

# 基于改进 3D 卷积神经网络的疼痛检测<sup>①</sup>

黄伟聪, 周卓沂, 李雄彬, 梁 艳

(华南师范大学 软件学院, 佛山 528225)

通信作者: 梁 艳, E-mail: liangyan@m.scnu.edu.cn



**摘 要:** 在临床实践中, 精确评估疼痛对于疼痛管理和诊断至关重要. 但传统的评估方法主观性高且依赖医生经验, 迫切需要更可靠客观的替代方法. 利用深度学习的方法实现基于面部表情的疼痛检测研究近年已取得显著进展, 但复杂的结构和高计算成本制约了其实际应用. 因此, 本文提出了一个改进的 3D 卷积神经网络, 采用轻量级的 3D 卷积神经网络 L3D 作为骨干网络, 并结合改进的 SE 注意力机制, 把多个不同尺度的特征进行融合, 捕捉疼痛序列中具有较强辨别能力的时空特征. 在 UNBC-McMaster 和 BioVid 数据集上进行评估, 与最新方法相比, 该方法在疼痛检测性能以及计算复杂度上取得了优势.

**关键词:** 疼痛检测; 疼痛表情; 3D 卷积神经网络; 轻量级; 注意力机制; 特征融合

引用格式: 黄伟聪, 周卓沂, 李雄彬, 梁艳. 基于改进 3D 卷积神经网络的疼痛检测. 计算机系统应用, 2024, 33(2): 291-298. <http://www.c-s-a.org.cn/1003-3254/9425.html>

## Pain Detection Based on Improved 3D Convolutional Neural Network

HUANG Wei-Cong, ZHOU Zhuo-Yi, LI Xiong-Bin, LIANG Yan

(School of Software, South China Normal University, Foshan 528225, China)

**Abstract:** In clinical practice, accurate pain assessment is crucial for pain management and diagnosis. However, traditional assessment methods are highly subjective and reliant on the expertise of medical professionals, highlighting the urgent need for more reliable and objective alternatives. The research on pain detection based on facial expression by deep learning has made remarkable progress in recent years, whereas the complex structure and high computational cost restrict its practical application. Therefore, this study proposes an improved 3D convolutional neural network (CNN) that utilizes a lightweight 3D CNN named L3D as the backbone network. It also incorporates an enhanced SE attention mechanism to fuse multiple features of different scales, capturing spatiotemporal characteristics with strong discriminative power in pain sequences. The proposed method is evaluated on UNBC-McMaster and BioVid datasets. Compared with the state-of-the-art methods, the proposed method achieves superior performance in pain detection and computational complexity.

**Key words:** pain detection; pain expression; 3D convolutional neural network (CNN); lightweight; attention mechanism; feature fusion

## 1 引言

疼痛是一种不愉快的感觉和情绪体验, 与实际或潜在的组织损伤有关, 常作为感知相关伤害的信号出现, 一般由疾病或伤害引起<sup>[1]</sup>. 疼痛会带来多种病理生

理风险, 其评估对患者的医疗诊断和治疗至关重要, 是医学诊断和治疗的重要指标之一. 它的精确评估有助于重症监护病房和术后程序中有效的疼痛控制或患者愈合管理<sup>[2]</sup>.

① 基金项目: 国家自然科学基金面上项目 (62076103); 广东省普通高校特色创新项目 (2022KTSCX035)

收稿时间: 2023-08-16; 修改时间: 2023-09-15; 采用时间: 2023-10-20; csa 在线出版时间: 2023-12-25

CNKI 网络首发时间: 2023-12-27

目前,疼痛的识别主要依靠医生的主观判断以及病人的自述描述,然而,这些方法存在较强的主观性,而且不适用于非交流性患者,如婴幼儿、意识障碍患者等。考虑到面部表情与疼痛程度的高度相关性,近年来,有较多学者开展基于面部表情的疼痛检测方法研究<sup>[3,4]</sup>。这类方法使用非接触式记录设备(如数码相机)采集数据,更为便捷友好<sup>[5]</sup>,进行长时间的持续监测也不会引起患者不适。

研究初期,大多使用传统的机器学习方法,通过面部图像进行疼痛评估。Lucey等人<sup>[6]</sup>利用主动外观模型(active appearance model, AAM)<sup>[7]</sup>实现了第1个用于二元疼痛评估的基线模型。Kaltwang等人<sup>[8]</sup>使用外观和形状特征,并结合支持向量机(support vector machine, SVM),用于持续疼痛强度评估。Irani等人<sup>[9]</sup>使用了一种基于面部肌肉释放能量的疼痛强度识别技术。Pedersen<sup>[10]</sup>提出了一种利用RGB值、热人脸图像和深度信息等特征组合的方法。

近年来,随着深度学习技术的快速发展,更多的学者采用基于深度学习的方法进行人脸表情疼痛评估<sup>[11,12]</sup>。Zhou等人<sup>[13]</sup>利用基于递归卷积神经网络(recurrent convolutional neural network, RCNN)的分类器来提取时间特征。Rodriguez等人<sup>[14]</sup>采用了一个经过预训练的长-短期记忆(long short-term memory, LSTM)的VGG-16模型来提取疼痛强度识别的空间和时间特征。除此之外,最近还有学者尝试使用基于其他类型的递归神经网络(recurrent neural network, RNN)模型<sup>[15]</sup>解决视频数据的疼痛检测任务,但时序网络权重重要通过时间反向传播进行更新,模型存在梯度消失问题<sup>[16]</sup>。Fan等人<sup>[17]</sup>使用Vision Transformer进行疼痛检测,但复杂的结构增加了计算的负担。Wu等人<sup>[18]</sup>在疼痛检测上使用注意力机制,验证了注意力对于疼痛检测的有效性。然而这些方法模型复杂,参数过多,导致计算成本高,容易过拟合<sup>[19]</sup>。

以往的疼痛检测大多基于帧级别,即针对单个序列帧进行疼痛评估,主要采用基于卷积神经网络(convolutional neural network, CNN)的方法。然而,单帧识别中,被试可能已经经历过疼痛的过程,但表情还未消退,因而导致误判。考虑到疼痛是一个持续的过程,序列级别的疼痛检测更符合人经历疼痛的过程,更具有可解释性,因而逐渐成为研究的主流。目前大多基于序列级别的疼痛检测研究使用CNN+RNN的结构来处

理数据,先通过CNN来提取图像中的特征,获得帧内的空间信息,再通过RNN处理序列数据中的长期依赖关系,对序列数据建模。这样的处理方式忽略了时空信息之间的关联性。3D卷积(convolutional 3D, C3D)<sup>[20]</sup>作为一种用于处理视频数据等三维信息的深度学习模型,在疼痛检测中将CNN扩展到三维,在卷积过程同时考虑疼痛视频输入的空间维度和时序维度,能够更好地捕获时空关系。比如,Tavakolian等人<sup>[21]</sup>提出了一种基于C3D网络的方法进行疼痛强度评估。但是,传统的C3D卷积参数量和计算量过大,不利于实际应用。

针对目前已有研究存在的问题,本文提出了一种改进的3D卷积神经网络用于疼痛检测,该网络采用轻量级的C3D网络L3D作为骨干网络,并结合了改进的SE注意力机制,能有效提取疼痛序列中的时空信息,融合多尺度特征,从而实现对疼痛程度的回归预测。

## 2 模型结构

### 2.1 改进的3D卷积神经网络

为了解决传统C3D网络参数量大、难以收敛等问题,本文针对疼痛检测任务,提出了一种用于疼痛检测的改进3D卷积神经网络。该网络模型主要由特征提取模块(L3D模型)和多尺度特征融合模块构成,如图1所示。其中,该模型的输入为原始数据集经过MTCNN处理后的人脸图像序列,该序列由L3D模型提取3种不同尺度的特征,包括大尺度特征B、中尺度特征M、小尺度特征S,随后分别经过改进的SE注意力块,提高疼痛相关特征的表达,接着对这些特征使用三线插值等比例上采样为相同大小的特征,并进行拼接。拼接后的特征再次通过注意力块对特征进行加权,得到融合特征B'M'S',经过全连接层FC,最终输出疼痛的强度值。

### 2.2 轻量级三维卷积骨干网络L3D

C3D是一种经典的3D卷积神经网络,它通过使用3D卷积核和池化核来同时考虑空间和时间两个维度上的特征,从而实现了对视频或者三维图像数据的高效表征和理解。C3D模型利用了卷积神经网络在空间和时间上共享权重的特点,使得模型在时间轴上对输入视频进行了建模。但是在计算视频的时间序列特征时,C3D模型需要大量的计算量,训练和应用处在计算资源限制的场景下时,可能需要很长时间完成,甚至无法直接使用。

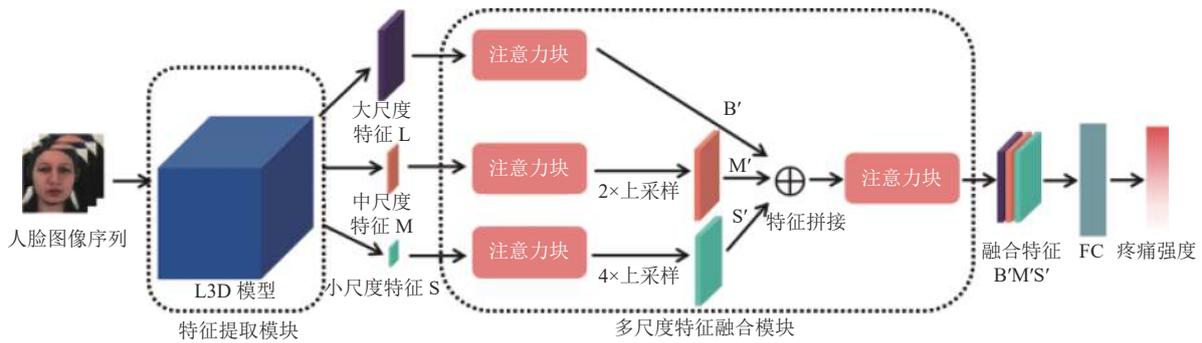


图1 改进的3D卷积神经网络模型

为了更好地解决时序疼痛检测场景下的三维卷积神经网络参数过大,训练时长过久,难以收敛的问题,本文提出一种轻量级的三维卷积骨干网络L3D,在C3D的基础上进行了以下改进。

(1) 重新设计了三维卷积结构:我们对C3D模型进行了改进,具体来说,我们减少了模型各层的输出特征数量,这意味着在每个卷积层中,我们使用了更少的卷积核或者降低了它们的维度.这个改进的效果在于减小了模型的参数量,使得模型更加轻量化,同时降低了计算复杂度。

(2) 添加BN层和最大池化层:我们在每个卷积操作后面添加了一层批归一化(batch normalization, BN)层,用于加速网络的训练和收敛速度. BN层有助于稳定训练过程,减小了梯度消失问题.此外,我们还在模型的最后添加了两层最大池化层,以减少特征数量.这有助于进一步减小模型的复杂度,降低了过拟合的风险。

(3) 简化了全连接层:为了防止模型过拟合并提高训练效率,我们还对全连接层进行了简化.我们减少了全连接层的层数以及特征大小.这个改进意味着我们在模型的最后阶段使用了更少的神经元,减小了参数数量,从而提高了模型的泛化能力。

图2展示了L3D与C3D的网络结构对比.从图2(b)可以看到, L3D整体结构包括8个卷积层、8个BN层、6个最大池化层以及2个全连接层.全部3D卷积核为 $3 \times 3 \times 3$ ,空间和时间上的步长为1. BN层的输入输出特征数相同,对应上一卷积核输出特征数.6个池化层中,除了Pool1池化核为 $1 \times 2 \times 2$ ,其他池化层的池化核都是 $2 \times 2 \times 2$ .全连接层有2层,第1层输出特征数为512,第2层输出特征数为1,对应任务中疼痛检测出来的值。

与图2(a)的C3D网络结构进行比较,可以看出

L3D在结构上增加了BN层,对其他各层参数进行了调参精简,使得模型在不损失过多精度的前提下,具有更低的复杂度,计算量更少,泛化能力更强。

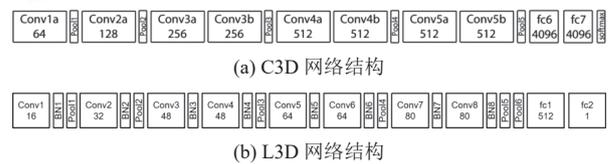


图2 L3D与C3D的网络结构对比

### 2.3 改进的SE注意力机制

在疼痛视频序列中,帧与帧之间比较相似,可能存在通道间的冗余和过度拟合的情况.为了避免一些通道可能会过度拟合训练数据,导致网络在新数据上的泛化性能较差,使用注意力机制方法为特征图不同区域赋予不同的权重,可以让网络强化表达对疼痛检测有效的特征,提升网络的性能和精度。

传统的SE注意力机制的引入可以在不增加网络复杂性的情况下,显著提升模型性能<sup>[22]</sup>,许多学者把SE注意力机制用于不同的计算机视觉任务,取得了优异的结果.但传统的SE注意力机制只使用平均池化的单分支对输入特征进行处理,仅关注特征图中通道上的平均特征形态,它更偏重于疼痛时表情的整体特征,难以同时关注到疼痛时表情的局部显著特征.而在疼痛检测中,人脸图像的局部表情区域,比如眼睛、嘴巴、鼻子等周围的区域与疼痛程度有极大的相关性,更细致地检测出面部区域的细微表情和痛苦迹象有助于提升疼痛检测的准确率。

因此,本文针对疼痛检测任务,对SE注意力机制进行改进,强调与面部表情相关的整体区域和局部区域相关联系,提高网络对面部特征的表达能力.改进的SE注意力模块如图3所示。

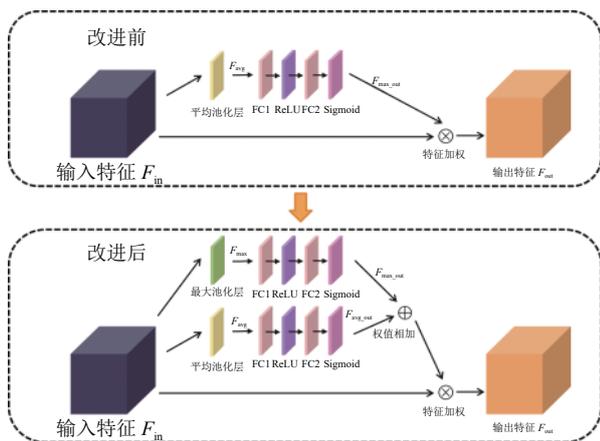


图3 改进的SE注意力模块

首先,改进的SE注意力机制使用平均池化和最大池化的双分支对输入特征进行处理,分别获得平均池化和最大池化的特征,使得网络不仅关注到疼痛时表情的整体特征,更能细致化关注到表情的局部显著特征。

其次,改进的SE注意力机制能自适应地调整每个通道的权重来减少冗余信息和增强有用信息。这一步是通过两个分支里面分别含有的两层全连接层实现的,全连接层的参数会在模型训练中被学习到。最终使得输出的对应输入特征中每个通道的权值不同。而将两个分支的权值相加,可以获得包含整体特征和局部特征相关性的关于输入特征通道的权重信息,确定输入特征通道的注意力权重。通过对注意力权重与输入特征点乘,最终获得带有注意力信息的特征。

改进的SE注意力模块是一种用于提取输入特征中重要信息的神经网络组件。该模块采用了最大池化和平均池化来捕捉输入特征  $F_{in}$  中的不同方面的信息。首先,为了强调输入特征中的显著性信息,即提取那些在整体中最为突出的特征,我们通过最大池化函数对输入特征  $F_{in}$  进行汇聚,得到最大池化输出特征  $F_{max}$ :

$$F_{max} = \text{Max}(F_{in}) \tag{1}$$

其中,Max表示最大池化函数。

其次,为了平滑输入特征并提取整体的一般性特征,我们利用平均池化函数将输入特征  $F_{in}$  进行汇聚,得到平均池化输出特征  $F_{avg}$ :

$$F_{avg} = \text{Avg}(F_{in}) \tag{2}$$

其中,Avg表示平均池化函数。

随后,采用两个全连接层 FC1 和 FC2,分别对

$F_{max}$  和  $F_{avg}$  特征进行处理。首先,通过全连接层 FC1 分别对  $F_{max}$ 、 $F_{avg}$  进行线性变换,并经过 ReLU 函数来引入非线性激活,得到转换后的特征。然后,再将这些转换后的特征通过全连接层 FC2 进行进一步的线性变换。最终,对这些经过处理的特征应用 Sigmoid 函数进行归一化,分别得到两个注意力权重:最大注意力权重  $F_{max\_out}$  和平均注意力权重  $F_{avg\_out}$ ,计算公式如下:

$$F_{max\_out} = \sigma(\text{FC2}(\text{R}(\text{FC1}(F_{max})))) \tag{3}$$

$$F_{avg\_out} = \sigma(\text{FC2}(\text{R}(\text{FC1}(F_{avg})))) \tag{4}$$

其中,FC1、FC2表示两个全连接层,R为ReLU函数, $\sigma$ 为Sigmoid函数。

接下来,对输入特征  $F_{in}$  中的每个元素按照注意力权重进行加权,以强调重要的特征,即:将两个注意力权重分布  $F_{max\_out}$  和  $F_{avg\_out}$  与输入特征  $F_{in}$  进行点乘操作。

$$F_{out} = (F_{max\_out} + F_{avg\_out}) \odot F_{in} \tag{5}$$

其中, $F_{out}$ 为输出特征, $\odot$ 表示按元素进行点乘操作。

### 2.4 多尺度特征融合

浅层特征具有较高的分辨率和丰富的位置信息,适合于检测小尺寸目标或具有细粒度结构的目标。而深层特征具有更广阔的感受野和更丰富的语义信息,能够更好地捕捉大尺寸目标或具有复杂语义的目标。在网络中设计多尺度结构,可以在不同的特征层上进行目标检测,并融合多尺度的特征信息。这样可以充分利用特征图的信息,提高检测器对相关特征的感知能力和准确性。

因此,本文通过L3D中不同尺度的三维卷积核,获取大、中、小3种不同尺度的图像特征信息,在多尺度融合的基础上结合改进的SE注意力机制,设计了一种多尺度特征融合模块,对L3D模型输出的多尺度特征进行融合。多尺度融合模块结构如图4所示。

为了按一定比例选用不同的尺度特征,本文选用L3D中Conv4、Conv6、Conv8这3个卷积对应的3个尺度特征作为大、中、小尺度特征,分别记为B、M、S。这3个卷积的核逐次经过相同尺寸的池化层,所以他们对应的特征成4:2:1的等比关系。大尺度特征B经过注意力块后得到B'。中尺度特征M和小尺度特征S分别经过注意力块后,进行上采样得到M'和S'。上采样的目的是为了将不同的尺度图像特征信息调整

为相同尺寸,有利于后续的拼接操作.本文具体使用了三线性插值方法进行上采样.随后,把 B'、M'和 S'以拼

接的方式融合起来,并经过注意力块,最终得到一个融合了3种尺度特征信息的疼痛表情融合特征 B'M'S'.

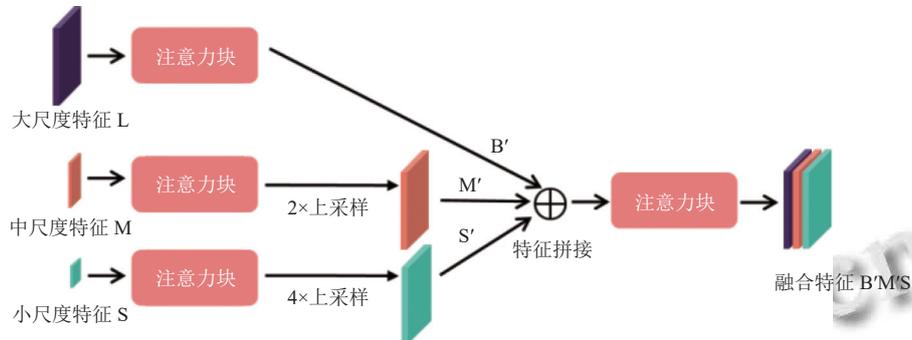


图4 多尺度融合模块

### 3 实验

#### 3.1 实验数据集

本文采用 UNBC-McMaster 和 BioVid 两个公开数据集对改进的 3D 卷积神经网络进行测试.

UNBC-McMaster 数据集<sup>[23]</sup>广泛用于疼痛检测,包含来自 25 位有肩痛的参与者的 200 个视频.数据集中的视频样本是在患者正常肩部和患病肩部的伸展、屈曲以及内旋和外旋期间收集的.在本文实验中,我们遵循基于视频的疼痛评估中常用的做法,采用自我报告的视觉模拟量表 (VAS)<sup>[24]</sup>作为标签来指示疼痛强度评分. VAS 将疼痛分为 11 个级别,范围从 0 (无痛) 到 10 (难以忍受的疼痛).

BioVid 数据集是一个多模态热痛数据集<sup>[25]</sup>,本文仅使用与疼痛相关的 Part A 部分,包括 87 个受试者的 8700 个视频.受试者在严格控制条件下受到 5 个级别的人工诱导的热刺激,分别对应 5 个疼痛强度级别,即无痛 (0 级)、轻微疼痛 (1 级)、剧烈疼痛 (4 级) 和两个中等疼痛强度 (2 级和 3 级).

#### 3.2 实现细节

本文实验设备配置: Linux 系统, GPU 为 RTX3090 显卡,模型基于 PyTorch 深度学习框架进行训练,在 PyCharm 上进行编译测试.

网络模型训练时采用的批处理大小为 100. 使用 Adam 优化器,设置的学习率和动量分别为 0.001 和 0.9. 使用早停法,训练周期并非固定,当模型收敛,效果不再提升后自动停止训练过程.

考虑到本网络用于疼痛强度检测,不同强度之间存在关联性,视频序列疼痛强度数值并非独立标签,而

是连续相关的,因此,本网络的任务为回归任务,采用 HuberLoss 作为损失函数, delta 设置为 0.7.

在训练及测试模型前,本文首先对视频序列进行预处理.使用多任务卷积神经网络 (multi-task convolutional neural network, MTCNN)<sup>[26]</sup>对视频序列每一帧图像进行人脸定位和分割,把人脸区域归一化为 224×224 像素,并且按视频序列的顺序保存.

本文使用 5 折交叉验证,将数据集分为 5 份,每次使用其中 4 份训练,另 1 份作为测试,依次重复 5 次,取 5 次结果的平均值作为最终的测试结果.

#### 3.3 评价指标

本文提出的网络需要解决回归任务,因此,我们采用平均绝对误差 (mean absolute error, MAE)、均方误差 (mean square error, MSE) 和均方根误差 (root mean squared error, RMSE) 作为评价指标来衡量模型的性能.

MAE、MSE、RMSE 的计算公式分别为:

$$MAE = \frac{\sum |y_{\text{pred}} - y_{\text{true}}|}{n} \quad (6)$$

$$MSE = \frac{\sum (y_{\text{pred}} - y_{\text{true}})^2}{n} \quad (7)$$

$$RMSE = \sqrt{\frac{\sum (y_{\text{pred}} - y_{\text{true}})^2}{n}} \quad (8)$$

其中,  $y_{\text{pred}}$  表示预测值,  $y_{\text{true}}$  表示真实值,  $n$  表示样本数. 对于 MAE、MSE 和 RMSE, 其值越小表示模型预测越准确.

#### 3.4 消融实验

为了验证多尺度融合方法和改进的注意力模块对

网络模型性能的影响,我们在 UNBC-McMaster 数据集和 BioVid 数据集进行了消融实验,结果如表 1 所示。

表 1 改进的 3D 卷积神经网络模型消融实验

方法	MAE	MSE	RMSE
L3D (基准)	2.30/1.25	8.02/2.13	2.83/1.45
L3D+多尺度特征融合	2.24/1.23	<u>7.55</u> /2.01	<u>2.75</u> /1.42
L3D+改进的SE注意力	<u>2.21</u> / <u>1.21</u>	7.65/ <u>1.95</u>	2.77/ <u>1.40</u>
L3D+多尺度特征融合+改进的SE注意力(本文方法)	<b>2.15/1.18</b>	<b>6.94/1.90</b>	<b>2.60/1.38</b>

注:加粗数据为最优结果,下划线数据为次优结果;斜杠前、后的数据分别为在 UNBC-McMaster 和 BioVid 数据集测试的结果

从表 1 的数据可以看出,与基准模型 L3D 相比, L3D+多尺度特征融合方法(不包含改进的 SE 注意力机制)在两个数据集上的评价指标均有所改善。其中, MSE 的值改善最为明显,在 UNBC-McMaster 和 BioVid 数据集上,分别减小了 0.47 和 0.12。实验结果表明,多尺度特征融合有助于提升疼痛检测的准确性和精度。

此外, L3D+改进的 SE 注意力模型与基准模型 L3D 相比,前者在两个数据集上的评价指标均有显著改善。证明改进的 SE 注意力机制的引入对于提升疼痛检测的准确性和精度也具有积极的影响。

在表 1 中的 4 种方法, L3D+多尺度特征融合方法+改进的 SE 注意力(即本文提出的模型)方法效果最好,在两个数据集上的 3 个评估指标都取得了最优结果,证明了多尺度特征融合和改进的 SE 注意力的协同效应。通过结合这两个关键组件, 3D 卷积神经网络模型能够更好地捕捉疼痛信号中的特征信息,并取得比基准模型更好的疼痛检测结果。

### 3.5 方法对比实验

为了进一步验证改进的 3D 卷积神经网络模型在疼痛检测任务中的有效性和优越性,本文把该模型与近年来的方法在 UNBC-McMaster 和 BioVid 数据集上进行对比,结果如表 2、表 3 所示。其中,表 2 中的 C3D 方法以及表 3 中的对比方法均采用复现的结果,表 2 其余的对比方法的结果来自文献。

从表 2 数据可以看出,在 UNBC-McMaster 数据集上,与 RNN-HCRF、Manifold trajectories、ML、CNN-RNN 和 DeepFaceLift 等方法相比,本文提出的网络模型在疼痛回归预测中取得了最优结果。此外,该模型比传统的 C3D 模型具有明显优势, MAE、MSE 和 RMSE 值分别减小了 0.41、0.95、0.21。

表 2 在 UNBC-McMaster 数据集的对比结果

方法	MAE	MSE	RMSE
RNN-HCRF <sup>[27]</sup>	2.46	—	—
Manifold trajectories <sup>[28]</sup>	2.44	9.92	3.15
ML <sup>[24]</sup>	2.39	—	—
CNN-RNN <sup>[29]</sup>	2.34	—	—
DeepFaceLift <sup>[30]</sup>	2.30	—	—
C3D	2.56	7.89	2.81
本文方法	2.15	6.94	2.60

表 3 在 BioVid 数据集的对比结果

方法	MAE	MSE	RMSE
ResNet18-LSTM	1.21	2.02	1.41
ResNet18-GRU	1.20	2.21	1.48
C3D	1.20	1.97	1.40
本文方法	1.18	1.90	1.37

我们也在 BioVid 数据集上把本文提出的网络与 ResNet18-LSTM、ResNet18-GRU 和 C3D 进行比较(见表 3)。本文提出的网络在疼痛回归预测方面表现出更好的性能,它的 MAE、MSE 和 RMSE 值分别为 1.18、1.90 和 1.37,对比结果证明了该网络的先进性。

总体而言,在 UNBC-McMaster 和 BioVid 两个数据集的实验结果表明,本文提出的改进 3D 卷积神经网络模型在疼痛检测任务中具有显著的先进性。与近年来的疼痛检测方法相比,该模型在疼痛回归预测的准确性和泛化能力方面表现出色。

### 3.6 计算复杂度比较

在深度学习中,模型的参数量和计算量是评估模型复杂度和性能的重要指标。较大的模型通常具有更高的参数量和计算量,这可以帮助模型更好地拟合数据,并提高其准确性和泛化能力。然而,较大的模型也会导致更多的计算资源和时间消耗,并且容易过拟合。

我们比较了本文方法以及疼痛检测常用的时空模型的参数量与计算量,如表 4 所示。本文提出的网络参数量仅为 3.04M,计算量仅为 12.96G,与常规处理时序的 ResNet18-LSTM、ResNet18-GRU 网络相比,参数量及计算量约为它们的 1/5。与传统的 C3D 网络相比,参数量不足它的 1/25,而计算量大小则仅为它的 1/47,远小于 C3D 网络。结果表明,本文提出的网络模型在参数量和计算量方面相对较小,具有更高的效率和轻量级的特点,在实际应用中具有较大优势。

## 4 结论

本文提出了一种改进的 3D 卷积神经网络,结合轻

量级 3D 卷积骨干网络和改进的 SE 注意力机制,用于疼痛检测任务. 该模型通过融合多尺度特征和引入改进的 SE 注意力机制,能高效地提取疼痛序列中的时空信息,并实现对疼痛程度的回归预测.

表 4 时空模型参数量与计算量比较

方法	参数量 (M)	计算量 (G)
ResNet18-LSTM	15.03	58.79
ResNet18-GRU	14.86	58.79
C3D	77.98	617.84
L3D	1.07	10.62
本文方法	3.04	12.96

在 UNBC-McMaster 和 BioVid 两个公开的疼痛检测数据集进行评估,本文提出的网络模型在两个数据集上均取得了最优结果. 此外,该模型在参数量和计算量方面具有明显的优势. 实验结果证明了该模型在疼痛检测任务中具有较好的性能和泛化能力.

本文提出的网络模型简单且高效的设计思路为轻便的疼痛检测方案提供了新的研究思路和参考价值. 然而,本文方法仍有改进的空间. 例如,多尺度融合模型可以进一步研究尺度数量和不同尺度间的融合方式. 此外,疼痛检测的性能仍有提升的潜力.

### 参考文献

- Gholami B, Haddad WM, Tannenbaum AR. Relevance vector machine learning for neonate pain intensity assessment using digital imaging. *IEEE Transactions on Biomedical Engineering*, 2010, 57(6): 1457–1466. [doi: [10.1109/TBME.2009.2039214](https://doi.org/10.1109/TBME.2009.2039214)]
- Paganelli AI, Mondéjar AG, da Silva AC, *et al.* Real-time data analysis in health monitoring systems: A comprehensive systematic literature review. *Journal of Biomedical Informatics*, 2022, 127: 104009. [doi: [10.1016/j.jbi.2022.104009](https://doi.org/10.1016/j.jbi.2022.104009)]
- Nagireddi JN, Vyas AK, Sanapati MR, *et al.* The analysis of pain research through the lens of artificial intelligence and machine learning. *Pain Physician*, 2022, 25(2): E211–E243.
- De Sario GD, Haider CR, Maita KC, *et al.* Using AI to detect pain through facial expressions: A review. *Bioengineering*, 2023, 10(5): 548. [doi: [10.3390/bioengineering10050548](https://doi.org/10.3390/bioengineering10050548)]
- Alghamdi T, Alagband G. Facial expressions based automatic pain assessment system. *Applied Sciences*, 2022, 12(13): 6423. [doi: [10.3390/app12136423](https://doi.org/10.3390/app12136423)]
- Lucey P, Cohn JF, Prkachin KM, *et al.* Painful monitoring: Automatic pain monitoring using the UNBC-McMaster shoulder pain expression archive database. *Image and Vision Computing*, 2012, 30(3): 197–205. [doi: [10.1016/j.imavis.2011.12.003](https://doi.org/10.1016/j.imavis.2011.12.003)]
- Cootes TF, Edwards GJ, Taylor CJ. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001, 23(6): 681–685. [doi: [10.1109/34.927467](https://doi.org/10.1109/34.927467)]
- Kaltwang S, Rudovic O, Pantic M. Continuous pain intensity estimation from facial expressions. *Proceedings of the 2012 International Symposium on Visual Computing*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012. 368–377. [doi: [10.1007/978-3-642-33191-6\\_36](https://doi.org/10.1007/978-3-642-33191-6_36)]
- Irani R, Nasrollahi K, Moeslund TB. Pain recognition using spatiotemporal oriented energy of facial muscles. *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops*. Boston: IEEE, 2015. 80–87.
- Pedersen H. Learning appearance features for pain detection using the UNBC-McMaster shoulder pain expression archive database. *Proceedings of the 10th International Conference on Computer Vision Systems*. Copenhagen: Springer, 2015. 128–136.
- Lucey P, Cohn JF, Prkachin KM, *et al.* Painful data: The UNBC-McMaster shoulder pain expression archive database. *Proceedings of the 2011 IEEE International Conference on Automatic Face & Gesture Recognition*. Santa Barbara: IEEE, 2011. 57–64.
- Gkikas S, Tsiknakis M. Automatic assessment of pain based on deep learning methods: A systematic review. *Computer Methods and Programs in Biomedicine*, 2023, 231: 107365. [doi: [10.1016/j.cmpb.2023.107365](https://doi.org/10.1016/j.cmpb.2023.107365)]
- Zhōu J, Hong XP, Su F, *et al.* Recurrent convolutional neural network regression for continuous pain intensity estimation in video. *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops*. Las Vegas: IEEE, 2016. 84–92.
- Rodríguez P, Cucurull G, González J, *et al.* Deep pain: Exploiting long short-term memory networks for facial expression classification. *IEEE Transactions on Cybernetics*, 2022, 52(5): 3314–3324. [doi: [10.1109/TCYB.2017.2662199](https://doi.org/10.1109/TCYB.2017.2662199)]
- Allen-Zhu Z, Li YZ, Song Z. On the convergence rate of training recurrent neural networks. *Proceedings of the 33rd International Conference on Neural Information Processing Systems*. Vancouver: Curran Associates Inc., 2019. 599.
- Hochreiter S. The vanishing gradient problem during learning recurrent neural nets and problem solutions. *International Journal of Uncertainty, Fuzziness and*

- Knowledge-based Systems, 1998, 6(2): 107–116. [doi: [10.1142/S0218488598000094](https://doi.org/10.1142/S0218488598000094)]
- 17 Fan YM, Wang HW, Zhu XY, *et al.* FER-PCVT: Facial expression recognition with patch-convolutional vision transformer for stroke patients. *Brain Sciences*, 2022, 12(12): 1626. [doi: [10.3390/brainsci12121626](https://doi.org/10.3390/brainsci12121626)]
- 18 Wu J, Shi Y, Tang Y, *et al.* Global-local combined features to detect pain intensity from facial expression images with attention mechanism. 2023. [doi: [10.21203/rs.3.rs-2627259/v1](https://doi.org/10.21203/rs.3.rs-2627259/v1)]
- 19 Ayril T, Pedersoli M, Bacon S, *et al.* Temporal stochastic Softmax for 3D CNNs: An application in facial expression recognition. *Proceedings of the 2021 IEEE Winter Conference on Applications of Computer Vision*. Waikoloa: IEEE, 2021. 3029–3038.
- 20 Tran D, Bourdev L, Fergus R, *et al.* Learning spatiotemporal features with 3D convolutional networks. *Proceedings of the 2015 IEEE International Conference on Computer Vision*. Santiago: IEEE, 2015. 4489–4497.
- 21 Tavakolian M, Hadid A. A spatiotemporal convolutional neural network for automatic pain intensity estimation from facial dynamics. *International Journal of Computer Vision*, 2019, 127(10): 1413–1425. [doi: [10.1007/s11263-019-01191-3](https://doi.org/10.1007/s11263-019-01191-3)]
- 22 Hu J, Shen L, Sun G. Squeeze-and-excitation networks. *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018. 7132–7141.
- 23 El Morabit S, Rivenq A, Zighem MEN, *et al.* Automatic pain estimation from facial expressions: A comparative analysis using off-the-shelf CNN architectures. *Electronics*, 2021, 10(16): 1926. [doi: [10.3390/electronics10161926](https://doi.org/10.3390/electronics10161926)]
- 24 Casti P, Mencattini A, Filippi J, *et al.* Metrological characterization of a pain detection system based on transfer entropy of facial landmarks. *IEEE Transactions on Instrumentation and Measurement*, 2021, 70: 5008308.
- 25 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. *Proceedings of the 2013 IEEE International Conference on Cybernetics*. Lausanne: IEEE, 2013. 128–131.
- 26 Kumar KK, Kasiviswanadham Y, Indira DVSNV, *et al.* Criminal face identification system using deep learning algorithm multi-task cascade neural network (MTCNN). *Materials Today: Proceedings*, 2023, 80: 2406–2410. [doi: [10.1016/j.matpr.2021.06.373](https://doi.org/10.1016/j.matpr.2021.06.373)]
- 27 Lopez Martinez D, Rudovic O, Picard R. Personalized automatic estimation of self-reported pain intensity from facial expressions. *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops*. Honolulu: IEEE, 2017. 70–79.
- 28 Szczapa B, Daoudi M, Berretti S, *et al.* Automatic estimation of self-reported pain by interpretable representations of motion dynamics. *Proceedings of the 25th International Conference on Pattern Recognition*. Milan: IEEE, 2021. 2544–2550.
- 29 Erekat D, Hammal Z, Siddiqui M, *et al.* Enforcing multilabel consistency for automatic spatio-temporal assessment of shoulder pain intensity. *Proceedings of the 2020 Companion Publication of the 2020 International Conference on Multimodal Interaction*. ACM, 2020. 156–164.
- 30 Liu DB, Peng FJ, Rudovic O, *et al.* DeepFaceLIFT: Interpretable personalized models for automatic estimation of self-reported pain. *Proceedings of the 1st IJCAI Workshop on Artificial Intelligence in Affective Computing*. Melbourne: PMLR, 2017. 1–16.

(校对责编: 牛欣悦)