

李志鹏, 李磊, 马柯. 基于时空分布的面部特征点云的疼痛诊断模型研究[J]. 智能计算机与应用, 2025, 15(2): 144–152.
DOI:10.20169/j.issn.2095-2163.24081802

基于时空分布的面部特征点云的疼痛诊断模型研究

李志鹏¹, 李磊^{1,2}, 马柯¹

(1 上海理工大学 健康科学与工程学院, 上海 200093; 2 上海理工大学 机器智能研究院, 上海 200093)

摘要: 疼痛是一类复杂的生理现象,在临床上正确的疼痛强度诊断至关重要。面部表情包含着丰富的疼痛信息,通过识别面部表情的动作变化可以有效识别疼痛强度。基于此,本文提出了以显示疼痛的面部视频为样本的疼痛诊断深度学习模型:通过人脸面部特征点提取模型构建时间尺度上的面部特征点的三维点集,并进行归一化处理;再利用点云分类模型对归一化的面部特征点时空信息进行特征向量的提取,并通过分类器进行分类;最终实现疼痛与否的二分类任务和疼痛等级五分类任务。本文所提模型在 BioVid Heat Pain 数据集上进行训练与测试,最终在二分类疼痛识别任务中达到 84.98% 的准确率,在五分类任务中准确率达到 37.65%,均优于现有模型,特别是在识别四级疼痛时表现出较高的精度,达到 72%。结果表明,本文所提出的基于时空分布的面部特征点云的疼痛诊断模型具有潜在的应用价值。

关键词: 疼痛识别; 面部特征点; 点云分类; 深度学习; Mediapipe; PointNet

中图分类号: TP391.4

文献标志码: A

文章编号: 2095-2163(2025)02-0144-09

Research on pain diagnosis model based on spatiotemporal distribution of facial feature point clouds

LI Zhipeng¹, LI Lei^{1,2}, MA Ke¹

(1 School of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China;

2 Institute of Machine Intelligence, University of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract: Pain is a complex physiological phenomenon, and accurate estimation of pain intensity is crucial in clinical treatment. Facial expressions contain rich information related to pain, and changes in facial expressions can effectively indicate pain intensity. Based on this, this paper proposes a deep learning model for pain diagnosis using facial video samples. Firstly, the facial feature point extraction model is utilized to construct a three-dimensional point set of facial feature points across time scales. This three-dimensional point set data is then normalized. The point cloud classification model is subsequently used to extract feature vectors from the normalized spatiotemporal information of the facial feature points, which are then classified to achieve binary classification of pain recognition and five-level classification of pain intensity. The proposed model is trained and tested on the BioVid Heat Pain dataset, achieving an accuracy of 84.98% in the binary pain recognition task and 37.65% in the five-level classification task, both outperforming existing models. Notably, the model demonstrates a high precision of 72% in recognizing level four pain. The results indicate that the proposed pain diagnosis model based on spatiotemporal distribution of facial feature point clouds has potential application value.

Key words: pain recognition; facial landmarks; point cloud classification; deep learning; Mediapipe; PointNet

0 引言

疼痛是一种复杂的多维现象,涉及生理、行为、情感、认知和发育等方面^[1]。其是一种通过警告有机体面临内部或外部攻击时的一种保护机制。在临床上,有效的疼痛管理是至关重要的,因为未经治疗的疼痛

会对器官系统或愈合过程产生负面影响^[2]。此外,疼痛会增加严重的健康问题风险,如心肌梗死、充血性心力衰竭、感染(通过抑制免疫系统),并可导致心动过速、高血压、心脏需氧量增加等问题^[2]。

由于疼痛是一种主观体验,不同的人对疼痛的反应差别很大,这种差异的部分原因在于大脑本身

作者简介: 李志鹏(1998—),男,硕士研究生,主要研究方向:疼痛研究,图像处理;李磊(1997—),男,博士研究生,主要研究方向:点云配准,深度学习。

通信作者: 马柯(1972—),男,博士,主任医师,主要研究方向:疼痛学。Email:marke72@163.com。

收稿日期: 2024-08-18

哈尔滨工业大学主办 ◆ 专题设计与应用

有能力通过激活疼痛控制系统来抑制疼痛信号输入到神经系统中^[3]。因此,目前疼痛评估常用的方法主要是患者的自我报告。然而,对于昏迷患者或者新生儿患者,则无法通过自我报告来进行疼痛评估^[4]。其他方式,诸如生理信号评估以及功能影像学评估,则需要昂贵的专业设备和专业知识。因此,研究一个客观的、自动的疼痛评估方法极具临床应用价值。

心理学研究表明,面部表情在情感表达中占据约55%的比例,而疼痛恰恰是一种情感体验^[5]。此外,Prkachin和Solomon研究了疼痛与面部活动之间的关系,并提出了疼痛程度的度量标准(Prkachin and Solomon Pain Intensity scale, PSPI)^[6]。因此,通过面部表情进行疼痛评估是一种有效的方法。

基于面部表情进行疼痛识别的传统方法主要通过机器学习来实现,实现方法简单,但存在识别效果受个体差异和环境影响,无法适应面部表情变化的瞬时性和快速性等缺点。随着深度学习在计算机视觉领域取得突破,卷积神经网络(Convolutional Neural Network, CNN)、循环神经网络(Recurrent Neural Network, RNN)、长短期记忆网络(Long Short-Term Memory, LSTM)以及transformer等深度学习模型,已逐渐用于基于面部表情的疼痛识别。相比于传统的机器学习,基于深度学习的疼痛识别具有更高的准确度和更强的适应性,但也存在一些问题。如:算法复杂、需要消耗大量的运算资源、出现过拟合,以及对数据集质量要求高等问题。同时,基于面部表情的疼痛识别往往受到头部姿势、照明和遮挡等因素的影响而无法准确识别疼痛强度。

Mediapipe的FaceMesh模型^[7]是谷歌通过迁移学习开发的面部特征识别机器学习模型,专门用于在三维空间中识别用户面部特征点。该模型能够在各种条件下运行,如不同的头部姿势、照明和背景,并且可以处理所有年龄和肤色的人脸图像。本研究中,利用Mediapipe模型的Face Mesh模块来获取视频中人脸面部的特征点。鉴于疼痛引起的人脸面部变化是一个包含时间动态特性的过程,动态的面部表情更加有利于疼痛识别。因此,本文创造性地构建了一个包含时空特征的点云模型,并使用PointNet模型进行特征提取。具体来讲,首先通过Mediapipe的FaceMesh模型获取连续视频中特定帧的人脸面部特征点,并将这些特征点按照时间序列排列,转化为空间点集合,经过归一化后形成一个包含人脸时空特征的点云模型;然后通过PointNet模

型提取点云特征向量;最后通过多层感知机(MLP)降维,并通过分类算法来实现疼痛识别的二分类和五分类。通过将本文模型在BioVid Heat Pain数据集^[8]上进行训练与测试,并与现有的其他先进模型算法进行比较的实验结果表明,本文所提出的模型具有更高的准确率,模型简化了数据预处理步骤,提高了模型的鲁棒性。PointNet的轻量级特性,使其能够在移动设备上部署,有助于实时监测患者的疼痛状态,为临床决策提供即时支持。

1 相关工作

面部动作编码系统(FACS)是一种用于面部表情自动识别的系统,其将面部表情划分为不同的动作单元(Action Units, AUs)组合^[9],AUs是指与面部表情相关的特定面部肌肉的运动。在此基础上,众多研究者提出了多种基于多特征向量组合的疼痛评估方法。疼痛评估有两种模式:确定疼痛的存在和评估疼痛强度。第一种模式通常被认为是一个二元分类问题,而第二种模式被认为是一个多分类问题。目前,通过面部表情来评估疼痛的工作由两个主要部分组成:特征提取和疼痛的自动评估或分类。在早期的工作中,用于疼痛识别的面部表情的特征提取主要采用传统的基于几何和外观纹理的手工提取方式。Wang等^[10]提出了一种方法,通过计算每帧图像中归一化面部特征点与平均面部形状之间的距离来提取几何特征,并使用每帧图像中归一化面部特征点周围的梯度方向直方图(HOG)来提取外观特征,最后结合这两种特征进行疼痛评估。Zafar等^[11]利用表示不同程度疼痛的22个面部特征点的距离和角度来计算面部表情的几何特征,并通过得到的几何特征来进行疼痛评估。Meng等^[12]则采用局部二值模式(Local Binary Patterns, LBP)来提取面部疼痛特征。

随着深度学习在图像识别领域中取得突破,面部表情图像的特征通过深度神经网络自动提取并用于疼痛识别。相比于单一帧的面部表情图像,连续的面部视频由于包含时间动态特性而被用于疼痛识别。Zhou等^[13]提出了一种递归卷积神经网络(RCNN),在卷积层中加入递归路径,从面部视频序列中评估疼痛强度。Rodriguez等^[14]使用预训练的VGG-Face模型提取视频序列的面部疼痛特征,再将该特征输入到长短期记忆网络(Long Short-Term Memory, LSTM)中,通过提取视频前后帧的时间信息来进行疼痛识别。2019年,Tavakolian等^[15]提出

了一种时空三维卷积神经网络来学习人脸视频的动态时空特征,该模型由几个具有不同时间深度的卷积层组成,用于自动捕获面部的时空变化来进行疼痛评估。Thomas^[16]团队采用 CNN 从每一帧视频中提取特征,然后将这些特征放入长短时记忆网络中获取视频特征向量,并对特征向量进行分类。Zhao 等^[17]则提出了一种动态面部表情识别变换器,该变换器将卷积空间变换器与时间变换器相结合,能够有效提取面部表情的时空特征。Huang 等^[18]采用一个带有空间注意力模块的空间子网络来提取图像帧的空间注意力特征,这些特征随后输入到一个带有门控循环单元(GRU)和时间注意力模块的时间子网络中,用于学习和组合时间特征以进行疼痛估计。为了提高疼痛估计的准确性,Huang 等^[19]又提出通过 2D 空间卷积后接 1D 时间卷积来学习时空特征,并在 1D 卷积之后添加注意力机制,在融合时空特征时同时学习空间和时间的注意力权重,从而提高疼痛估计的准确性。Xu 等^[20]提出了一种深度注意力变换器网络(PET),PET 由两个不同的子网络组成:图像编码器网络和视频变换器子网络。在图像编码器子网络中,ResNet 与瓶颈注意力块相结合,用于学习面部图像的特征。在视频变换器子网络中,使用变换器编码器来捕捉视频帧之间的时序关系。最终,时空特征通过多层感知机(MLP)进行融合实现疼痛强度的评估。Liao 等^[21]则采用了一个时序变换器编码器网络来提取时间特征,有效地捕捉了面部表情中的局部和全局变化。

2 诊断模型

本文提出的基于面部特征点和点云分类算法的疼痛诊断模型整体框架如图 1 所示。首先,对视频进行降采样,并通过 Mediapipe 中的 Face mesh 模型提取人脸的连续帧的面部特征点,将其按照时间序列排列,从而得到三维空间点集合,再进行归一化处理得到点云集合;然后通过 PointNet 模型对点云进行编码得到全局特征向量;最后通过多层感知机降维并通过分类算法对这些特征向量进行分类,从而得到疼痛诊断结果。

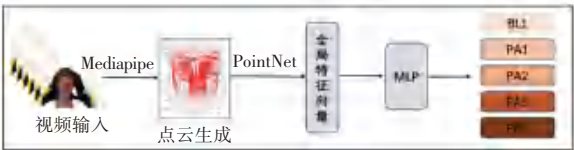


图 1 疼痛识别整体框架图

Fig. 1 Frame diagram of pain recognition

2.1 面部特征点提取

2.1.1 MediaPipe 模型

MediaPipe 是一款基于 Apache 2.0 许可(开源)发行的跨平台机器学习解决方案,专为直播和流媒体设计。其中包括面部检测、手势检测、物体跟踪以及人体骨骼检测等多种实用功能,本文主要利用 MediaPipe 的 Facemesh 模块。FaceMesh 是一种机器学习模型,专门用于识别和捕捉三维图像中人脸的拓扑结构和形状。其通过结合 BlazeFace 模型确定人脸的具体位置,一旦人脸的位置确定,FaceMesh 就会定位和标注出如图 2 所示的总共 468 个面部关键点。此外,该模块还能实时地将这些关键点的位置反馈给计算机系统。

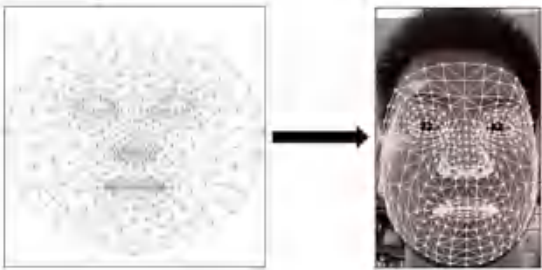


图 2 FaceMesh 采集面部 468 个特征点的拓扑模型

Fig. 2 Model employed by MediaPipe Facemesh, which represents the facial topology with 468 landmarks

2.1.2 基于 Mediapipe 的面部特征点提取

人体受到的内外界刺激达到产生疼痛的阈值后,神经末梢会受到触发并释放相关的介质,从而使大脑产生疼痛感知。面部肌肉组织与大脑中的疼痛调控中枢相连并受其调控,疼痛感知会促使大脑发出信号调节人的面部变化。例如,面部紧绷、眼眶扭曲、嘴角下弯等都可能是疼痛的表现。

理想状态下,Mediapipe 模型中的 FaceMesh 模块能够精准地从人脸图像中提取出 468 个关键特征点。然而,若提取全部面部特征点,不仅会因处理大量重复信息而造成计算资源的过度消耗,而且忽视了疼痛表情的核心变化——即由面部肌肉活动引发的嘴部、眼部、眉毛等部位的微妙变动。鉴于此,本文采取了一种精简策略,在面部信息特征点上只保留与疼痛表情强相关的嘴巴、眼睛、鼻子、瞳孔、眉毛以及脸部整体轮廓的特征点,总计 162 特征点,特征点的选择如图 3 所示。通过这一优化措施,不仅能显著降低数据处理的复杂度,同时也能更聚焦于那些对疼痛诊断至关重要的面部微表情,从而提升分析的准确性和效率。



图 3 疼痛特征点选取

Fig. 3 Facial pain landmark selection strategy

此外, BioVid 数据集中录制的人脸视频的帧率为 25 fps, 每个视频片段时长约为 5.44 s, 累计 136 帧。考虑到 25 Hz 的视频相邻两帧的面部信息差别不大, 因此每个样本视频从第一帧开始, 每隔三帧选取一帧, 将每个视频降采样为 34 张 RGB 图片, 在尽量不遗漏疼痛信息的前提下减少计算量。对视频降采样之后, 提取与疼痛相关的特征点, 然后将其按照时间序列排列, 便可将时间尺度上的面部信息特征点转化为三维空间点集, 如图 4 所示。

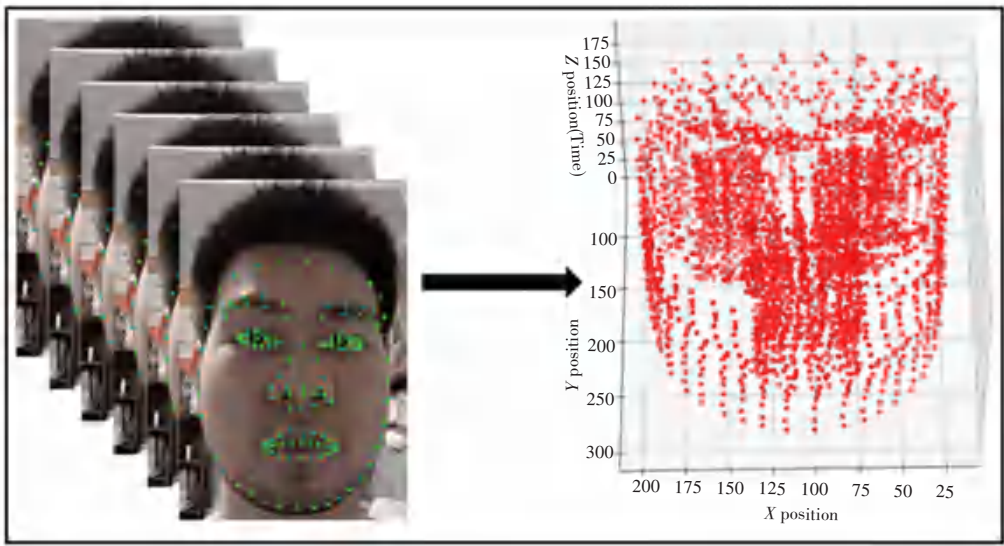


图 4 面部特征点云模型

Fig. 4 Point cloud model of facial features

将疼痛视频信息转换为三维空间点集后, 对数据进行归一化处理, 将所有特征统一缩放至 $[0, 1]$ 区间内。在机器学习模型训练过程中, 若特征数据各维度存在显著的量级差异且数据未经归一化处理, 模型可能会倾向于赋予量级较大的特征更高的权重, 从而导致学习过程的偏斜, 最终影响模型的预测性能和泛化能力。归一化处理可以确保模型在训练时平等地考虑所有特征, 从而减少偏差, 提升模型的整体表现。本文使用最小—最大归一化方法 (Min-Max Scaling), 原理如下式所示。

$$i_{\text{norm}} = \frac{i - \min(i)}{\max(i) - \min(i)} \quad (1)$$

其中, $\min(i)$ 和 $\max(i)$ 分别是维度 i 在所有数据点中的最小值和最大值。

具体操作时, 在人脸的宽度、高度及时间轴这 3 个维度上实施归一化。同时也保证了特征点集中任意面部信息特征点之间的相对位置在时间和空间上保持不变, 保留了关键的疼痛信息。此外, 数据归

一化的实施对于优化算法 (如梯度下降) 的效率具有显著提升作用, 有助于加快模型训练的收敛速度, 并提高算法在复杂数据集上的稳定表现。

2.2 点云编码与分类

完成归一化处理后的特征点集合可以视为一个三维点云模型, 为了实现有效的分类就要对其编码并转化为特征向量, 然后通过分类器对特征向量进行分类, 最终得到疼痛诊断的结果。本文采用 PointNet 模型从点云数据中获取三维点集特征向量, 再通过多层感知机降维分类。

2.2.1 PointNet 模型

PointNet 是一种深度学习模型, 其可以直接将点云数据作为输入并进行分类, 可以处理任意形状的点云数据, 且对输入点的顺序不敏感, 具有良好的鲁棒性和泛化能力。PointNet 的核心是将点云数据视为一个无序的点集, 然后通过多层感知机 (MLP) 和最大池化 (Max Pooling) 等操作来处理点集, 以获得点云数据的全局特征。PointNet 包含以下模块:

输入层用于接收点云数据,并将其转化为包含点云特征信息的固定大小矩阵;全连接层通过多层感知机,获得点云数据的局部特征;最大池化层对点云数据的局部特征进行最大池化操作,以获得点云数据的全局特征。PointNet 使用上述模块处理点云数据,以获得点云数据的全局特征表示,然后用于分类、分割、配准和识别等任务。图 5 展示了 PointNet 的网络结构,其原理如下式所示:

$$f(X) = \max\{h(x_1), h(x_2), \dots, h(x_n)\} \quad (2)$$

其中, $x_i \in R^3$, h 是应用于每个点的 MLP 网络, $\max$ 表示对所有点特征进行的最大池化操作,以提取全局特征。

络结构如图 6 所示,对于一个输入的三维点集合,将一系列共享权重的全连接层(即 MLP)应用于每个点的变换,从而确保处理过程对点的排列顺序不敏感,在全局上对每个点不作区别,从而增强其全局特征。从另一个角度来看,这种方法类似于对每个点使用相同的卷积核进行卷积操作,因此共享权重的全连接层在本质上表现为一种卷积形式。每个卷积层之间进行批量归一化并使用 ReLU 激活函数,将结果约束在学习效果良好范围内的同时又增加了其非线性特征。批量归一化是通过使用通道平均值 μ_c 和标准偏差 σ_c 来对前一层的输出进行归一化处理,减少内部协变量偏移来加速网络收敛从而稳定训练过程。批量归一化的表示如下式所示。

$$y_{\gamma,\beta,c} = \frac{x_{i,j,c} - \mu_c}{\sigma_c} \cdot \gamma + \beta \quad (3)$$

相关参数定义如下式所示:

$$\mu_c = \frac{1}{NL} \sum_{i,j} x_{i,j,c}$$
$$\sigma_c^2 = \frac{1}{NL} \sum_{i,j} (x_{i,j,c} - \mu_c)^2 \quad (4)$$

其中, i 和 j 是空间索引, c 是通道索引; μ_c 和 σ_c^2 分别表示本批次样本中通道 c 的平均值和方差。在上述假设中, γ 和 β 是引入的可学习参数,从而使神经网络在即使处于输入不是正态分布的情况下进行适当学习。而 ReLU 函数是一种在深度学习中广泛使用的激活函数,其原理如下式所示:

$$f(x) = \max(0, x) \quad (5)$$

同时,在前两层卷积神经网络之间,使用 T-Net 模块学习一个变换矩阵,该变换矩阵应用于输入点云集合,通过矩阵乘法增强模型对点云旋转的不变性。T-Net1 和 T-Net2 在架构上没有本质区别,只是由于与其作仿射变换的矩阵维度不同而有区别,都通过一系列全连接层和激活函数来实现。

图 5 PointNet 的模型结构

Fig. 5 Network model structure diagram of PointNet

2.2.2 特征向量提取

本文基于PointNet进行点云特征向量提取的网

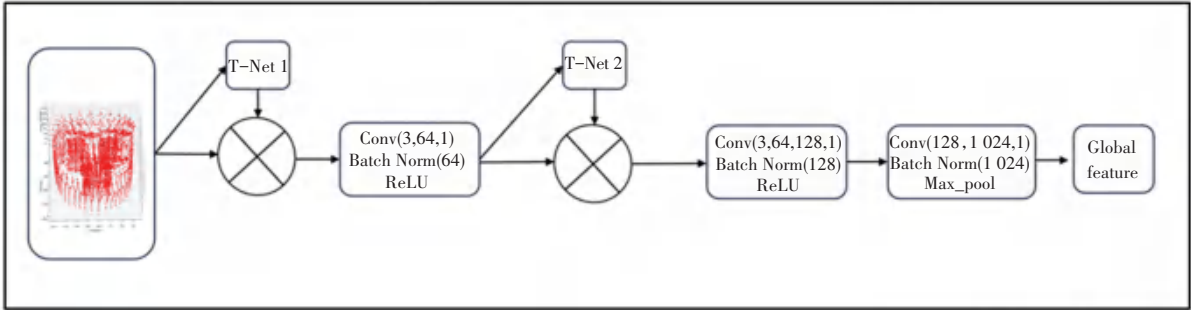


图 6 特征向量提取模型网络结构

Fig. 6 Structure of the model used to extract feature vectors

最终,通过一系列卷积操作、归一化处理、ReLU函数激活和仿射变换,将点集合非线性地映射到高维空间。通过最大池化操作进行降采样,从而减小特征图的大小,在降低后续计算的同时保留主要特征。这些步骤最终形成了描述视频疼痛特征的三维点集合特征向量,即视频的疼痛特征。

2.2.3 疼痛特征向量分类

在提取疼痛特征向量之后,需要对其进行分类。分类器的原理是通过多层神经网络对特征进行逐步降维,整体结构如图7所示,图中分别是特征二分类和五分类的分类器。两个分类器结构基本相同,主要区别在于输出层维度分别与分类类别数相对应。在特征逐步降维的过程中,使用ReLU函数激活增加非线性度以预防梯度消失,同时用Dropout做数据增强处理,提升模型泛化能力,用批量归一化层缩放数据到一个小的固定范围,以提高模型训练的稳定性效率。Dropout是一种正则化技术,核心思想是在训练过程中随机“丢弃”(即设置为0)神经网络中包含隐藏层和可见层在内的部分神经元,以避免过拟合^[22]。

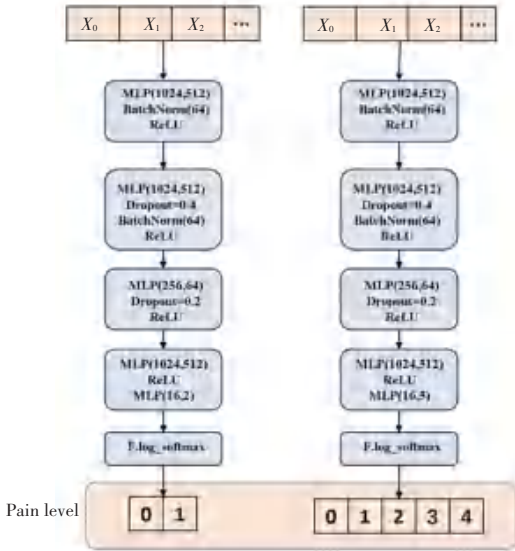


图7 面部疼痛信息特征分类器网络结构

Fig. 7 Network structure of the feature classifier of facial pain information

通过多个多层感知机降维数据到与类别数相同时,使用log_softmax函数进行转换。log_softmax可以认为是softmax变换和自然对数运算的结合。其中,softmax函数将输入向量“标准化”为概率分布的函数,其中每个元素的范围都在0到1之间,并且所有元素的总和为1。对于向量 $\mathbf{Z}$ 中的每个元素 z_i ,

log_softmax 数学表达式如下式所示:

$$\begin{aligned} \text{softmax}(z_i) &= \frac{e^{z_i}}{\sum_j e^{z_j}} \\ \log_softmax(z_i) &= \log\left(\frac{e^{z_i}}{\sum_{j=1}^n e^{z_j}}\right) = z_i - \log\left(\sum_{j=1}^n e^{z_j}\right) \end{aligned} \quad (6)$$

通常来讲,直接使用softmax后取对数可能在数学上不稳定,尤其是当softmax的输出极其接近0时。log_softmax通过在一个步骤中合并这两个运算,避免了因浮点数精度限制造成的下溢问题。使用log_softmax可以简化后续使用交叉熵损失的计算,因为交叉熵损失本质上是真实标签分布和预测概率分布的对数差的期望。这种简化可以帮助快速准确地计算梯度,加快模型的收敛速度。下式为交叉熵损失的定义,式中 y_c 表示实际的标签值, $\hat{y}$ 表示预测的结果。

$$\mathcal{L}(y, \hat{y}) = - \sum_{c=1}^C y_c \log(\hat{y}_c) \quad (7)$$

2.3 实验设置

2.3.1 数据集

为了训练和测试本文所提出的疼痛诊断模型,本文实验均在BioVid Heat Pain数据集上训练与验证。BioVid Heat Pain数据库是一个多模态数据集,包含视频(A部分)和生物医学信号(B部分)。该数据集共采集了90名年龄在18至65岁之间的健康受试者的数据,在数据收集之前,确定了每名受试者的疼痛阈值($T_p = T_1$,即温度刺激产生疼痛感时的温度)和耐受阈值($T_T = T_4$,即疼痛变得无法忍受时的温度),并根据个体阈值调整每位受试者的施加温度。受试者通过4种等级的温度刺激来量化4种等级的疼痛(T_1, T_2, T_3, T_4),温度与疼痛等级的对应关系如图8所示。每种疼痛强度被随机模拟了20次,每次持续4s,随后是8~12s的暂停,刺激与暂停的温度示意如图9所示。最终,成功收集了87名受试者的视频数据,每人100段视频,包含20个基线视频和80个疼痛视频,共计8700个样本。视频的分辨率为1388×1038像素,帧率为25Hz,每段视频为5.44s,136帧。由于本文的研究重点是通过疼痛高峰期的面部表情来确定疼痛强度,因此仅使用了8700个视频(A部分)。

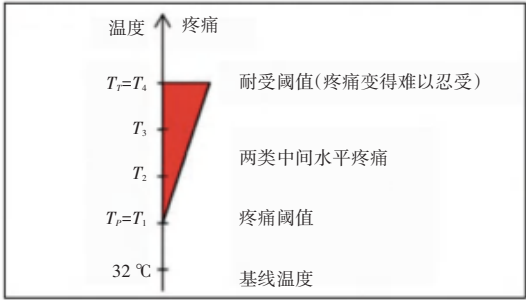


图 8 疼痛等级与个人温度感受对应关系

Fig. 8 Pain levels and according person-specific temperatures

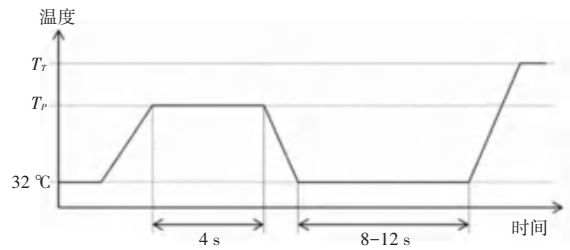


图 9 交替刺激与暂停的示例性温度曲线

Fig. 9 Exemplary temperature curve with alternating stimuli and pauses

2.3.2 实验细节

本文针对疼痛的评估主要包括对是否存在疼痛的评估以及对具体疼痛等级的评估。每段视频转化为一个点集合,形成一个样本,并与相应的热痛刺激等级标签对应。针对 BioVid Heat Pain 数据库(A 部分),本文进行了疼痛的二分类任务和五分类任务。在二分类实验中,数据被划分为无痛(BL1)和有痛(PA1, PA2, PA3, PA4)两类。在五分类实验中,数据按照 5 个疼痛等级标签进行分类。为了防止样本不平衡,在二分类任务中,将无痛类别的面部视频数据扩充为原来的 4 倍,以保持样本的平衡。由于无法转化为空间点集合的视频极少,因此可以忽略其对实验结果的影响。

在本文的实验中,为了充分利用 BioVid Heat Pain 数据库中的面部视频数据,本文的疼痛评估是通过留一交叉验证(Leave - One - Out Cross - Validation)的方式来进行训练和验证。将其中一组作为测试集,其余所有组作为训练集,从而验证分类方法的有效性。总共有 87 组实验,最终的评价标准是所有组测试集准确率的平均值。

在模型的训练和测试过程中,模型前向传播每批次取 16 个样本,模型的目标函数是输入样本的标签和对应样本模型预测值之间的交叉熵损失(式

7),并使用 Adam 优化器进行优化,学习率设置为 10^{-5} 。模型的训练逻辑采用早停策略结合分组训练的方法,以优化模型的泛化能力并避免过拟合。具体而言,如果模型在测试集上的性能在连续 16 个 epoch 内未见提升,则中断当前训练序列,重新初始化模型,并使用下一组数据进行训练。通过这种方式,可以有效地防止模型过拟合,并提高其在未见数据上的表现。

3 结果与分析

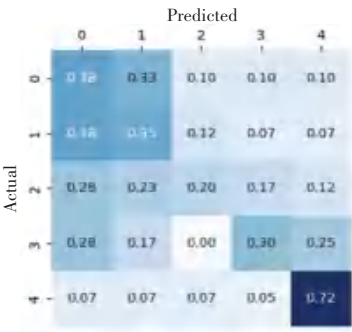
在模型训练和测试中,采用留一交叉验证的方法来训练本文提出的模型。在模型的评估方面,引入准确率(Accuracy)、召回率(Recall)、精确率(Precision)和 F1 - Score 作为模型的评价指标。最终,本文所提出的模型在留一交叉验证下的平均性能见表 1。

表 1 疼痛识别分类结果

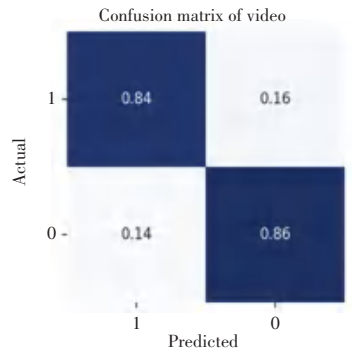
Table 1 Classification results for pain recognition %				
Model	Accuracy	Recall	Precision	F1 - score
二分类	84.98	84.09	85.71	84.89
五分类	37.65	33.29	34.34	33.81

视频信号二分类的准确度为 84.98%,说明本文提出的模型在识别疼痛与否有良好的准确度,而视频信号五分类的平均精度为 37.65%,在疼痛具体等级的识别中逊色于二分类。图 10 展示了疼痛识别五分类和二分类最佳模型在验证集上的混淆矩阵。可以发现,在疼痛识别五分类的任务中,模型在四级疼痛的识别中精度最高,达到了 72%的准确度,这可能与受试者在受到剧痛级别的热痛信号时面部动作较为明显有关;而二级疼痛和三级疼痛的识别率最低,可能是因为受试者在接受二级和三级的热痛信号时,面部表情的变化比较相似。

为了全面评估本文模型的性能水平,本研究将本文所提出的模型与其他基于 BioVid Heat Pain 数据集进行面部表情疼痛识别的先进模型进行比较。所有模型均使用 BioVid Heat Pain 数据集,并且执行相同的分类任务,采用准确率(Accuracy)作为模型性能指标。此外,为了进一步探讨基于面部表情和生理信号在疼痛识别中的优劣,本研究还将基于生理信号(皮肤电活动(EDA)、心电图(ECG))的疼痛识别模型纳入比较。表 2 展示的是疼痛识别二分类与其他模型的比较结果;表 3 展示的是疼痛识别五分类与其他模型的展示结果。



Confusion matrix of Video
(a) 五分类混淆矩阵



(b) 二分类混淆矩阵
图 10 混淆矩阵

Fig. 10 Confusion matrix

表 2 二分类模型结果与其他先进模型比较

Table 2 Binary classification results are compared with other advanced models

Model/Classifier	Modality	Level	Accuracy/ %
Chance	Video/RGB	2	50.00
RFc with FAD ^[23]	Video/RGB	2	65.80
MobileNetV2 ^[23]	Video/RGB	2	65.50
CNN+LSTM Layer ^[24]	ECG	2	76.68
CNN+LSTM Layer ^[24]	EDA	2	92.86
CNN+LSTM Layer ^[24]	EDA&ECG	2	97.27
本文	Video	2	84.98

表 3 五分类模型结果与其他先进模型比较

Table 3 Five-class classification results are compared with other advanced models

Model/Classifier	Modality	Level	Accuracy/ %
Chance	Video/RGB	5	20.00
ResNet 18 ^[25] 2020	Video/RGB	5	36.60
SVM ^[26] 2017	Video/RGB	5	27.30
FAD—RF ^[26] 2017	Video/RGB	5	30.80
SLSTM ^[27] 2019	Video/RGB	5	29.70
SAFEPA ^[28] 2023	Video/RGB	5	33.28
DDCAE ^[29] 2021	EDA、EMG、ECG	5	35.44
本文	Video	5	37.65

从表 2 可以看出,本文提出的模型在疼痛识别的二分类任务中表现优异,显著优于文献[23]所展示的 RFc with FAD 模型和 MobileNetV2 模型,也高于文献[24]提出的基于 ECG 单模态的疼痛识别二分类模型。然而,本文的疼痛识别二分类模型在准确率上低于基于 EDA 单模态以及基于 EDA 和 ECG 多模态的疼痛识别二分类模型。可见,基于生理信号的模型在某些特定情况下可能具有独特的优势,面部表情和生理信号的结合可能会带来更为全面的疼痛识别方案。

在疼痛识别五分类任务中,本文所提出模型取得了最高 37.65% 的准确度,相比文献[25]提出的 ResNet 18 模型提高了 1.05%。综上所述,本文基于时空分布的面部特征点云的疼痛诊断模型是有效的。

4 结束语

根据疼痛与面部表情动作的关系,本文提出了一种新颖的基于时空分布的面部特征点云的疼痛诊断深度学习模型,通过对疼痛面部视频数据的分析,将面部特征点按时序转化为三维点集数据,通过专用的三维点云分类算法进行特征提取和分类,实现对疼痛强度的评估。有效降低了视频诊断疼痛模型的复杂度,并且所提模型对头部姿势和背景具有良好的鲁棒性,模型的轻量级特性使其能够在移动设备上部署进行实时疼痛监测,为疼痛管理和评估提供了一种全新的方法。将本文模型在 BioVid Heat Pain 数据集上训练与测试,在疼痛识别二分类和疼痛等级五分类的任务中分别取得了 84.98% 和 37.65% 的准确度。

尽管本研究在疼痛诊断方面取得了一定的成果,但仍存在一些局限性。首先,数据集的多样性和规模有限,未来研究可以考虑引入更多种类和更大规模的数据集,以验证模型的泛化能力;其次,本文选取的面部特征点可能较为主观,各部分特征点对于疼痛信息的反映程度没有进一步研究,可以预见的是,通过选取更能反映疼痛信息的特征点一定可以更好地诊断疼痛。未来也可以考虑结合其他生理信号(如心电图、脑电图等)进行多模态融合研究,以期获得更加全面和准确的疼痛评估结果。

参考文献

[1] ZELTZER L K, BARR R G, MCGRATH P A, et al. Pediatric pain: Interacting behavioral and physical factors. [J]. Pediatrics,

- 1992, 90(5 Pt 2): 816–821.
- [2] STEIMAN I. Weiner's pain management; A practical guide for clinicians[J]. Journal of the Canadian Chiropractic Association, 2006, 50(4): 289.
 - [3] SEYEDALI K. Guyton and Hall: Textbook of medical physiology [J]. Surgical Neurology I, 2017, 8(1): 275.
 - [4] WERNER P, AI – HAMADI A, LIMBRECHT E K, et al. Automatic pain assessment with facial activity descriptors [J]. IEEE Transactions on Affective Computing, 2017, 8(3): 286–299.
 - [5] WERNER P, LOPEZ M D, WALTER S, et al. Automatic recognition methods supporting pain assessment; A survey [J]. IEEE Transactions on Affective Computing, 2022, 13(1): 530–552. DOI:10.1109/TAFFC.2019.2946774
 - [6] PRKACHIN K M. The consistency of facial expressions of pain; A comparison across modalities[J]. Pain, 1992, 51(3): 297–306.
 - [7] KARTYNNIK Y, ABLAVATSIK A, GRISHCHENKO I, et al. Real-time facial surface geometry from monocular video on mobile GPUs[J]. arXiv preprint arXiv, 1907.06724, 2019.
 - [8] WALTER S, GRUSS S, EHLEITER H, et al. The biovid heat pain database data for the advancement and systematic validation of an automated pain recognition system [C]//Proceedings of the 2013 IEEE International Conference on Cybernetics. Piscataway, NJ: IEEE, 2013; 128–131.
 - [9] SZAJNBERG N M. What the face reveals: Basic and applied studies of spontaneous expression using the facial action coding system [J]. Journal of the American Psychoanalytic Association, 2022, 70(3): 591–595.
 - [10] WANG Jinwei, SUN Huazhi. Pain intensity estimation using deep spatiotemporal and handcrafted features [J]. IEICE Transactions on Information and Systems, 2018, E101-D(6): 1572–1580.
 - [11] ZAFAR Z, KHAN N A. Pain intensity evaluation through facial action units [C]// Proceedings of the 2014 22nd International Conference on Pattern Recognition. Piscataway, NJ: IEEE, 2014; 4696–4701.
 - [12] MENG H Y, BIANCHI B N. Affective state level recognition in naturalistic facial and vocal expressions[J]. IEEE Transactions on Cybernetics, 2014, 44(3): 315–328.
 - [13] ZHOU Jing, HONG Xiaopeng, SU Fei, et al. Recurrent convolutional neural network regression for continuous pain intensity estimation in video [C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Piscataway, NJ: IEEE, 2016; 1535–1543.
 - [14] RODRIGUEZ P, CUCURULL G, GONZALEZ J, et al. Deep pain; Exploiting Long Short-Term Memory Networks for facial expression classification [J]. IEEE Transactions on Cybernetics, 2022, 52(5): 3314–3324.
 - [15] TAVAKOLIAN M, HADID A. A spatiotemporal convolutional neural network for automatic pain intensity estimation from facial dynamics [J]. International Journal of Computer Vision, 2019, 127(10): 1413–1425.
 - [16] THOMAS B. LSTM neural networks for detection and assessment of back pain risk in manual lifting [D]. Cincinnati: University of Cincinnati, 2021.
 - [17] ZHAO Zhenqun, LIU Qingshan. Former-DFER: Dynamic facial expression recognition transformer [C]// Proceedings of the 29th ACM International Conference on Multimedia. New York: ACM, 2021; 1553–1561.
 - [18] HUANG D, XIA Z Q, MWESIGYE J, et al. Pain – attentive network; A deep spatio – temporal attention model for pain estimation [J]. Multimedia Tools and Applications, 2020, 79(37–38): 28329–28354.
 - [19] HUANG Dong, FENG Xiaoyi, ZHANG Haixi, et al. Spatio – temporal pain estimation network with measuring pseudo heart rate gain [J]. IEEE Transactions on Multimedia, 2021, 24: 3300–3313.
 - [20] XU Haochen, LIU Manhua. A deep attention transformer network for pain estimation with facial expression video [C]// Proceedings of the Chinese Conference on Biometric Recognition. Cham: Springer, 2021; 112–119.
 - [21] LIAO Jiachen, HAO Yan, ZHOU Zhuoyi, et al. Sequence-level affective level estimation based on pyramidal facial expression features [J]. Pattern Recognition, 2024, 145: 109958.
 - [22] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, et al. Dropout: A simple way to prevent neural networks from overfitting [J]. Journal of Machine Learning Research, 2014, 15: 1929–1958.
 - [23] OTHMAN E, WERNER P, SAXEN F, et al. Cross – database evaluation of pain recognition from facial video [C]// Proceedings of the 11th International Symposium on Image and Signal Processing and Analysis. 2019. DOI: 10.1109/ISPA.2019.8868562
 - [24] SUBRAMANIAM S D, DASS B. Automated nociceptive pain assessment using physiological signals and a hybrid deep learning network [J]. IEEE Sensors Journal, 2021, 21(3): 3335–3343.
 - [25] DRAGOMIR M C, FLOREA C, PUPEZESCU V. Automatic subject independent pain intensity estimation using a deep learning approach [C]// Proceedings of the 2020 International Conference on e-Health and Bioengineering. Piscataway, NJ: IEEE, 2020; 1–4.
 - [26] FEGHOUL K, BOUAZIZI M, MAIA D S. Facial action unit detection using 3D face landmarks for pain detection [C]// Proceedings of the 2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society. Piscataway, NJ: IEEE, 2023; 1–5.
 - [27] ZHI Ruicong, WAN Ming. Dynamic facial expression feature learning based on sparse RNN [C]// IEEE Joint International Information Technology and Artificial Intelligence Conference. Piscataway, NJ: IEEE, 2019; 24–26.
 - [28] ALGHAMDI T, ALAGHBAND G. SAFEPA: An expandable multi-pose facial expressions pain assessment method [J]. Applied Sciences, 2023, 13(12): 7206.
 - [29] THIAM P, HIEHN H, BRAUN D A, et al. Multi – modal pain intensity assessment based on physiological signals; A deep learning perspective [J]. Frontiers in Physiology, 2021, 12: 720464.