

基于YOLOv11与DeepSeek的课堂专注度智能分析系统设计

尚靖玲, 黄欣雨, 郭彦伟, 关雪敏*

广西民族师范学院数学与计算机科学学院, 广西 崇左

收稿日期: 2026年9月8日; 录用日期: 2026年9月30日; 发布日期: 2026年10月15日

摘要

课堂专注度评估长期以来都是高校教学质量监控中比较薄弱的环节, 在广西民族地区的高校, 这一情况表现得更加明显。人工巡课能够覆盖到的班级比较有限, 评价结果主要依靠督导个人的经验, 缺少量化的标准; 教室里的监控设备每天都在录制视频, 但这些视频数据长期没有得到有效的利用; 即使引入了行为检测算法, 输出的内容也大多停留在行为计数这一层面, 与教师真正需要的教学诊断之间还存在一定的差距。针对上述情况, 本文设计了一套课堂专注度智能分析系统, 检测部分采用YOLOv11模型, 诊断部分接入DeepSeek大模型。系统按照视觉感知、量化评估、语义诊断三个层次进行组织: 感知层以YOLOv11n作为基线模型, 在C3K2特征输出之后加入通道-空间注意力模块, 同时在颈部网络增加一条浅层特征通向检测层的融合支路, 用来减少教室人员密集遮挡和后排小目标所带来的漏检; 评估层采用课堂参与度指数(EPI)对专注度进行量化, 行为权重通过层次分析法并结合广西民族院校课堂的实际情况进行标定, 输出0到100分的时序评分; 诊断层通过三层Prompt调用DeepSeek模型, 把时序行为统计转换为教学改进报告。从实验结果来看, 基线模型的参数量为2.62M, mAP@0.5为0.842, 单帧推理时间约为18.3 ms, 五类行为中低头类别的检测效果最好, mAP@0.5达到0.952。系统在校内多个班级进行了试用, EPI评分与教师主观评价的相关系数为0.87。系统可以运行在普通办公电脑上, 直接使用教室已有的监控设备, 推理过程在本地完成, 不需要额外采购硬件, 比较适合经费有限的民族地区院校使用。

关键词

YOLOv11, DeepSeek大模型, 课堂专注度, 行为目标检测, EPI量化指数

Design of an Intelligent Classroom Attentiveness Analysis System Based on YOLOv11 and DeepSeek

Jingling Shang, Xinyu Huang, Yanwei Guo, Xuemin Guan*

*通讯作者。

文章引用: 尚靖玲, 黄欣雨, 郭彦伟, 关雪敏. 基于YOLOv11与DeepSeek的课堂专注度智能分析系统设计[J]. 软件工程与应用, 2026, 15(5): 688-695. DOI: 10.12677/sea.2026.155064

Abstract

Classroom attentiveness evaluation has long been a relatively weak part of teaching quality monitoring in universities, and this problem is more noticeable at universities in the ethnic areas of Guangxi. Manual patrols can cover only a limited number of classes, and the evaluation mainly depends on the personal experience of supervisors, with no quantitative standard; surveillance cameras in classrooms record videos every day, but these video data have not been effectively used for a long time; even when behavior detection algorithms are applied, their output mostly stays at the level of behavior counting, which still falls short of the teaching diagnosis that teachers need. In view of the above situation, this paper designs an intelligent classroom attentiveness analysis system, in which YOLOv11 is used for behavior detection and the DeepSeek large language model is integrated for diagnosis. The system is organized into three layers, namely visual perception, quantitative evaluation, and semantic diagnosis. The perception layer takes YOLOv11n as the baseline model; a channel-spatial attention module is added after the C3K2 feature output, and a fusion branch from shallow features to the detection layer is added in the neck network, so as to reduce missed detections caused by dense occlusion and small targets in back rows. The evaluation layer uses the Engagement Participation Index (EPI) to quantify attentiveness; the behavior weights are calibrated through the analytic hierarchy process based on the actual conditions of classrooms in Guangxi ethnic universities, and time-series scores from 0 to 100 are produced. The diagnosis layer invokes the DeepSeek model through three-layer prompts, and converts time-series behavior statistics into teaching improvement reports. The experimental results show that the baseline model has 2.62M parameters, an mAP@0.5 of 0.842, and a single-frame inference time of about 18.3 ms; among the five behavior categories, the head-down category achieves the best detection result, with an mAP@0.5 of 0.952. The system has been tried in several classes on campus, and the correlation coefficient between EPI scores and teachers' subjective evaluation is 0.87. The system can run on ordinary office computers and directly use the existing surveillance equipment in classrooms; the inference is completed locally, and no additional hardware purchase is required, which makes it relatively suitable for ethnic-area universities with limited budgets.

Keywords

YOLOv11, DeepSeek Large Language Model, Classroom Attentiveness, Behavioral Object Detection, EPI Quantitative Index

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

近年来,教育数字化转型不断推进,课堂教学质量评估也受到了越来越多的关注。2025年,《教育部等九部门关于加快推进教育数字化的意见》¹提出要利用智能技术感知课堂教学状态,开展数据驱动的

¹https://www.gov.cn/zhengce/zhengceku/202504/content_7019045.htm

精准教研；《基础教育课程教学改革深化行动方案》²也对教学评价改革提出了要求，鼓励探索过程性评价和增值性评价的新模式。2026 年，广西全区教育工作会议把加速人工智能与教育深度融合列为当年的重点任务，提出要面向民族地区院校开发成本较低、容易落地的智慧教学工具。广西民族师范学院是桂西南地区唯一的本科师范院校，学生来自壮、汉、瑶、苗等多个民族，班级规模的差别也比较大，小班只有 30 人左右，大班在阶梯教室上合班课，人数达到 120 人，教学硬件方面的预算同样比较有限。从实际运行的情况来看，目前的课堂专注度评估主要依靠人工巡课，这种方式存在几个比较突出的问题：第一，教学督导的人数有限，很难做到每节课、每个班级都巡查到，坐在后排和教室边缘位置的学生经常不在巡查范围之内；第二，对课堂情况的判断主要依靠个人经验，不同督导听同一位教师的课，给出的评价有时会比较明显的差别；第三，教室基本上都安装了监控设备，每天都会产生大量视频，由于缺少自动化的分析工具，这些视频长期处于闲置状态；第四，现有的一些技术方案中，行为数据与教学诊断是相互脱节的，结果大多只停留在行为计数的层面，不能直接转化为教师可以使用的教学建议。围绕课堂专注度与学习行为分析，已有研究从时序多特征融合[1]、在线学习注意力检测[2]、学生行为目标检测[3]、举手行为检测[4]、智慧课堂行为识别[5]、教学手势与专注度关系[6]以及课堂专注度监测系统[7]等方面进行了探索，但多数成果仍停留在行为识别或统计层面，缺少与教学诊断的直接衔接。在这样的背景下，开发一套符合民族地区院校办学条件、成本较低并且可以复制推广的课堂专注度智能分析系统，具有比较现实的意义。

本文的研究工作依托 2026 年国家级大学生创新创业训练计划项目开展，设计并实现了一套基于 YOLOv11 与 DeepSeek 的课堂专注度智能分析系统。系统的主要工作包括四个部分，分别是对 YOLOv11 检测模型进行改进、建立 EPI 量化评估模型、通过三层 Prompt 策略驱动 DeepSeek 大模型完成教学诊断，以及开发 Web 端的可视化系统；这四个部分依次衔接，把视觉感知、量化评估和语义诊断连接成一条完整的技术链条，为民族地区高校的课堂过程性评价提供了一种成本较低的解决方式。

2. 基于 YOLOv11 与 DeepSeek 的课堂专注度智能分析系统

2.1. 系统整体架构设计

系统整体采用三层结构，从下到上分别是视觉感知层、量化评估层和语义诊断层。最下层是视觉感知层，核心部分为改进后的 YOLOv11 模型，支持三种数据输入方式，分别是图片拖拽上传、本地视频导入和教室监控摄像头实时取流；模型可以自动识别五类学生行为，包括学习、写字、玩手机、低头和趴桌睡觉，并逐帧输出行为类别、坐标框、置信度以及时序频次等结构化数据。模型的推理过程在本地 PC 上完成，视频数据不上传至云端，以符合校园学生隐私管理方面的要求。中间层是量化评估层，主要包括课堂参与度指数(EPI)计算引擎和时序分析模块；EPI 的作用是把离散的单帧行为统计转换为连续的、可以相互比较的专注度评分，时序模块则按照分钟滑动窗口对整节课的数据进行汇总，生成行为分布柱状图和 EPI 走势曲线，负向行为比较集中的时段会被自动标记出来。最上层是语义诊断层，该层与 DeepSeek 大模型 API 进行对接，把中层输出的时序量化数据全部导入，再通过三层 Prompt 加入场景约束和教育专业知识，最终输出三段式的学情报告，内容包括课堂整体专注度评价、关键时段注意力波动的原因分析，以及可以执行的教学优化建议，从而把机器识别出的数据转化为教师能够参考的教学信息。

2.2. 核心关键技术实现

(1) 改进 YOLOv11 检测模型，基线网络选用 YOLOv11n 轻量化版本，参数量约为 2.6M，推理速度

²http://www.moe.gov.cn/srcsite/A26/jcj_kcjggh/202306/t20230601_1062380.html

较快,能够满足普通 PC 和教室多媒体终端的本地部署需要。教室场景中存在两个比较突出的检测难点,一个是人员密集时相互遮挡,另一个是后排学生在画面中的目标较小;针对这两个问题,本文在骨干网络 C3K2 的特征输出层之后加入了一个通道-空间注意力模块, (Convolutional Block Attention Module, CBAM) [8]其中通道注意力分支通过全局池化对贡献较大的特征通道进行筛选,使模型对头部姿态、手部动作等具有区分度的行为区域响应更强,空间注意力分支则对画面中行为关键区域的权重进行提取,同时把桌椅、墙面等背景内容的干扰抑制下去;除此之外,本文还对颈部网络的多尺度特征融合方式进行了调整,增加了一条浅层高分辨率特征图通向检测层的传递通路,浅层特征中包含的空间细节信息比较多,这样可以让模型更充分地利用这些信息,后排小尺寸学生目标的检测能力也随之得到提高。改进模型的网络结构如图 1 所示。

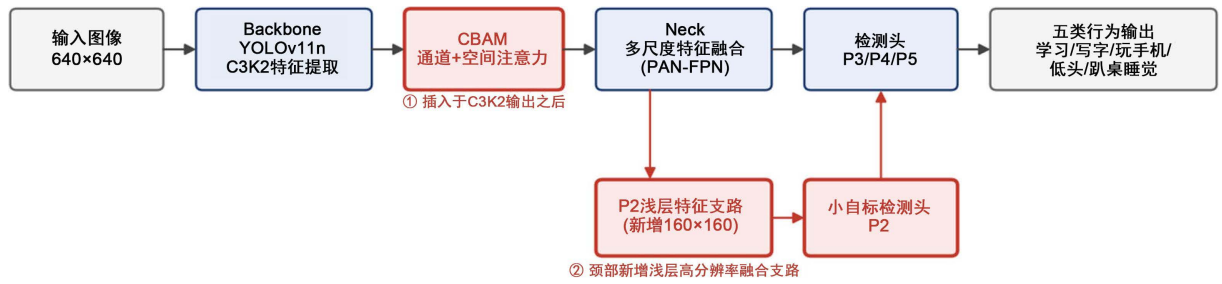


Figure 1. Structure of the improved YOLOv11n network (red modules indicate newly added components)

图 1. 改进 YOLOv11n 网络结构图(红色模块为新增部分)

(2) EPI 量化模型:

$$EPI = \frac{\alpha P_{\text{pos}}}{\beta P_{\text{neg}} + \varepsilon} \times 100 \quad (1)$$

$$P_{\text{pos}} = \frac{1.0 f_{\text{study}} + 0.8 f_{\text{write}}}{\sum_{k=1}^5 f_k}, P_{\text{neg}} = \frac{1.0 f_{\text{phone}} + 1.0 f_{\text{head-down}} + 1.5 f_{\text{sleep}}}{\sum_{k=1}^5 f_k} \quad (2)$$

本文采用课堂参与度指数(Engagement Participation Index, EPI)对专注度进行量化,分数归一化到[0, 100]区间。式(2)中, f_k 表示第 k 类行为在单位时间窗内的检测帧数,五类行为的权重依次取 1.0、0.8、1.0、1.0 和 1.5 (趴桌睡觉按 1.5 倍从重计罚); P_{pos} 、 P_{neg} 分别为加权后的正向行为归一化频率与负向行为归一化时长占比。 α 取值为 0.6, β 取值为 0.4, 各权重通过层次分析法并结合广西民族地区高校的课堂样本标定得到, EPI 分数低于 40 分时作为专注度预警处理。

从物理意义上看, P_{pos} 反映学习、写字等专注行为的相对强弱,取值越大表示专注程度越高; P_{neg} 反映玩手机、低头、趴桌睡觉等走神行为的相对强弱,取值越大表示专注程度越低。在边界条件处理上, ε 为数据平滑因子,当负向行为极少、 P_{neg} 趋近于 0 时,可避免分母为零而导致评分发散,并对得分进行平滑;当某一时间窗内未检测到有效目标时,该窗不参与 EPI 计算,沿用上一个有效窗评分;计算结果统一截断到[0, 100]区间。

(3) DeepSeek 诊断引擎, Prompt 按照三层进行引导:第一层是事实层,把完整的结构化时序数据输入模型,包括全课平均 EPI、各时段的波动情况和负向行为集中的时间;第二层是知识层,向模型注入教育学注意力曲线理论以及民族院校的教学特点,避免诊断内容过于空泛;第三层是任务层,要求模型以高校教学数据分析师的身份,按照 STAR 原则输出三部分内容,分别是整体评价、时段问题归因,以及三条可以直接落实的课堂优化建议。

(4) Web 可视化系统, 前端采用 React 进行开发, 后端采用 Node.js 进行开发, 图表部分使用 ECharts 库进行渲染; 图片检测、视频分析和摄像头实时监控三种模式可以在页面上通过一键操作进行切换。系统的功能模块主要包括数据源上传、实时检测、数据可视化、学情报告和历史归档五个部分, 其中实时检测可以同步显示检测框和 EPI 分值, 数据可视化包括柱状图和走势曲线, 学情报告由 DeepSeek 生成并支持 PDF 格式导出, 历史归档可以按照班级、课程和日期进行检索; 教师不需要具备编程方面的知识, 就可以独立完成整个流程的操作。

2.3. 三层 Prompt 诊断示例与学情报告样本

为说明诊断层工作方式, 下面给出一组匿名三层 Prompt 及 DeepSeek 生成的学情报告样本(班级、课程等信息已隐去)。

事实层 Prompt: “本节课平均 EPI 为 100 分, 玩手机 2 次, 趴桌睡觉 1 次, 各时段 EPI 均为 100 分, 请仅据此分析。”

知识层 Prompt: “请结合广西民族地区高校课堂特点: 多民族合班, 需区分整体表现与个别学生短暂分心, 避免简单归因于态度, 可适当考虑民族文化元素。”

任务层 Prompt: “以高校教学数据分析师身份, 输出三部分: 整体专注度评价、问题行为及时段分析、可落地改进建议; 语气客观, 不针对个人。”

在上述三层 Prompt 的引导下, DeepSeek 生成的学情报告样本如下:

1) 整体课堂专注度评价: 平均 EPI 为 100, 表明学生课堂参与度极高, 全程保持高度专注, 无注意力分散迹象。这一数据在民族地区高校中属于优秀水平, 反映出学生对课程内容兴趣浓厚, 教师教学设计有效。

2) 问题行为及时段分析: 尽管整体表现优异, 但仍存在少量问题行为: 玩手机出现 2 次, 趴桌睡觉 1 次。时段数据显示, 所有 EPI 值均为 100, 说明问题行为发生在全局高专注背景下, 可能是个别学生短暂分心, 未对整体课堂造成显著影响。

3) 针对性改进教学建议: 针对民族地区学生文化背景, 建议教师融入本地民族文化元素(如案例、民歌、手工艺), 提升课堂亲和力。对玩手机和趴桌行为, 可增加互动环节(如小组讨论、抢答), 并设定明确的课堂行为规范。同时, 结合学生多元智能特点(如擅长口头表达), 设计更多参与式任务, 强化个体参与感。

3. 实验与结果分析

3.1. 实验设置与评价标准

实验使用的显卡为 NVIDIA RTX 4060 Laptop, 显存为 8GB, 深度学习框架采用 Ultralytics YOLOv11 搭配 PyTorch 2.0。数据集面向教室场景构建, 采集主要依托学校教室现有的监控摄像头, 同时使用移动拍摄设备进行补充, 场地涵盖阶梯教室、普通平面教室和计算机机房三种类型, 光照条件包括上午强光、阴天散射光和室内灯光等多种情况, 班级人数从 30 人到 120 人不等。标注工作使用 LabelImg 工具逐帧进行, 标注目标框经过精细化处理, 类别一共 5 类, 分别是学习、写字、玩手机、低头和趴桌睡觉。数据增强采用 Mosaic 拼接、随机尺度缩放和亮度对比度调节几种方式组合进行。数据集总量约为 5000 张图片, 其中训练集约 4000 张, 验证集 990 张, 大致按照 8:2 的比例划分。训练参数的设置如下, 优化器选用 SGD, 初始学习率为 0.01, 批量大小为 16, 输入图像尺寸为 640×640 , 训练轮次上限为 100 轮, 并采用 $\text{patience} = 20$ 的早停策略, 同时加载 COCO 数据集的预训练权重来加快收敛速度。损失函数由 CIoU 边界框回归损失、DFL 分布焦点损失和 BCE 分类损失加权组合构成。评价指标选用 $\text{mAP}@0.5$ 、

mAP@0.5:0.95、Precision、Recall、参数量和推理速度,另外还通过混淆矩阵对各类别的误检情况进行分析。

3.2. 实验结果与对比分析

基线 YOLOv11n 模型在验证集上的检测结果如下,总参数量为 2.62M, mAP@0.5 为 0.842, Precision 为 0.813, Recall 为 0.775, mAP@0.5:0.95 为 0.640, 单帧推理时间约为 18.3 ms。各个类别的 mAP@0.5 分别为,学习 0.800, 写字 0.781, 玩手机 0.787, 低头 0.952, 趴桌睡觉 0.891; 其中低头类别的检测精度最高, 主要原因是该类别的姿态特征比较明显, 样本数量也比较充足, 写字类别的精度相对偏低, 原因是写字与玩手机、低头在动作形态上比较相似。从混淆矩阵的结果来看, 模型的误检主要集中在写字和玩手机两个类别之间, 表现为两类行为的相互混淆, 其中写字被误判为玩手机的有 199 例, 玩手机被误判为写字的有 224 例, 此外还有 640 个学习样本被误判为背景; 出现这些误检, 主要是因为相关行为在头部姿态和手部区域上的视觉特征比较接近。训练过程中的损失与精度曲线如图 2 所示。

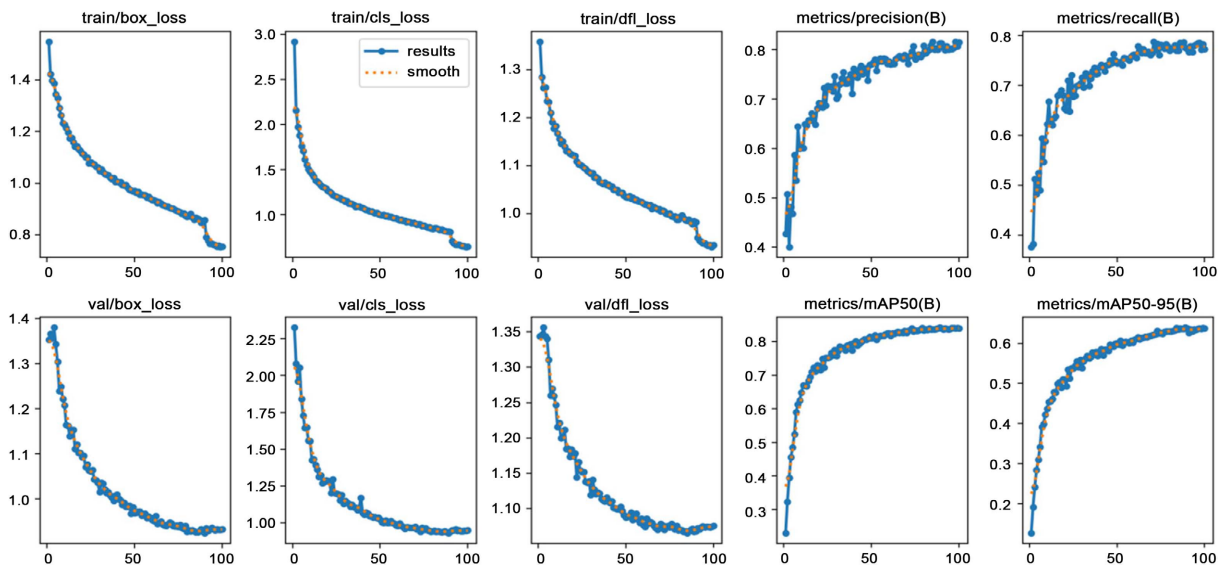


Figure 2. Loss and accuracy curves during model training

图 2. 模型训练过程损失与精度曲线

为了对各个改进模块的作用进行验证, 本文设计了三组消融实验, 实验结果如表 1 所示。加入注意力模块之后, 参数量从 2.62M 增加到 2.71M, mAP@0.5 从 0.842 提高到 0.868, Recall 从 0.775 提高到 0.804, 推理时间从 18.3 ms 增加到 22.6 ms; 在此基础上加入多尺度优化之后, 参数量增加到 2.89M, mAP@0.5 进一步提高到 0.883, Recall 达到 0.821; 两项改进全部叠加之后, 模型参数量为 2.98M, mAP@0.5 达到 0.908, Recall 达到 0.847, 推理时间为 32.7 ms。从上述结果可以看出, 注意力模块主要通过加强关键区域的特征响应来提高模型的判别能力, 多尺度优化主要通过扩大特征覆盖的尺度范围来改善小目标的检测效果, 两种策略的作用点不同, 结合使用时能够形成互补。

系统已经在多个班级完成了试点验证, 将 EPI 评分与教师的主观评价进行对比, 二者的皮尔逊相关系数为 0.87, 说明 EPI 量化指数具有较高的教学参考价值; 在学情报告的实用性方面, 参与试点的教师中有 85% 表示认可。整套系统对使用者的要求不高, 教师不需要掌握编程知识, 在网页端进行简单操作即可使用, 系统在普通办公 PC 上能够稳定运行, 并且直接复用校园现有的监控硬件, 不需要采购专用的 GPU 服务器或者高清行为捕捉摄像头, 与民族院校经费有限的实际情况比较符合。

Table 1. Comparison of ablation experiment results
表 1. 消融实验结果对比

模型配置	参数量(M)	mAP@0.5	推理速度(ms)	Recall
YOLOv11n 基线	2.62	0.842	18.3	0.775
+注意力模块	2.71	0.868	22.6	0.804
+多尺度优化	2.89	0.883	28.4	0.821
改进模型(两项)	2.98	0.908	32.7	0.847

4. 结论

本文针对广西民族地区高校课堂专注度评估中存在的几个问题，包括人工巡查覆盖范围有限、主观评价缺少量化标准、监控视频数据闲置，以及行为检测与教学诊断之间存在语义断层，设计并实现了一套基于 YOLOv11 与 DeepSeek 的课堂专注度智能分析系统。系统采用视觉感知、量化评估、语义诊断三层递进的技术架构，通过加入通道 - 空间注意力模块和优化多尺度浅层特征融合支路，使密集遮挡和小目标场景下的检测精度得到了提高；提出的课堂参与度指数(EPI)量化模型，实现了 0 到 100 分的标准化专注度时序评分；设计的三层分层 Prompt 工程可以驱动 DeepSeek 大模型把行为数据转化为能够直接使用的教学改进报告；基于 React 和 Node.js 开发的可视化系统对校园现有监控设备进行复用，实现了低成本部署和隐私保护。从实验结果来看，YOLOv11n 基线模型的参数量为 2.62M，mAP@0.5 为 0.842，单帧推理时间约为 18.3 ms，改进模型的 mAP@0.5 提高到 0.908，五类课堂行为中低头类别的检测精度最高，为 0.952；EPI 评分与教师主观评价的相关系数为 0.87，学情报告的实用性得到 85%参与试点教师的认可。系统已经在多个班级完成试点验证，运行情况稳定，可以为民族地区高校的课堂过程性评价提供一种可复制的低成本 AI 解决方案。

本研究仍然存在一定的局限性，消融实验中各个改进模块在不同光照条件和座位布局下的泛化性能还需要进一步验证，更大规模的跨校、跨场景测试还没有开展，DeepSeek 诊断报告对教学的长期效果也需要通过持续积累数据来进行追踪。后续的工作主要围绕以下几个方向展开：第一，把数据集扩展到 10,000 张以上，覆盖更多民族地区院校的课堂场景，使模型对多样化课堂环境的适应能力得到增强；第二，对模型量化与剪枝技术进行探索，进一步降低推理延迟，以适应边缘计算设备的部署需要；第三，引入面部表情、视线方向等多模态信息进行融合，使专注度评估更加全面和可靠；第四，建立长期追踪机制，通过跨学期、跨课程的持续数据采集，对系统改进教学质量的持续效果进行验证。

基金项目

广西教育科学“十四五”规划 2023 年度专项课题《地方应用型高校创新创业教育与网络工程专业融合发展研究》(2023ZJY1900)；国家级大学生创新创业训练计划项目“基于 YOLOv11 与 DeepSeek 的课堂专注度智能分析系统设计”(202610604010)。

参考文献

[1] 黎先平. 基于时序多特征融合的课堂专注度评估研究[D]: [硕士学位论文]. 南昌: 江西财经大学, 2024.

[2] 何飞. 在线学习场景中学生学习注意力检测方法研究[D]: [硕士学位论文]. 秦皇岛: 燕山大学, 2025.

[3] 李美. 基于改进 YOLOv11 的学生课堂行为检测算法[J]. 信息技术与信息化, 2026(3): 47-50.

[4] 程蓉, 王怡臻. 基于 YOLOv11 与人脸识别的课堂举手行为智能检测系统研究[J]. 现代信息科技, 2025, 9(24): 33-38.

[5] 贾士钦. 基于深度学习的智慧课堂学习行为识别研究[D]: [硕士学位论文]. 曲阜: 曲阜师范大学, 2025.

-
- [6] 施剑阳. 基于计算机视觉技术的教学手势对课堂专注度的影响研究[D]: [硕士学位论文]. 金华: 浙江师范大学, 2024.
 - [7] 陈洋洋, 汤凯文, 沈鑫, 等. 基于 AI 的课堂专注度智能监测和分析系统[J]. 物联网技术, 2026, 16(5): 43-45.
 - [8] Woo, S., Park, J., Lee, J. and Kweon, I.S. (2018) CBAM: Convolutional Block Attention Module. In: Ferrari, V., Hebert, M., Sminchisescu, C. and Weiss, Y., Eds., *Computer Vision—ECCV 2018*, Springer, 3-19.
https://doi.org/10.1007/978-3-030-01234-2_1