

基于 YOLOv7-SSW 模型的智慧课堂行为识别算法

徐 韬¹, 陈孟元^{2,3}, 程云麟¹

1. 安徽工业大学 商学院, 安徽 马鞍山 243032
2. 安徽工程大学 电气工程学院, 安徽 芜湖 241000
3. 高端装备先进感知与智能控制教育部重点实验室, 安徽 芜湖 241000

摘要:目的 现有的学生行为识别模型存在检测精度较低且预测速度不足的问题, 基于此, 提出一种基于 YOLOv7-SSW 模型的智慧课堂行为识别算法, 用于提升检测精度的同时保障算法的实时性。方法 首先, 在 YOLOv7 的主干网络引入 SE(Squeeze-and-Excitation Module) 注意力机制, 建立通道相关性增强算法对信息的敏感性, 进而提升算法的检测精度; 其次, 为了增强算法的实时性, 将原始采用串行通道设计的 SPPCSPS 模块改进为并行通道设计, 有效提升了算法的检测速度; 最后, 引入 Wise-IoU 损失函数增强算法处理训练数据中的低质量样本能力, 进一步提升模型检测精度。结果 所提方法在 STBD-08 数据集上的 mAP(mean Average Precision) 达到 91.9%, 与 CBPH-Net、YOLOv7 算法相比, 分别提升了 4.4%、5.6%, 且单帧推理时间仅 45.1 ms。结论 所提算法实现了学生课堂行为的实时准确识别, 为推动思政教育从知识灌输向行为素养培育模式转变提供了技术支撑, 对推动教育数字化创新发展具有重要意义。

关键词: 高等教育; 思政教育; 学生行为识别; YOLOv7

中图分类号: G641; TP391; G434 **文献标识码:** A **doi:** 10.16055/j.issn.1672-058X.2026.0001.008

Behavior Recognition Algorithm for Smart Classroom Based on YOLOv7-SSW Model

XU Tao¹, CHEN Mengyuan^{2,3}, CHENG Yunlin¹

1. School of Business, Anhui University of Technology, Ma' anshan 243032, Anhui, China
2. School of Electrical Engineering, Anhui Polytechnic University, Wuhu 241000, Anhui, China
3. Key Laboratory of Advanced Perception and Intelligent Control of High-end Equipment, Wuhu 241000, Anhui, China

Abstract: Objective Existing student behavior recognition models suffer from low detection accuracy and insufficient prediction speed. To address this, this paper proposes a smart classroom behavior recognition algorithm based on the YOLOv7-SSW model, aiming to improve detection accuracy while ensuring real-time performance. **Methods** First, the Squeeze-and-Excitation (SE) attention mechanism was integrated into the backbone network of YOLOv7. This integration strengthened the modeling of inter-channel correlations and improved feature sensitivity, thereby improving the detection accuracy of the algorithm. Second, to enhance real-time performance, the original serial-channel Spatial Pyramid Pooling with Cross Stage Partial Connections Structure (SPPCSPS) module was redesigned into a parallel architecture, significantly accelerating the detection speed. Finally, the Wise-IoU loss function was integrated to strengthen the algorithm's capability in handling low-quality samples in training data, further boosting detection accuracy. **Results** The

收稿日期: 2024-04-12 **修回日期:** 2024-06-27 **文章编号:** 1672-058X(2026)01-0064-08

基金项目: 国家自然科学基金项目(61903002)资助; 安徽工业大学“沃土计划”项目(WTZ202409)资助; 安徽工业大学思想政治研究会专项课题(2024SYZM01)资助; 安徽工业大学大学生科研训练计划项目(202510360147, S202510360447)资助。

作者简介: 徐韬(1995—), 男, 安徽安庆人, 硕士, 助教, 从事思政数字化研究。

通信作者: 陈孟元(1984—), 男, 安徽芜湖人, 博士, 教授, 从事目标跟踪与识别研究。Email: mychen@ahpu.edu.cn。

引用格式: 徐韬, 陈孟元, 程云麟. 基于 YOLOv7-SSW 模型的智慧课堂行为识别算法[J]. 重庆工商大学学报(自然科学版), 2026, 43(2): 64-71.

XU Tao, CHEN Mengyuan, CHENG Yunlin. Behavior recognition algorithm for smart classroom based on YOLOv7-SSW model [J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2026, 43(2): 64-71.

proposed method achieved a mean Average Precision (mAP) of 91.9% on the STBD-08 dataset, representing improvements of 4.4% and 5.6% over the CBPH-Net and YOLOv7 algorithms, respectively. The single-frame inference time was only 45.1 ms. **Conclusion** The proposed algorithm enables real-time and accurate recognition of student classroom behaviors, providing technical support for transforming ideological and political education from knowledge indoctrination to behavioral literacy cultivation. It holds significant importance for advancing digital innovation and development in education.

Keywords: higher education; ideological and political education; student behavior recognition; YOLOv7

习近平总书记指出:“中国高度重视人工智能对教育的深刻影响,积极推动人工智能和教育深度融合,促进教育变革创新”^[1],这一重要论断为高校思想政治工作的数字化转型和创新发展指明了方向。《教育强国建设规划纲要(2024—2035年)》文件要求“促进人工智能助力教育变革”^[2],并强调要构建“基于大数据和人工智能支持的教育评价和科学决策制度”^[3]。在此背景下,高校急需运用课堂行为识别技术重构思政教育生态,通过精准化、动态化的教学监测手段,推动“三全育人”理念的实质性落地。聚焦课堂教学场域,课堂行为识别技术是指利用深度学习等人工智能算法对学生的课堂微表情和行为动作进行解析,通过生成动态的课堂参与度热力图,量化学风分数、专注度、互动频率等指标^[4],为打造数据驱动的精准确思政教育体系提供智能化工具。

在传统教学模式中,教师往往通过人工观测手段对学生课堂行为实施主观评估,这种评价方式不仅存在效率瓶颈,更因缺乏量化指标而难以形成科学的教学评价。部分高校虽引入视频监控系统进行教学影像采集,但后期仍需要依靠人工回溯视频内容实施行为解码,这种滞后的分析模式无法实现教学过程的动态监测^[5]。为了克服以上问题,人们利用目标检测等人工智能(Artificial Intelligence, AI)技术来分析课堂视频中的学生行为,并展开了深入研究^[6-9]。早期人们提出了模糊广义学习系统(BLS),应用于课堂行为识别,该系统将多类模糊支持向量机(Multi-class Fuzzy Support Matrix Machines, MCF-SMMs)等机器学习方法与BLS相结合以提高分类器的性能。然而,这些方法通常依赖预定义或手工制作的特征,无法有效捕获复杂特征。随着深度学习技术的发展,课堂行为识别技术也取得了巨大的进展。Wu等^[10]提出一种基于感兴趣区域和人脸跟踪的课堂学生行为识别运动目标检测算法,能有效克服课堂环境的复杂性,识别举手、站立两种课堂行为,但该算法准确性不高且识别类别较少;为了提升检测精度,Pang等^[11]改进了传统的聚类分析算法与随机森林算法,并结合人体骨骼模型提升了学生课堂行

为的检测精度;Suzuki等^[12]为了提升检测精度并检测更多的课堂行为,基于OpenPose框架提出了GM-AR系统,对学生的13种典型动作进行姿势估计;王等^[13]提出一种融合人体姿态估计和目标检测的学生课堂行为识别方法,该方法先利用OpenPose提取人体姿态全局特征,然后融合YOLOv3算法提取的交互物体局部特征,最终提高了系统识别的准确性。然而此类人体姿势估计方法对人类关键点的复杂细节处理能力有待提升。为了克服这一问题,Jia等^[14]将YOLOv5与情境注意机制和OpenPose算法相结合进行学生行为识别,提升了对人体关键点的检测精度。尽管以上基于骨骼关键点的人体姿势估计方法能对学生的课堂行为进行识别,但此类方法依赖高精度关节点标注的数据预处理机制,导致标注资源投入呈几何级数增长,耗费大量的人力和时间成本;同时,将骨骼特征提取与目标检测结合,增加了系统的复杂性,极易造成系统运行过程中出现特征解析延迟与多模态数据流失同步等问题,严重影响系统的实时性。为了克服这一问题,Samkari等^[15]利用多尺度特征融合增强高复杂背景关键点的检测;Zhao等^[16]利用不同大小的卷积核并解析多尺度特征来增模型的多尺度特征信息识别能力;陈晨等^[17]融合注意力机制与深度卷积,增强了模型的检测精度。然而,以上方法均存在难以平衡检测精度与实时性的问题。

因此,本文提出一种基于YOLOv7-SSW(YOLOv7-SE+SPPFCSPC+Wise-IoU)模型的智慧课堂行为识别方法,选用YOLOv7模型进行改进。为了提升模型的检测能力,通过在YOLOv7的主干网络引入SE注意力机制^[18],建立通道相关性增强算法对信息的敏感性,进而提升算法的检测精度;与此同时,为了保证系统的实时性,将原始采用串行通道设计的SPPCSPS模块改进为并行通道设计,有效提升了算法的检测速度;最后,引入Wise-IoU损失函数增强算法处理训练数据中的低

质量样本能力,进一步提升模型检测精度。

1 YOLOv7 模型及其改进

1.1 YOLOv7 简介

针对现有学生行为识别模型存在检测精度较低且预测速度不足的问题,提出一种基于 YOLOv7-SSW 模型的智慧课堂行为识别算法。该算法基于 YOLOv7 模型进行改进。YOLOv7 是由 Alexey Bochkovskiy 等开发者在 2022 年提出的 YOLO(You Only Look Once) 系列目标检测算法的版本之一,在速度和精度上实现了显著优化^[19-22]。YOLOv7 模型包括 6 个架构:YOLOv7-Tiny、YOLOv7、YOLOv7x、YOLOv7-E6、YOLOv7-D6 和 YOLOv7-E6E,主要区别是卷积层的数量和辅助探针的数量。YOLOv7 包含输入(Input)、主干网络(Backbone)、头部(Head)和颈部(Neck)4 个模块。YOLOv7 的 Input 主要用于计算图像的原生大小与输入大小之间的缩放比例,以获得缩放后的图像,再执行自适应填充以获得最终的输入图像。图像输入模型后,将其归一化并转换为多层 640×640×3 的图像输入主干

网络中。主干网络通过对原始图像进行特征提取,将其转换为特征图用于后续的目标检测。颈部作为主干网络和头部的桥梁,在目标检测中起着至关重要的作用。Neck 能将来自主干网络中输入的多个特征图进行融合,以便模型能够同时处理多个尺度的信息。目标检测模型的 Head 负责生成最终的目标检测结果,包含边界框回归、类分类、对象置信度估计等功能。

1.2 YOLOv7-SSW 模型

为了增强智慧课堂中学生课堂行为识别准确性和实时性,本文基于 YOLOv7 模型进行改进,提出一种 YOLOv7-SSW 课堂行为识别网络模型。YOLOv7-SSW 模型包含 4 个部分:输入(Input)、主干网络(Backbone)、头部(Head)和颈部(Neck)。在 Backbone 中添加 SE 注意力机制,建立通道相关性,增加算法对信息的敏感性;其次,在 Neck 中使用并行通道的 SPPFCSPC 模块替换原有的 SPPCSPC 模块,减少模型的参数量,在不减少感受野的同时,加快模型处理速度,增强算法的实时性;最后,引入 Wise-IoU 损失函数,进一步提升模型精度。YOLOv7-SSW 模型结构如图 1 所示。

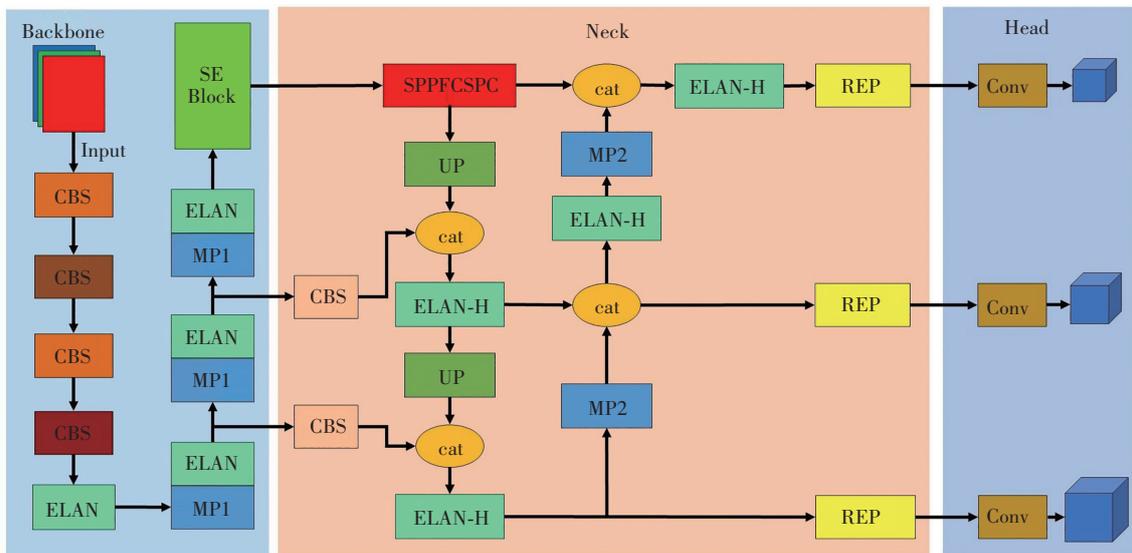


图 1 YOLOv7-SSW 模型结构图

Fig. 1 Structure diagram of YOLOv7 model

1.2.1 SE 注意力机制

注意力机制启发于人类对视觉信息的选择性关注,通过这种方式实现对视觉处理资源的合理利用。而在深度学习中,引入注意力机制可以模拟人类的选择性关注,对关注的信息赋予较大的权重,从而避免信息冗余。挤压和激励网络模块(Squeeze-and-Excitation Module, SE)是一种即插即用注意力模块,能显式地建立通道相关性,以增加模型对通道信息的敏感性,提高模型的鲁棒性。SE 注意力模块结构如图 2 所示。

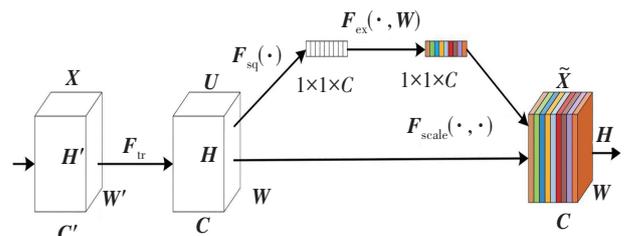


图 2 SE 注意力模块结构图

Fig. 2 SE attention module structure diagram

假设给定输入 I ,第 k 个通道的挤压步骤为

$$z_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W x_c(i, j) \quad (1)$$

式(1)中, X 表示输入, c 是第 c 次迭代。接着, 激励可以被表示为

$$\tilde{X} = X \cdot \text{sigmoid}(T_2(\text{ReLU}(T_1(z)))) \quad (2)$$

式(2)中, “ \cdot ”表示通道乘法; sigmoid 表示非线性激活函数; T_1 和 T_2 是两个线性变换, 用以捕获每个通道的信息; ReLU 表示另一非线性激活函数。

1.2.2 SPPFCSPC 模块

YOLOv7 采用 SPPCSPC 空间金字塔池化方法 (SPP), 其集成了 3 个并行最大池化 MaxPool 操作, 内核大小为 5×5 、 9×9 和 13×13 卷积, 如图 3 所示。该方法具有大量的参数, 并且处理速度较慢, 其计算公式为

$$R = \text{MaxPool}_{k=5}^{p=2}(F) \odot \text{MaxPool}_{k=9}^{p=4}(F) \text{MaxPool}_{k=13}^{p=6}(F) \quad (3)$$

式(3)中, R 表示输出结果, F 表示输入特征图, \odot 表示张量连接。

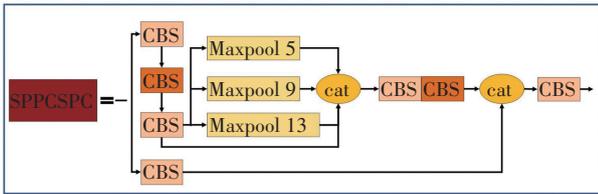


图 3 SPPCSPC 模块结构图

Fig. 3 Structure diagram of SPPCSPC module

本文通过将 SPPCSPC 中的最大池化层转换为 SPPFCSPC 中的形式。3 个并行最大池化 MaxPool 操作, 内核大小均为 5×5 卷积, 如图 4 所示。在保持相同感受野的同时, 提高了模型对数据的特征提取能力, 最终提升了网络准确性。SPPFCSPC 计算公式为

$$\begin{cases} R_1 = \text{MaxPool}_{k=5}^{p=2}(F) \\ R_2 = \text{MaxPool}_{k=5}^{p=2}(R_1) \\ R_3 = \text{MaxPool}_{k=5}^{p=2}(R_2) \\ R_4 = R_1 \odot R_2 \odot R_3 \end{cases} \quad (4)$$

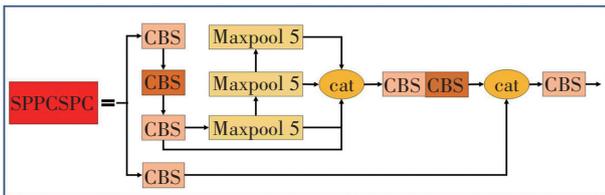


图 4 SPPFCSPC 模块结构图

Fig. 4 Structure diagram of SPPFCSPC module

1.2.3 损失函数优化

边界框损失函数是目标检测损失函数的重要组成部分, 对增强目标检测模型的性能具有重要作用。

YOLOv7 网络采用 GIoU (Generalized Intersection over Union) 损失函数, 虽然该方法对边界框与真实边界框的宽高比进行了有效的关注, 但是对于训练数据中存在的低质量样本却未能有效解决。为了克服这一问题, 提升课堂行为识别精度, 本文采用 Wise-IoU 损失函数替换 GIoU (Generalized Intersection over Union) 损失函数, 通过在锚框与目标框重合度高时, 减少集合度量的惩罚, 从而降低对预训练的干预, 进而提升模型的泛化能力。Wise-IoU 计算公式为

$$\begin{cases} L_{WIoU} = rR_{WIoU}L_{IoU}, r = \frac{\beta}{\delta\alpha^{\beta-\delta}} \\ R_{WIoU} = \exp\left(\frac{(x-x_{gt})^2 + (y-y_{gt})^2}{(W_g^2 + H_g^2)^*}\right) \in [1, e) \\ \beta = \frac{L_{IoU}^*}{L_{IoU}} \in [0, +\infty) \end{cases} \quad (5)$$

式(5)中, r 表示非单调聚焦系数, R_{WIoU} 表示采用 Wise-IoU 计算的惩罚项, L_{IoU} 表示 IoU Loss, $L_{IoU} \in [0, 1]$, β 为离群度, 与锚框质量成反比, δ 和 α 为超参数, (x, y) 和 $y(x_{gt}, y_{gt})$ 分别为预测框中心点坐标和真实框中心点坐标, W_g 和 H_g 是预测框与真实框的最小包围框宽高, 上标 * 表示将 W_g 和 H_g 从计算图中分离。

2 实验与结果分析

2.1 数据集介绍

本文采用文献[16]提出的 STBD-08 数据集, 该数据集由 4 432 幅图像和 151 574 个边界框组成。随后以 8 : 2 的比例划分为训练集和测试集, 包含写作 (writing)、阅读 (reading)、听课 (listening)、转身 (turning around)、举手 (raising hand)、站立 (standing)、讨论 (discussing)、教师指导 (teacher guiding) 8 类行为类别。示例如图 5 所示, 课堂行为分类标准如表 1 所示。



图 5 课堂行为类别示例

Fig. 5 Examples of classroom behavior categories

表 1 课堂行为标准分类

Table 1 Standard classification of classroom behaviors

课堂行为类别	行为描述
书写(writing)	学生低头在记笔记
阅读(reading)	学生看书本内容
听课(listening)	学生抬头认真听课
转身(turning around)	学生转身与他人交谈
举手(raising hand)	学生举起手回答问题
站立(standing)	学生站起来回答问题
讨论(discussing)	学生一起探讨问题
教师指导(teacher guiding)	教师引导学生回答

2.2 实验环境配置与评价指标

本文中的实验在具有 Ubuntu 18.04 操作系统的计算机上执行。该计算机配置有 Intel(R) Core(TM) i5-11400F 的 CPU, NVIDIA RTX 3060 GPU 和 16GB 显存。计算统一设备架构(CUDA)版本为 11.0, 深度学习框架使用 PyTorch 1.7.0, 编译器为 python 3.8。参数设置如下:模型初始学习率为 0.01, 优化器为 SGD, 训练周期为 100。

本文采用准确率 Precision ($f_{\text{Precision}}$)、召回率 Recall (f_{Recall}) 以及平均精度 mAP (mean Average Precision, f_{mAP}) 用于评估网络模型性能。计算如式(6)~式(9)所示:

$$f_{\text{Precision}} = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FP}}} \times 100\% \quad (6)$$

$$f_{\text{Recall}} = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FN}}} \times 100\% \quad (7)$$

$$f_{\text{AP}} = \int_0^1 P(R) dR \quad (8)$$

$$f_{\text{mAP}} = \frac{\sum_{i=1}^M f_{\text{AP}}(i)}{M} \times 100\% \quad (9)$$

其中, N_{TP} 为真阳性, N_{FP} 为假阳性, N_{FN} 为假阴性, $f_{\text{AP}}(i)$ 表示第 i 个类别的平均精确度, M 表示类别的数量。

2.3 训练结果分析

本文算法训练获得的混淆矩阵如图 6 所示。其中, 行(True)表示真实类别, 列(Predicted)表示预测类别, 对角线上的值表示正确预测所占比例, 其余元素为错误预测的比例。由图 6 可知本文算法各类别都保持了高水平的识别精度。

本文算法训练所得准确率-召回率曲线如图 7 所示。由图 7 可知, 本文算法平均精度达到 91.9%, 其他各个类别也取得了较高精度。

本文算法训练过程的准确率、召回率以及 mAP 如图 8 所示。从图 8 中可以看出: 本模型训练在第 89 轮训练时达到了良好的结果, 显示出良好的收敛性和稳定性。

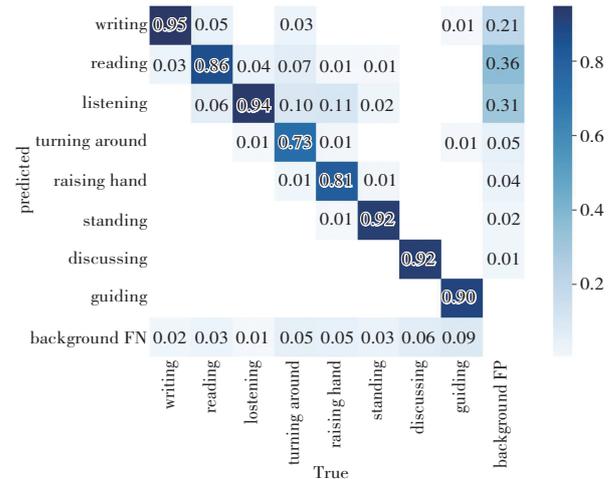


图 6 混淆矩阵

Fig. 6 Confusion matrix

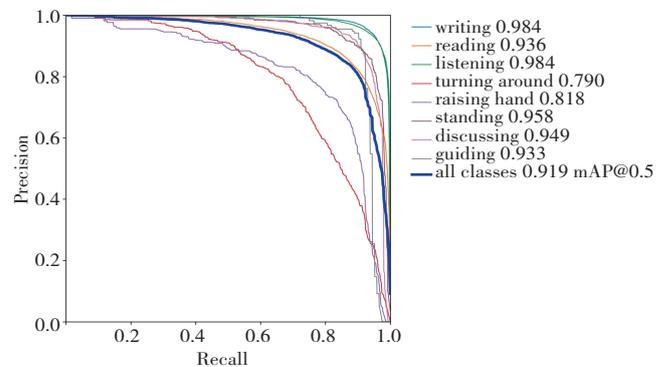


图 7 准确率-召回率曲线

Fig. 7 Comparison of accuracy-recall curves

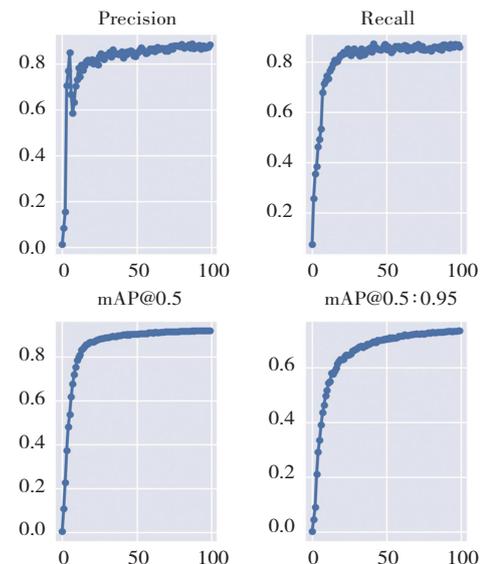


图 8 YOLOv7-SSW 训练过程(准确率、召回率以及 mAP)

Fig. 8 Training process of YOLOv7-SSW

2.4 课堂行为识别对比实验

为了验证本文算法的有效性, 将本文算法与基线算法 YOLOv7 进行课堂行为识别对比, 如图 9 所示。图 9 中第 1 列为输入图像, 第 2 列为 YOLOv7 检测结

果,第 3 列为本文算法识别结果。由图 9 第 1 行可知,本文算法可以准确识别右边第 2 列第 3 排学生为阅读(reading)状态,而 YOLOv7 识别错误。由图 9 第 2 行可知,图 9 中学生正处于相互讨论状态,本文算法准确识别了所有结果,而 YOLOv7 存在部分识别错误。由图 9 第 3 行可知,YOLOv7 错误地将教师指导(teacher guiding)识别为讨论(discussing),同时错误

地将中间第 4 排右边学生阅读(reading)状态识别为听课(listening)状态,而本文算法能准确识别。由图 9 第 4 行可知,YOLOv7 错误地将举手(raising hand)识别为听课(discussing),甚至是站立(standing),而本文算法准确识别了所有行为类别。因此,本文算法与 YOLOv7 相比识别准确率更高,能对学生课堂行为进行有效识别。



图 9 课堂行为识别效果对比

Fig. 9 Comparison of classroom behavior recognition effects

2.5 不同算法对比实验

将本文算法与其他课堂行为算法进行对比,结果如表 2 所示。由表 2 数据可知,本文算法的 mAP 与 CBPH-Net、YOLOv7 相比分别提升了 4.4%、5.6%,且本文算法单帧推理时间为 45.1 ms,在保持较高检测精度的同时保证了系统实时性。

表 2 本文算法与其他课堂行为识别方法对比

Table 2 Comparison of the proposed algorithm and other classroom behavior recognition methods

模型	mAP/%	单帧推理时间/ms
FasterR-NN ^[23]	79.9	64.1
TPHOLov5 ^[24]	86.4	32.3
YOLOX	86.7	42.8
CBPH-Net ^[16]	87.5	31.3
YOLOv7	86.3	36.6
YOLOv7-SSW	91.9	45.1

2.6 消融实验

为了进一步验证本文算法各个模块的有效性,将 SE、SPPFCSPC 以及 Wise-IoU 模块添加到基线 YOLOv7 模型进行消融实验。第 1 组为原始 YOLOv7 模型,第 2 组将 SE 模块添加到 YOLOv7 中,第 3 组将 SE 和 SPPFCSPC 模块同时添加到 YOLOv7 中,第 4 组将 SE、SPPFCSPC 以及 Wise-IoU 模块同时添加到 YOLOv7 中,实验结果如表 3 示。

表 3 消融实验对比

Table 3 Comparison of ablation experiments

组	SE	SPPFCSPC	Wise-IoU	mAP/%	单帧推理时间/ms
1	—	—	—	86.3	36.6
2	✓	—	—	89.2	52.5
3	✓	✓	—	89.7	43.7
4	✓	✓	✓	91.9	45.1

其中“√”表示使用了相应模块,“—”表示未使用该模块。由表 4 可知,第 2 组添加 SE 模块后,mAP 提升了 2.9%,说明 SE 模块通过建立通道相关性,增加了模型对信息的敏感性,但此时推理时间增加了 15.9 ms;第 3 组在第 2 组基础上添加 SPPFCSPC 模块后,虽然 mAP 仅提升了 0.5%,但是单帧推理时间降低了 8.8 ms,说明 SPPFCSPC 模块在增强特征提取能力的同时,加快了网络的速度;第 4 组在第 3 组基础上继续添加 Wise-IoU 损失函数,模型精度进一步得到提升。

为了进一步验证 SE 注意力机制的有效性,进行注意力机制热力图对比实验,如图 10 所示。图 10 中颜色深浅表示算法对该区域的关注程度,颜色越深,关注程度越高。由图 10 可知,添加 SE 注意力机制后模型对图像中学生所在区域更加关注,验证了算法的有效性。



(a) 输入图像 (b) 无 SE (c) 添加 SE

图 10 注意力机制对比

Fig. 10 Comparison of attention mechanisms

实验结果表明:YOLOv7-SSW 达到了精度与速度的均衡,实现了智慧课堂学生专注、消极以及互动等行为的准确识别,为推动精准思政教育体系提供了可行性的技术方案。

3 结 论

面向智慧课堂提出一种 YOLOv7-SSW 课堂行为识别算法。创新之处在于:通过添加 SE 注意力机制建立通道相关性,增加算法对信息的敏感性,提升算法检测精度;引入 SPPFCSPC 模型,降低原有算法的计算量,在保持算法准确性的同时提升算法实时性;引入 Wise-IoU 损失函数,增强算法处理训练数据中的低质量样本能力,进一步提升算法课堂行为识别的准确性。本文算法可部署在教室终端,赋能以下教育场景:构建多维度课堂评估体系,通过行为指标量化学生专注度与互动质量,为教学评价体系革新提供动态数据支持;依托细粒度个体特征分析引擎,形成个性化学习画像,支撑精准化分层教学实践;建立区域性课堂行为数据库,重

点监测欠发达地区教学特征差异,助力教育均衡发展。

鉴于学生课堂除本文所提阅读、书写等 8 种行为外,还存在使用手机、睡觉等行为,并且学生的肢体和表情特征十分丰富。因此,未来将针对高教场景建立多模态行为数据库,通过迁移学习提升模型在高等教育、职业教育等多元场景的泛化能力,并融合 VR/AR 技术开展教学,构建线上与线下、虚拟与现实协通联动的智教生态体系,推动形成“数据驱动-智能反馈-沉浸交互”的教育数字化转型闭环,助力教育强国建设。

参考文献(References):

- [1] 朱唯星, 陶磊. 人工智能赋能思政课: 教学样态、问题及治理路径[J]. 学校党建与思想教育, 2025(4): 57-60.
ZHU Wei-xing, TAO Lei. Artificial intelligence empowering ideological and political course: Teaching patterns, problems and governance paths[J]. The Party Building and Ideological Education in Schools, 2025(4): 57-60.
- [2] 袁磊, 徐济远, 刘沃奇. 数智教育生态下人机协同教学模式转型[J]. 开放教育研究, 2025, 31(2): 108-117.
YUAN Lei, XU Ji-yuan, LIU Wo-qi. Human-machine collaborative teaching paradigm transformation in digital-intelligent education ecosystem[J]. Open Education Research, 2025, 31(2): 108-117.
- [3] 卢致杰. 立足数智时代新起点推进高考综合改革新实践: 上海第八届教育考试与评价研讨会会议综述[J]. 中国考试, 2025(3): 105-110.
LU Zhi-jie. Standing at the new starting point of the digital intelligence era, advancing the comprehensive reform of Gaokao: A summary of the 8th seminar on education examination and evaluation[J]. Journal of China Examinations, 2025(4): 105-110.
- [4] 陈建新. 基于 YOLOv8-SCB 模型的思政课堂行为识别算法[J]. 山西能源学院学报, 2025, 38(1): 23-25.
CHEN Jian-xin. Identification algorithm of ideological and political classroom behavior based on YOLOv8-SCB model[J]. Journal of Shanxi Institute of Energy, 2025, 38(1): 23-25.
- [5] ZHAO J, ZHU H, NIU L. BiTNet: A lightweight object detection network for real-time classroom behavior recognition with transformer and bi-directional pyramid network[J]. Journal of King Saud University-Computer and Information Sciences, 2023, 35(8): 101670.
- [6] TANG L, XIE T, YANG Y, et al. Classroom behavior

- detection based on improved YOLOv5 algorithm combining multi-scale feature fusion and attention mechanism[J]. *Applied Sciences*, 2022, 12(13): 6790.
- [7] XIAO T, HE X, WU J. Student classroom behavior recognition and evaluation system based on YOLOX[C]// *Proceedings of the 2nd International Conference on Signal Image Processing and Communication*. SPIE, 2022: 100.
- [8] YANG F, WANG T, WANG X. Student classroom behavior detection based on YOLOv7+BRA and multi-model fusion[C]// *Image and Graphics*. Cham: Springer Nature Switzerland, 2023: 41-52.
- [9] WEI Y, LEI F, GAO J, et al. Student action recognition based on fuzzy broad learning system[C]// *Proceedings of the International Conference on Intelligent Education and Intelligent Research*. Piscataway: IEEE Press, 2023: 128-135.
- [10] WU B, WANG C, HUANG W, et al. Recognition of student classroom behaviors based on moving target detection[J]. *Traitement Du Signal*, 2021, 38(1): 215-220.
- [11] PANG C. Simulation of student classroom behavior recognition based on cluster analysis and random forest algorithm[J]. *Journal of Intelligent & Fuzzy Systems*, 2021, 40(2): 2421-2431.
- [12] SUZUKI S, AMEMIYA Y, SATO M. Enhancement of gross-motor action recognition for children by CNN with OpenPose[C]// *Proceedings of the IECON 2019-45th Annual Conference of the IEEE Industrial Electronics Society*. New York: ACM, 2019: 5382-5387.
- [13] 王泽杰, 沈超敏, 赵春, 等. 融合人体姿态估计和目标检测的学生课堂行为识别[J]. *华东师范大学学报(自然科学版)*, 2022(2): 55-66.
WANG Ze-jie, SHEN Chao-min, ZHAO Chun, et al. Recognition of classroom learning behaviors based on the fusion of human pose estimation and object detection[J]. *Journal of East China Normal University (Natural Science)*, 2022(2): 55-66.
- [14] JIA Q, HE J. Student behavior recognition in classroom based on deep learning[J]. *Applied Sciences*, 2024, 14(17): 7981.
- [15] SAMKARI E, ARIF M, ALGHAMDI M, et al. Human pose estimation using deep learning: A systematic literature review[J]. *Machine Learning and Knowledge Extraction*, 2023, 5(4): 1612-1659.
- [16] ZHAO J, ZHU H. CBPH-net: A small object detector for behavior recognition in classroom scenarios[J]. *IEEE Transactions on Instrumentation Measurement*, 2023, 72: 2521112.
- [17] 陈晨, 保文星, 陈旭, 等. 改进 YOLOv8 的学生课堂行为识别算法: DMS-YOLOv8[J]. *计算机工程与应用*, 2024, 60(24): 222-234.
CHEN Chen, BAO Wen-xing, CHEN Xu, et al. Improved YOLOv8-based algorithm for classroom behavior recognition of students: DMS-YOLOv8[J]. *Computer Engineering and Applications*, 2024, 60(24): 222-234.
- [18] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-excitation networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42(8): 2011-2023.
- [19] 王国明, 宋健. 基于改进的 YOLOv7 河道漂浮物小目标检测模型[J/OL]. *重庆工商大学学报(自然科学版)*. <https://link.cnki.net/urlid/50.1155.N.20240711.1424.002>.
WANG Guo-ming, SONG Jian. Small target detection model of floating objects in YOLOv7 river based on improvement[J/OL]. *Journal of Chongqing Technology and Business University (Natural Science Edition)*. <https://link.cnki.net/urlid/50.1155.N.20240711.1424.002>.
- [20] SHI B, ZHANG L, TANG J, et al. Research on underwater object detection algorithm based on YOLOv7[C]// *Proceedings of the Cross Strait Radio Science and Wireless Technology Conference*. Piscataway: IEEE Press, 2024: 1-6.
- [21] CHEN J, LIU H, ZHANG Y, et al. A multiscale lightweight and efficient model based on YOLOv7: Applied to citrus orchard[J]. *Plants*, 2022, 11(23): 3260.
- [22] HUSSAIN M, KHANAM R. In-depth review of YOLOv1 to YOLOv10 variants for enhanced photovoltaic defect detection[J]. *Solar*, 2024, 4(3): 351-386.
- [23] GAO C, YE S, TIAN H, et al. Multi-scale single-stage pose detection with adaptive sample training in the classroom scene[J]. *Knowledge-Based Systems*, 2021, 222: 107008.
- [24] ZHU X, LYU S, WANG X, et al. TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios[C]// *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*. Piscataway: IEEE Press, 2021: 2778-2788.

责任编辑:李翠薇