

基于计算机视觉的课堂专注度检测研究

邓子豪, 梁 艳

(华南师范大学软件学院, 广东 广州 510006)

摘要: 为了检测课堂学习时班级整体专注程度, 设计了一套基于计算机视觉的课堂专注度检测系统。该系统根据面部关键点位置计算头部偏转角度, 进而判断视线是否位于专注区域, 采用 ViViT 模型对学习者的专注度进行四分类, 最后综合人脸检测信息、视线信息、专注度分类信息计算班级整体专注度。实验结果证明, 该系统能根据监控视频快速、有效地评估班级在某时刻的整体专注度, 为教师提供客观、准确的专注度数据, 有助于教师有针对性地及时调整教学方式, 提高教学质量。

关键词: 线下教学; 专注度检测; 计算机视觉; 人脸检测; ViViT

DOI: 10.11907/rjdk.231759

中图分类号: TP391.41

文献标识码: A

开放科学(资源服务)标识码(OSID):

文章编号: 1672-7800(2024)008-0254-07



Engagement Detection for Offline Education Based on Computer Vision

DENG Zihao, LIANG Yan

(School of Software, South China Normal University, Guangzhou 510006, China)

Abstract: In order to detect the overall engagement level of the class during classroom learning, a computer vision-based classroom engagement detection system is designed. The head deflection angle is calculated based on the location of key points on the face, which is used to determine whether the line of sight is located in the engagement area. The ViViT model is adopted to perform four classifications of learners' engagement. Face detection information, line of sight information, and engagement classification information are finally combined to calculate the overall engagement of the class. The experimental results prove that the system can quickly and effectively evaluate the overall engagement of the class at a certain moment based on the monitoring video. It can provide teachers with objective and accurate engagement data, which can help teachers adjust teaching methods in a targeted and timely manner to improve teaching quality.

Key Words: offline teaching; engagement detection; computer vision; face detection; ViViT

0 引言

学生课堂专注程度与学习效果之间存在正相关关系^[1]。在传统课堂环境中, 教师可以直接通过观察学生的表情、目视方向等信息, 判断学生专注程度。然而, 教师精力有限, 无法同时兼顾课堂教学与学生专注度的全面把控。

近年来, 计算机视觉技术的飞速发展教育现代化提供了一种新的途径。基于计算机视觉的课堂专注度检测系统可以实时监测学生的专注程度, 帮助教师实时了解学生在课堂上的学习情况, 并有针对性地调整课堂教学内容

和方式。此外, 学生的专注度也可以作为一种教学评价指标, 用于评估教师的教学效果^[2]。

基于计算机视觉的方法利用摄像头获取学习者的图像或视频, 通过分析面部表情、头部姿态等因素进行专注度判断^[3]。该类方法可以非接触式地采集数据, 不会对学生的学习造成影响, 而且成本较低, 易于推广, 因此成为近年来专注度检测研究的主流方法。Yang 等^[4]提出一个基于多实例机制和 LSTM 的多模态回归模型, 预测学习者的专注程度。Liao 等^[5]提出一种深度时空网络 DFSTN 提取视频中的空间特征和时序特征, 更准确地感知学习者的专注程度。Mehta 等^[6]提出 DenseAttNet 模型以实现学生专注度检测, 在 DAiSEE 数据集上的四分类实验表明, 其准确率

收稿日期: 2023-08-16

基金项目: 国家科技创新 2030——“脑科学与类脑智能技术”重点项目(2022ZD0208900); 国家自然科学基金项目(62076103)

作者简介: 邓子豪(2001-), 男, 华南师范大学软件学院学生, 研究方向为计算机视觉; 梁艳(1979-), 女, 博士, 华南师范大学软件学院讲师、硕士生导师, 研究方向为图像处理与模式识别。本文通讯作者: 梁艳。

达 63.59%。但这些方法主要针对在线教育场景,处理的图像中只有一位学习者的面部,图像清晰,背景比较简单。对于线下课堂环境,采集到的图像通常存在较多的人脸,而且每张人脸的尺寸较小,有些面部存在遮挡现象。

因此,针对线下课堂的特点,部分学者展开了相关研究。早期学者们主要采用传统的机器学习方法,如:Zaletelj 等^[7]利用 Kinect 传感器获取每个学生的骨架节点和人脸数据,利用学生的面部和身体特征进行专注度分类。然而,该方法忽略了线下课堂环境中学生身体容易被遮挡的问题。因此,有部分学者仅从头部提取相关信息进行判断^[8]。Thomas 等^[9]根据头部姿态、眼部凝视、动作单元等数据构建特征向量,用支持向量机分类,最高获得 90% 的准确率。Lam 等^[10]提出了一种数据融合算法,嵌入脸部、眼部和位置信息以构成一个行为实体,同一个人的所有行为实体组合成一个序列,所有人的序列构成一个集合,通过 argmax 分类实现学生整体专注度计算。

近年来,学者们尝试采用深度学习方法对线下学习的学生进行专注度检测。唐康等^[11]采用优化的人脸检测级联 CNN 模型对包含多人的课堂学习图片进行人脸检测,然

后根据学习者的抬头率量化课堂专注度和时间分布。但该方法的评价指标过于单一,没有考虑人脸特征、眼睛注视点等信息与专注度的关系。郭晓旭等^[12]利用 CNN 模型进行人脸检测,通过分析学习者动作单元、效价、唤醒空间的关联以计算课堂专注度,对整节课以及各时间段的学生整体专注度进行评分。利用自采的视频对系统进行测试,准确率超过 91%。但该系统未在更多的视频数据中进行测试。

针对现有研究的局限性,本文开发了一个基于计算机视觉的课堂注意力检测系统。考虑到线下教学环境,学生的躯干容易被桌椅或其他学生遮挡^[13]。因此,仅利用学生的面部信息进行判断,并计算出全体学生的整体专注度,为教师课堂授课提供理论依据和数据支撑。

1 系统设计

本文设计了一个基于计算机视觉的课堂专注度检测系统,该系统主要包括人脸检测模块、视线计算模块、专注度检测模块和专注度评分模块,整体架构如图 1 所示。

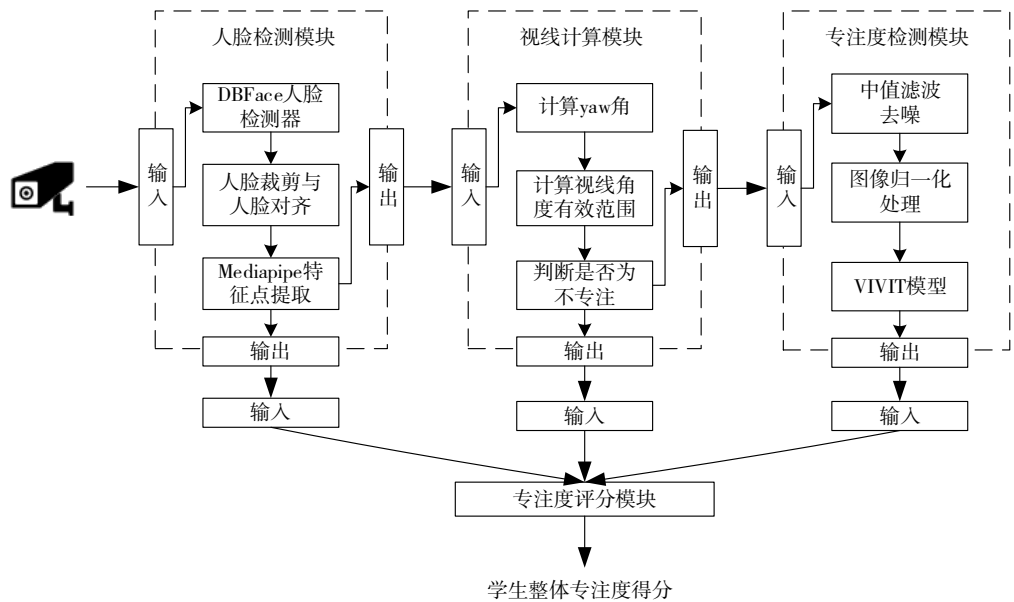


Fig. 1 Architecture diagram of the proposed system

图1 本文系统架构

人脸检测模块的目的是检测出人脸区域及其关键点的位置;视线计算模块用于计算学生的有效视线范围,并进行标注;专注度检测模块则主要利用 ViViT 模型进行注意力分类^[14];在专注度评分模块中,结合前 3 个模块中获得的数据,综合计算出学生的整体专注度得分。该系统通过引入视线计算模块,对视线信息进行建模,从而提高 ViViT 模型对专注程度的判断能力。

1.1 人脸检测模块

由于教室内拍摄的视频通常包含多位学生及背景信息,因此,视频首先需要进行人脸检测模块处理,主要包

括:人脸检测、人脸裁剪、人脸中心点记录、人脸关键点检测。

考虑到不同的摄像头采集的视频帧率有所不同,本文采取等时间间隔选取关键帧的操作,统一每分钟等间隔抽取 16 帧照片作为当前时刻的数据输入,并对图像进行灰度化处理,降低后续人脸运算复杂度,然后使用 DBFace 模型对图像进行人脸检测^[15]。

课堂环境中,学习者的面部可能会因为各种原因被遮挡,导致不同的时间点检测到的总人数发生变化。为了解决该问题,本文提出人脸中心点匹配法(人脸中心点为人

脸裁剪时矩形框的中心点),通过比较不同时刻的人脸中心点位置变化而判断某位学习者在当前时刻是否被正确检测到。具体操作如下:在完成当前时刻的人脸裁剪之后,将模块保存的前一时刻人脸中心点数组与当前人脸中心点数组进行比较,判断两中心点是否相等,判断公式如式(1)所示。

$$\begin{cases} \sqrt{(x_i^{pre} - x_j^{now})^2 + (y_i^{pre} - y_j^{now})^2} \leq eps, \text{相等} \\ \sqrt{(x_i^{pre} - x_j^{now})^2 + (y_i^{pre} - y_j^{now})^2} > eps, \text{不相等} \end{cases} \quad (1)$$

其中, x_i^{pre} 和 y_i^{pre} 分别为前一时刻第 i 个人脸中心点 P_i^{pre}

的 x 坐标和 y 坐标, x_j^{now} 和 y_j^{now} 分别为当前时刻第 j 个人脸中心点 P_j^{now} 的 x 坐标和 y 坐标, eps 为绝对误差。

根据 P_i^{pre} 与 P_j^{now} 的关系,采取表1所示策略更新学习者的位置。经过人脸中心点比对,即可确定当前时刻检测到的每个学生的人脸位置。

随后,使用 Mediapipe 库进行人脸关键点检测,确定左眼角、右眼角、鼻尖、左嘴角、右嘴角和下巴共6个关键点的位置。对于不能检测到6个关键点的图像,由于人脸信息不完整而无法判定其状态,因此将这类图像删除,不再传递到下一个模块处理,同时标记为“无法检测到人脸”。

Table 1 Update strategy for face center point arrays

表1 人脸中心点数组更新策略

关系	策略
P_i^{pre} 与 P_j^{now} 相等	用 P_j^{now} 更新 P_i^{pre}
P_i^{pre} 与所有 P_j^{now} 不相等,即前一时刻存在的学习者当前时刻不存在	将 P_i^{pre} 从数组删除并将其标记为无法检测到人脸
P_j^{now} 与所有 P_i^{pre} 不相等,即当前时刻存在的学习者前一时刻不存在	将 P_j^{now} 新增到数组中

1.2 视线计算模块

在授课环境中,学习者大部分时间会注视黑板,因此可以根据视线所在区域对学习者的专注情况作出初步判断,从而减少送入专注度检测模块的图像数量,降低系统运行时长。

本文利用 OpenCV 库封装的 solvePNP 函数、Rodrigues 函数、norm 函数实现视线计算。solvePNP 函数用于求解给

$$R = \begin{bmatrix} \cos\theta\cos\phi & \sin\psi\sin\theta\cos\phi - \cos\psi\sin\phi & \cos\psi\sin\theta\cos\phi + \sin\psi\sin\phi \\ \cos\theta\sin\phi & \sin\psi\sin\theta\sin\phi + \cos\psi\cos\phi & \cos\psi\sin\theta\sin\phi - \sin\psi\cos\phi \\ -\sin\theta & \sin\psi\cos\theta & \cos\psi\cos\theta \end{bmatrix} \quad (2)$$

根据旋转矩阵 R ,可以计算出欧拉角^[16]中的 θ 角,如式(3)所示,用于判断学生的视线。

$$\theta = -\arcsin(R_{(3,1)}) \quad (3)$$

其中, ψ 、 θ 、 ϕ 分别表示头部绕相机设定的 x 轴进行俯仰旋转(Pitch Rotation)、 y 轴进行左右旋转(Yaw Rotation)、 z 轴进行平面旋转(Roll Rotation)的角度,如图2所示。

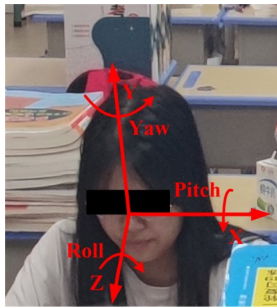


Fig. 2 Schematic diagram of Euler angle

图2 欧拉角示意图

为了计算出学生视线的有效范围,本文对线下教室的总体环境进行建模,如图3所示。其中,摄像头位于黑板正上方,处于教室、黑板的中轴线上。利用人脸检测模块提供的人脸中心点数据,可以区分学生所在的行与列。当

定对象相对于相机的姿态估计,它返回旋转向量和平移向量。solvePNP 函数的一个重要参数是平面关键点数组,本文选取左眼角、右眼角、鼻尖、左嘴角、右嘴角和下巴6个人脸关键点组成关键点数组。获得旋转向量后,通过 Rodrigues 函数与 norm 函数可以实现从旋转向量到旋转矩阵 R 的计算。旋转矩阵 R 的展开式如式(2)所示。

人脸中心点数据中坐标 x 差值小于 eps_x (x 轴绝对误差)则认为其属于同一列,同理坐标 y 差值小于 eps_y (y 轴绝对误差)则认为其所属同一行。

根据学习者所处位置,位于第 i 行第 j 列的学习者 yaw 角有效范围 $[yaw_{min}, yaw_{max}]$ 的计算公式如式(4)、式(5)

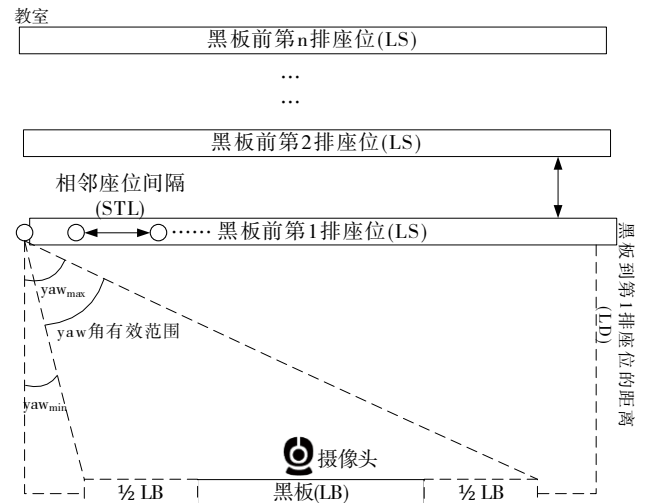


Fig. 3 Schematic of the effective range of classroom and yaw angle

图3 教室与 yaw 角有效范围示意图

所示。

$$yaw_{min} = \frac{\frac{LS}{2} - LB - i \times STL}{\sqrt{\left(\frac{LS}{2} - LB - i \times STL\right)^2 + (LD + j \times SIL)^2}} \quad (4)$$

$$yaw_{max} = \frac{\frac{LS}{2} + LB - i \times STL}{\sqrt{\left(\frac{LS}{2} + LB - i \times STL\right)^2 + (LD + j \times SIL)^2}} \quad (5)$$

其中,LS表示教室横排座位长度,LB为黑板长度,LD为黑板到第一排座位的距离,SIL为每排座位之间的间隔,STL为相邻座位之间的间隔, yaw_{min} 和 yaw_{max} 分别表示yaw角的最小值和最大值。

根据学生的旋转角度 θ 与yaw角有效范围之间的关系,可以对学习者专注度状态进行初步判断:

$$\begin{cases} yaw_{(i,j)}^{min} \leq \theta_{(i,j)} \leq yaw_{(i,j)}^{max}, \text{需要进一步判断} \\ \theta_{(i,j)} < yaw_{(i,j)}^{min}, \text{非常不专注} \\ \theta_{(i,j)} > yaw_{(i,j)}^{max}, \text{非常不专注} \end{cases} \quad (6)$$

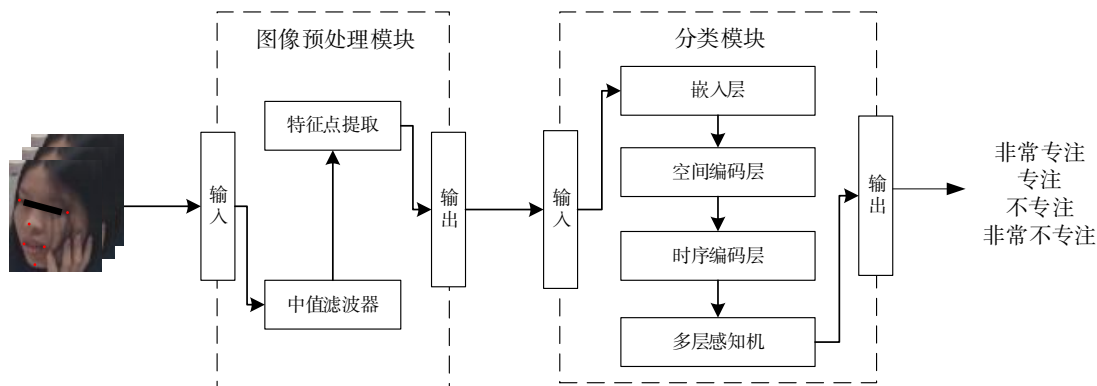


Fig. 4 Architecture of the engagement detection algorithm

图4 专注度检测算法架构

1.4 专注度评分模块

专注度评分模块的目标是量化课室内所有学生的整体专注度,目前学术界未有公认的量化标准。受文献[12]的学习状态评分标准启发,本文综合人脸检测信息、视线信息、专注度分类信息计算班级整体专注度。该模块的输入包括3部分:①在人脸检测模块中标志为无法检测到人脸的学生信息;②在视线检测模块中标志为非常不专注的学生信息;③在专注度检测模块中确定的每位学习者各专注度类型的分类概率。

首先计算每位学习者的专注度分数:

$$F_{score} = -2 \times p_1 - 1 \times p_2 + 0 \times p_3 + 1 \times p_4 + 2 \times p_5 \quad (7)$$

其中, p_1, p_2, p_3, p_4, p_5 分别表示单位时间内学生状态为非常专注、专注、无法检测到人脸、不专注、非常不专注的概率。对于人脸检测模块中标志为无法检测到人脸的学生, p_3 为1,其余状态的概率为0;对于视线检测模块中标志为非常不专注的学生, p_1 为1,其余状态概率为0;对于来自

当学生视线位于yaw角有效范围内时,将人脸图像输入到专注度检测模块作进一步判断。对于判断为非常不专注状态的学习者,统计其数量并输入到专注度评分模块。

1.3 专注度检测模块

专注度检测模块包括图像预处理和专注度分类两个子模块,如图4所示。在图像预处理子模块中,首先对图像进行中值滤波处理,去除噪声干扰,随后对图像进行大小归一化处理,将图像统一为 224×224 的标准。

在专注度分类子模块中,本文搭建了一个ViViT^[14]神经网络模型,该模型分为4层。第一层为嵌入层,对每帧输入的图像进行切片,并对每一帧图像切片获得的序列进行拼接得到视频V(连续多帧图像)的映射序列 z ;第二、三层为空间编码层和时序编码层,分别用于提取序列 z 的空间特征和时序特征;最后一层为多层感知机,根据空间和时序特征得到最终专注度分类结果。

在该模块中,将学生的专注度分类为非常专注、专注、不专注、非常不专注4种类型,模块的输出为各类型的分类概率。

专注度检测模块的输入, p_3 设为0。 F_{score} 取值范围为 $[W_{min}, W_{max}]$,在该系统中, W_{min} 和 W_{max} 分别取值为-2和2。

随后计算当前时刻班级整体专注度分数。假设有 n 位学习者,根据式(7)计算出每位学习者的专注度数值 F_{score}^i ($i = 1, 2, 3 \dots n$),对所有学习者专注度数值求平均,公式如式(8)所示。

$$F_{avg} = \frac{\sum_{i=1}^n F_{score}^i}{n} \quad (8)$$

将均值 F_{avg} 归一化到0~100之间的数值,即可得到当前时刻班级整体专注度分数如式(9)所示。

$$Score = \frac{F_{avg} - W_{min}}{W_{max} - W_{min}} \times 100 \quad (9)$$

Score值越高,说明班级整体专注程度越高。

2 系统测试

2.1 实现细节

系统采用 Python 语言在 Windows 10 平台上进行构建,运行环境为 Python 3.9.12。其中专注度检测模型基于 Pytorch 框架进行搭建,使用 DAiSEE 数据集进行训练^[17]。

DAiSEE 数据集包含 112 位学生的 9 068 个 10 s 长的视频片段,采集的频率为 30 帧/s,分辨率为 1 920×1 080。根据专注程度的不同,数据被划分为 4 种级别:非常不专注、不专注、专注和非常专注,如图 5 所示。

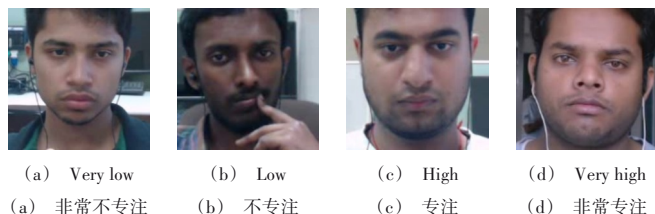


Fig. 5 Examples of four different states of engagement in the DAiSEE dataset

图 5 DAiSEE 数据集 4 种不同专注度状态的例子

由于 DAiSEE 数据集样本数据较少,直接使用原始数据集进行训练会导致模型泛化能力差且容易过拟合。因此,本文首先采用数据增强方法对数据集进行扩充,即对每个片段所截取出的图像分别采用平移、旋转、均衡化 3 种方法生成新的图像。然后使用扩充后的数据集进行模型预训练,训练集:测试集=4:1。优化算法选用随机梯度下降算法,初始学习率为 0.001,动量为 0.9。训练总次数为 100 次,每 25 个 epoch 学习率衰减 80%。

考虑到 DAiSEE 数据集存在样本标签不均衡的问题,本文使用 Focal Loss 作为损失函数,引入一个可调节的平衡参数,对易分类样本(分类概率高的样本)施加一个较小的权重,而对于困难样本(分类概率低的样本)则施加较大的权重,使模型更加关注困难样本的学习,从而提高整体分类性能^[18]。

模型中每个 Transformer 编码器^[19]的深度、头数和单个 patch 的尺寸分别是 4、3、192。训练损失曲线如图 6 所示,曲线总体呈现下降趋势。在前 20 轮中,损失值从 0.323 急剧下降到 0.129;在第 20~90 轮中,损失值下降速度放缓;在第 90~100 轮中损失值在 0.072 上下波动并收敛。在训练完成后模型准确率达到 65.5%。

2.2 系统性能测试

2.2.1 在 DAiSEE 数据集上的测试

本文采用的模型既能获取群体的专注度评分,也能计算个体专注程度。为了验证模型性能,本文采用回归和分类实验分别检测模型的群体检测性能和个体检测性能。在群体检测中,以 10 个个体数据(每个视频只包含一位学习者)作为一个输入,离散的分类输出经过式(7)转化为数

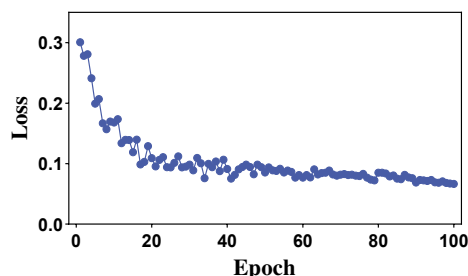


Fig. 6 Model training loss curve convergence graph

图 6 模型训练损失曲线收敛图

值。实验比较了本文模型与传统的 ViViT 模型在 DAiSEE 数据集的回归和分类测试结果,分别采用均方误差(Mean-Square Error, MSE)、决定系数 R^2 及准确率作为评估指标^[20]。实验结果如表 2、表 3 所示。

Table 2 Comparison of model assessment indicator values

表 2 模型评估指标值比较

模型	MSE	R^2
ViViT	0.766	0.387
本文模型	0.411	0.671

Table 3 Comparison of accuracy of each engagement category

表 3 各专注度类别准确率比较

模型	非常不专注	不专注	专注	非常专注	平均值
ViViT	86.7	8.8	56.4	100	62.9
本文模型	85.7	10.4	78.9	96.3	67.8

由表 2 可以看出,本文使用的模型获得 0.411 的 MSE,小于传统的 ViViT 模型,表明本文模型在预测值与实际观测值之间的差异更小。本文模型的 R^2 值比 ViViT 模型的更大,证明本文模型在捕捉目标变量的变异性上优于 ViViT 模型。

表 3 展示了本文模型和 ViViT 模型在分类任务下的实验结果。本文模型在非常不专注、非常专注两类别的准确率上略低于传统 ViViT 模型,但另外两类别的准确率均明显优于 ViViT 模型。整体比较可知,本文模型的平均准确率比 ViViT 模型提升了 4.9%。这种显著的性能差异表明本文模型在检测学生课堂专注度上具有更强的能力,证明视线计算模块的引入,有效地提升了模型对于专注度的识别能力。

2.2.2 在自采数据集上的测试

为了进一步测试该系统在实际场景中应用的性能,本文拍摄了 10 段 1min 包含 10 位受试者上课时的视频用于实验,如图 7 所示。由于在实际场景中难以准确为 4 种类别打标签,因此在拍摄视频时让 10 位受试者随机地做 3 种动作,包括:注视黑板(非常专注)、遮挡脸部(无法检测到人脸)、望向窗外(非常不专注),模拟真实授课场景中不同的学习状态。此外,为了减少个体差异带来的影响,对自

采视频进行了 3 次检测,结果如表 4 所示。



Fig. 7 An example of self-collected video
图 7 自采视频例子

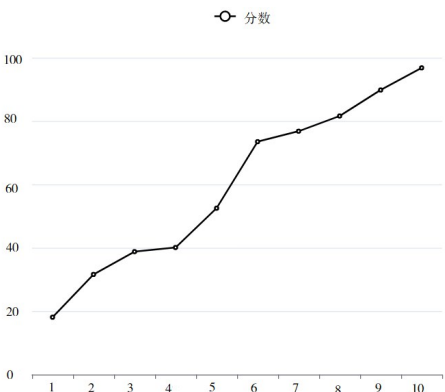
Table 4 Systematic scoring of subjects under different performances
表 4 受试者不同表现下的系统评分

序号	注视黑板 (人数)	遮挡脸部 (人数)	望向窗外 (人数)	第一次 系统评 分	第二次 系统评 分	第三次 系统评 分	平均系 统评分
1	0	0	10	18.1	21.4	22.9	20.8
2	0	2	8	31.6	28.9	27.2	29.2
3	0	4	6	38.8	37.5	39.5	38.6
4	2	2	6	40.1	42.0	39.5	40.5
5	4	0	6	52.5	54.6	53.9	53.7
6	6	0	4	73.5	74.0	73.0	73.5
7	6	2	2	76.8	74.7	78.6	76.7
8	8	0	2	81.6	83.3	80.8	81.9
9	8	2	0	89.8	93.2	90.5	91.2
10	10	0	0	96.8	87.9	92.2	92.3

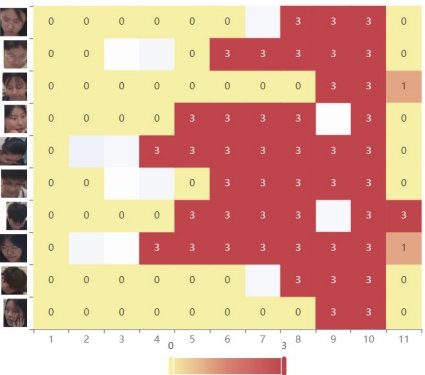
可以看出,当班级学生整体呈现出某种状态时,3 次不同时刻获得的系统评分差异并不大,说明系统受不同个体的影响较小。当所有人都望向窗外时,系统给出了最低分数 20.8 分。随着注视黑板的人数逐渐增加,系统评分也逐步提高。当 5 位学生都注视黑板时,系统平均评分为 92.3 分。系统评分与实际情况相符。

值得注意的是,当所有人望向窗外时,学习者的视线不在专注区内,学习者状态属于非常不专注,按本文公式计算,理论分数应为 0,但 3 次系统评分均超过了 18,实验结果偏差较大,其原因可能是受试者头部偏转幅度不足或过分偏转导致面部识别不全。此外,从班级状态 3 和 5 可以看出,注视黑板和遮挡脸部对于系统分数的影响不同,注视黑板的行为更能体现出专注的状态,因此对分数产生的影响更大,分数提升得更加明显。

该系统可通过折线图和热力图的形式分别展现班级整体以及单个学习者的专注程度,如图 8 所示(彩图扫 OSID 码可见)。教师可以根据折线图直观地了解班级整体的专注度趋势,而热力图则展示了各学生在不同时间点的专注程度,颜色越红代表越专注,颜色越黄则代表越不专注,图中数字 0、1、2、3 分别代表非常不专注、不专注、专注、非常专注,空白则代表没有检测到人脸。



(a) Line graph of overall class engagement
(a) 班级整体专注度折线图



(b) Heat map of individuals' engagement
(b) 个体专注度热力图

Fig. 8 Engagement detection interface
图 8 专注度检测界面

2.2.3 系统响应时间测试

为了测试系统的实时性,本文分别采集了 5~40 人的线下课堂视频,每个视频进行时长为 30 s 的系统响应时间测试。为了避免实验的偶然性,对每个视频测试了 10 次,取 10 次的平均响应时间作为该视频的测试结果,如图 9 所示。

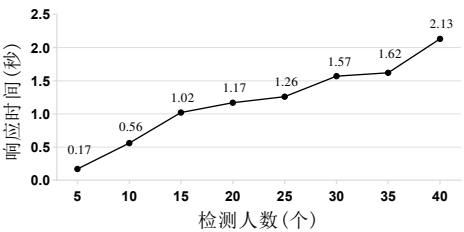


Fig. 9 System response time for different number of people detected
图 9 不同检测人数下系统响应时间

从图 9 可以看出,当学生人数为 10 和 20 时,系统响应时间分别为 0.56 s 和 1.17 s;当学生人数达到 40 时,响应时间也仅为 2.13 s。实验数据表明,该系统能满足人数在 40 人左右的线下教学环境的实时性检测需求。

3 结语

为了对线下课堂教学中班级整体专注度进行实时自动检测,本文设计了一套基于计算机视觉的课堂专注度检测系统,利用DBFace模型检测视频序列中的人脸区域及关键点位置,根据学生的头部偏转角度判断视线范围,利用ViViT模型进行专注度分类,通过融合人脸检测信息、视线信息和专注度分类信息,最终将班级在某时刻的整体专注度量化为一个位于0~100区间的评分。实验结果证明,该系统能够准确、实时地检测线下课堂教学环境中的班级整体专注程度,提供客观有效的专注度评分,帮助教师快速获取课堂中学习者的整体专注状态。

由于ViViT模型结构较为复杂,当运行系统的设备硬件性能较差、算力不足时,会导致系统响应时间过长,无法达到实时检测效果。此外,对于模糊不清的人脸图像,系统难以进行有效的关键点检测,从而降低了系统识别准确率。在未来工作中,将改进关键点检测模型和专注度检测模型,提高关键点检测模型对于模糊图像的检测性能,降低专注度检测模型复杂度,提高系统实时性。

参考文献:

- [1] BOND M, BUNTINS K, BEDENLIER S, et al. Mapping research in student engagement and educational technology in higher education: a systematic evidence map[J]. International Journal of Educational Technology In Higher Education, 2020, 17(1): 1-30.
- [2] SHI Y. Research on evaluation model of classroom attention of students based on face recognition technology[D]. Wuhan: Central China Normal University, 2020.
石奕. 基于人脸识别技术的学生课堂专注度评价模型研究[D]. 武汉: 华中师范大学, 2020.
- [3] WERLANG P S, JAUQUES P A. Student engagement recognition from videos: a comparison between deep learning neural network architectures[J]. Bulletin of the Technical Committee on Learning Technology, 2021, 21(3): 7-12.
- [4] YANG J, WANG K, PENG X, et al. Deep recurrent multi-instance learning with spatio-temporal features for engagement intensity prediction[C]// Proceedings of the 20th ACM International Conference on Multimodal Interaction, 2018: 594-598.
- [5] LIAO J, LIANG Y, PAN J. Deep facial spatiotemporal network for engagement prediction in online learning[J]. Applied Intelligence, 2021, 51: 6609-6621.
- [6] MEHTA N K, PRASAD S S, SAURAV S, et al. Three-dimensional DenseNet self-attention neural network for automatic detection of student's engagement[J]. Applied Intelligence, 2022, 52(12): 13803-13823.
- [7] ZALETELJ J, KOŠIR A. Predicting students' attention in the classroom from Kinect facial and body features[J]. EURASIP Journal on Image And Video Processing, 2017(1): 1-12.
- [8] ZHONG M C, ZHANG J L, LAN Y B, et al. Research on online education concentration based on face detection and fuzzy comprehensive evaluation[J]. Computer Science, 2020, 47(S2): 196-203.
钟马驰, 张俊朗, 蓝扬波, 等. 基于人脸检测和模糊综合评判的在线教育专注度研究[J]. 计算机科学, 2020, 47(S2): 196-203.
- [9] THOMAS C, JAYAGOPI D B. Predicting student engagement in classrooms using facial behavioral cues[C]//Proceedings of ACM SIGCHI International Workshop, 2017: 33-40.
- [10] LAM P T, CHI L P, TUAN N H, et al. A computer-vision based application for student behavior monitoring in classroom[J]. Applied Sciences, 2019, 9(22): 4729.
- [11] TANG K, XIAN Q, LI M Y. The research of college classroom focus based on face detection[J]. Journal of Chongqing Normal University (Natural Science Edition), 2019, 36(5): 123-129.
唐康, 先强, 李明勇. 基于人脸检测的大学课堂关注度研究[J]. 重庆师范大学学报(自然科学版), 2019, 36(5): 123-129.
- [12] GUO X X. Research on the analysis system of students' classroom attention analysis based on micro-expression recognition[D]. Kunming: Yunnan Normal University, 2019.
郭晓旭. 基于微表情识别的学生课堂专注度分析系统研究[D]. 昆明: 云南师范大学, 2019.
- [13] HE X L, GAO Q, LI Y Y, et al. Spontaneous learning facial expression recognition based on deep learning[J]. Computer Applications and Software, 2019, 36(3): 180-186.
何秀玲, 高倩, 李洋洋, 等. 基于深度学习模型的自发学习表情识别方法研究[J]. 计算机应用与软件, 2019, 36(3): 180-186.
- [14] ARNAB A, DEGHANI M, HEIGOLD G, et al. Vivit: a video vision transformer[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 6836-6846.
- [15] Github. DBFace[EB/OL]. <https://github.com/dlunion/DBFace>.
- [16] FERTIG A, WEIGOLD M, CHEN Y. Machine learning based quality prediction for milling processes using internal machine tool data[J]. Advances in Industrial and Manufacturing Engineering, 2022, 4: 100074.
- [17] MOHAMAD N O, DRAS M, HAMEY L, et al. Automatic recognition of student engagement using deep learning and facial expression[C]//Proceedings of Joint European Conference on Machine Learning and Knowledge Discovery in Databases, 2020: 273-289.
- [18] ZAIDI S S A, ANSARI M S, ASLAM A, et al. A survey of modern deep learning based object detection models[J]. Digital Signal Processing, 2022, 126: 103514.
- [19] GUO M H, XU T X, LIU J J, et al. Attention mechanisms in computer vision: a survey[J]. Computational Visual Media, 2022, 8(3): 331-368.
- [20] TANG D, YANG X, WANG X. Improving the transferability of the crash prediction model using the TrAdaBoost. R2 algorithm[J]. Accident Analysis & Prevention, 2020, 141: 105551.

(责任编辑: 孙娟)