

摘 要

近年来,人工智能已逐步融入教育领域,为教学评价体系带来了革新。传统教学评价体系主要以教师为评价主体,其中存在诸多问题,包括评价手段的单一性、评价结果的不全面性、反馈的滞后性以及具有主观性。同时,较少关注学生学习过程中的专注度。然而,研究表明学生在课堂上的专注度与其学习效果和理解程度密切相关。鉴于此,本论文深入研究在学生已抬头行为的场景下识别其专注度,致力于将专注度指标引入教学评价体系中,更准确地评估学生的学习状态,提高评价体系的可靠性和有效性。本论文的具体研究工作与主要贡献如下:

(1) 针对当前多数专注度数据集在标注方式上采用人工标注,普遍存在主观性过高和资源消耗较大等问题,本文设计了一种专注度情感诱发的心理实验范式 (Experimental Paradigm)。该范式能够有效诱发出不同的专注度情感状态,而且在标注过程中依据范式实现了更高的客观性和自动化,提升了标注的准确性和效率。并利用此范式构建了一个学生心理专注度视频数据集 (SPEV),涵盖了 47 名参与者的数据。经过数据筛选和处理后,数据集最终包含了 579 个时长为 6 秒的视频,且进行了“专注”和“不专注”两类标签的标注。

(2) 针对现有基于视频的专注度识别模型存在信息丢失、长距离特征捕获能力差、未能利用局部和整体的特征以及未能充分考虑视频中空域和时域信息间关联性等,本文提出了一种基于时序人脸表征的学生专注度识别模型 (TFR-SER)。该模型由人脸表征提取模块、时序多头自注意力模块 (TMSA) 以及时序表征注意力模块 (TRA) 组成。人脸表征提取模块结合掩码自编码器和知识蒸馏的方法进行预训练和微调,从而使其能够提取出具有代表性的空间面部表征。TMSA 模块利用多头自注意力机制对时序卷积网络进行改进,使其能够更专注于时间序列中关键特征,以捕获时间序列数据中局部和全局信息的联系,增强长距离依赖捕获能力。TRA 模块通过引入多头注意力机制,该机制具备动态调整权重的能力,能有效融合空间和时间维度上的信息。使其学习到多层次、多尺度的时空特征表示,提高识别的准确性和鲁棒性。TFR-SER 模型在 SPEV 数据集上达到 89.57% 的准确度,并在 DAiSEE 数据集上达到 61.43% 的准确度。

(3) 基于 TFR-SER 模型,本文设计并实现了一个学生专注度识别系统。该系统集成了用户模块、数据采集、专注度识别、学生考勤以及数据展示等功能模块,实现了对学生专注度的系统化管理、流程化操作和直观展示。它能够实时监测课堂中每位学生的专注度状态,并提供专注度与考勤数据的简单统计。

目 录

摘 要	I
Abstract	III
目 录	V
1 引 言	1
1.1 研究背景和意义	1
1.1.1 研究背景	1
1.1.2 研究意义	2
1.2 国内外研究现状及分析	2
1.2.1 基于几何的专注度识别方法	3
1.2.2 基于外观的专注度识别方法	3
1.2.3 基于行为的专注度识别方法	4
1.3 本文工作内容	5
1.4 本文组织结构	6
2 相关理论与技术介绍	8
2.1 掩码自编码器	8
2.1.1 MAE 编码器	8
2.1.2 MAE 解码器	9
2.2 知识蒸馏算法	10
2.3 时序卷积网络	11
2.3.1 因果卷积	11
2.3.2 膨胀卷积	12
2.3.3 残差模块	13
2.4 多头注意力机制	13
2.4.1 注意力机制	13
2.4.2 多头注意力机制	14
2.5 本章小结	16
3 专注度情感诱发的心理实验范式及数据集构建	17
3.1 心理实验范式中材料设计	17
3.1.1 专注情感诱发视频材料设计	17
3.1.2 不专注情感诱发视频材料设计	18
3.1.3 教学视频中测试问卷材料设计	19
3.2 专注度情感诱发的心理实验范式设计	19
3.3 学生心理专注度视频数据集的构建	21
3.3.1 数据采集实验前期准备	22
3.3.2 数据采集实验的实施	23
3.3.3 数据集标注	24
3.3.4 数据处理	25
3.4 本章小结	26
4 基于时序人脸表征的学生专注度识别模型	27

4.1 TFR-SER 模型的网络架构	27
4.2 人脸表征提取模块	28
4.2.1 模型预训练阶段	29
4.2.2 模型微调阶段	30
4.3 时序多头自注意力模块	31
4.4 时序表征注意力模块	32
4.5 实验设置	33
4.5.1 学生心理专注度视频数据集 (SPEV)	33
4.5.2 DAiSEE 数据集	34
4.5.3 数据集预处理	35
4.5.4 实验环境配置	36
4.5.5 模型超参数设置	36
4.6 实验结果与分析	36
4.6.1 评估指标	36
4.6.2 Backbone 对比实验	37
4.6.3 结合知识蒸馏方法的微调对比实验	38
4.6.4 学生心理专注度视频数据集上的实验结果	39
4.6.5 消融实验	39
4.6.6 DAiSEE 数据集上算法性能对比	40
4.7 本章小结	41
5 基于 TFR-SER 的学生专注度识别系统的设计与实现	42
5.1 系统需求分析	42
5.1.1 系统功能性需求分析	42
5.1.2 系统非功能性需求分析	43
5.2 系统总体设计	43
5.2.1 系统功能模块设计	43
5.2.2 系统架构设计	44
5.2.3 系统核心流程设计	45
5.2.4 数据库设计	46
5.2.5 运行环境	48
5.3 系统详细设计与实现	49
5.3.1 用户模块	49
5.3.2 数据采集模块	51
5.3.3 专注度识别模块	52
5.3.4 学生考勤模块	54
5.3.5 数据展示模块	56
5.4 本章小结	58
6 总结与展望	59
6.1 总结	59
6.2 展望	60
参考文献	61
致 谢	67
在读期间公开发表论文（著）及科研情况	69

1 引言

1.1 研究背景和意义

1.1.1 研究背景

近年来,科技迅速发展和人工智能持续进步使得信息技术与各行各业,特别是教育领域的融合日益显著。教育行业作为社会发展的关键组成部分,正在积极借助科技和人工智能推动进步。政府与教育机构为提升教育质量和改革,制定了多项政策方案。2023年,中共中央教育部办公厅在发布了《基础教育课程教学改革深化行动方案》^①,着重课程改革和机制创新,旨在加速教学改进,提升教育质量,确保学生均衡发展。教学评价改革,特别是过程评价的引入,是教育改革的核心方向。

传统教学评价主要依据学生的最终成果,常忽略学生学习过程。在课堂教学中,教师主要通过课堂观察和提问等方式来了解学生的学习状况。然而,班级规模导致教师难以全面观察每位学生的学习状况,引发信息反馈的滞后性与片面性。因此,需采用更全面、客观的评价方式以克服传统方法的局限。

科技进步使得人工智能能够与教育融合,刷新教学评价体系,增强其全面性和科学性。课堂教学是提升教学质量和实效的关键。研究表明学生在课堂中专注度水平与学习成效密切相关,高专注学生往往学习效率更佳^[1,2]。因此,学生专注度的重要性不言而喻,它反映了注意力和参与度^[3]。因此,将学生专注度纳入评价体系能够丰富创新教学评价手段。实现以评促教、以评促学,以此来促进学生全面发展。

以往研究^[4-6]多数依赖于课堂抬头率这一简单指标对学生专注度进行量化。然而,仅凭学生的抬头行为并不能准确地反映其是否专注于教学内容,学生的思绪可能已经偏离了教学活动。因此,单纯依靠抬头率这一指标来判断学生的专注状态并不全面,需要在学生已抬头的行为下进行更细致的专注度判断。此外,以往数据集标注的方式存在主观性和资源消耗等问题^[7-9]。为了更科学和更准确地量化学生的专注程度,本文设计了专注度情感诱发的心理实验范式并基于此构建了一个学生心理专注度视频数据集,以及提出了一种基于时序人脸表征的学生专注度识别模型,并在构建的数据集上验证了该模型的有效性,最后基于此模型实现了一个学生专注度识别系统。

^① 教育部办公厅关于印发《基础教育课程教学改革深化行动方案》的通知

1.1.2 研究意义

传统教学评价体系主要是以教师为评价主体，其中存在诸多问题，包括评价手段的简单性、评价结果的不全面性、反馈的滞后性以及具有一定的主观性。因此，本文所研究的学生专注度识别技术及其系统的设计与实现，旨在监测学生的在课堂上的学习状态，使专注度评估更加科学、全面和可靠。该技术的应用将为教师、学生和教学评价体系提供诸多益处。具体如下：

对教师而言，通过实时监测学生的专注度来及时掌握学生的学习状况，并据此调整教学策略。同时，通过班级中所有学生专注度的变化趋势有助于发现教学中的潜在问题。如若学生普遍专注度低下，可能反映教学方法的不足。因此，教师可以持续改进教学方法，以提升教学效果和学生参与度。这有助于提高教学效果，使教学更加个性化和针对性

对学生而言，能够根据实时反馈调整学习态度和注意力，培养良好的学习习惯。若学生在课堂上长时间专注度显著低下，则可推测其未能充分集中于学习。这助于学生识别其早学习过程中的知识缺失。针对知识短板，可以自我调整学习策略，提高学习效率和成果。此外，通过与师生交流分享，学生能相互交流观点和反思，从而促进共同进步。

对于教学评价体系，量化学生专注度的方法提供了一种创新的途径，从而为教学过程评估带来数据支持。以往学生学习的评价依靠教师的观察、作业完成和考试成绩，易于受到教师主观因素（例如情绪和偏好）的干扰，从而影响评价的客观性和准确性。故此，利用计算机技术来收集和分析学生的专注度数据，可以为教学评价提供一个坚实且客观的数据基础，确保评估过程更具科学性和客观性。

1.2 国内外研究现状及分析

教育学家 Tyler 在 20 世纪 30 年代就已对学生的课堂学习进行研究，引入了“time on task”这一概念。指出学生在课堂上投入的专注时间越长，获取的知识量也越大。Smallwood 等人^[10]也阐述了专注力的缺失显著影响学生的学习效率。因此，许多研究人员致力于提高学生专注度，并对识别方法表示出浓厚兴趣^[11]，开启了一系列关于学生专注度的实证研究^[12]。Kevin 等人^[13]描述专注度为学生在课堂上的注意力集中能力，特别是对教学和学习材料的关注，视其为专注度的关键指标。Catherine 等人^[14]认为，专注度可通过学生的姿态、任务完成速度和抑制非学习行为来衡量。O'Brien 等人^[15]提出通过学生填写问卷来评估他们的专注度水平。虽然关于学生专注度的定义在以上各项研究中有所不同，但共同之处在

于都未对专注度给出具体化描述。随着计算机视觉技术的进步,目前可以用更客观和可靠的方式判断学生专注度。

国内外现有大量研究文献已探讨学生专注度,其中 Blasco 等人^[16]于 2013 年提出的定义,包括情绪专注、行为专注和认知专注三大维度,获得普遍认可。其中认知专注主要适用于课后评价,不适合实时识别学生的专注度^[17]。情绪专注主要通过面部表情反映学习者在学习过程中的情绪变化,这与专注度水平直接相关^[18-20]。专注度识别方法根据面部信息主要分两类:基于几何的识别方法、基于外观的识别方法。行为专注可将专注度识别方法归纳为:基于行为的识别方法。

1.2.1 基于几何的专注度识别方法

基于几何的专注度识别方法主要是通过通过分析人脸的各个部位(眼睛、嘴巴、鼻子等)提取局部特征来识别学生的专注度。Ekman 和 Friesen 等人^[19]设计了面部动作编码系统(FACS)。FACS 分析面部表情,将面部肌肉运动定义为表情动作单元(AU)。近年来,FACS 广泛应用于面部表情识别及学生专注度识别中。McDaniel 等人^[21]通过收集学习者的面部表情,利用 FACS 将面部表情进行编码,得出学习者的情绪状态影响学习效率的结论。Deng 等人^[22]提出了基于机器学习的方法来识别学生学习过程中的专注度。此方法使用 Gabor 小波分析提取眼睛状态特征,并用支持向量机(SVM)分类。He 等人^[23]提出一种结合局部几何与全局特征的学习表情自动识别方法,包括核主成分分析降维的局部二值模式和 CNN 提取的深度特征来进行识别。

结合表情动作单元(AU)和传感器数据的专注度识别方法得到了广泛应用。Saneiro 等人^[24]通过 Kinect 相机分析人脸关键点和头部姿势等,包括下巴降低和眉毛动作。采用机器学习技术分类五级专注度:兴奋、放松、决心、感兴趣、专注。Psaltis 等人^[25]结合 AU、Kinect 传感器数据和游戏事件来识别和量化学习者的专注度。该研究使用人工神经网络(ANN)进行分类,取得了良好的专注度分类效果。Monkaresi 等人^[26]则结合了几何、纹理和生理特征(如心率)来识别学生的专注度水平。尽管基于 AU 结合传感器的专注度识别方法有理论支持,但其在实际教室环境中的应用由于需要佩戴设备来采集数据而受到限制,阻碍了其广泛推广。相较而言,本文提出的专注度识别方法规避了这一障碍,仅通过教室内的监控摄像头进行数据采集,无需佩戴任何额外设备。

1.2.2 基于外观的专注度识别方法

基于外观的专注度识别方法中利用整个面部区域提取的特征来识别学生的专注度。早期,局部二值模式(LBP)和定向梯度直方图(HOG)是专注度识别中常用的特征提取技术。Monkaresi 等人^[26]结合 LBP 提取的面部特征和心率来评

估学习者在教育活动中的专注度，该研究发现面部表情的识别准确性优于心率。Happy 等人^[27]使用从整体面部外观提取的 LBP 特征来识别基本情绪，并结合眼睛、头部姿势等信息来识别学生专注度。Kamath 等人^[28]开发了一种实例加权多核学习 SVM 模型用于在线学习中识别学生专注度。Prabin 等人^[29]则通过 Microsoft Azure 情感 API 评估学生面部情绪得分并分配权重，以此评估学生专注度。后来计算机算力的增强使得深度学习技术迅速发展，学生专注度识别领域也引入了深度学习技术。

深度学习技术已成为专注度识别研究的核心方向。Gupta 等人^[7]使用卷积神经网络 (CNN) 模型根据人脸面部特征将学生专注度分为四类。研究^[30,31]则是通过多种改进后的 CNN 模型来识别在课堂中的学生专注度，且比较了不同 CNN 模型的实验结果。研究^[32,33]都旨在提取人脸面部特征并建立一个系统流水线以识别学生专注度。Ling 等人^[34]应用 YOLO 模型和基于自注意力的 ViT 模型^[35]在视频中检测并识别面部表情，以此评估学生专注度。Mandia 等人^[36]通过 ViT 模型以及多头注意力机制端到端学习面部图像特征以自动评估学生专注度。Abedi 等人^[37]利用 ResNet 和 TCN 网络的组合对专注度进行识别，通过 ResNet 提取面部特征并采用时序卷积网络 (TCN) 进行分类。章文松^[38]则提出了一种改进的 ResNet 模型和自注意力机制进行学生专注度识别的方法。Mehta 等人^[39]提出一种三维 DenseNet 自注意力神经网络，有效识别并强调重要的时空特征，用于自动识别在线学习中的学生专注度。

现有的学生专注度识别研究中，常见方法主要是依赖于面部图像的单一特征，或是对视频数据中的空间和时间信息分别处理。前者忽略了数据在时空维度方面的信息，而后者可能会忽视时域和空域两者之间的内在联系，导致关键信息遗漏。分离处理限制了模型捕获视觉相关性的能力，影响了其整体性能。故此，本文提出基于时序人脸表征的专注度识别模型，能够有效融合空间与时间信息，使模型学习到多层次、多尺度的特征表示，从而强化对专注度关键特征的识别。

1.2.3 基于行为的专注度识别方法

基于行为的专注度识别方法主要依据人体的头部姿态、手部姿势等，利用身体语言作为非语言交流手段，这也是情绪表达的关键要素。越来越多的研究将姿态估计技术应用用于教育，通过计算机技术提取学习者姿态特征，以识别学生专注度^[40]。Zhong 等人^[41]开发了一个分析大学生课堂行为的深度学习模型和系统，重点在于学习和专注度的评估。Burnik 等人^[42]利用 Kinect 2 采集骨骼关键点，分析头部姿态和肢体方向以评估学生的专注度。Grafsgaard 等人^[43]探讨了计算机辅助教学中学习者手势与姿势的影响，尤其是其与对话及导师互动的关系，用于识别学习者专注度。Zaletelj 等人^[44]使用 Kinect One 传感器自动评估课堂学生

专注度，构建了包含凝视点和身体姿势的特征集，之后通过机器学习分类器识别学生专注度的变化。

研究表明，通过观察头部姿态，可以有效地判断一个人的专注点^[45]。因此，许多专注度识别的研究开始将头部姿态作为一个关键参数加以应用。Xiaoyang 等人^[46]提取学生的眼睛特征、AU 特征、身体和头部姿势多种特征，结合神经图灵机思想的框架学习并融合权重，用于识别学生在线课程中的参与度。Ma 等人^[47]则结合生物启发特征与局部二值模式评估头部姿态，识别学生课堂专注度。Chang 等^[48]提出三模块系统：头部姿态检测、疲劳评分（眼睛/嘴巴闭合）和情感评分（面部表情），综合评估学生专注度。研究^[49]中采用 Retinaface 模型检测人脸和 ViT 模型提取学生头部姿态参数，从课堂视频和音频中提取数据，综合头部姿态与班级动态识别学生专注度。余绕东^[50]开发智能检测系统收集脑波及头部姿态数据评估学生状态，尽管高效但成本限制了其普及。

上述多数相关研究依赖于特定的实验工具进行，限制了其在实际教学环境中的应用，且在实际课堂环境中，学生的身体和头部的移动幅度通常较小，难于准确判断学生专注度。本文提出了一种创新方案，通过时序人脸表征精细量化学生的专注度。该方法是捕捉面部的变化，这些变化通常反映了学生的专注状态，无需依靠明显的身体或头部动作。

1.3 本文工作内容

学生在课堂上的专注度与其学习效果和理解程度密切相关，因此对学生专注度的评估尤为重要。准确评估学生专注度能为教师提供客观的学生学习状态反馈，帮助教师更好了解学生在课堂中的表现，并及时调整教学策略。以往研究通过课堂抬头率进行专注度识别，但仅凭抬头率这一简单指标并不能准确地反映学生专注状态。且现有基于视频的专注度识别模型存在信息丢失、长距离特征捕获能力差、没有采用局部和整体特征以及未能充分考虑视频中空域和时域信息间关联性等。另外，当前多数专注度数据集在标注方式上普遍存在主观性过高和资源消耗较大等问题以及学生专注度领域的公开数据集相对缺乏的问题。

针对以上问题，本文设计了专注度情感诱发的心理范式并利用该范式构建了一个学生心理专注度视频数据集，以及提出了一种基于时序人脸表征的学生专注度识别模型，并在构建的数据集上验证了该模型的有效性，在该模型的基础上实现了一个学生专注度识别系统。本文的主要工作内容如下：

(1) 设计了一种专注度情感诱发的心理实验范式。该范式能够有效诱发出不同的专注度情感状态，而且在标注过程中实现了更高的客观性和自动化，提升了标注的准确性和效率。并基于此范式构建了一个学生心理专注度视频数据集

(SPEV), 涵盖了 47 名参与者的数据。经过数据筛选和处理后, 数据集最终包含了 579 个时长为 6 秒的视频片段。此外, 本文对所收集的数据集进行了“专注”和“不专注”两类标签的标注。

(2) 提出了一种基于时序人脸表征的学生专注度识别模型 (TFR-SER)。该模型包含三个模块, 即人脸表征提取模块、时序多头自注意力模块 (TMSA) 以及时序表征注意力模块 (TRA)。人脸表征提取模块采用掩码自编码器和知识蒸馏的方法进行预训练和微调, 从而使其能够提取出具有代表性的空间面部表征。TMSA 模块利用多头自注意力机制对时序卷积网络进行改进, 使其能够更专注于时间序列中关键特征, 以捕获时间序列数据中局部和全局信息的联系, 增强长距离依赖捕获能力。TRA 模块通过引入多头注意力机制, 该机制具备动态调整权重的能力, 能有效融合空间和时间维度上的信息。使其学习到多层次、多尺度的时空特征表示, 从而提高分类的准确性和鲁棒性。并在学生专注度数据集上进行实验, 验证了此模型的有效。

(3) 设计并实现了一套基于 TFR-SER 模型的学生专注度识别系统。该系统集成了用户、数据采集、专注度识别、学生考勤以及数据展示等功能模块。实现了对学生专注度的系统化管理、流程化操作和直观展示。其中提出了小教室场景下学生专注度识别的方案。该方案整合了人脸检测、人脸对齐、人脸识别、TFR-SER 等算法。构建了从采集的课堂视频中进行学生专注度识别的算法 workflow。课堂视频传入后, 首先通过人脸检测算法检测出人脸区域, 然后进行人脸对齐操作, 对齐后的人脸图像通过人脸识别得到人脸身份信息。同时将对齐后的视频帧序列输入到 TFR-SER 模型进行专注度识别, 最后可输出视频帧中人脸的身份信息和专注度水平, 使该系统能够实时监测课堂中每位学生的专注度状态。

1.4 本文组织结构

本文一共包含了六个章节, 具体每个章节组织结构如下:

第一章为引言。此章首先概述了在当今社会发展背景下, 学生专注度识别技术的研究背景和意义。其次, 介绍了专注度识别领域的国内外研究发展现状。最后阐述了本文主要研究内容。

第二章为相关理论与技术介绍。此章对本研究所涉及的相关理论及技术进行了介绍, 包括掩码自编码器、知识蒸馏、时序卷积网络和多头注意力机制。

第三章为专注度情感诱发的心理实验范式及数据集构建。此章首先介绍了心理实验范式中材料的设计, 之后介绍了专注度诱发的心理实验范式的设计, 最后, 详细阐述利用设计的心理实验范式构建的学生心理专注度视频数据集。

第四章为基于时序人脸表征的学生专注度识别模型。此章依次介绍了模型的总体网络架构、模型中各模块的详细设计、实验设置、模型的实验结果和分析。起初，通过介绍模型的总体网络架构，提出了学生专注度识别模型 (TFR-SER)，该模型基于时序人脸表征。随后，逐一详细介绍了模型架构中的三个关键模块，并描述了模型训练的过程。本章最后通过实验结果和深入分析，验证了该模型的有效性。

第五章为基于 TFR-SER 的学生专注度识别系统的设计与实现。此章阐述了在线下教室环境下学生专注度识别系统的设计与实现流程。主要从系统的需求分析、总体设计、系统功能模块详细设计与实现等多个层面进行了系统性阐释。

第六章为总结与展望。此章主要对总结本文的主要内容，并对专注度识别方法以及学生专注度识别系统尚存的不足之处进行剖析与反思，同时对未来的研究做出展望。

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/325343032230012012>