

基于伪节点交叉注意力的远程步态情绪识别

卢亮宇, 周成菊

(华南师范大学人工智能学院, 广东 佛山 528000)

摘要: 近年来, 情绪识别在心理计算、人机交互和精神状态监测中的应用引起了广泛关注。与面部情绪识别和脑电情绪识别(EEG)等其他方式相比, 步态情绪识别所使用的采集无需高精度拍摄, 且可以不用佩戴专门的采集设备进行远距离采集。尽管该领域已经开展了一系列研究并取得了相应进展, 但目前仍面临两个主要挑战。一是现有大多数基于步态情绪识别的工作都侧重于通过图卷积网络(GCN)从骨骼图像中探索人体关节的局部相关性, 而忽略了人体关节的全局相关性; 二是使用了人体自然连接关节骨架图, 原有的固定连接会限制网络捕捉远距离关节之间相互作用的能力。为了解决这些问题, 提出了一种基于伪节点交叉注意力的图卷积网络, 通过伪节点的方法有效地实现全局和局部关节节点的信息及时传递, 并使用交叉注意力方法捕获有效和高效的步态表示以进行情绪状态识别。将所提出的方法在情绪步态数据集 Emotion-Gait 上进行评估, 准确率达到 88.63%, 与已有经典先进模型相比性能更优。

关键词: 步态情绪识别; 交叉注意力; 图卷积神经网络; 关节邻接矩阵

DOI: 10.11907/rjdk.241870

中图分类号: TP391.41

文献标识码: A

文章编号: 1672-7800(2026)001-0047-07

扫描二维码阅读全文:



Long Distance Gait Emotion Recognition Based on Pseudo Node Cross Attention

LU Liangyu, ZHOU Chengju

(School of Artificial Intelligence, South China Normal University, Foshan 528000, China)

Abstract: In recent years, the application of emotion recognition in mental computing, human-computer interaction, and mental state monitoring has attracted considerable attention. Compared with other methods such as facial and electroencephalogram (EEG) emotion recognition, gait data does not require high-precision shooting equipment and can be collected at a long distance. Although some research has been conducted in this field and corresponding progress has been made, there are still two problems that remain unsolved. First, most of the existing work on gait-based emotion recognition focuses on exploring the local correlation of human joints from skeleton images through graph convolutional networks (GCNs), while ignoring the global correlation of human joints. Second, the human body's naturally connected joint skeleton graph is used, and the original fixed connection will limit the network's ability to capture the interaction between distant joints. To solve these problems, this paper proposes a graph convolutional network based on pseudo-node cross-attention, which realizes the timely transmission of information of all joints through the pseudo-node method and uses the cross-attention method to capture efficient gait representation for emotional state recognition. The proposed method is evaluated on the dataset Emotion-Gait, and the accuracy rate reaches 88.63%, it has better performance compared with the existing classic advanced models.

Key Words: gait emotion recognition; cross attention; graph convolutional networks; joint adjacency matrix

0 引言

情绪在人类日常生活中扮演着重要角色, 对人类的思维方式、行为模式和社交关系有着深远影响, 具有极高的应用潜力。目前, 情绪识别在行为预测^[1]、安防监控^[2]、治

疗康复^[3]、患者心理健康监测^[4]等方面得到了广泛应用, 因而远距离准确高效且简易可行地识别人类情绪尤为重要。通过识别人类情绪, 可以更深入地了解自身和他人的情感状态, 人机交互、人与人的交流更加顺畅, 社会也因此更加和谐。

目前, 感知人类情绪的主流方式有面部表情、言语、生

收稿日期: 2024-11-18

作者简介: 卢亮宇(2000-), 男, 华南师范大学人工智能学院硕士研究生, 研究方向为步态识别与情绪识别; 周成菊(1986-), 女, 博士, 华南师范大学人工智能学院副研究员、硕士生导师, 研究方向为步态识别、计算机视觉。本文通讯作者: 卢亮宇。

理信号和肢体动作等。通过面部表情进行情绪识别,一方面需要高精度的面部图像,对拍摄设备以及被测人员与拍摄设备的距离有很高要求;另一方面,面部表情容易伪装,有时候无法反映被测人员内心的真实情感。通过言语进行情绪识别,同样存在无法反映被测人员内心真实情感的问题,因而并非总是可靠。通过脑电信号进行情绪识别需要被测人员配戴专业的脑电感知设备,这给在生活中通过脑电方式感知人类情绪带来了一定限制。基于步态的动作识别已被证明对于人类动作识别有效,研究人员开始意识到步态在情绪识别中广泛应用的可能性。本文重点分析了人类日常生活中最常见的动作——步行,步行可以从远距离观察到被测人的情绪状态,同时具有较高的情绪识别精度。

近年来,心理学研究表明,步态还可以反映人的情绪和心理状态^[5-7]。通过步态进行情绪识别引起了越来越多的关注,特别是,随着人体姿势估计算法和深度传感器技术的不断发展^[8-9],使用步态以感知情绪已成为当前研究热点^[10-11]。情感状态可以通过步态加以识别,处于不同情绪状态的人会因为步幅、步频等形成不一样的走路方式,因此可以直观地从日常生活中的步态或姿势感知他人的情绪。现有的基于步态的情绪识别方法通常使用RGB视频或人体骨骼数据。RGB视频包含直观的场景信息,但很容易受到光照、服饰遮挡等因素影响。人体骨骼数据是人体关节和骨骼的拓扑表示,骨骼关节数据表示视频每一帧的人体关节点三维坐标,包括骨骼空间上的结构信息和时序动态信息。使用骨骼进行步态情绪识别的优势是可以进行远距离的步态数据采集,可避免光线、遮挡物和复杂背景干扰,因而本文主要基于人体骨骼开展步态情绪识别。

根据数据格式不同,可将基于人体骨骼的研究方法分为两类:基于图像的方法和基于骨架图的方法^[12]。基于图像的方法根据不同的规则将骨骼关节数据联合坐标数据编码为图像,并设计卷积神经网络(Convolutional Neural Networks, CNN)对图像进行分类。这种方法的缺点是由于卷积核的局部感知能力,在较长的几何距离下挖掘空间中不相邻的关节关系效率低下。基于骨架图的方法是将人类骨骼连接的关节看作是非欧几里得几何空间中的图,并且可以用图卷积网络(Graph Convolutional Network, GCN)进行处理。基于GCN的方法发挥着不可替代的作用,提高了基于骨骼的步态情绪识别能力。这种方法有效地挖掘了关节之间的关系,但该方法有两个隐藏的缺点。首先,区分复杂情绪需要捕捉步态序列中关节的高阶相关性。而大多数基于GCN的方法使用现有的骨架图(固定的物理连接骨架结构),这限制了人体关节之间高阶信息的捕获。例如,虽然在行走过程中手和脚之间的相关性很明显,但在图卷积扩散过程中,这种相关性可能会减弱。文献[13]提出使用时空图卷积网络(Spatial Temporal Graph Convolutional Network, ST-GCN)进行步态情绪识别。该方法是将

人体骨骼的自然连接结构当作一个固定不变的空间上的图结构。但是该方法存在的问题是:仅靠人体关节自然连接图只能捕获关节之间在空间上的局部相关性,无法捕获全局的非相邻关节之间的关系,全局关节关系对于情绪识别是很重要的一环。文献[14]提出一种改进的ST-GCN,使用全局连接以补充人体自然连接结构。但该方法并没有妥善解决上述问题,其选择了一个中心关节与其他所有关节连接,在多层GCN迭代后,这种连接方式会导致中心关节具有比边缘关节更大的感受范围,所接受的信息会更多,进而会导致信息不平衡,从而忽略了边缘关节对情绪识别的作用。文献[15]提出一种基于伪节点的策略以获取关节全局信息和局部信息,通过添加伪节点并将其与其他自然节点结合起来,以捕获关节全局信息和局部信息,同时缓解信息不平衡问题。但该方法所捕获的关节全局信息和关节局部信息没有进行交互,会导致关节信息存在冗余。同时,也没有对全局信息与局部信息的权重进行分配,可能存在单一动作权重过高从而对步态整体情绪影响过高的现象。

本文针对上述问题,首先对关节拓扑结构进行优化,不再设定固定的人体自然连接关节,关节拓扑结构由数据驱动,让关节之间的信息传递不再受限;其次,本文在原有关节节点的基础上添加伪节点,以获取关节全局信息;最后,设计了一个基于伪节点的交叉注意力模块,利用伪节点获取的全局信息与关节局部信息进行交互,聚焦于对情绪识别有帮助的关节连接。

本文主要贡献如下:①优化关节拓扑结构,使用数据驱动的骨架图结构,关节拓扑结构由数据决定,让关节之间的信息传递不再受到原有固定连接的限制;②设计了一个基于伪节点的交叉注意力模块(Cross Attention Module, CAM),利用注意力机制以捕捉关节之间以及伪节点与关节之间蕴含的情绪信息;③在Emotion-Gait数据集上进行实验,情绪识别准确率达到了88.63%,优于当前经典先进模型性能。

1 相关工作

1.1 拓扑改进方法

不同的关节拓扑结构会影响基于骨骼的动作和步态识别准确性,在动作识别中,研究者们首次探索了更灵活的骨骼连接方案。文献[16]提出时空路由的概念,使用子群聚类以探索空间维度上的连接关系,并通过测量节点轨迹的相关程度,在时间维度上找到结构信息。文献[17]提出一个编码—解码模型以捕捉特定动作的潜在依赖关系,通过动作链接和结构链接扩展了现有骨骼图,以表示更高阶的依赖关系。文献[18]提出一种自适应图卷积网络(Adaptive Graph Convolutional Network, AGCN),通过3个子图自适应调整拓扑结构。AGCN使得图拓扑结构与图卷积

网络一并优化,极大提高了模型灵活性。文献[19]提出通道级别的拓扑,动态地学习不同通道的拓扑结构。然而,在行走过程中,不同关节的轨迹相似,这使得上述模型难以探索有价值的拓扑结构。因此,需要提出更加明确和通用的关节拓扑结构。在基于步态的情感识别中,文献[14]采用了全局链接,即以根关节为起点连接其他关节的星型结构,以捕捉步态的整体状态。以上方法为寻找新的骨架图拓扑结构进行了有价值的探索,但图拓扑结构仍有优化空间。本文通过完全依靠数据生成骨架图拓扑结构,并添加交叉注意力机制对情绪信息丰富的关节连接进行加强,并开展相关实验证明其有效性,弥补了原有方法的骨架拓扑图存在固定连接限制关节连接关系探索的问题。

1.2 注意力机制方法

人类通过关注情感丰富的内容以辨别目标情绪^[20]。例如,人们通常集中在人肢体的关键区域,如面部的眼睛和嘴巴,以迅速解读他人的情绪状态^[21]。同样,在语音情绪感知中,个体倾向于分析声音响亮的部分或那些浸润着更高浓度情感线索的部分^[22]。在这些观察的指导下,文献[16]提出一种卷积神经网络(CNN)架构,其表层网络层定位于富含情感信息的面部区域,该方法使模型学习到比以往方法更具有区分性的特征。文献[23]设计了一种创新的3D残差注意力机制,以强调富含表情的区域。文献[24]提出一种基于面部的注意图并将其嵌入到选定区域的局部情感特征学习方法。此外,文献[25]引入深度区域和多标签学习方法(DRML),通过区域层专注于最具情感信息的区域。在语音情绪识别领域,文献[26]开发了一种

Transformer模型,通过显著性图衡量情感信息,在时间维度上整合了以帧为单位的情感信息。文献[27]利用相对属性为每一帧分配情感强度。此外,文献[28]提出一种基于注意力的三层卷积循环神经网络,以增强情感信息的表达。在通过步态识别情绪的任务中,文献[13]通过每个关节反向传播的数值展示了不同情绪下(如愤怒、快乐和悲伤)关节的重要性。以上方法仅展示了重要关节,还无法有效捕捉重要关节之间的关系,也没有将这些信息应用于网络训练过程。为了让模型像人类一样关注情感丰富的内容,本文提出了一个交叉注意力模块(Cross Attention Module, CAM),该方法使模型能够自适应地集中关注富含情感信息的关节,从而更有效地提取有区分性的情感特征。

2 模型方法

2.1 模型总体框架

本文模型总体框架如图1所示,主要由ST-GCN网络和交叉注意力模块(CAM)组成。首先,为了同时捕获整个骨架上相邻关节与非相邻关节之间的拓扑关系,在关节图中添加伪节点,并为ST-GCN构建一个表示伪节点的扩展邻接矩阵;其次,步态数据经过三层ST-GCN网络训练,将训练所得的步态特征传递给CAM模块,CAM中的多个交叉注意力子模块分别用于检测不同身体块的局部关节信息和全局关节信息;最后,通过MLP层和Softmax层输出情感分类结果。

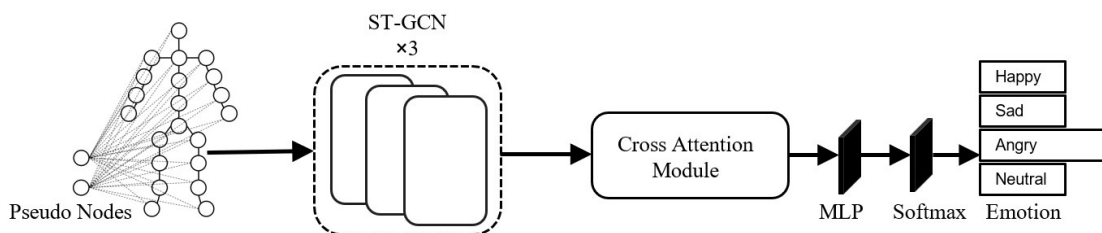


Fig. 1 The overall framework of the proposed model

图1 模型总体框架

2.2 基于伪节点的交叉注意力模块

为解决基于骨骼的步态情绪识别中,关节全局信息与关节局部信息交互不足的问题,本文提出了交叉注意力模块CAM。交叉注意力机制能够突出重要关节关系,同时抑制一般关节关系,为模型带来全局关节信息。CAM具体结构如图2所示,主要由4个子模块构成,分别是手部交叉注意力模块(Hand Cross Attention, HCA)、腿部交叉注意力模块(Leg Cross Attention, LCA)、手腿交叉注意力模块(Hand-Leg Cross Attention, HLCA)和伪节点交叉注意力模块(Pseudo Nodes Cross Attention, PNCA)。HCA模块通过对手部关节和剩余部位的关节使用交叉注意力,计算每个手部关节的特征注意力分数,利用手部关节信息以提取有用情感信息。LCA模块通

过对腿部关节和其余部位的关节使用交叉注意力,计算每个腿部关节的特征注意力分数,利用腿部关节信息提取有用情感信息。HLCA模块将HCA模块和LCA模块提取出的手部、腿部关节情感信息再使用一次注意力机制,计算手部关节和腿部关节的特征注意力分数,使模型关注到手部和腿部关节中联系紧密的关节对,提取到更深层的信息,并对两个模块所提取的信息差异进行比较,以更有效地提取出上肢关节和下肢关节具有区分性的情感特征。PNCA模块对伪节点和所有人体关节节点使用交叉注意力,将伪节点中的关节全局信息与其余节点中的信息进行比较,从而关注到非相邻关节之间有用的情感特征并将其提取出来。最后,CAM模块聚合HLCA子模块和PNCA子模块中有用的情感特征并将其输出。

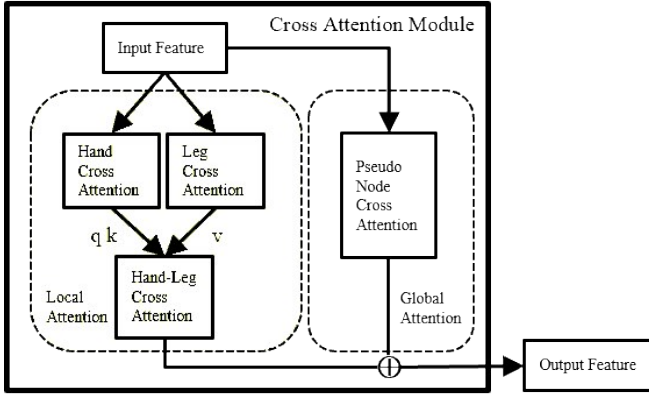


Fig. 2 The architecture of Cross Attention Module

图2 Cross Attention Module 结构

本文对 CAM 模块的运行过程进行描述。首先,接收 ST-GCN 网络的输出特征 $X_1 \in \mathbb{R}^{N \times T \times C}$ 。 X_1 按照上肢与下肢关节进行划分与分割,并传递到两个单一身体部分交叉注意力(Hand Cross Attention, Leg Cross Attention)分支。HCA 自适应地区分手关节与其余关节的关系,产生特征 $\{X_2^H \in \mathbb{R}^{H \times T \times C}\}$,其中 H 表示手部关节。LCA 区分腿脚关节与其余关节的关系,生成特征 $\{X_2^F \in \mathbb{R}^{F \times T \times C}\}$,其中 F 表示腿部关节。下一步,分别经过 HCA 和 LCA 处理后的特征 X_2^H 和 X_2^F 通过 Hand-Leg Cross Attention 自适应地区分上肢关节与下肢关节的关系,产生特征 $X_3^{UD} \in \mathbb{R}^{N \times T \times C}$ 。PNCA 用来区分伪节点与全身关节之间的关系,生成特征 $X_3^P \in \mathbb{R}^{N \times T \times C}$,其中 N 是全身关节数量。将 X_3^{UD} 和 X_3^P 逐元素相加后使用 MLP 进行处理以产生特征,最终通过 Softmax 层输出情绪分类结果。

为了获得局部关节信息,进而对细微的关节动作所蕴含的情绪信息加以利用,本文使用手部交叉注意力(HCA)和腿部交叉注意力(LCA)分别在手部关节和腿部关节找出情绪信息丰富的关节点。在此基础上,进一步通过手腿交叉注意力(HLCA)获取手部和腿部关节的空间信息。伪节点交叉注意力(PNCA)用于获取全局关节信息。HCA、LCA 和 PNCA 具有相同的内部结构,区别在于输入以及关节的划分方式。通过关注身体不同部位并使用不同的注意力机制这种方法使模型能够从步态序列中捕捉局部和全局的情绪信息,从而增强模型的情绪识别能力。

为了让信息在关节和身体部位之间扩散,引入一个带有交叉注意力子块的单个身体部位交叉注意力块。每个交叉注意力块使用多头交叉注意力机制以交互、扩散主分支(L 分支)和小分支(S 分支)的特征。本文对 LCA 的交叉注意力块的运行过程进行描述: L 分支的输入为 $\{X_2^F \in \mathbb{R}^{F \times T \times C}\}$,其中 F 表示通过 1×1 卷积和空间变换提取的腿部关节; S 分支的输入为 $\{X_2^0 \in \mathbb{R}^{O \times T \times C}\}$,其中 O 表示通过 1×1 卷积和空间变换提取的除腿部关节外的其余关节。对于分支 L ,它从分支 S 收集 Token 并将它们连接起来。其计算公式如式(1)所示。

$$(x^F)_{cls}^l = f^l((x^F)_{cls}^l), (x^F)^l = [(x^F)_{cls}^l] \parallel [(x^0)^s]_{d'-cls}^s \quad (1)$$

其中, $f^l()$ 表示尺寸对齐的投影函数。在数学上,交叉注意力可表示为如式(2)所示。

$$A = \text{softmax} \left(\frac{qk^T}{\sqrt{\frac{C}{h}}} \right), CA((x^F)^l) = Av \quad (2)$$

其中, $q = (x^F)_{cls}^l W_q, k = (x^F)^l W_k, v = (x^F)^l W_v$ 。

$W_q, W_k, W_v \in \mathbb{R}^{C \times (Ch)}$ 是可学习的参数, C 和 h 分别是嵌入的维度和注意力头的数量, A 表示线性的交叉注意力图。

L 分支交叉注意力模块的输出 $(y^F)_2$ 定义如式(3)所示。

$$(y^F)_2 = [g^l((y^F)_{cls}^l) \parallel (x^F)_{d'-cls}^l] \quad (3)$$

其中, $f^l()$ 和 $g^l()$ 分别为 n 维对齐的投影和反投影函数。最后,如式(4)所示,经过 MLP 的对齐和 FFN 多层感知器后,得到 L 分支的输出 X_2^F , S 分支结构与 L 分支相同但参数不共享。

$$z_2^F = f_l[(y^F)_2] \parallel f_s[(y^0)_2], X_2^F = z_2^F + \text{FFN}(z_2^F) \quad (4)$$

HLCA 子模块的作用是为了获得上肢关节与下肢关节所蕴含情感信息的差异,探索上下肢关节非相邻关节之间的联系。HLCA 接受来自 HCA 与 LCA 两个子模块的输出 X_2^H, X_2^F 作为输入,与上述 LCA 一样,模块内具有 L 分支与 S 分支,也使用了多头交叉注意力对 L 分支与 S 分支进行交互并找出两部分的特征差异。对于 L 分支,输入为 $\{X_3^{UD} \in \mathbb{R}^{UD \times T \times C}\}$,其中 U 所代表的关节特征来自于 HCA, D 所代表的关节特征来自于 LCA,即 X_3^{UD} 是由两个输入拼凑而成。同理对于 S 分支,输入也为 $\{X_3^{UD} \in \mathbb{R}^{UD \times T \times C}\}$,其中 U 所代表的关节特征来自于 LCA, D 所代表的关节特征来自于 HCA。最终,与 LCA 类似,经过 MLP、FFN 层处理后,输出特征 X_3^{UD} 。

PNCA 子模块的作用是为了获得伪节点与全身自然关节节点所蕴含关节全局信息的差异,探索全局关节更深层的联系。PNCA 内部结构与 HCA、LCA 类似,但将输入分为 $\{X_2^P \in \mathbb{R}^{P \times T \times C}\}$ 和 $\{X_2^B \in \mathbb{R}^{B \times T \times C}\}$,分别传入 L 分支和 S 分支中,最后经过一系列与 LCA 相同的处理,输出特征 X_3^{PB} 。

最终, CAM 模块聚合特征 X_3^{UD} 和特征 X_3^{PB} ,经过平均池化后进行输出,传入 MLP 层和 Softmax 层并输出情感分类结果。

3 实验与结果分析

3.1 数据集

本文使用的 Emotion-Gait 数据集^[23]是由数个 3D 姿势序列数据集收集的步态组成,包括经典数据集 BML^[30]、Human3.6M^[31]、ICT^[32]、CMU-MoCap^[33]和 ELMD^[34],以及新采集的步态数据。为了保持不同来源的姿势模型关节集的统一,便于训练网络,Emotion-Gait 对各种步态数据进行

全面且统一的处理,将视频、图片和图序列步态数据都转换为ELMD中使用的21关节姿势模型,使数据集具有更高的标准化程度和可比性^[34]。每一个步态样本的时间长度均为240帧。在输入中每5帧采样一次,并将生成的48帧作为输入数据传递给图卷积神经网络,即所有数据的 $T=48$ 。该数据集总共由3 924个步态序列组成,其中有1 835个步态序列带有情绪标签,其余2 089个步态序列未有情绪标签。在带标签数据中,大约58%的步态为快乐情感,32%的步态为悲伤情感,23%的步态为愤怒情感,只有14%的步态为中性情感。

3.2 训练策略

优化器能在训练中持续调整网络参数、加速训练过程。其中,Adam由于自适应学习率使得参数调整更加灵活高效,其实现简单且性能优良。因此,本文在训练中采用Adam优化网络模型。模型训练超参数配置如表1所示。

Table 1 Hyperparameter configuration for model training

表1 模型训练的超参数配置

训练超参数		参数值
Epoch	训练周期	500
Batch-size	批尺寸	8
optimizer	优化器	Adam
momentum	动量	0.9
Learning rate	初始学习率	0.1
Weight decay	权重衰减	1×10^{-4}
p	Dropout 舍弃概率	0.2

3.3 评估方法

本文采用评估指标准确率(Accuracy)以评估所提出模型的性能,如式(5)所示。

$$ACC = (TP + TN) / TD \tag{5}$$

其中,TP表示预测真阳性的步态样本数量,TN表示预测真阴性的样本数量,TD表示步态样本总数量。

3.3 实验结果

3.3.1 不同拓扑结构的关节邻接矩阵比较

本文在Emotion-Gait数据集上进行了一系列实验,评估以数据驱动的关节拓扑结构结合伪节点引入ST-GCN网络的有效性。首先,以单一关节拓扑结构进行情绪识别的准确率衡量网络模型性能;其次,将多个关节拓扑结构组合进行比较,以衡量网络模型性能。实验结果如表2所示。

其中,Ak、Bk、Ck均为邻接矩阵,Ak表示固定连接的人体骨架拓扑图,Bk表示没有固定连接的由数据驱动的骨架拓扑图,Ck表示为每一个步态样本生成一个独有的由该样本决定的骨架拓扑图,PN表示为关节邻接矩阵添加伪节点。各项实验均在Emotion-Gait数据集中使用三层ST-GCN图卷积神经网络,除邻接矩阵外,其余参数保持一致。

由表2结果可知,只使用由数据驱动的骨架拓扑图具

Table 2 Joint adjacency matrices comparison of different topological structures

表2 不同拓扑结构的关节邻接矩阵比较

关节拓扑结构	准确率/%	Train 训练时间/s
Ak + PN	81.02	122
Bk + PN	83.21	261
Ck + PN	77.01	180
Ak + Bk + PN	80.16	297
Ak + Ck + PN	78.38	201
Bk + Ck + PN	79.25	340
Ak + Bk + Ck + PN	82.54	384

有更高的情绪识别准确率;只使用Bk相较于只使用Ak,在Emotion-Gait数据集上的情绪识别准确率提升2.19%;只使用Bk相较于只使用Ck的情绪识别准确率提升6.20%。Ak因为固定的关节连接限制了模型对关节关系的探索,Ck因数据集样本数量少,其对每个样本生成的骨架拓扑图泛化性不足,从而影响了情绪识别准确率。而本文使用带有Bk的模型具有更灵活的关节拓扑图生成方式,所生成的关节间关系不仅仅局限于原有的关节连接,而是基于步态数据去探索对情绪识别有帮助的关节对,并在邻接矩阵中对这些关节对的权重进行自适应调整。

3.3.2 CAM模块的有效性验证

CAM模块由HLCA和PNCA两个子模块构成,HLCA表示Hand-Leg Cross Attention子模块,PNCA表示Pseudo Node Cross Attention子模块。本文为了评估在模型中添加基于伪节点的交叉注意力模块的有效性,将ST-GCN和CTR-GCN作为Baseline,依次插入HLCA子模块、PNCA子模块以及两个子模块的结合,结果如表3所示。由此可知,相较于原始Baseline,添加了注意力模块的模型在Emotion-Gait数据集上的识别准确率有所提升。具体而言,ST-GCN融入HLCA子模块后,准确率提升至87.67%;融入PNCA子模块后,准确率提升至86.55%;融入HLCA、PNCA子模块后,准确率提升至82.77%;融入PNCA子模块后,准确率提升至83.93%;融入HLCA、PNCA子模块后,准确率提升至85.12%。

实验结果表明,在两个Baseline,即ST-GCN和CTR-GCN中,CAM各子模块以及子模块的组合均对模型在

Table 3 Ablation experiments of CAM module on Emotion-Gait dataset

表3 CAM模块在Emotion-Gait数据集上的消融实验

网络模型	准确率/%	Train 训练时间/s
ST-GCN	81.02	261
CTR-GCN	81.66	320
ST-GCN + HLCA	87.67	627
ST-GCN + PNCA	86.55	591
CTR-GCN + HLCA	82.77	722
CTR-GCN + PNCA	83.93	718
CTR-GCN + HLCA + PNCA	85.12	1 504
ST-GCN + HLCA + PNCA	88.63	1 250

Emotion-Gait 上的情绪识别准确率有所提升,这表明 HLCA 子模块和 PNCA 子模块都有助于更好地识别步态情绪。对手关节和腿关节进行交叉注意力,能够有效提取出手和腿的局部关节特征,表明 HLCA 子模块可有效地挖掘上肢关节与下肢关节之间的关系,并能够利用这些不相邻关节关系提取出相应的情绪特征,有助于模型进行情绪识别。而 PNCA 子模块则利用伪节点帮助全身关节进行情绪信息传递,从而帮助非相邻关节更好地进行信息交换,进而帮助模型挖掘更深层的关节全局关系。同时,实验结果也表明,当 HLCA 子模块和 PNCA 子模块组合使用时,其对于 ST-GCN 和 CTR-GCN 模型性能均有所提升。通过添加伪节点,能够更加有效地帮助手关节和腿关节进行情绪信息交互。

3.3.3 与先进步态情绪识别方法比较

为了进一步验证本文模型在步态情绪识别任务上的优越性,在 Emotion-Gait 数据集上与近年来经典网络模型进行比较,结果如表 4 所示。由此可知,本文模型的情绪识别准确率相较于经典模型 STEP、LSTM 分别提升了 10.98%、13.25%,证明了本文模型在步态情绪识别任务上具有一定优势。

Table 4 Comparison with state-of-the-art methods on Emotion-Gait dataset

表 4 与先进方法在 Emotion-Gait 数据集上的比较

网络模型	准确率/%	参数量/M
STEP ^[13] (2020, AAAI)	77.65	0.72
G-GCSN ^[14] (2021, ACCV)	80.31	0.73
LSTM ^[35] (2020, ECCV)	75.38	0.31
TAEW ^[36] (2020, ECCV)	81.89	40.44
ProxEmo ^[37] (2020, IROS)	80.33	0.67
CTR-GCN ^[19] (2021, ICCV)	83.10	3.55
TNTC ^[38] (2022, ICASSP)	85.97	60.21
DDG ^[39] (2023, Cognitive Systems Research)	85.67	36.71
STAGCN ^[40] (2023, ICME)	87.90	0.91
Ours	88.63	34.46

还可以看出,本文模型参数量达到了 34.46 MB,这是因为 CAM 模块中有 4 个交叉注意力子模块,故模型参数量会相应增多。但由于本文模型网络层数删减为三层,并对关节邻接矩阵作了优化,因而使用了以数据驱动的关节邻接矩阵 Bk 和添加基于伪节点的交叉注意力模块 CAM 后的 ST-GCN 网络训练时间为 1 250 s。

4 结语

本文针对骨架拓扑结构限制关节关系探索的问题、关节全局信息不充分的问题,提出了一种新的基于伪节点的交叉注意力情绪识别模型。该模型从整个空间维度以及与人关节局部特征全局关节特征融合出发,使用 CAM 以提高情绪类别的识别性能。CAM 使用手关节和腿部关节通过交叉注意力以学习空间中的局部关节信息,并使用

伪关节节点通过交叉注意力学习空间中的全局关节信息。实验结果表明,改进后的网络在 Emotion-Gait 数据集上的情绪识别准确率达到 88.63%,与已有经典先进模型相比性能更优。对于以数据驱动的关节邻接矩阵 Bk, CAM 模块对网络的情绪识别准确率具有重要作用。同时,本文模型也存在一些不足,比如:需进一步提升模型识别准确率、减少网络参数量,同时可将模型应用于具体工程实践。

参考文献:

- [1] DENHAM S A, WORKMAN E, COLE P M, et al. Prediction of externalizing behavior problems from early to middle childhood: the role of parental socialization and emotion expression[J]. *Development and Psychopathology*, 2012, 12(1): 23-45.
- [2] ARUNNEHRU J, KALAISELVI G M. Automatic human emotion recognition in surveillance video[J]. *Intelligent Techniques in Signal Processing for Multimedia Security*, 2017, 10: 321-342.
- [3] RIVAS J J, ORIHUELA-ESPINA F, ENRIQUE S L, et al. Detecting affective states in virtual rehabilitation[C]//*Proceedings of the 9th International Conference on Pervasive Computing Technologies for Healthcare*, 2015: 287-292.
- [4] AWAIS M, RAZA M, SINGH N, et al. LSTM-based emotion detection using physiological signals: IoT framework for healthcare and distance learning in COVID-19[J]. *IEEE Internet of Things Journal*, 2020, 8(23): 16863-16871.
- [5] MONTEPARE J M, GOLDSTEIN S B, CLAUSEN A. The identification of emotions from gait information[J]. *Journal of Nonverbal Behavior*, 1987, 11: 33-42.
- [6] MICHALAK J, TROJE N F, FISCHER J, et al. Embodiment of sadness and depression—gait patterns associated with dysphoric mood[J]. *Psychosomatic Medicine*, 2009, 71(5): 580-587.
- [7] DELIGIANI F, GUO Y, YANG G Z. From emotions to mood disorders: a survey on gait analysis methodology[J]. *IEEE Journal of Biomedical and Health Informatics*, 2019, 23(6): 2302-2316.
- [8] CAO Z, SIMON T, WEI S E, et al. Realtime multi-person 2d pose estimation using part affinity fields[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2017: 7291-7299.
- [9] ZHANG Z. Microsoft kinect sensor and its effect[J]. *IEEE Multimedia*, 2012, 19(2): 4-10.
- [10] LI B, ZHU C, LI S, et al. Identifying emotions from non-contact gaits information based on microsoft kinects[J]. *IEEE Transactions on Affective Computing*, 2016, 9(4): 585-591.
- [11] CRENN A, KHAN R A, MEYER A, et al. Body expression recognition from animated 3d skeleton[C]//*Proceedings of the International Conference on 3D Imaging*, 2016: 1-7.
- [12] ZHANG J B, GAO J, HUANG Z Y, et al. Gait emotion recognition based on a multi-scale partitioning directed spatio-temporal graph[J]. *Journal of Electronics & Information Technology*, 2024, 46(3): 1069-1078.
张家波,高洁,黄钟玉,等.基于多尺度分区有向时空图的步态情绪识别[J]. *电子与信息学报*, 2024, 46(3): 1069-1078.
- [13] BHATTACHARYA U, MITTAL T, CHANDRA R, et al. Step: spatial temporal graph convolutional networks for emotion perception from gaits[C]//*Proceedings of the AAAI Conference on Artificial Intelligence*, 2020: 1342-1350.
- [14] ZHUANG Y, LIN L F, TONG R F, et al. G-GCSN: global graph convolution shrinkage network for emotion perception from gait[C]//*ACCV 2020*

- Workshops: 15th Asian Conference on Computer Vision, 2020:46–57.
- [15] CHAI S, LIU J Q, JAIN R K, et al. A multi-head pseudo nodes based spatial-temporal graph convolutional network for emotion perception from GAIT[J]. *Neurocomputing*, 2022, 511:437–447.
- [16] LI B, LI X, ZHANG Z, et al. Spatio-temporal graph routing for skeleton-based action recognition[C]//*Proceedings of the AAAI Conference on Artificial Intelligence*, 2019:8561–8568.
- [17] LI M, CHEN S, CHEN X, et al. Actionalstructural graph convolutional networks for skeleton-based action recognition[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019:3595–3603.
- [18] SHI L, ZHANG Y, CHENG J, et al. Two-stream adaptive graph convolutional networks for skeleton-based action recognition[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019:26–35.
- [19] CHEN Y, ZHANG Z, YUAN C, et al. Channelwise topology refinement graph convolution for skeleton based action recognition[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021:359–368.
- [20] DE SANTANA C A, COLOMBINI E L. Attention, please! a survey of neural attention models in deep learning[J]. *Artificial Intelligence Review*, 2022, 55(8):6037–6124.
- [21] KUMAR A J R, BHANU B. Micro-expression classification based on landmark relations with graph attention convolutional network[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021:1511–1520.
- [22] CHEN M, HE X, YANG J, et al. 3-D convolutional recurrent neural networks with attention model for speech emotion recognition[J]. *IEEE Signal Processing Letters*, 2018, 25(10):1440–1444.
- [23] YOLCU G, OZTEL I, KAZAN S, et al. Facial expression recognition for monitoring neurological disorders based on convolutional neural network[J]. *Multimedia Tools and Applications*, 2019, 78:31581–31603.
- [24] GAJJALA V R, REDDY S P T, MUKHERJEE S, et al. MERANet: facial micro-expression recognition using 3D residual attention network[C]//*Proceedings of the twelfth Indian Conference on Computer Vision, Graphics and Image Processing*, 2021:1–10.
- [25] LI W, ABTAHI F, ZHU Z. Eac-net: deep nets with enhancing and cropping for facial action unit detection[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(11):2583–2596.
- [26] ZHAO K, CHU W S, ZHANG H. Deep region and multi label learning for facial action unit detection[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 3391–3399.
- [27] SCHNELL B, GARNER P N. Improving emotional tts with an emotion intensity input from unsupervised extraction[C]//*Proceedings of the 11th ISCA Speech Synthesis Workshop*, 2021:60–65.
- [28] ZHU X, YANG S, YANG G. Controlling emotion strength with relative attribute for end-to-end speech synthesis[C]//*2019 IEEE Automatic Speech Recognition and Understanding Workshop*, 2019:192–199.
- [29] RANDHAVANE T, BHATTACHARYA U, KABRA P, et al. Learning gait emotions using affective and deep features[C]//*Proceedings of the 15th ACM SIGGRAPH Conference on Motion, Interaction and Games*, 2022: 1–10.
- [30] MA Y L, PATERSON H M, POLLICK F E. A motion capture library for the study of identity, gender, and emotion perception from biological motion[J]. *Behavior Research Methods*, 2006, 38(1):134–141.
- [31] IONESCU C, PAPAVALA D, OLARU V, et al. Human3.6m: large scale datasets and predictive methods for 3d human sensing in natural environments[J]. *IEEE Transactions On Pattern Analysis and Machine Intelligence*, 2013, 36(7):1325–1339.
- [32] NARANG S, BEST A, FENG A, et al. Motion recognition of self and others on realistic 3d avatars[J]. *Computer Animation and Virtual Worlds*, 2017, 28(3–4):e1762.
- [33] CMN graphics lab. CMN graphics lab motion capture database[EB/OL]. <http://mocap.cs.cmu.edu/>.
- [34] HABIBIE I, HOLDEN D, SCHWARZ J, et al. A recurrent variational autoencoder for human motion synthesis[C]//*London: Proceedings of the British Machine Vision Conference (BMVC)*, 2017.
- [35] RANDHAVANE T, BHATTACHARYA U, KAPSASKIS K, et al. Identifying emotions from walking using affective and deep features[DB/OL]. <https://arxiv.org/abs/1906.11884>, 2019.
- [36] BHATTACHARYA U, RONCAL C, MITTAL T, et al. Take an emotion walk: perceiving emotions from gaits using hierarchical attention pooling and affective mapping[C]//*European Conference on Computer Vision*, 2020:145–163.
- [37] NARAYANAN V, MANOGHAR B M, DORBALA V S, et al. Proximo: gait-based emotion learning and multi-view proxemic fusion for socially-aware robot navigation[C]//*2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020:8200–8207.
- [38] HU C F, SHENG W J, DONG B, et al. TNTC: two-stream network with transformer-based complementarity for gait-based emotion recognition[C]//*ICASSP IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022:3229–3233.
- [39] CHEN X, LIU Z, XIAO J, et al. DDG: dependency-difference gait based on emotional information attention for perceiving emotions from gait[J]. *Cognitive Systems Research*, 2023, 82:101150.
- [40] CHEN C, SUN X. STA-GCN: spatial temporal adaptive graph convolutional network for gait emotion recognition[C]//*IEEE International Conference on Multimedia and Expo(ICME)*, 2023:1385–1390.

(责任编辑:孙 娟)