队手语数据集上的实验证明了本文方法的有效性.

2 基础知识

2.1 问题描述

基于多传感器融合的手势识别可以看作多分类问题,其目的是从原始传感器数据中提取具有强表征能力的信息,进而识别手势的类型. 给定一系列预处理的传感器数据$\{\boldsymbol{D}_i\}_{i=1}^{T}$,$T$表示手势数据的采样数,则$t$时刻的手势数据$\boldsymbol{D}_t \in \boldsymbol{R}^{N\times S}$,如式(1)所示:

$$\boldsymbol{D}_t=\begin{pmatrix} d_1^1 & \cdots & d_1^S \\ \vdots & \ddots & \vdots \\ d_N^1 & \cdots & d_N^S \end{pmatrix} \tag{1}$$

其中,N表示嵌入在人体上肢传感器的数量;S表示每个传感器的属性维度. 根据人体自然连接性和传感器的空间位置,将传感器数据$\boldsymbol{D}$转换为图结构数据$G=(V,E,F)$,其中V表示人体关节点,E表示人体上肢自然连接性,F表示各关节点处传感器特征,转换如式(2)所示:

$$\boldsymbol{G}_t=\boldsymbol{H}\boldsymbol{D}_t \tag{2}$$

其中,$\boldsymbol{H}$为转换矩阵,其功能为将传感器数据构造为图结构数据. 假设φ表示特征提取模型,给定t时刻预处理后的图结构G_t,则特征向量$\boldsymbol{X}_t$可以按式(3)的方式提取:

$$\boldsymbol{X}_t=\varphi(G_t) \tag{3}$$

使用手势识别模型ϕ,计算每个手势$Y=\{y_1, y_2, \cdots, y_c\}$的置信度得分$P$,即条件概率分布,如式(4)所示:

$$P(y_i|\boldsymbol{X}_t,\theta)=\phi(\boldsymbol{X}_t,\theta),\quad i=1,2,\cdots,c \tag{4}$$

其中,θ表示手势识别模型ϕ中所有的参数;c为手势种类的数量. 已知手势的置信度得分P,则识别模型的预测值y_t^p为概率最大的手势数据标签,计算方式如式(5)所示:

$$y_t^p=\operatorname{argmax} P(y_i|\boldsymbol{X}_t,\theta),\quad i=1,2,\cdots,c \tag{5}$$

2.2 基于空间关联的手势图构建

为了描述嵌入在关节点处传感器的空间分布信息,STGNN-HGR引入图结构$G=(V,E,F)$对嵌入在人体上肢传感器的空间位置进行建模,即对上臂、前臂、手掌以及五指进行关节点抽象建模[23]. 图的顶点表示各个关节点,顶点集$V=\{v_1, v_2, \cdots, v_M\}$,$M$表示关节点数量. 图的边表示肢体连接特性,边集$E=\{e_1, e_2, \cdots, e_n\}$,$n$表示边的数量. 图的输入特征为各关节点处的传感器

数值,组成特征集合$F=\{f_1,f_2,\cdots,f_T\}$,T表示采样数. 手势建模如图1所示.

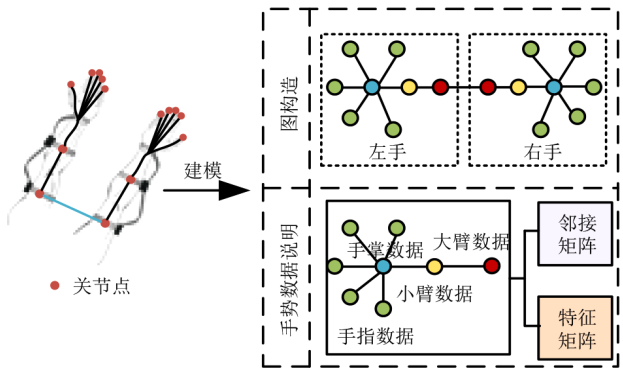


图1　手势建模

传感器的空间信息由邻接矩阵和特征矩阵组成. 邻接矩阵$\boldsymbol{A}\in\boldsymbol{R}^{M\times M}$表示传感器空间关联特性,邻接矩阵$\boldsymbol{a}_{ij}\in\boldsymbol{A}$描述如式(6):

$$\boldsymbol{a}_{ij}=\begin{cases}1, & \text{if } \exists\ e_{ij}\in E \quad i\neq j\\ 0, & \text{otherwise}\end{cases} \tag{6}$$

其中,当不同关节点v_i和v_j存在自然连接时,邻接矩阵$\boldsymbol{a}_{ij}$的取值为1,$\boldsymbol{a}_{ij}=1$的数量为$2n$;否则,邻接矩阵$\boldsymbol{a}_{ij}$的取值为0.

特征矩阵$\boldsymbol{F}\in\boldsymbol{R}^{M\times S}$由嵌入在人体上肢关节点的传感器数据组成,$M$表示关节点数量,$S$表示关节点的特征维度. $\{f_{ij}\}_{j=1}^{S}$表示关节点v_i的特征属性,$\{f_{ij}\}_{i=1}^{M}$表示不同关节点处特征f_j的传感器数值.

3　基于时空图神经网络的手势识别

本文提出的基于时空图神经网络的手势识别模型(STGNN-HGR)主要包含以下三部分:数据建模、特征提取、手势识别. 图2展示了STGNN-HGR的框架. 首先,为了充分利用关节点之间的连通性,STGNN-HGR使用图结构表示嵌入在人体上肢的传感器空间信息. 空间信息主要由邻接矩阵和特征矩阵组成,邻接矩阵表示人体上肢关节点的自然连接性,特征矩阵表示手部运动及姿态信息即各关节点处的传感器数值. 然后,STGNN-HGR结合图卷积神经网络(Graph Neural Networks,GCN)聚合相互连通关节点的运动信息. 最后,按照时序信息,将同一关节点不同采样的空间特征转换为时间序列,并借助入门控制循环单元(Gated Recurrent Unit,GRU)提取手势数据的时序特征,解决动态手势的时序性和长期依赖问题,从而完成对手势动作的识别.

3.1　图卷积神经网络

为了解决非欧氏空间数据的特征提取问题,STGNN-HGR采用图卷积经神经网络聚合相互连通关节点的手势姿态信息. 首先,为了聚合各关节点的自身运动特征,形成自环,将邻接矩阵$\boldsymbol{A}$与单位矩阵$\boldsymbol{I}\in\boldsymbol{R}^{M\times M}$相加,即$\boldsymbol{A}+\boldsymbol{I}$. 然后,由于各个节点的度的差异性会导致梯度消失或梯度爆炸现象,因此,为了加强模型学习时的数值稳定性,需要将$\boldsymbol{A}+\boldsymbol{I}$进行归一化处理,即$\boldsymbol{\Lambda}^{-1/2}(\boldsymbol{A}+\boldsymbol{I})\boldsymbol{\Lambda}^{-1/2}$. 最后,$t\in T$时刻的空间特征$\boldsymbol{f}_{\text{space}}^{t}\in\boldsymbol{R}^{M\times S}$由式(7)提取:

$$\boldsymbol{f}_{\text{space}}^{t}=\text{LeakyReLU}(\boldsymbol{\Lambda}^{-\frac{1}{2}}(\boldsymbol{A}+\boldsymbol{I})\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{F}\boldsymbol{W}) \tag{7}$$

其中,$\boldsymbol{W}$表示权重矩阵;$\boldsymbol{\Lambda}$表示节点的对角度矩阵;LeakyReLU(·)为非线性激活函数,用来解决常用的ReLU(·)激活函数由于单侧抑制导致神经元无法有效更新的问题,如式(8)所示:

$$\text{LeakyReLU}(x)=\begin{cases}x, & \text{if } x\geqslant 0\\ \lambda x, & \text{if } x<0\end{cases} \tag{8}$$

其中,参数λ为取值在[0,1]上的超参数,用于控制变量x为负值时映射函数的斜率大小,避免在变量x小于0时梯度消失的情况.

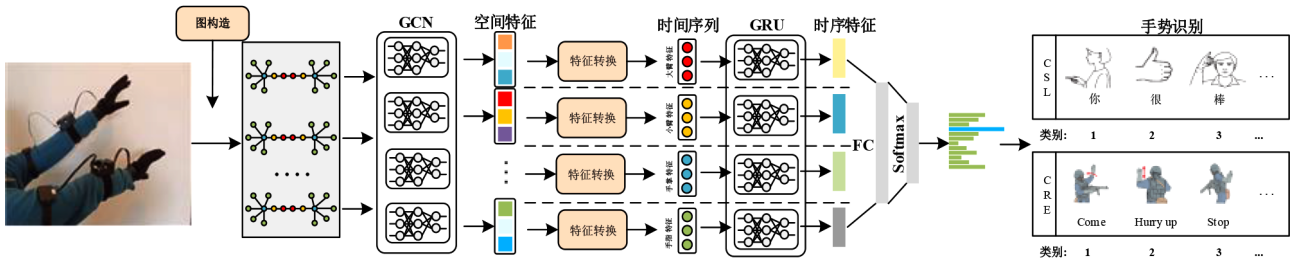


图2　STGNN-HGR框架

3.2　门控循环单元

动态手势数据不仅具有时序性,而且存在着长距离依赖问题,这意味着手势数据不仅随时间不断变化,还受到之前某一时刻输入数据的影响. 为了充分提取手势数据的时序信息,STGNN-HGR将同一关节点不同采样处的空间特征转化为时间序列,如图3所示.

已知手势数据空间特征表示为$\{\boldsymbol{f}_{\text{space}}^{t}\}_{t=1}^{T}$,且$\{\boldsymbol{f}_{\text{space}}^{t}\}_{t=1}^{T}\in\boldsymbol{R}^{M\times S\times T}$,$T$表示采样数. 按照时间序列,将同一关节点$v_i$不同采样的空间特征转化为时间特征序列$\{\boldsymbol{f}_{\text{time}}^{i}\}_{i=1}^{M}\in\boldsymbol{R}^{S\times T\times M}$,转化过程如式(9)所示:

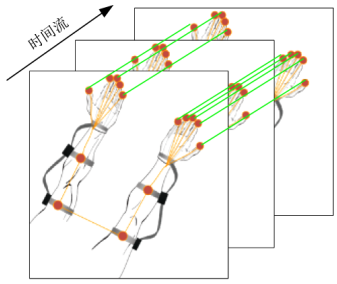


图3 手势时序特征

$$\{\boldsymbol{f}_{\text{time}}^{i}\}_{i=1}^{M}=\boldsymbol{H}\{\boldsymbol{f}_{\text{space}}^{t}\}_{t=1}^{T} \tag{9}$$

其中，$\boldsymbol{H}$为转化矩阵，用于将空间特征按照时间流顺序转化为时序信息.

传统机器学习模型(如HMM)能够有效地识别序列数据，但无法解决手势数据的长距离依赖关系，即在较长手势序列中，当前位置的手势类别可能依赖输入开始时的数据. 为了表征节点时序信息，解决动态手势数据的长距离依赖问题，STGNN-HGR采用GRU利用门控机制控制输入与记忆等信息，不仅解决了循环神经网络的梯度消失问题，又简化了长短期记忆单元的计算机制，提高了手势识别的效率和精度. GRU的结构如图4所示.

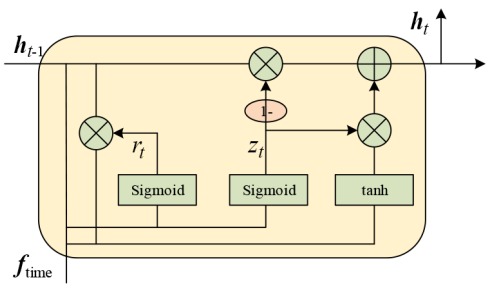


图4 GRU结构图

首先，时间序列特征$\boldsymbol{f}_{\text{time}}$与上一时刻隐层状态$\boldsymbol{h}_{t-1}$构成更新门，得到更新状态$z_t$，用于控制前一刻时间步和当前时间步传递信息量，其计算如公式(10)所示：

$$z_t=\sigma(\boldsymbol{W}_z\boldsymbol{f}_{\text{time}}+\boldsymbol{U}_z\boldsymbol{h}_{t-1}+\boldsymbol{b}_z) \tag{10}$$

其中，$\boldsymbol{W}_z$与$\boldsymbol{U}_z$分别为更新门中对特征和上一时刻隐层状态的权重矩阵，$\boldsymbol{b}_z$为偏置量. 然后，设置重置门决定当前时刻遗忘多少上一时刻的手势信息，如式(11)所示：

$$r_t=\sigma(\boldsymbol{W}_r\boldsymbol{f}_{\text{time}}+\boldsymbol{U}_r\boldsymbol{h}_{t-1}+\boldsymbol{b}_r) \tag{11}$$

其中，$\boldsymbol{W}_r$与$\boldsymbol{U}_r$分别为重置门中对特征和上一时刻隐层状态的权重矩阵，$\boldsymbol{b}_r$为偏置量. 接着，计算新记忆状态$\widetilde{\boldsymbol{h}}_t$，其得到的上一时刻手势信息量由重置门$r_t$决定，如式(12)所示：

$$\tilde{\boldsymbol{h}}_t=\tanh(r_t\odot\boldsymbol{U}_h\boldsymbol{h}_{t-1}+\boldsymbol{W}_h\boldsymbol{f}_{\text{time}}+\boldsymbol{b}_h) \tag{12}$$

与式(10)、式(11)类似，$\boldsymbol{W}_h$与$\boldsymbol{U}_h$分别为新记忆状态计算中对特征和上一时刻隐层状态的权重矩阵，$\boldsymbol{b}_h$为偏置量. 最后，得到当前时间步的最终信息量，即保留当前单元的信息量并传递到下一个单元，如式(13)所示：

$$\boldsymbol{h}_t=(1-z_t)\odot\boldsymbol{h}_{t-1}+z_t\odot\tilde{\boldsymbol{h}}_t \tag{13}$$

其中，$\odot$表示数乘，主要用于计算不同时刻信息的保留与传递情况.

3.3 时空图神经网络手势识别

本文提出的STGNN-HGR模型将关节点处的传感器数据转化图结构数据，结合图卷积神经网络聚合邻接节点信息，解决了手势数据的空间关联性问题. 此外，STGNN-HGR通过GRU提取手势数据的时间序列信息，解决了动态手势的时序性和长距离依赖问题. STGNN-HGR的算法流程如算法1所示.

算法1 时空图神经网络手势识别STGNN-HGR

输入：手势数据集$D=\{d_1, d_2, \cdots, d_T\}$与对应标签$Y=\{y_1, y_2, \cdots, y_T\}$

输出：手势识别的精确度acc

创建手势邻接矩阵$\boldsymbol{A}\in\boldsymbol{R}^{M\times M}$;

创建特征矩阵$\{\boldsymbol{F}_t\}_{t=1}^{T}\in\boldsymbol{R}^{M\times S}$;

FOREACH $d_t\in D=\{d_1, d_2, \cdots, d_T\}$:

$\boldsymbol{F}_t=\text{reshape}(d_t)$; //将传感器数据转化为特征矩阵

END FOR

FOREACH 采样数$t\in T$:

$\boldsymbol{f}_{\text{space}}^{t}=\text{LeakyReLu}(\boldsymbol{\Lambda}^{-\frac{1}{2}}(\boldsymbol{A}+\boldsymbol{I})\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{F}_t\boldsymbol{W})$; //提取手势数据的空间特征

END FOR

$\{\boldsymbol{f}_{\text{time}}^{i}\}_{i=1}^{M}=\boldsymbol{H}\{\boldsymbol{f}_{\text{space}}^{t}\}_{t=1}^{T}$; //空间特征转换时间序列

FOREACH $v_i\in V=\{v_1, v_2, \cdots, v_m\}$: //对于每个节点时间流

$\boldsymbol{f}_{ts}^{i}=\text{GRU}(\boldsymbol{f}_{\text{time}}^{i})$; //提取手势数据的时序特征

END FOR

$\boldsymbol{f}_{ts}=\text{concat}\{\boldsymbol{f}_{ts}^{1}, \boldsymbol{f}_{ts}^{2}, \cdots, \boldsymbol{f}_{ts}^{M}\}$; //特征融合

$\tilde{y}=\text{softmax}(\boldsymbol{f}_{ts})$; //获得手势数据预测值

$\text{loss}=-\sum_{i=1}^{T}y_i\log(\widehat{y_i})$; //使用损失函数训练模型

$\text{acc}=\text{STGNN}(D_{\text{test}})$; //计算模型测试集识别准确率

RETURN acc //返回准确率

4 实验与性能分析

为了验证本文方法的有效性，在中国手语数据集与标准军队动态手势数据集上设计对比实验. 选择由基于统计模型与基于深度学习模型组成的5种不同的手势识别算法进行对比分析.

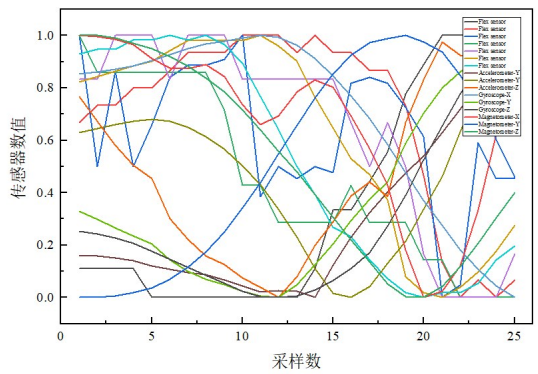
4.1 实验数据分析

本文设计实现了集成双手臂环的数据手套作为数据采集平台，创建了中国手语数据集(Chinese Sign Language，CSL). CSL由6名志愿者(3名男性和3名女性)完成，他们按照日常中国手语的动作协议[24]采集了的手

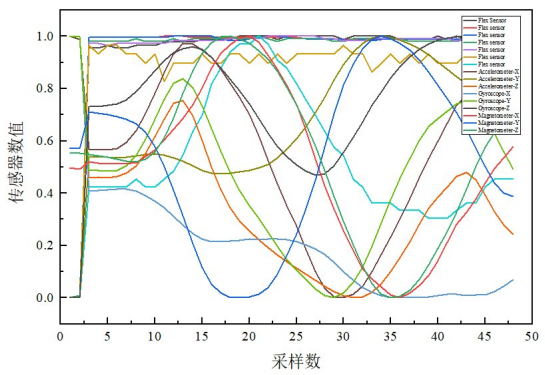
势数据. 所有的志愿者在实验过程中佩戴数据手套. 以20Hz的固定频率,利用嵌入在数据手套中的弯曲传感器、陀螺仪以及加速度计,采集手指的弯曲信息、手掌以及大小臂的姿态信息,并对手势数据进行手动标记. CSL被随机分为两组,其中70%的数据用作训练集,30%用作测试集. CSL数据集主要包含复合动态手势(如"你好"复合手势由"你"和"好"两个基本手势组成)和基本动态手势(如再见、惊讶等手势).

公开数据集选用标准军队动态手势数据集[25](Standardized Hand Signals for Close Range Engagement Operations,CRE). CRE包含6类动态手势,每类采集81次,每次采样数据是可变长度的时间序列. 采样过程中,弯曲传感器采集手指、手腕以及肘部的弯曲信息,使用惯性测量单元(IMU-MPU-9250)捕捉手臂的姿态信息. 图5为"Come"手势和"I don't understand"手势采集过程中的传感器数值.

CRE与CSL两种数据集的统计如表1所示.



(a)　"Come"手势



(b)　"I don't understand"手势

图5　CRE传感器数据图

4.2　评价指标

评价标准是衡量模型鲁棒性、泛化性的主要方式. 手势识别作为典型的多分类问题,在本文中使用准确率作为评价指标对识别结果做出评价. 混淆矩阵是一个$N\times N$矩阵,N表示手势类别标签数量,行表示识别模型的预测值,列表示实例对应的真实值,单元格C_{ij}则表示真实值y_j被识别为y_i的次数.

表1　实验数据集

数据集	样本总数	手势种类	采集设备	说明
CRE	22 595	6	数据手套	单词级手势
CSL	95 040	72	数据手套	单词级手势

手势识别过程中会出现以下情况:真实的手势种类被正确预测为正手势种类的情况,记为TP(True Positive);不相关的手势种类被正确地预测为负手势种类的情况,记为TN(True Negative);不相关的手势种类被错误地预测为正手势种类的情况,记为FP(False Positive);真实的手势数据被错误预测为负手势种类的情况,记为FN(False Negative).

准确率(Accuracy)是指对于给定的测试手势数据集,识别模型正确分类器的样本数占总样本数的比值. 准确率的计算如式(14)所示:

$$\text{Accuracy}=\frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}} \tag{14}$$

4.3　实验结果与分析

为了验证STGNN-HGR能够自动提取具有强表征能力的深度特征,STGNN-HGR与基于统计模型的识别方法进行对比,包含K最近邻(K-Nearest Neighbor,KNN)[26]、SVM[27]等. 此外,为了验证STGNN-HGR能够提取手势数据的空间关联信息,解决动态手势的时序性以及长距离依赖问题,将STGNN-HGR与CNN[28],GRU[29],CNN-LSTM[18]进行对比. 对比算法的具体描述如表2所示.

4.3.1　参数分析

(1)图卷积神经网络层数分析

图卷积神经网络能够聚合相互连通关节点的手部运动信息,借助不同粒度的空间特征来表征不同层次的语义信息. 浅层特征包含更多的原始信息,但语义歧义的问题突出;深层特征具有较高的语义性,能够有效表征原始数据,但会丢失原始数据的特性. 因此图卷积神经网络的层数影响着手势识别的性能. 经过多次实验,图6给出了不同网络层数下的手势识别精度. 随着图卷积神经网络层数增加,识别准确率也越高;但当层数增加到3层时,准确率逐渐趋于平稳,甚至出现下降趋势. 因此本文使用深度为3的图卷积神经网络提取手势的空间特征.

(2)维度填充方式分析

本文的数据集是基于异构多传感器融合的数据集,在构建图特征矩阵过程中,手臂、手掌的特征维度和手指的特征维度出现不一致现象(如惯性测量单元

表2 对比算法

算法名称	特点	说明
K最近邻(KNN)	利用周围有限的邻近样本对类域进行判定	验证深度特征表示能力强于手动提取特征
支持向量机(SVM)	利用内积核函数解决非线性分类问题	验证深度特征表示能力强于手动提取特征
卷积神经网络(CNN)	能够提取手势数据具有高语义的深度特征	验证时序性对手势识别效果的影响
门控循环单元(GRU)	能够解决动态手势的时序性和长距离依赖问题	验证手势数据空间特征能够更好地表征手势数据
CNN-LSTM	CNN提取手势数据的深度特征,结合LSTM解决时序性和长距离依赖问题	验证STGNN能够更好地表征手势数据,提高手势识别准确率,增强手势识别性能
STGNN-HGR	使用GCN提取数据的空间特征,并结合GRU解决时序性和长距离依赖问题.	STGNN能够提取数据空间特征,解决动态手势数据的时序性及长距离依赖问题

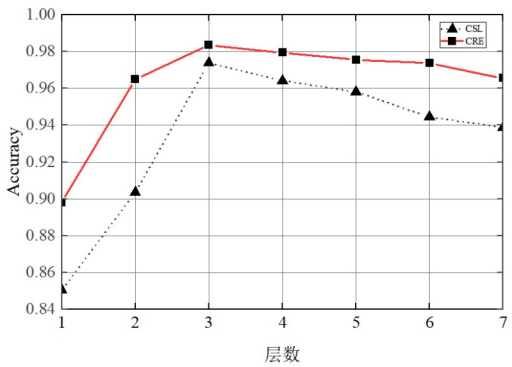


图6 图卷积层数分析

的特征维度与弯曲传感器的特征维度不一致),影响了图的构造. 本文采用3种方法对传感器缺失的特征维度进行填充.

(a) 0填充. 为了减少负相关特征对识别精度的影响,每次经过图卷积运算迭代更新后,得到的空间图特征中缺失的运动信息用0填充.

(b) 统计特征填充. 为了充分挖掘动态手势数据的空间特征,将弯曲传感器缺少的特征信息,用统计特征进行填充,包含均值特征、最值特征、方差特征等,具体描述如表3所示.

表3 常用统计特征

统计特征	描述	计算公式
均值特征	窗口中信号的平均值	$\text{mean}=\frac{1}{n}\sum_{i=1}^{n}a_i$
最值特征	窗口中信号的最大值和最小值	$\max=\max(a_i),i=1,2,\cdots,n$ $\min=\min(a_i),i=1,2,\cdots,n$
范围特征	窗口中信号最大值与最小值之差	$\text{range}=\lvert\max-\min\rvert$
方差特征	窗口中一组数据的分散程度度量	$s^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\text{mean})^2$

(c)自由填充. 为了聚合相互连通关节点的动态特征信息,使用自由填充方法对缺失的特征值进行填充. 首先将弯曲传感器处缺少的特征信息用0填充的方式填充. 经过每次图卷积运算迭代更新后,得到的图特征向量数值维持不变,此时缺失的弯曲传感器特征值为周边关节点特征的聚合值.

图7给出了在不同维度填充方法下,时空图神经网络模型的识别精度图.

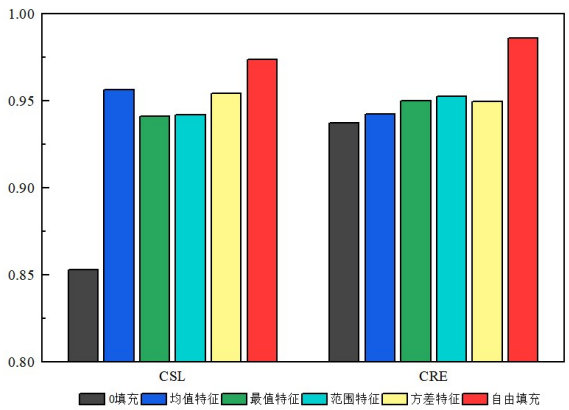


图7 维度填充方式分析

从图7可知,自由填充的识别准确率最高,主要原因在于0填充方法和统计填充方法带来了人工干扰问题,从而影响手势识别的精度;而自由填充的方法充分聚合周边各关节点传感器运动特征,从而增强了动态手势识别的性能. 因此,本文采用自由填充方式填充空缺维度值.

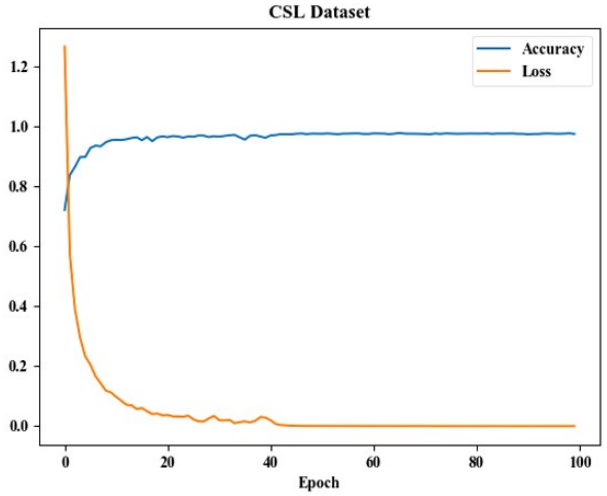
(3)超参数分析

经过多次实验,将模型参数设置为以下参数值时,模型具有最优效果:中国手语数据集训练过程中,学习率设为0.005,批次大小设为100,迭代次数设为100;标准军队动态手势数据集训练过程中,学习率设为0.001,批次大小设为100,迭代次数设为250. 结果如图8所示.

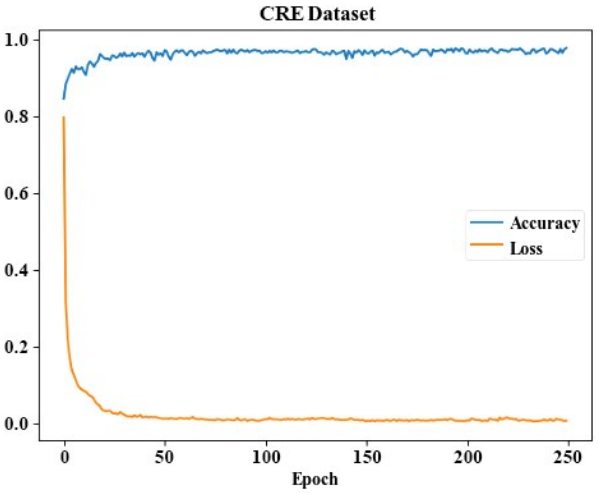
图8的损失和正确率具有相同的坐标量纲,其中图8(a)为中国手语基于以上参数设置的训练过程,图8(b)为标准军队动态手势基于以上参数设置的训练过程. 从图8中可以看出,模型训练过程收敛速度较快,且收敛平稳. 因此,训练出来的模型具有强鲁棒性、高适应性的优点.

4.3.2 实验结果分析

中国手语数据集数据由手臂姿态信息、手掌姿态信息以及手指弯曲信息构成,能够实现全面描述手势



(a)　中国手语数据集



(b)　标准军队动态手势数据集

图 8　参数选择

的变化轨迹. 而标准军队动态手势数据集关注手指、腕部、肘部弯曲信息以及手臂姿态信息. 图 9 为对比算法实验结果图.

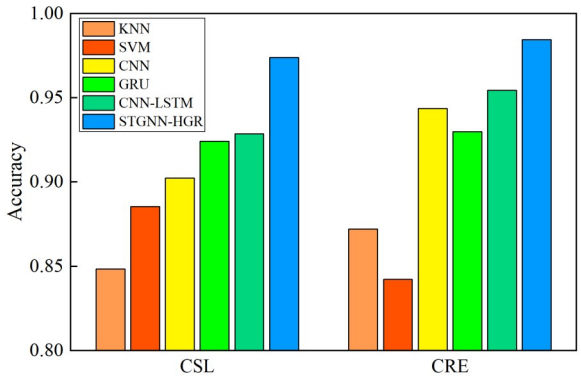
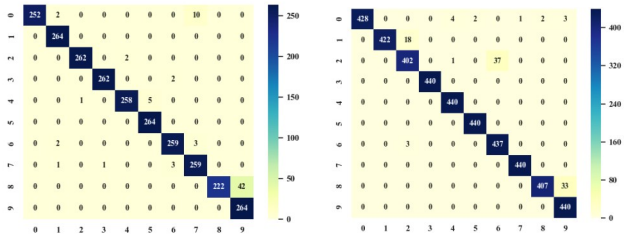


图 9　对比算法准确率分析

从图中看出,基于深度学习的方法手势识别效果高于基于统计模型的方法,主要原因为深度特征解决了统计特征的人工干扰问题,更能够精确地表征手势数据. 此外,由于中国手语数据集更具动态性、时序性,因此 GRU 识别效果优于其他对比算法. 而标准军队动态手势数据集主要关注手势的空间特性,因此,CNN 的识别准确率略优于其他算法. 整体上,STGNN-HGR 的识别效果优于其他对比算法,中国手语数据集的识别准确率为 97. 20%,平均高于对比算法 5%. 标准军队动态手势数据集识别准确率为 98. 63%,平均高于对比算法 4%.

同时,为了精确评价手势识别模型的性能,图 10 给出了 CSL 数据集和 CRE 手势数据集识别过程中的混淆矩阵,其中各手势标签含义如表 4 所示. 基于混淆矩阵,结合基于加权特征增强的手势识别结果,可以发现时空图神经网络模型和加权特征增强模型识别精度相似,易混淆的高层次复杂手势相同,即分别从数据特征和传感器空间分布角度,实现了对动态复杂手势的有效识别.



(a)　CSL 数据集　　(b)　CRE 数据集

图 10　混淆矩阵

此外,表 4 还给出了两种数据集中各个手势的识别准确率. 从表 4 可以看出,STGNN-HGR 对中国手语数据集和标准军队动态手势数据集的每个手势都达到了良好的识别效果. 其中,在中国手语数据集识别过程中,“再见”手势、“翻”等手势识别准确率达到了 100%;在标准军队动态手势数据集识别过程中,“I don’t understand”等手势识别准确率也达到了 100.00%.

通过深入分析表 4 和图 10 发现,在中国手语的识别过程中,一部分标签为“惊讶”的手势数据被误识别为标签为“谁”的手势. 可能原因在于使用欧拉角描述手部的姿态变化信息存在一定的局限性,即当手势数据发生小角度变化时,传感器数据耦合性降低,导致三轴的传感器数值相互独立,无法协同全方位表示现实手势的运动轨迹,从而影响了手势识别的效果.

图 11 展示了“惊讶”手势和“谁”手势的欧拉角数据图.

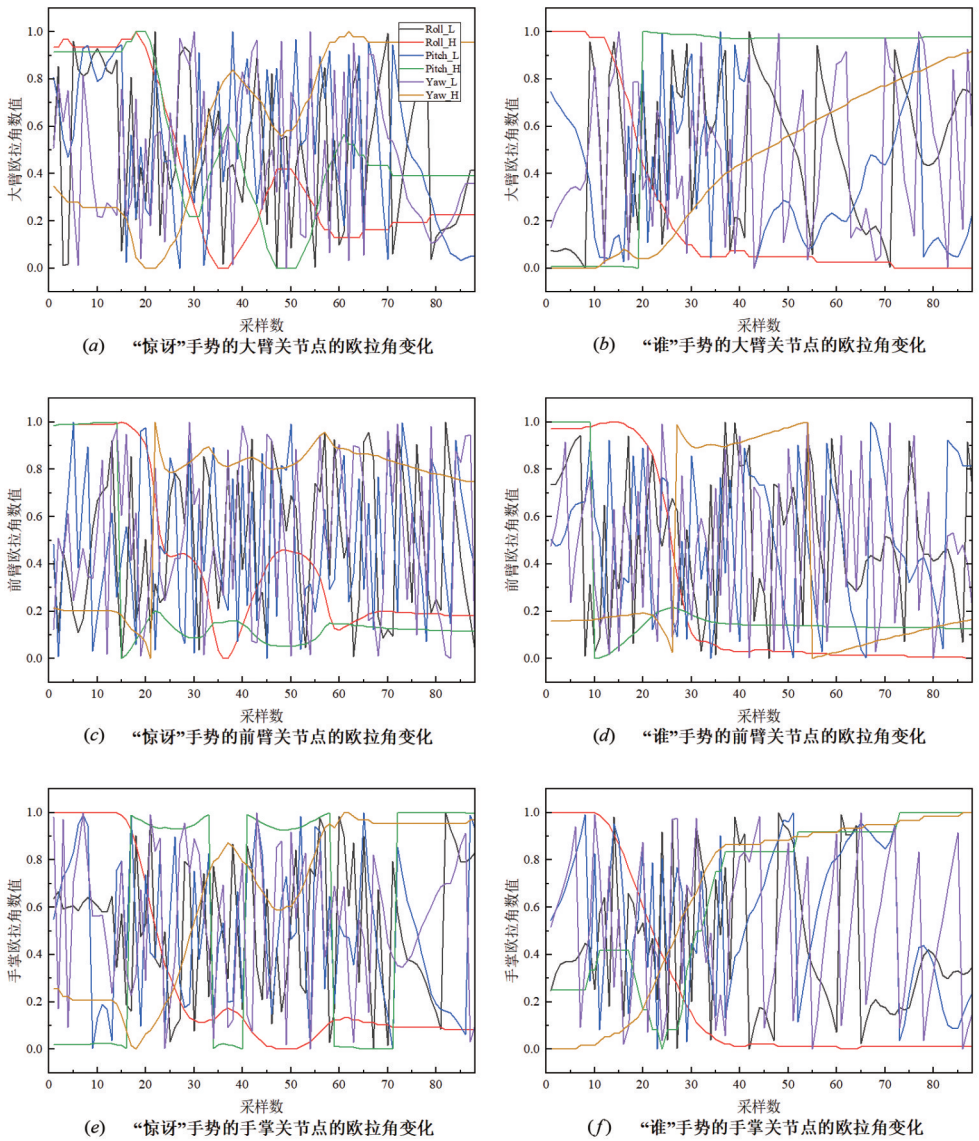
表4 标签信息

标签	CSL	Acc(%)	CRE	Acc(%)
0	你好	99.45	Come	97.27
1	再见	100.00	Hurry up	95.92
2	擦	99.24	Rally point	93.64
3	吃	99.24	Vehicle	100.00
4	打	97.73	I don't understand	100.00
5	翻	100.00	Door	100.00
6	喝	98.10	Go here	99.32
7	很好	98.10	Ammunition	100.00
8	惊讶	84.09	Go Prone	97.04
9	谁	100.00	Point of Entry	100.00

图11(*a*)和图11(*b*)分别为“惊讶”手势与“谁”手势的大臂关节点的欧拉角变化图,图11(*c*)和图11(*d*)分别为“惊讶”手势与“谁”手势的前臂关节点的欧拉角变化图,图11(*e*)和图11(*f*)分别为“惊讶”手势与“谁”手势的手掌关节点的欧拉角角变化图.从图11可以看出,“惊讶”手势和“谁”手势的欧拉角变化具有一定的相似性,导致手势识别的精度下降.

此外,在军队手势识别过程中,“Vehicle”手势和“Hurry up”手势容易产生混淆.从数据采集的过程可知,军队手势主要关注的是手臂的姿态信息.因此,导致“Vehicle”手势和“Hurry up”手势混淆的可能原因在于佩戴在手臂上的惯性测量单元捕捉到的手势数据具有很高的相似度,包含加速度计数值、陀螺仪数值、磁力计数值等.

图12为“Vehicle”手势和“Hurry up”手势的手臂姿态与运动信息.从图12中可以看出,两个手势的运动



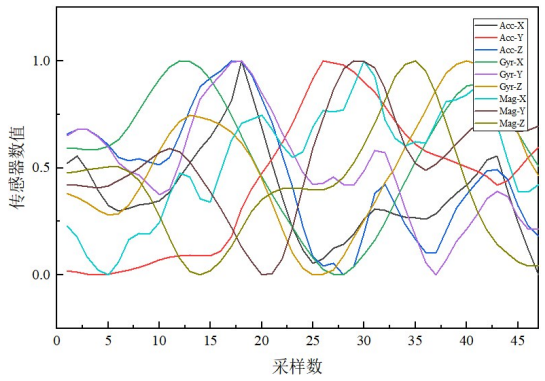
(*a*) “惊讶”手势的大臂关节点的欧拉角变化 (*b*) “谁”手势的大臂关节点的欧拉角变化

(*c*) “惊讶”手势的前臂关节点的欧拉角变化 (*d*) “谁”手势的前臂关节点的欧拉角变化

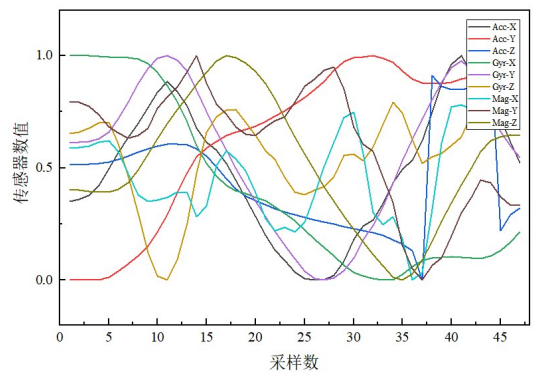
(*e*) “惊讶”手势的手掌关节点的欧拉角变化 (*f*) “谁”手势的手掌关节点的欧拉角变化

图11 CSL数据集手势欧拉角

轨迹具有一定的重叠性,手臂摆动的幅度和变化的频率存在一定的相似性.



(a)　“Hurry up”手势



(b)　“Vehicle”手势

图 12　CRE 数据集

4.4　模型消融实验

本文提出的 STGNN-HGR 模型主要包括了时空图神经网络与 GRU 模块. 其中,时空图神经网络用于提取手臂、手掌上关节点的空间连接特征;GRU 用于捕获手势数据的时序特征. 为了验证 STGNN-HGR 模型的表征能力与动作识别的有效性,本节设计了消融实验来分析模型组成对手势识别结果的影响,在 CRE 与 CSL 两种数据集上进行了验证,在消融实验中,设置学习率为 0.001,迭代次数为 200. 实验结果如表 5 所示,STGNN-HGR-w/o GCN 表示在 STGNN-HGR 基础上去除 GCN 模块,只保留 GRU 模块用于手势识别,用于验证 GCN 模块是否能提升手势识别的准确率. 由于本文识别的手势还拥有长距离依赖的特征,无法直接去除 GRU 模块进行分析,因此,采用长短时记忆单元(LSTM)替换 STGNN-HGR 中的 GRU 模块,表示为 STGNN-HGR-LSTM,通过这种方式验证 GRU 模块是否能更好地提取时序化手势数据特征.

手势空间具有连接关系,图结构能够更好地表达这种空间连接关系,借助图卷积网络能够更好地识别

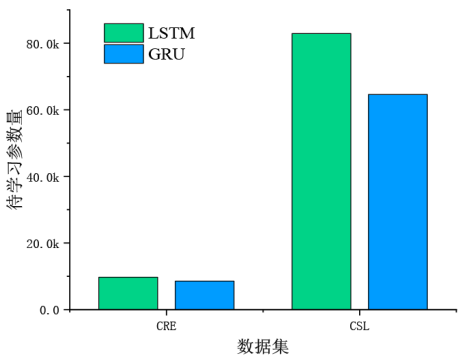
表 5　消融实验结果

算法名称	CRE 数据集	CSL 数据集
STGNN-HGR-w/o GCN	93.79	94.18
STGNN-HGR-LSTM	96.18	96.13
STGNN-HGR	96.82	97.27

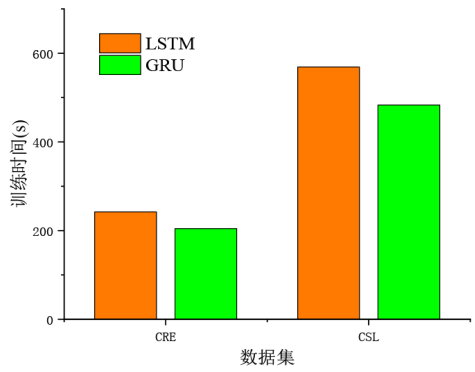
这种复杂手势. 因此,在手势识别模型 STGNN-HGR 中将用于空间连接关系计算的 GCN 消除以后,会导致明显的精度降低,如表 5 中的第一行所示,在两种数据集上均有显著的识别精度损失. 尽管 LSTM 也能够表达复杂手势的时序信息,但是在使用 LSTM 替换 GRU 时也带来了轻微精度损失,如表 5 的第二行所示,其主要原因为 GRU 单元中的参数相比于 LSTM 更少,在长时序的手势数据上训练得到的模型泛化能力更优,因此在未知的测试集上识别效果高于 LSTM.

此外,GRU 相比于 LSTM,具有待学习参数少、训练时间短的特性,因此本文还比较了 GRU 与 LSTM 在模型参数量以及训练时间上的差异,实验结果如图 13 所示.

图 13(a)与图 13(b)分别给出了 STGNN-HGR 使用 GRU 与 LSTM 进行手势识别时,待学习参数量与训练时间上的差异. 从图 13(a)中可以看出,相比于 LSTM,在手势识别中使用 GRU 可以有效地减少模型中待学习的



(a)　GRU 与 LSTM 参数量对比



(b)　GRU 与 LSTM 训练时间对比

图 13　GRU 与 LSTM 性能对比

参数以提升模型的泛化能力. 此外,由图13(*b*)中的训练时间对比可知,使用GRU去捕获时序手势特征可以显著降低模型训练所需的时间.

综上所述,由于STGNN-HGR的手势识别准确率均高于两种消融对比方法,证明了本文综合考虑复杂手势的空间连接性以及手势的时序特征,利用图结构、图神经网络提取手势的空间特征信息,且使用GRU捕获手势动作的时序特征,并将两种特征信息用于手势识别,能有效地提升手势识别的准确率,取得了较好的识别效果.

5 结论

本文提出了基于时空图神经网络的手势识别模型STGNN-HGR,通过将关节点建模为图结构并使用时空图神经网络,有效地提取手势数据的空间关联信息,表征了人体关节点的空间关联性,充分挖掘关节点之间的依赖关系. 此外,STGNN-HGR利用门控循环单元解决了手势的时序性和长距离依赖问题. 实验结果表明,STGNN-HGR能够有效地识别手势,在中国手语数据集识别准确率为97.2%,在标准军队动态手势数据集上的识别准确率为98.63%,均优于对比算法的识别效果.

参考文献

[1] 李艳德. 基于穿戴传感感知的手势识别模型与应用研究[D]. 兰州: 兰州大学, 2019.
LI Y D. Research on Gesture Recognition Model and Its Application Based on Wear Sensing Perception[D]. Lanzhou: Lanzhou University, 2019.

[2] 王勇, 王沙沙, 田增山, 等. 基于FMCW雷达的双流融合神经网络手势识别方法[J]. 电子学报, 2019, 47(7): 1408-1415.
WANG Y, WANG S S, TIAN Z S, et al. Two-stream fusion neural network approach for hand gesture recognition based on fmcw radar[J]. Acta Electronica Sinica, 2019, 47(7): 1408-1415. (in Chinese)

[3] 冯志全, 杨学文, 徐涛, 等. 结合手势二进制编码和类-Hausdorff距离的手势识别[J]. 电子学报, 2017, 45(9): 2281-2291.
FENG Z Q, YANG X W, XU T, et al. Gesture recognition based on combining gesture binary descriptor and hausdorff-like distance[J]. Acta Electronica Sinica, 2017, 45(9): 2281-2291. (in Chinese)

[4] KIM M, JEON C, KIM J, et al. A study on immersion and presence of a portable hand haptic system for immersive virtual reality[J]. Sensors, 2017, 17(5): 1141-1158.

[5] LI R Q, ZHENG D W, HAN Z Y, et al. MHealth: A smartphone-controlled, wearable platform for tumour treatment [J]. Materials Today, 2020, 40(11): 91-100.

[6] ISON M, VUJAKLIJA I, WHITSELL B, et al. High-density electromyography and motor skill learning for robust long-term control of a 7-DoF robot arm[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2015, 24(4): 424-433.

[7] FANG B, SUN F, LIU H, et al. A novel data glove using inertial and magnetic sensors for motion capture and robotic arm-hand teleoperation[J]. Industrial Robot: An International Journal, 2017, 44(2): 155-165.

[8] CHOU P H, HSU Y L, LEE W L, et al. Development of a smart home system based on multi-sensor data fusion technology[C]//Proceedings of IEEE International Conference on Applied System Innovation. Sapporo: IEEE, 2017: 690-693.

[9] MA B, CHEN B, ZHANG Z, et al. Combat gesture classification using through-the-wall radar based on multi-domain features association[C]//Proceedings of IEEE Radar Conference. Florence: IEEE, 2020: 1-5.

[10] 李愚, 柴国钟, 卢纯福, 等. 基于增量自适应学习的在线肌电手势识别[J]. 计算机科学, 2019, 46(4): 274-279.
LI Y, CHAI G Z, LU C F, et al. On-line sEMG hand gesture recognition based on incremental adaptive learning [J]. Computer Science, 2019, 46(4): 274-279. (in Chinese)

[11] 陈国良, 葛凯凯, 李聪浩. 基于多特征HMM融合的复杂动态手势识别[J]. 华中科技大学学报(自然科学版), 2018, 46(12): 42-47.
CHEN G L, GE K K, LI C H. Complex dynamic gesture recognition based on multiple features and HMM fusion [J]. Journal of Huazhong University of Science and Technology(Natural Science Edition), 2018, 46(12): 42-47. (in Chinese)

[12] CHEN W, ZHANG Z. Hand gesture recognition using sEMG signals based on support vector machine[C]//Proceedings of the 8th Joint International Information Technology and Artificial Intelligence Conference. Chongqing: IEEE, 2019: 230-234.

[13] KUMAR P, GAUBA H, ROY P P, et al. Coupled HMM-based multi-sensor data fusion for sign language recognition[J]. Pattern Recognition Letters, 2017, 86(1): 1-8.

[14] PURUSHOTHAMAN A, PALANISWAMY S. Development of smart home using gesture recognition for elderly and disabled[J]. Journal of Computational and Theoretical Nanoscience, 2020, 17(1): 177-181.

[15] CHEN L, FU J, WU Y, et al. Hand gesture recognition using compact CNN via surface electromyography signals [J]. Sensors, 2020, 20(3): 672-680.
[16] SHIN S, KIM W Y. Skeleton-based dynamic hand gesture recognition using a part-based gru-rnn for gesture-based interface[J]. IEEE Access, 2020, 8(3): 50236-50243.
[17] CHEN Y, MA G, YUAN C, et al. Graph convolutional network with structure pooling and joint-wise channel attention for action recognition[J]. Pattern Recognition, 2020, 103(7): 107321-103731.
[18] NUNEZ J C, CABIDO R, PANTRIGO J J, et al. Convolutional neural networks and long short-term memory for skeleton-based human activity and hand gesture recognition[J]. Pattern Recognition, 2018, 76(4): 80-94.
[19] CHEN H, TONG R, CHEN M, et al. A hybrid cnn-svm classifier for hand gesture recognition with surface emg signals[C]//Proceedings of the 17th International Conference on Machine Learning and Cybernetics. Chengdu: IEEE, 2018: 619-624.
[20] 刘肖, 袁冠, 张艳梅, 等. 基于自适应多分类器融合的手势识别[J]. 计算机科学, 2020, 47(7): 103-110.
LIU X, YUAN G, ZHANG Y M, et al. Hand gesture recognition based on self-adaptive multi-classifiers fusion[J]. Computer Science, 2020, 47(7): 103-110. (in Chinese)
[21] ZHU G, ZHANG L, YANG L, et al. Redundancy and attention in convolutional LSTM for gesture recognition[J]. IEEE Transactions on Neural Networks and Learning Systems, 2019, 31(4): 1323-1335.
[22] YAN S, XIONG Y, LIN D. Spatial temporal graph convolutional networks for skeleton-based action recognition [C]//Thirty-second AAAI conference on artificial intelligence. New Orleans: AAAI, 2018: 1-10.
[23] 白铂, 刘玉婷, 马驰骋, 等. 图神经网络[J]. 中国科学: 数学, 2020, 50(3): 367-384.
BAI B, LIU Y T, MA C C, et al. Graph neural networks [J]. Scientia Sinica: Mathematica, 2020, 50(3): 367-384. (in Chinese)
[24] ZHANG X, CHEN X, LI Y. A framework for hand gesture recognition based on accelerometer and EMG sensors [J]. IEEE Transactions on Systems Man & Cybernetics Part A Systems & Humans, 2011, 41(6): 1064-1076.
[25] TIDWELL R, AKUMALLA S, KARLAPUTI S, et al. Evaluating the feasibility of EMG and bend sensors for classifying hand gestures[C]//Proceedings of the International Conference on Multimedia and Human Computer Interaction. Toronto: ISSN, 2013: 1-8.
[26] WAHID M F, TAFRESHI R, AL-SOWAIDI M, et al. Subject-independent hand gesture recognition using normalization and machine learning algorithms[J]. Journal of Computational Science, 2018, 27(7): 69-76.
[27] SHARMA S, MODI S, RANA P S, et al. Hand gesture recognition using Gaussian threshold and different SVM kernels[C]//Proceedings of International Conference on Advances in Computing and Data Sciences. Dehradun: Springer, 2018: 138-147.
[28] LI G, TANG H, SUN Y, et al. Hand gesture recognition based on convolution neural network[J]. Cluster Computing, 2019, 22(2): 2719-2729.
[29] KHODABANDELOU G, JUNG P G, AMIRAT Y, et al. Attention-based gated recurrent unit for gesture recognition[J]. IEEE Transactions on Automation Science and Engineering, 2020, 18(2): 495-507.

作者简介

袁　冠　男, 1982 年生, 江苏睢宁人. 现为中国矿业大学计算机科学与技术学院教授. 主要研究方向为时空大数据技术以及计算智能.
E-mail: yuanguan@cumt.edu.cn

邴　睿　男, 1994 年生, 甘肃兰州人. 现为中国矿业大学博士生. 主要研究方向为图数据挖掘.
E-mail: bingrui@cumt.edu.cn

刘　肖　男, 1994 年生, 江苏徐州人. 主要研究方向为模式识别与感知计算.
E-mail: liuxiaocumt2018@163.com