DOI：10.20271/j.cnki.1673-9833.2026.6008

面向学生课堂行为检测的算法优化研究

高雅文，黄贤明，梁爱南，余鸿翔，司鹏举

（湖南工业大学 计算机与人工智能学院，湖南 株洲 412007）

摘　要：针对学生课堂行为检测中后排遮挡和小目标检测难题，提出一种基于改进 YOLOv8 的深度学习优化算法。首先，将模型颈部的 C2f 模块替换为融合分组混洗卷积（GSConv）、坐标注意力机制（coordinate attention，CA）的改进结构，在降低模型复杂度的同时增强通道与位置信息感知；其次，引入空间深度转换卷积（SPD-Conv）模块缓解细粒度信息丢失问题；最后，在检测头中融合混合局部通道注意力（mixed local channel attention，MLCA）并增加小目标检测层，提升小目标捕获能力。实验表明，优化模型在 SCB-Dataset3 数据集上 *mAP*@0.5 和 *mAP*@0.5:0.95 分别提升 2.3% 和 3.6%，在 PASCAL VOC2012 上精确率提高 2.9%，验证了其有效性和泛化性。

关键词：学生课堂行为；智慧课堂；目标检测；深度学习

中图分类号：TP391　　**文献标志码**：A　　**文章编号**：1673-9833(2026)06-0056-08

引文格式：高雅文，黄贤明，梁爱南，等 . 面向学生课堂行为检测的算法优化研究 [J]. 湖南工业大学学报，2026，40(6)：56-63.

Research on Algorithm Optimization for Students' Classroom Performance Detection

GAO Yawen，HUANG Xianming，LIANG Ainan，YU Hongxiang，SI Pengju

（School of Computer Science and Artificial Intelligence，Hunan University of Technology，Zhuzhou Hunan 412007，China）

Abstract：In view of the challenges of rear-seat occlusion and small-object detection in student classroom performance monitoring, an improved deep learning optimization algorithm has been proposed based on YOLOv8. First, the C2f module in the model's neck is to be replaced with an improved structure incorporating group shuffle convolution (GSConv) and Coordinate Attention, thus reducing model complexity while enhancing channel and positional information perception. Secondly, an SPD-Conv module is introduced to alleviate the problem of fine-grained information loss. Finally, mixed local channel attention (MLCA) is integrated into the detection head, with a small-object detection layer incorporated to improve small target recognition performance. Experimental results demonstrate that the optimization of the model on the SCB-Dataset3 dataset *mAP*@0.5 and *mAP*@0.5:0.95 increases by 2.3% and 3.6% respectively, with an accuracy improvement of 2.9% on PASCAL VOC2012, verifying its effectiveness and generalization.

Keywords：students' classroom behavior；smart classroom；object detection；deep learning

收稿日期：2025-04-30

基金项目：教育部中国高校产学研创新基金资助项目（2023DT002）

作者简介：高雅文，女，湖南工业大学硕士生，主要研究方向为计算机视觉，E-mail：yawen_gao2001@163.com

通信作者：黄贤明，男，湖南工业大学副教授，主要研究方向为大数据技术，知识图谱，教育信息化，E-mail：841619296@qq.com

1　背景知识

随着教育信息化的推进，智慧教育成为当代教育信息化的新境界、新诉求[1]。传统的教学模式中，教师难以全面、实时地关注每个学生的学习状态和行为。智慧课堂作为智慧教育的重要组成部分，借助技术手段优化教学过程，可有效提升教学效果。其中，学生课堂行为检测技术能够帮助教师精准了解学生的课堂表现，调整教学策略，提高教学成效。

基于深度学习的目标检测算法极大地推动了计算机视觉领域的发展。2014 年，R. Girshick 等[2]提出了区域的卷积神经网络（region-based convolutional neural network，R-CNN）算法，首次将深度学习引入目标检测领域。尽管该算法提高了目标检测的精度，但计算量较大、速度较慢。针对 R-CNN 显现的弊端，R. Girshick[3] 提出了 Fast R-CNN 算法。同年，Ren S. Q. 等[4]提出了 Faster R-CNN 算法。随后，单阶段目标检测算法兴起，代表性算法有 YOLO（You Only Look Once）[5−6] 系列、SSD（Single Shot Multibox Detector）[7] 等。近年来，Transformer[8] 架构在自然语言处理领域取得了巨大成功，并逐渐被应用于计算机视觉领域。因此涌现出基于 Transformer 的目标检测算法。与基于卷积神经网络（CNN）的检测算法不同，基于 Transformer 的检测算法将目标检测任务视为序列到序列的预测问题，并通过引入自注意力机制推理目标与全局图像上下文关系，如 DETR（detection transformer）[9]。

目标检测作为深度学习的典型算法之一，具有优异的目标识别性能，使其成为学生课堂行为检测的首选方法。众多学者对此进行了大量研究，白捷等[10]提出了基于多路特征融合的 Faster R-CNN 模型，通过融合不同层次的特征信息，同时利用迁移学习在数据集上实现了较高的检测性能，减少了数据依赖。但是由于数据集规模较小，在一定程度上会影响模型的泛化能力。Wang Z. F. 等[11]提出了一种全新的融合检测模型，其将 DETR、Swin Transformer 和轻量级特征金字塔网络（FPN）有效结合，在私人数据集中取得优异性能，为相关研究提供了宝贵的资源。然而在模型复杂度和训练时间上仍有进步空间。Zhao J. H. 等[12]提出一种名为 CBPH-Net 的检测器，利用 CBPH 检测头、同构多分支 FEM 以及椭圆边界框实现了有效应用于课堂行为的识别任务。尽管 CBPH-Net 在特定数据集上取得了良好结果，但仍需要在更多样化的课堂环境中进行进一步验证。

除此之外，学生课堂行为检测还存在一些其他方面的问题。例如教室后排的学生在图像中常表现为小目标，这些学生图像的有效信息较少，给特征提取带来了困难，容易导致漏检或误检[13]；教室范围内学生分布密集，每个学生的行为可能各不相同，这给目标框的定位和行为分类带来了极大的挑战。此外，同种课堂行为在不同学生之间可能表现出多种形态，增加了行为定义难度，导致检测准确率的降低和检测效率的下降。因此，为提升学生课堂行为检测的准确率和检测结果可靠性，本文提出 GSCA 模块，将 C2f 模块中的 Bottleneck 的标准卷积替换为 GSConv，并加入 CA 注意力机制组成 GSCA。该模型能够在保持较低计算复杂度的同时，关注表现出不同形态的同种行为，保留更多的特征信息，从而提升特征表达能力。其次，引入 SPD-Conv 模块，以有效减少在处理低分辨率图像和小物体时的细粒度信息丢失问题。最后设计 SD_MLCA 模块并增加小目标检测层，提升对小目标的检测能力。

2　相关研究工作

2.1　基线模型 YOLOv8n

YOLOv8 的网络结构由骨干网络（backbone）、颈部结构（neck）和检测头（head）3 部分构成。骨干网络的主要作用是提取图像特征，其中的 C2f 模块借鉴了 YOLOv7 中高效层聚合网络（ELAN）[14] 思想，通过并行更多的梯度流分支，让 YOLOv8 可以在保证轻量化的同时获得更加丰富的梯度流信息。在骨干网络末尾使用空间金字塔池化快速（SPPF）模块，增强特征提取能力。颈部采用 PAN-FPN[15−16] 结构通过自底向上和自顶向下的路径，将不同尺度的特征图进行融合。检测头部分与 YOLOv5 不同，采用解耦头（decoupled head）[17] 结构同时采用无锚框（anchor-free）策略，进一步简化模型结构。

YOLOv8 是 Ultralytics 公司经过多代优化的成熟模型，性能均衡。无锚点设计与灵活骨干网使其在小目标检测方面表现出色，而 YOLOv8n 作为轻量级版本，更适合学生课堂行为检测需求。因此本研究选择 YOLOv8n 作为基线模型。

2.2　优化模型 GCSM

如图 1 所示，GCSM 在 YOLOv8n 的基础上进行了优化。由于实际教育场景中，学生分布密集，存在学生相互遮挡以及同种学生行为可能存在多种表现形态的问题，因此提出以 GSCA 模块替代 C2f 模块，模型能够在保持较低计算复杂度的同时关注表现出不同形态的同种行为，提高检测的准确率。随后引入 SPD-Conv 模块，通过将特征图的空间维度转换为深

度维度，增强了特征表示，弥补低分辨率图像和小物体图像信息不足的缺陷。最后设计 SD_MLCA 检测头并增加一个小目标检测层，提升小目标检测能力。

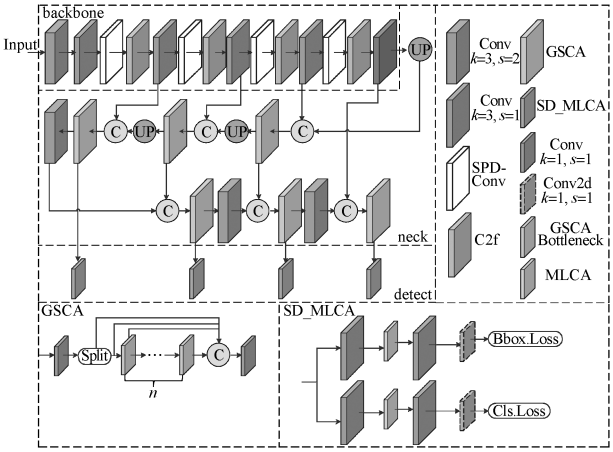


图 1 GCSM 模型结构

Fig. 1 Structure of the GCSM model

2.3 GSCA 模块

YOLOv8n 中的 C2f 模块是用于特征提取和融合的关键组件，该模块在保留 YOLOv5 中 C3 模块基础结构的同时，整合了 YOLOv7 的 ELAN 模块设计理念，主要增加了并行梯度传播路径的数量，使得网络能够捕获更丰富的特征表示。而 Bottleneck 模块是 C2f 模块的核心组成部分，通过一系列卷积、归一化和激活函数对输入特征图进行处理。本文提出的 GSCA 就是针对其中的 Bottleneck 模块进行改进得到的。将原 Bottleneck 模块中的标准卷积替换成更加轻量级的 GSConv，并引入了能将位置信息嵌入到通道注意力中的坐标注意力机制，得到 GSCABottleneck，如图 2 所示。

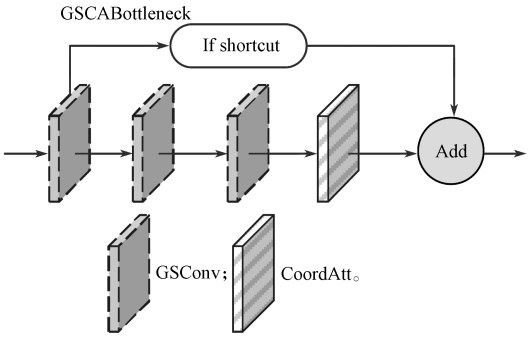


图 2 GSCABottleneck 结构图

Fig. 2 Structure of GSCABottleneck

本文将 GSCA 模块用于替换颈部的 C2f 模块，主要原因是如果在模型的所有阶段都使用 GSConv，模型的网络层会加深，这可能会加剧对数据流的阻力，显著增加推理时间。而在 neck 阶段使用 GSConv，则可以在不显著增加推理时间的情况下，提升模型的性能。在 neck 阶段，特征图已经被足够地压缩，即通道维度达到最大，而宽度和高度维度减至最小，且转换变得适度。此时，利用 GSConv 处理拼接的特征图恰到好处，可以减少冗余重复信息，无需进一步压缩。

2.3.1 轻量级卷积技术 GSConv

GSConv 是一种结合了分组卷积（grouped convolution）、深度可分离卷积（depthwise separable convolution，DWConv）和混洗（shuffle）机制的轻量级卷积技术 [18]。首先将输入特征图分成多个组，对每个组分别进行独立的卷积操作。其次在分组卷积的基础上，深度可分离卷积将卷积操作分解为深度卷积（depthwise convolution）和逐点卷积（pointwise convolution）。深度卷积对每个输入通道独立进行卷积操作，而逐点卷积则用于整合深度卷积的输出。经过分组卷积和深度可分离卷积操作后的特征图被连接起来。针对分组卷积操作可能引发的特征通道间信息交互受限的问题，GSConv 引入了 shuffle 机制，该机制通过对卷积输出的特征张量实施通道维度的置换操作，建立跨组间的特征连接，确保不同分组间的信息能够充分交互。GSConv 结构如图 3 所示。

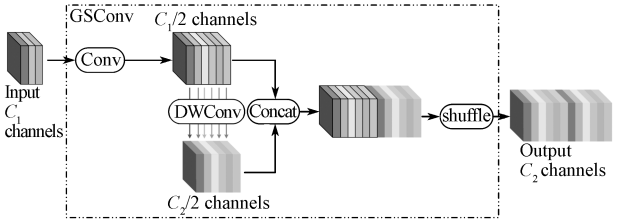


图 3 GSConv 结构图

Fig. 3 Structure of GSConv

2.3.2 坐标注意力机制

坐标注意力机制 [19] 是一种将位置信息嵌入通道注意力中的注意力机制。压缩与激励注意力机制（squeeze-and-excitation，SE）[20] 只考虑编码通道间信息而忽略了位置信息，而卷积块注意力机制（convolutional block attention module，CBAM）[21] 通过引入大尺度的卷积核提取空间特征却未能有效解决长程依赖问题。而 CA 注意力机制不仅可以捕获跨通道信息，还可以捕获方向感知和位置敏感信息，有助于模型更准确地定位和识别感兴趣的对象。

坐标注意力通过坐标信息嵌入和坐标注意力生成两个步骤同时编码具有精确位置信息的通道关系和长距离依赖关系，它的具体结构见图 4。对于输入特征图 $\boldsymbol{X}$，设其高度为 H、宽度为 W，先使用尺寸 $(H, 1)$ 和 $(1, W)$ 的池化核分别沿着水平坐标方向和竖直坐标方向对每个通道进行编码，以使注意力能够捕获一个空间方向上的长距离依赖关系，并保留另一个空间方向上的精确位置信息。因此，高度为 h 的第 c 个通道

和宽度方向坐标为 w 的第 c 个通道的输出表述分别见式（1）和（2）。

$$z_c^h(h)=\frac{1}{W}\sum_{0\leqslant i\leqslant W}x_c(h,i)\,, \tag{1}$$

$$z_c^w(w)=\frac{1}{H}\sum_{0\leqslant j\leqslant H}x_c(j,w)\,。 \tag{2}$$

式（1）（2）中，x_c 为第 c 个通道的输入特征图。

为了有效利用坐标信息嵌入所提取的全局感受野及精确位置特征，后续需执行注意力权重的生成。具体而言，首先将坐标嵌入模块输出的两个方向的特征张量沿空间维度拼接，随后通过一个共享的 1×1 卷积层进行特征变换，记为 F_1，其计算过程如式（3）所示。

$$\boldsymbol{f}=\delta(F_1([\boldsymbol{z}^h,\boldsymbol{z}^w]))\,。 \tag{3}$$

式中：符号 [,] 为沿空间维度的串联操作；δ 为非线性激活函数；$\boldsymbol{f}\in\mathbb{R}^{C/r\times(H+W)}$ 为水平与垂直方向的中间特征表示，其中，参数 r 用于调节特征通道的压缩比率；$\boldsymbol{z}^h$和$\boldsymbol{z}^w$为式（1）和式（2）输出的特征张量。

坐标注意力机制的处理流程中，特征张量 $\boldsymbol{f}$ 沿空间维度被分解为两个独立的分量 $\boldsymbol{f}^h\in\mathbb{R}^{C/r\times H}$ 和 $\boldsymbol{f}^w\in\mathbb{R}^{C/r\times W}$。通过两个独立的 1×1 卷积变换 F_h 和 F_w 分别对这两个分量进行处理，使其通道维度与原始输入特征 $\boldsymbol{X}$ 保持一致。这一变换过程可由式（4）和式（5）表示，其中 σ 代表 Sigmoid 激活函数，用于对输出特征进行非线性归一化处理。

$$\boldsymbol{g}^h=\sigma(F_h(\boldsymbol{f}^h))\,, \tag{4}$$

$$\boldsymbol{g}^w=\sigma(F_w(\boldsymbol{f}^w))\,。 \tag{5}$$

最后将作为注意力权重的 $\boldsymbol{g}^h$ 和 $\boldsymbol{g}^w$ 进行扩展得到坐标注意力，具体见式（6）。

$$y_c(i,j)=x_c(i,j)\times g_c^h(i)\times g_c^w(j)\,。 \tag{6}$$

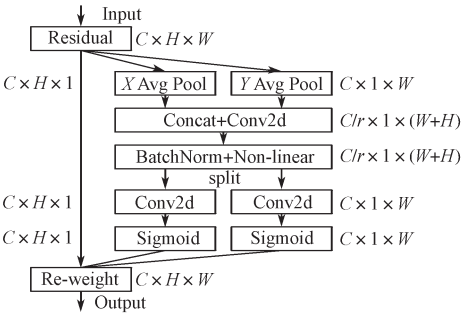


图 4　Coordinate Attention 结构图

Fig. 4　Structure of Coordinate Attention

2.4　SPD-Conv 模块

在当前的卷积神经网络架构设计中，存在一个普遍存在且待优化的缺陷：由于跨步卷积或池化层的应用，容易出现细粒度信息的丢失和有效特征表示学习不到位的问题。而新的 CNN 构建 SPD-Conv 模块 [22] 可以代替每个跨步卷积层和每个池化层，在保留更多空间细节信息的同时，提升对小目标的检测能力。

SPD-Conv 包含一个空间到深度（SPD）层，该层通过重排特征图的元素，将空间信息转换到深度通道维度，从而实现下采样而不损失信息。这种方法保留了通道维度中的所有信息，避免了传统下采样方法中的信息丢失。在 SPD 层之后，SPD-Conv 使用非步长，即步长为 1 的卷积层进一步处理特征图。这种设计允许网络在不丢失空间信息的前提下，通过可学习的参数精细调整特征表示，有助于提高模型对小物体和低分辨率图像的处理能力。图 5 是 *scale*=2 时 SPD-Conv 的结构。首先从大小为 $S\times S\times C_1$ 的中间特征图 $\boldsymbol{X}$ 开始，使用 *scale* 因子进行下采样，将 $\boldsymbol{X}$ 划分为 $scale^2$ 个子特征图。每个子特征图是 $\boldsymbol{X}$ 的一部分，其中 $\boldsymbol{X}(i,j)$ 中的索引 i 和 j 被 *scale* 整除。经过尺度变换处理后的多个特征子图按照通道轴向进行拼接组合，生成的特征图 $\boldsymbol{X}_0$ 在空间维度上缩减为输入特征的 1/*scale*，通道数量则扩展至原始输入的 $scale^2$ 倍。接着对 $\boldsymbol{X}_0$ 应用 C_2 滤波器添加一个步长为 1 的卷积层，其中 C_2 少于 $\boldsymbol{X}_0$ 的通道数 $scale^2C_1$。结果得到特征图 $\boldsymbol{X}_{00}$，其空间维度保持不变，但通道维度减少到 C_2。

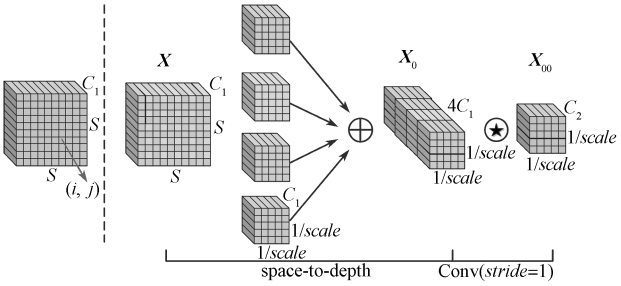


图 5　*scale*=2 时 SPD-Conv 的结构

Fig. 5　Structure of SPD-Conv with *scale*=2

2.5　轻量化检测头 SD_MLCA

YOLOv8n 的 head 部分是目标检测的最后阶段，负责将网络学习到的特征转换为具体的检测结果。为了帮助模型更高效地融合不同层次的特征，且保持较低的参数量，课题组引入混合局部通道注意力构建 SD_MLCA 模块，并增加了 P2 层即增加 160×160 的检测特征图，用于检测 4×4 像素以上的目标。

混合局部通道注意力通过综合考虑通道、空间、局部和全局信息来增强特征的表达能力 [23]。图 6 为 MLCA 结构图。MLCA 通过局部平均池化（LAP）和全局平均池化（GAP）分别提取特征图的局部空间细节和全局上下文信息。接着，利用一维卷积对这些特征进行转换和压缩，再通过反池化（UNAP）和加法操作将局部与全局信息融合，最后通过乘法操作

对原始特征进行加权选择，从而聚合多尺度的特征信息，增强特征的表达能力，提升模型在教室场景中的识别精度。

而 P2 层位于网络的较浅层，能够捕捉到更多的细粒度特征，P2 层的加入也进一步强化了模型的多尺度信息融合能力。

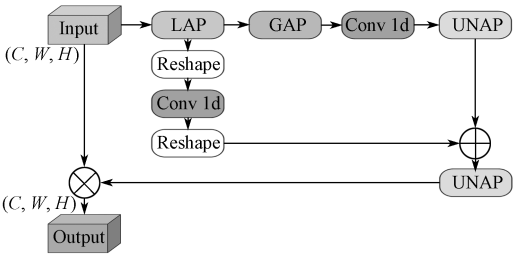


图 6 MLCA 结构图

Fig. 6 Structure of MLCA

3 实验结果与分析

3.1 数据集

本文使用的数据集是 SCB-Dataset3[24]，该数据集涵盖了从幼儿园到大学的不同场景，包括 5 686 张图像，45 578 个标签，一共有 6 种行为，即举手、阅读、书写、使用手机、低头和俯卧在桌子上。其中举手行为多出现在幼儿园、初中、高中场景，使用手机和俯卧在桌上多出现在大学场景。图 7 中绿色虚线框表示“举手”样本，红色虚线框表示“阅读”样本，白色虚线框表示“书写”样本，蓝色虚线框表示“使用手机”样本，黄色虚线框表示“低头”样本，紫色虚线框表示“俯卧在桌上”样本。



图 7 数据集图像展示

Fig. 7 Dataset image presentation

为验证GCSM模型在非特定场景下的泛化能力，选取 PASCAL VOC2012 数据集 [25] 作为泛化实验数据集。该数据集是计算机视觉领域极具代表性的通用目标检测基准数据集，被广泛应用于模型性能评估、跨域迁移能力验证，涵盖日常生活中 20 类常见目标，包含行人、车辆、动物、家具及日常物品等典型检测目标，数据集包含约 11 540 张标注图像，样本场景丰富、标注规范，能够有效反映模型在非课堂专用数据下的检测迁移能力。

3.2 实验环境

本文的实验均使用以下环境，CPU 为 15 vCPU AMD EPYC 7642 48-Core Processor，GPU 为 NVIDIA Geforce RTX 3090 24 GB，深度学习环境为 PyTorch 1.9.0，Cuda 11.1，编程语言为 Python 3.8，开发工具为 VSCode。在训练过程中，输入图像的分辨率为 640 × 640 像素，批量大小设为 32，模型训练轮次为 200。模型使用默认的参数设置，即初始学习率为 0.01，动量为 0.937，权重衰减率为 0.000 5，随机种子为 0。

3.3 实验评价指标

为了客观验证学生课堂行为检测优化模型的性能，课题组选用精确率 *P*、召回率 *R*、平均精度 *mAP*、参数量 *Parameters* 和十亿次浮点运算量 *GFLOPs* 作为评价指标。

3.4 消融实验

为了验证各个模块对学生课堂行为检测优化模型的性能影响，并进一步评估各个模块的贡献，以 YOLOv8n 为基线模型，利用 GSCA、SPD-Conv、SD_MLCA 构建 GCSM。消融实验结果见表 1。

表 1 消融实验结果

Table 1 Results of ablation experiments %

实验	*P*	*R*	*mAP*@0.5	*mAP*@0.5:0.95
Baseline	80.6	77.9	83.9	63.3
+ GSCA	80.3	78.9	84.4	62.6
+ SPD-Conv	82.4	78.6	84.9	63.7
+ SD_MLCA	82.9	81.4	86.2	66.9

表 1 中，实验 1 为 YOLOv8n 基线模型实验，实验 2 在 YOLOv8n 基础上加入 GSCA 模块，结果表明召回率提高了 1%，这表示 GSCA 模块的加入增强了模型感知目标的能力，减少了漏检率。在第 3 组实验中，再加入 SPD-Conv 模块，相较于实验 2，尽管召回率有小幅度的下降，但是 *P*、*mAP*@0.5、*mAP*@0.5:0.95 均有一定程度的上升，其中精确率 *P* 较基线模型上升了 1.8%，表明该模块的加入有效抑制了误检。最后，引入 SD_MLCA 模块，至此本文提出的优化模型的 *P* 达到 82.9%，相比于基线模型提升了 2.3%，召回率 *R* 达到 81.4%，提升了 3.5%，*mAP*@0.5 达到 86.2%，提升了 2.3%，*mAP*@0.5:0.95 达到 66.9%，提升了 3.6%。

3.5 性能分析

3.5.1 在 SCB-Dataset3 上的对比实验

为了客观地验证优化模型对学生课堂行为检测的有效性，课题组选取 EfficientDet[26]、YOLOv5、YOLOv7、RT-DETR-l[27]、YOLOv10[28] 等优秀模型进行对比实验，分别在 SCB-Dataset3 数据集上进行训练。实验结果如表 2 所示。

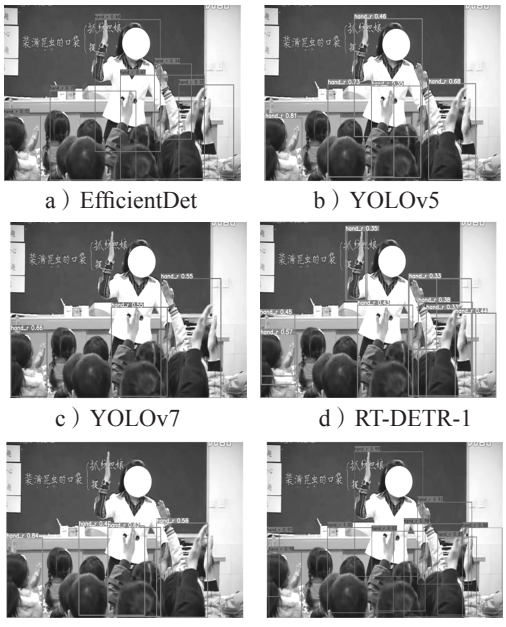
表 2　对比实验结果

Table 2　Comparison of experimental results

模型	*P*/%	*R*/%	*mAP*@0.5/%	*mAP*@0.5:0.95/%	*Parameters*/10^6	*GFLOPs*
EfficientDet	78.9	60.4	75.0	59.1	3.8	4.8
YOLOv5	83.5	76.7	82.7	57.8	1.8	4.2
YOLOv7	78.6	80.6	84.1	62.1	37.2	105.2
YOLOv8n	80.6	77.9	83.9	63.3	3.0	8.2
RT-DETR-l	76.9	72.8	76.0	56.2	32.0	103.5
YOLOv10	80.7	78.8	84.3	62.9	2.6	7.8
GCSM	82.9	81.4	86.2	66.9	2.8	15.1

与其他模型相比较，不同目标检测模型呈现了各自的优势和局限性。EfficientDet 和 YOLOv5 尽管具有较低的参数量和 GFLOPs，但是其他各项评价指标均与其他模型存在一定差距；而 YOLOv7 以及 RT-DETR-1 则因模型较为复杂，相比其他算法，需要更多的计算资源来运行，且训练时间较长。即使与 YOLOv10 相比，本文提出的模型也具有一定的优越性。因此得出结论，本文的优化模型综合检测性能最好，该模型的 *mAP*@0.5、*mAP*@0.5:0.95 和召回率 *R* 最高；参数量相较于 YOLOv8n 模型减少了约 6.7%，计算复杂度保持在可接受的范围内。

从课堂行为数据集中随机选取 2 张典型场景图片，其中含前排清晰目标与后排小目标各 1 张。为了确保检测结果的准确性和可靠性，选择所有模型在训练过程中表现最佳的权重，并在统一测试条件下对图片进行检测。图 8 和图 9 分别为各模型在 SCB-Dataset3 数据集上的学生课堂行为检测结果图。



a）EfficientDet　b）YOLOv5

c）YOLOv7　d）RT-DETR-1

e）YOLOv10　f）Ours

图 8　各模型对前排清晰目标图片的预测结果

Fig. 8　Prediction results of each model for front-row clear target images



a）EfficientDet　b）YOLOv5

c）YOLOv7　d）RT-DETR-1

e）YOLOv10　f）Ours

图 9　各模型对后排小目标图片的预测结果

Fig. 9　Prediction results of each model for back-row small target images

由图 8 和 9 可以看出本文算法能更全面正确地检测出前排学生课堂行为，也能正确地检测出后排学生行为，基本不存在误检漏检。

3.5.2　在 PASCAL VOC2012 上的对比实验

为全面评估本文算法的泛化能力，在 PASCAL VOC2012 数据集上进行对比实验。与专注于实际教学场景的 SCB-Dataset3 数据集不同，PASCAL VOC2012 数据集包含更具多样性的日常物体，图像具有更显著的尺度变化和更频繁的遮挡情况，以及更复杂的背景干扰。因此，在本次实验中选择使用该数据集。实验结果见表 3。

表 3　在 PASCAL VOC2012 数据集上的实验结果

Table 3　Experimental results on the PASCAL VOC2012 dataset

%

模型	*P*	*R*	*mAP*@0.5	*mAP*@0.5:0.95
YOLOv8n	68.8	54.6	60.3	41.6
GCSM	71.7	56.0	62.5	42.9

从表 3 中可知，本文算法的召回率、精确率、*mAP*@0.5 和 *mAP*@0.5:0.95 与 YOLOv8n 相比均有所提高，精确率更是达到了 71.7%，相比 YOLOv8n 提高了 2.9%，说明本文算法能有效降低误检率。

4 结语

针对学生课堂行为检测中存在的教室后排学生行为容易漏检或误检以及教室中学生分布密集导致检测难度增加的问题，本文提出一种面向学生课堂行为检测的优化算法。将模型颈部的 C2f 模块替换成 GSCA 模块，保留更多特征信息，提升特征表达能力；针对后排学生的课堂行为像素不足以及后排学生相对于整个图像而言较小，引入 SPD-Conv 模块和 SD_MLCA 模块，有效减少在处理低分辨率图像和小物体时的细粒度信息丢失问题，提升小目标检测能力。在 SCB-Dataset3 数据集上的实验结果表明，GCSM 模型相较于基线模型，精确率 *P* 提升了 2.3%，召回率 *R* 提升了 3.5%，*mAP*@0.5 提升了 2.3%，*mAP*@0.5:0.95 提升了 3.6%。在主流目标检测模型的对比实验中，也验证了 GCSM 模型的有效性。在 PASCAL VOC2012 数据集上，GCSM 模型同样达到了较高的检测精度，表明该模型具较高泛化性能。

参考文献：

[1] 廖 剑，刘 选，刘革平 . 理感联通：人工智能赋能智慧教育新范式 [J]. 中国电化教育，2023(6)：18−24.
LIAO Jian，LIU Xuan，LIU Geping. Rational-Perceptual Interconnectionism：The New Paradigm of Artifical Intelligence in Education[J]. China Educational Technology，2023(6)：18−24.

[2] GIRSHICK R，DONAHUE J，DARRELL T，et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus：IEEE，2014：580−587.

[3] GIRSHICK R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision (ICCV). Santiago：IEEE，2016：1440−1448.

[4] REN S Q，HE K M，GIRSHICK R，et al. Faster R-CNN：Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2017，39(6)：1137−1149.

[5] REDMON J，DIVVALA S，GIRSHICK R，et al. You Only Look Once：Unified，Real-Time Object Detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas：IEEE，2016：779−788.

[6] 王 宁，智 敏 . 深度学习下的单阶段通用目标检测算法研究综述 [J]. 计算机科学与探索，2025，19(5)：1115−1140.
WANG Ning，ZHI Min. Review of One-Stage Universal Object Detection Algorithms in Deep Learning[J]. Journal of Frontiers of Computer Science and Technology，2025，19(5)：1115−1140.

[7] LIU W，ANGUELOV D，ERHAN D，et al. SSD：Single Shot MultiBox Detector[C]//Computer Vision：ECCV 2016. Cham：Springer，2016：21−37.

[8] VASWANI A，SHAZEER N，PARMAR N，et al. Attention Is All You Need[C]// Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach：NIPS，2017：6000−6010.

[9] CARION N，MASSA F，SYNNAEVE G，et al. End-to-End Object Detection with Transformers[C]//Computer Vision：ECCV 2020. Cham：Springer，2020：213−229.

[10] 白 捷，高海力，王永众，等 . 基于多路特征融合的 Faster R-CNN 与迁移学习的学生课堂行为检测 [J]. 广西师范大学学报 (自然科学版)，2020，38(5)：1−11.
BAI Jie，GAO Haili，WANG Yongzhong，et al. Detection of Students’ Classroom Performance Based on Faster R-CNN and Transfer Learning with Multi-Channel Feature Fusion[J]. Journal of Guangxi Normal University (Natural Science Edition)，2020，38(5)：1−11.

[11] WANG Z F，YAO J L，ZENG C Y，et al. Students’ Classroom Behavior Detection System Incorporating Deformable DETR with Swin Transformer and Light-Weight Feature Pyramid Network[J]. Systems，2023，11(7)：372.

[12] ZHAO J H，ZHU H Y. CBPH-Net：A Small Object Detector for Behavior Recognition in Classroom Scenarios[J]. IEEE Transactions on Instrumentation and Measurement，2023，72：2521112.

[13] LIU Q T，JIANG R Y，XU Q，et al. YOLOv8n_BT：Research on Classroom Learning Behavior Recognition Algorithm Based on Improved YOLOv8n[J]. IEEE Access，2024，12：36391−36403.

[14] WANG C Y，BOCHKOVSKIY A，LIAO H M. YOLOv7：Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver：IEEE，2023：7464−7475.

[15] WANG W H，XIE E Z，SONG X G，et al. Efficient and Accurate Arbitrary-Shaped Text Detection with Pixel

Aggregation Network[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul：IEEE，2020：8439−8448.

[16] LIU S，QI L，QIN H F，et al. Path Aggregation Network for Instance Segmentation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City：IEEE，2018：8759−8768.

[17] GE Z，LIU S T，WANG F，et al. YOLOX：Exceeding YOLO Series in 2021[EB/OL].（2021−07−18）. 2021：arXiv：2107.08430. https://arxiv.org/abs/2107.08430.

[18] LI H L，LI J，WEI H B，et al. Slim-Neck by GSConv：A Lightweight-Design for Real-Time Detector Architectures[J]. Journal of Real-Time Image Processing，2024，21(3)：62.

[19] HOU Q B，ZHOU D Q，FENG J S. Coordinate Attention for Efficient Mobile Network Design[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville：IEEE，2021：13708−13717.

[20] HU J，SHEN L，SUN G. Squeeze-and-Excitation Networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City：IEEE，2018：7132−7141.

[21] WOO S，PARK J，LEE J Y，et al. CBAM：Convolutional Block Attention Module[C]//Computer Vision：ECCV 2018. Cham：Springer，2018：3−19.

[22] SUNKARA R，LUO T. No More Strided Convolutions or Pooling：A New CNN Building Block For Low-Resolution Images and Small Objects[C]//Machine Learning and Knowledge Discovery in Databases. Cham：Springer，2023：443−459.

[23] 唐梦雨，曾志林，文　勇 . 基于改进 YOLOv8 算法的垃圾分类识别方法研究 [J]. 环境工程，2025(4)：110−120.
TANG Mengyu，ZENG Zhilin，WEN Yong. An Improved Garbage Classification Detection Algorithm Based on YOLOv8[J]. Environmental Engineering，2025(4)：110−120.

[24] YANG F，WANG T. SCB-Dataset3：A Benchmark for Detecting Student Classroom Behavior[EB/OL].（2023−10−04）. 2023：arXiv：2310.02522. https://arxiv.org/abs/2310.02522.

[25] EVERINGHAM M，ALI ESLAMI S M，VAN GOOL L，et al. The Pascal Visual Object Classes Challenge：A Retrospective[J]. International Journal of Computer Vision，2015，111(1)：98−136.

[26] TAN M X，PANG R M，LE Q V. EfficientDet：Scalable and Efficient Object Detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle：IEEE，2020：10778−10787.

[27] ZHAO Y A，LÜ W Y，XU S L，et al. DETRs Beat YOLOs on Real-Time Object Detection[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle：IEEE，2024：16965−16974.

[28] WANG A，CHEN H，LIU L H，et al. YOLOv10：Real-Time End-to-End Object Detection[EB/OL].（2024−05−23）. 2024：arXiv：2405.14458. https://arxiv.org/abs/2405.14458.

（责任编辑：申　剑）