

基于 ResNet50+SE 的大学生抑郁症检测算法研究

董家钢

(浙江育英职业技术学院, 浙江 杭州 310018)

一、引言

(一)研究背景

抑郁症作为全球公共卫生领域的重要议题,在大学生群体中呈现高发趋势。流行病学研究表明,约 25% 的大学生存在抑郁症状,14% 的个体曾出现与自杀相关的心理或行为倾向^[1]。抑郁症不仅严重影响大学生的学业成就、社交互动和日常生活质量,甚至会导致自杀等极端后果^[2]。在抑郁症的诊断方面,传统的临床诊断和自评量表检测方法长期占据主导地位。临床诊断主要依赖医生对患者症状的观察和病史询问,而自评量表,如汉密尔顿抑郁量表(HAM-D)^[3]、症状自评量表(SCL-90-R)^[4]、抑郁自评量表(BDI)^[5]等,虽然操作相对简便,但具有明显的主观性缺陷,导致检测结果的准确性不高。随着技术的进步,新兴的生物学方法,如基于神经递质和组学技术的研究^[6],尽管在理论上具有较高的准确性,但面临着系统复杂、实验环境难以模拟人体真实情况以及伦理问题等困难,限制了其广泛应用。在此背景下,人工智能技术的快速发展为抑郁症的检测带来了新的契机。其中包括可穿戴 AI、移动应用程序、机器学习与核磁共振结合等技术^{[7][8][9]}可以在提供与焦虑和抑郁相关的心理健康服务方面提供巨大的前景。而其背后所支撑的深度学习技术,为更准确、高效的抑郁症检测提供了新的可能^[10]。

(二)研究现状

当前抑郁症的检测方法主要有以下几类:

1.抑郁量表检测:这是最为传统和广泛使用的方法,其核心是依据医生的观察和临床访谈,并借助各种抑郁量表进行评估。例如,汉密尔顿量表在临床研究中使用较为普遍,有研究对其 6 个版本(HAMD-6)结合简明精神病评定量表(BPRS-5)

进行了信度和效度的分析^[11],发现该量表在临床应用上有一定的优点,如对抑郁症的严重程度和治疗变化较为敏感,且操作简便,但在测量维度的全面性上有所欠缺,不具备理想的心理测试品质。这种方法的局限性促使研究人员寻求其他更优的检测手段。

2.生物学因素检测:近年来,众多生物学因素被纳入抑郁症的研究范畴,如多巴胺、去甲肾上腺素、5-羟色胺^[12]等神经递质被认为与抑郁症的发病机制密切相关,神经递质假说已成为较为公认的理论基础。然而,这些生物学研究也面临着诸多挑战,如机体代谢受多种因素干扰,生物标志物采集困难,以及在实验中难以完全模拟人体环境,导致实验数据可能存在偏差。

3.人工智能技术检测:人工智能技术在抑郁症检测领域展现出了新的活力。有研究提出了 1D-CNN-GRU-ATTN 模型,该模型结合了 CNN 和 GRU 的优点,在文本信息处理上表现出色,其性能相较于以往的深度学习方法有 10% 的提升,准确率达到 98%^[13]。斯坦福大学李飞飞团队利用 3D 面部表情和口语语言的多模态数据,采用因果卷积网络(C-CNN)技术,对抑郁症症状的严重程度进行评估,展现了多模态数据在抑郁症智能检测中的潜力^[14]。尽管如此,现有的人工智能检测方法仍存在诸如准确性待提高、对大学生群体的针对性多模态检测技术尚未成熟等问题。

(三)研究目的

基于上述研究,本文舍弃了局限性更大传统的抑郁量表检测方法和生物检测方法,选择使用人工智能技术检测方法。将 ResNet50 模型与 SE 模块相结合,提高了模型的性能。并将研究对象定为抑郁症发病率更高且表情更丰满,更立体的大学生群体。因此,本研究的贡献主要体现在以下两点:

(1). 本研究提出一种融合 ResNet50 残差网络与 SE(Squeeze-and-Excitation)模块的深度学习模型,通过嵌入 SE 模块的通道注意力机制解决了传统模型在关键特征提取能力上的不足。

(2). 首次针对大学生群体筛选 RAF 数据集,保留 7 类基本表情数据,并通过数据增强策略提升数据多样性,确保模型对大学生表情特征的精准适配。

(3). 为大学生抑郁症的早期筛查提供了一种非侵入式、高精度的智能化解决方案,弥补传统量表主观性强与生物检测复杂度高的缺陷,具有潜在的社会应用价值。

二、方法

本研究采用的是 RAF (Real-world Affective Faces)数据集^[15],由清华大学媒体与网络实验室(Tsinghua Media and Networking Group)发布,是面部表情识别研究的重要资源。该数据集包含约 29672 张真实世界的面部图像,涵盖 7 种基本表情(愤怒、厌恶、恐惧、快乐、悲伤、惊讶和中性)以及 11 种复合表情。这些图像具有丰富的多样性,涉及不同年龄、性别、种族、姿态和光照条件,并且每张图像都经过多名标注者的独立标注,以确保标注的准确性。在本研究中,为了使研究更贴合大学生群体,我们从 RAF 数据集中筛选出符合大学生特征的 7 种基本表情的数据图片。筛选过程基于图像中人物的年龄、性别、种族以及面部特征等标准,以确保所选样本能够代表大学生群体的面部表情特征。具体而言,我们依据图像中人物的外貌特征(如年龄在 18-25 岁之间、常见的大学生面部特征等)进行人工筛选,以确保所选样本与大学生群体的特征相匹配。最终确定训练集包含 3142 张图像,测试集包含 524 张图像,共计 2618 张图像。

通过这种方式,我们能够确保所使用的数据集不仅具有多样性,而且能够准确反映大学生群体的面部表情特征,从而为模型的训练和验证提供更加可靠的数据基础。

1. 数据预处理

为提升模型对复杂场景的适应性与特征学习能力,本研究设计了一套系统性的数据增强框架。该框架通过几何形变、色彩扰动和标准化处理的

协同作用,旨在缓解数据分布偏差与小样本过拟合问题,同时确保特征空间的一致性,为模型训练提供高质量的泛化样本。

首先,在几何空间变换方面,本研究通过多尺度随机裁剪(缩放区间为 0.8 - 1.0)和固定分辨率调整(目标尺寸为 112×112)模拟不同成像距离下的面部比例变化。这种设计不仅增强了模型对局部细节与全局结构的联合感知能力,还通过统一输入尺寸为后续特征提取提供了标准化的空间基础。此外,为消除模型对方向特征的偏好性依赖,本研究引入了基于 50% 概率的水平镜像翻转,并施加了±15°的随机旋转变换,以覆盖头部自然姿态变化。这一策略有效提升了模型对倾斜视角的鲁棒性,为复杂场景下的特征学习提供了更广泛的视角覆盖。

其次,在色彩扰动方面,本研究通过随机调整图像的亮度(幅度 0.2)、对比度(0.2)、饱和度(0.2)以及色调(偏移 0.1),模拟实际环境中光照不均与设备色差对表情特征的影响。这种色彩扰动策略避免了模型对特定色彩模式的过度依赖,同时增强了模型对复杂光照条件的适应性,为特征提取提供了更鲁棒的输入。

最后,为优化梯度传播效率与训练稳定性,本研究基于 ImageNet 预训练模型的统计参数(均值: 0.485, 0.456, 0.406; 标准差: 0.229, 0.224, 0.225)对图像进行归一化处理。这一标准化操作不仅消除了输入数据的尺度差异,还为模型训练提供了统一的特征空间,从而进一步提升了训练过程的稳定性和收敛速度。数据预处理流程如图 1 所示。

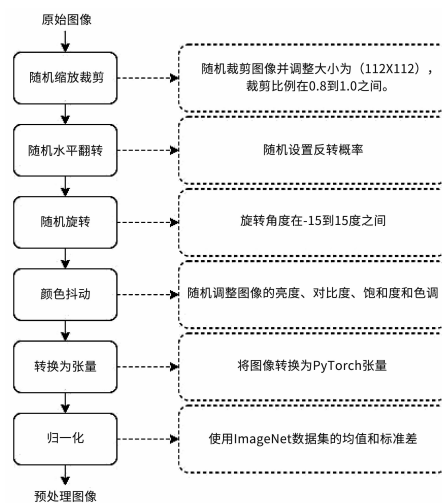


图1 数据预处理流程

2. 模型构建

在本研究中,我们提出了一种深度学习模型,即 ResNet50 与 SE(Squeeze-and-Excitation)模块的结合体。该模型通过在 ResNet50 架构的每个残差块中集成 SEBlock,实现了对通道特征重要性的动态调整,从而增强了模型对关键特征的识别能力。ResNet50 架构的核心优势在于其残差块设计,该设计通过引入 Shortcut 连接有效缓解了深层网络训练中的梯度消失问题,使得网络能够支持更深层次的训练。而 SEBlock 模块则通过执行 Squeeze、Excitation 和 Scale 三个操作,对特征图的通道间关系进行建模,实现了对通道特征响应的自适应重新校准。

在 ResNet50+SE 模型中,每个残差块均采用了 Bottleneck 结构设计。这种结构首先利用 1×1 卷积层对输入特征图的通道数进行压缩,以减少计算量;随后,通过 3×3 卷积层提取空间特征;最后,再次使用 1×1 卷积层恢复通道数至原始维度。在残差路径的末端,模型集成了 SEBlock 模块,该模块将经过校准的特征与 Shortcut 连接的输出相加,并通过 ReLU 激活函数生成最终的输出特征图。这种设计不仅保留了 ResNet50 在深度扩展方面的优势,而且通过 SEBlock 模块赋予了网络自适应的特征选择能力,从而显著提升了模型的识别精度。通过这种结合,我们能够在保持模型深度的同时,提高其对重要特征的敏感度,进而优化模型的整体性能。

(1) SEBlock 模型:

SEBlock(Squeeze-and-Excitation Block)是一种用于提升卷积神经网络性能的轻量级模块^[6]。它通过对特征图的通道间关系进行建模,自适应地重新校准通道特征响应,从而增强模型对重要特征的敏感性,并抑制不重要的特征。SEBlock 主要由三个操作部分组成:Squeeze,Excitation 和 Scale。其中 Squeeze 通过全局平均池化将每个通道的所有空间位置的值相加,然后除以空间位置的数量,得到一个与通道数量相同的一维向量,从而将输入特征图的空间维度进行压缩,得到一个通道描述符。Excitation 通过一个简单的全连接神经网络,根据 Squeeze 操作得到的通道描述符,为每个通道生成一个权重。权重的大小通常与输入特征图的通道数量相同,并且每个权重都对应于

一个通道的重要性因子。Scale 则是将 Excitation 操作得到的权重与输入特征图的每个通道相乘,从而实现对通道重要性的调整。

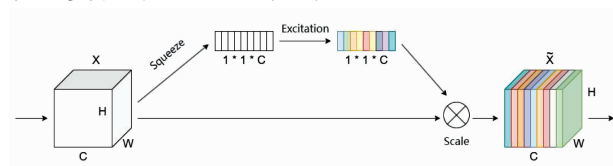


图2 SEBlock 结构图

(2)ResNet50 模型:

ResNet(Residual Network,残差网络)是一种深度神经网络架构,由微软研究院的研究员提出^[7]。它在 ImageNet 竞赛中取得了冠军,并显著提高了图像识别的准确率,成为深度学习领域的一个重要里程碑,其核心在于残差块(Residual Block)的设计。残差块允许网络在训练过程中学习恒等映射,有助于缓解深度神经网络中的梯度消失和梯度爆炸问题。残差结构如图 3 所示,其中使用了两种类型的残差块:同层使用的残差块如图 3(a)和跨层使用的残差块如图 3(b)。图中虚线部分为 Shortcut,即残差连接。图 3(a)中的残差块输入与输出的矩阵深度维度一致;图 3(b)的展示中,输入矩阵则可以改变其深度维度。残差学习的核心公式为:

$$(1)$$

图 3 中两种不同的残差块都是先输入原始矩阵 x ,通过复杂的卷积处理,得到特征矩阵 $F(x)$,最后将 $F(x)$ 与 x 相加得到最终的输出矩阵 $H(x)$ 。

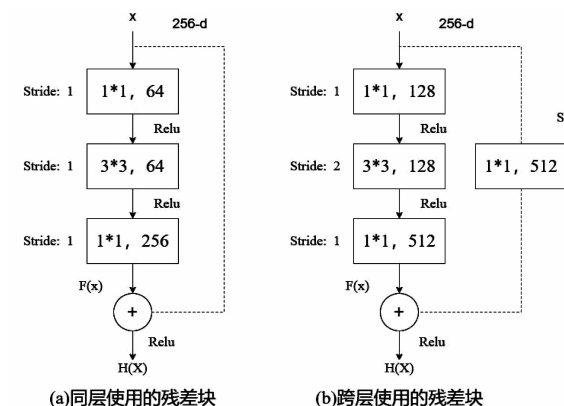


图3 残差块结构图

(3)模型训练策略

在模型训练过程中,本研究通过综合性的训练策略确保模型能够有效学习数据特征并泛化到

新的数据。首先,采用交叉熵损失函数 (Cross-Entropy Loss) 作为优化目标,该函数能够有效衡量模型预测概率分布与真实标签分布之间的差异,从而为模型参数调整提供明确的方向。其次,使用 Adam 优化器作为训练的核心算法,其初始学习率设置为 0.001,并在训练过程中根据验证集的性能动态调整学习率。Adam 优化器结合了动量和均方根传播的优点,能够根据模型参数的梯度自动调整学习率,从而提高训练效率并加速模型收敛。

为了防止模型过拟合,本研究在模型训练中加入了 L2 正则化,通过在损失函数中添加模型权重的平方和项,约束模型权重的大小,降低模型复杂度,增强其泛化能力。L2 正则化的系数设置为 0.0001,这一策略在训练过程中对模型权重进行了有效约束,使得模型在学习数据特征的同时保持了良好的泛化性能。

此外,训练轮数(epoch)设置为 50,这意味着模型会遍历整个训练数据集 50 次。在每个 epoch 结束后,模型在验证集上进行性能评估,并保存性能最优的模型权重。这种策略不仅确保了模型能够充分学习数据特征,还通过验证集的评估避免了过拟合的发生。

3. 评价指标

在本研究中,我们对基于 ResNet50 和 SE (Squeeze-and-Excitation) 模块的大学生抑郁症检测算法进行了全面的性能评估。评估的目的不仅在于验证模型的理论设计有效性,更在于确保其在实际筛选应用中的可靠性和稳健性。因此,本文采用混淆矩阵、准确率、精确率、召回率及 F1 分数等多维度评价指标,从分类错误类型、类别不平衡鲁棒性、模型综合性能等角度全面分析算法表现。这些指标不仅量化了 SE 模块对特征选择能力的提升,还为模型的实际部署提供了关键优化方向。

1. 混淆矩阵(Confusion Matrix)

混淆矩阵是用于评估分类模型性能的一种表格,通常用于二分类问题。在本实验中可用来进行模型的评估,其结构如下表:

表 1 混淆矩阵结构表

	预测为正类	预测为负类
实际为正类	TP (True Positive)	FN (False Negative)
实际为负类	FP (False Positive)	TN (True Negative)

其中,TP、TN、FP 和 FN 分别是真正类(True Positive)、真负类 (True Negative)、假正类(False Positive)和假负类(False Negative)。本研究中 TP 表示模型预测为患有抑郁症且实际为患有抑郁症的患者数,TN 表示模型预测未患抑郁症且实际也为未患抑郁症的患者数,FP 表示模型预测为患有抑郁症但实际未患抑郁症的患者数,FN 表示模型预测为未患抑郁症但实际患有抑郁症的患者数。

2. 准确率(accuracy)

准确率是分类模型预测正确的样本数占总样本数的比例,计算公式如下:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

在本研究中,准确率表明了模型正确预测抑郁症患者数占总患者数的比例,反映了模型整体预测的正确性。

3. 精确率(Precision)

精确率是模型预测为正类的样本中,实际为正类的比例。计算公式如下:

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

在本研究中,精确率表示模型预测为抑郁症的样本中有多少真正患有抑郁症,适用于关注误诊的场景,避免将健康人误诊为抑郁症患者。

4. 召回率(Recall)

召回率是实际为正类的样本中,被模型正确预测为正类的比例。计算公式如下:

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

在本研究中,召回率表示模型能够预测出到多少真正的抑郁症,适用于关注漏诊的场景,确保尽可能多地发现真正的抑郁症患者。

5. F1 分数(F1 Score)

F1 分数是精确率和召回率的调和平均数,用于平衡精确率和召回率。计算公式如下:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

在本研究中,F1 确保在精确率和召回率之间

取得平衡,特别适用于类别不平衡的情况,在抑郁症筛查中同时关注模型的精确性和覆盖率。

三、结果与分析

3.1 实验环境

本研究所采用的环境如表 2 所示。

表 2 实验环境

类型	参数
GPU	4070Ti
CUDA	12
内存	16GB
深度学习框架	Pytorch
学习率	0.001
epoch	50
优化器	Adam
Batch_size	10

3.2 性能对比

本研究选取了 ResNet50 +SE、ResNet50、VGG19 和 MoblieNetV2 这四种模型进行比较。表 3 中展示了四种模型在大学生抑郁症检测任务上的表现。

表 3 在不同模型上的性能对比

模型	准确率 (%)	Precision (%)	Recall (%)	F1 Score (%)
ResNet50+SE	78.483	72.173	69.688	70.91
ResNet50	74.687	71.228	68.454	69.81
VGG19	61.082	60.134	49.526	54.32
MoblieNetV2	76.695	72.080	63.502	67.52

从表 3 中可以看出,ResNet50+SE 模型在准确率、精确率、召回率和 F1 分数等方面均表现最佳,这表明该模型在大学生抑郁症检测任务上具有明显优势。具体而言,其准确率达到 78.483%,相较于传统的 ResNet50 模型(74.687%)有显著提升,说明 SE 模块的引入有效增强了模型对关键特征的捕捉能力,从而提高了分类的准确性。与 VGG19 模型相比,ResNet50+SE 模型的各项指标均有大幅提升,这主要是因为 VGG19 虽然网络结构简单,但在处理复杂的面部表情特征时,其特征提取能力相对有限。而与 MoblieNetV2 相比,ResNet50+SE 在保证模型精度的同时,并未在模型轻量化方面做出过多妥协,因此在性能表现上更胜一筹。

为了进一步验证 SE 模块在 ResNet50 模型中的有效性,本文将 ResNet50 与其他两种常见的注意力机制模块——CBAM 和 SKNet 相结合,进行性能对比实验。CBAM (Convolutional Block Attention Module) 通过通道和空间两个维度的注意力机制,能够综合考虑特征图的通道和空间信息,从而更全面地增强特征表达;SKNet(Selective Kernel Network)则通过选择性核融合的方式,使模型能够自适应地选择不同感受野的特征,增强了模型对不同尺度特征的捕捉能力。通过比较这三种注意力机制在 ResNet50 模型上的应用效果,可以更清晰地了解 SE 模块的独特优势,如表 4 所示。

表 4 不同注意力机制上的性能对比

模型	准确率 (%)	Precision (%)	Recall (%)	F1 Score (%)
ResNet50+SE	78.483	72.173	69.688	70.91
ResNet50+CBAM	77.106	71.467	67.934	69.66
ResNet50+SKNet	76.128	71.161	62.192	66.37

从表 4 中可以看出,ResNet50+SE 模型在准确率、精确率、召回率和 F1 分数等方面均优于 ResNet50+CBAM 和 ResNet50+SKNet 模型。具体而言,ResNet50+SE 的准确率比 ResNet50+CBAM 高 1.377%,比 ResNet50+SKNet 高 2.355%。这表明 SE 模块在提升 ResNet50 模型性能方面具有独特的优势。SE 模块通过自适应地重新校准通道特征响应,能够更精准地增强模型对关键特征的敏感性,从而在抑郁症检测任务上取得了更好的效果。相比之下,CBAM 虽然也考虑了通道和空间信息,但在实际应用中,其对模型性能的提升幅度相对较小;而 SKNet 由于主要关注不同感受野的特征融合,在处理复杂的面部表情特征时,其优势并未能充分发挥出来。

3.4 可解释性描述

由表 3 可以看出,ResNet50 模型在加入 SE 注意力机制后的准确率与原 ResNet50 模型相比,准确率提高了 12.796%。同时相较于 VGG19 模型,准确率提高了 17.401%,相较于 MoblieNetV2 模型,准确率提高了 1.788%。结果证明,经筛选过的 RAF 数据集在 ResNet50 + SE 模型上的准确率高 于 ResNet50、VGG19 和 MoblieNetV2 模型,证明 ResNet50 + SE 模型具有更好的性能。

由表4可以看出,ResNet50模型加入SE注意力机制的准确率与加入CBAM注意力机制相比,准确率提高了3.377%,与加入SKNet注意力机制相比,准确率提高了2.355%。结果证明,经筛选过的RAF数据集在加入了SE注意力机制的ResNet50模型上的准确率比加入CBAM和SKNet注意力机制的ResNet50模型上的准确率高。结果证明,SE注意力机制相较于CBAM和SKNet注意力机制,在加入ResNet50模型后对经筛选过的RAF数据集的性能更高。

四、结论

本研究设计并评估了一个基于ResNet50结

合SE模块的大学生抑郁症检测算法。通过利用RAF数据集和深度学习技术,实验结果表明,ResNet50+SE模型在性能上显著优于ResNet50,其准确率提升约13%,而对比其他传统模型,以及对比加入其他不同注意力机制的ResNet50模型,在准确率上也有不少提升,且在稳定性和收敛速度方面有显著改善。该模型结合了ResNet50的强大特征提取能力和SE模块的通道注意力机制,为大学生抑郁症的检测提供了一种创新的解决方案。然而,受到数据集规模、单一模态数据、模型可解释性和实际场景验证的限制,本研究的系统仍有待进一步完善和优化。

(责任编辑:张伟东)

(上接第25页)

化节、汉服秀等校园文化活动,展示文化礼仪风采,营造浓厚文化氛围,将课堂所学延展至校园生活各处,不断提升礼仪文化的吸引力、影响力和渗透力。鼓励学生进入社区开展传统礼仪文化普及志愿服务,在社会大舞台上传播礼仪文化,让学生在知行合一中坚定文化自信,成为传统文化与现代服务礼仪的传承者、践行者与传播者。例如,浙江育英职业技术学院参赛第十届长三角航空服务礼仪大赛,获团队一等奖。通过以赛促教、教赛融合,将职业技能竞赛的内容和要求融入课堂教学,以比赛促进学生提升自我技能,同时提供推动礼仪素养提升、弘扬文明新风尚的展示交流平台。

(四)提升教师素养,筑牢文化传承保障

教师作为服务礼仪课程教学的组织者和实施者,其自身的专业能力和文化素养直接影响课程教学质量 and 文化传承效果。首先,教师需要加强对中国传统文化的系统和深入研究。教师只有自身具备深厚的传统文化底蕴,才能更好地将传统文化融入教学之中。其次,教师应积极参加与传统文化相关的培训课程、学术研讨会和交流活动,了解行业最新动态和发展趋势,更新教育理念和教学方法,学习先进的文化传承经验和教学实践案例。再

者,教师还可以利用业余时间参加传统文化体验活动,亲身感受传统文化的魅力和价值,增强文化认同感和情感共鸣。教师还应具备一定的课程开发能力和资源整合能力,根据高职院校学生的特点和专业需求,开发具有传统文化特色的服务礼仪课程教材、课件、教案等教学资源。此外,教师还要以身作则,注重自身的言行举止和礼仪修养,以良好的文化形象和道德风范为学生树立榜样,潜移默化地影响学生对传统文化的认知和态度。

五、结语

本文在文化自信视域下,突破了高职院校礼仪教育重技能轻文化的窠臼,分析了中国传统文化与服务礼仪课程的内在联系与耦合机制,构建了“理念-内容-方法-师资”四位一体的融合路径,为中国传统文化在职业教育领域的创造性转化和创新性发展提供了可复制范式。其价值不仅在于为高职院校服务礼仪课程改革提供了新思路,更提出了产教融合背景下“文化+技能”复合人才培养的新路径,对提升学生综合素质、完善现代职业教育体系、建设社会主义文化强国具有重要意义。

(责任编辑:元小佩)