



课堂密集场景类内感知半监督学生行为检测

张睿, 张延春, 张佳琪, 白峭峰, 樊光瑞, 寇旭鹏

引用本文:

张睿, 张延春, 张佳琪, 等. 课堂密集场景类内感知半监督学生行为检测[J]. 智能系统学报, 2026, 21(4): 1013-1029.

ZHANG Rui, ZHANG Yanchun, ZHANG Jiaqi, et al. Class-wise distribution-aware semi-supervised student behavior detection in crowded classrooms[J]. *CAAI Transactions on Intelligent Systems*, 2026, 21(4): 1013-1029.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202510035>

您可能感兴趣的其他文章

基于改进的Faster RCNN面部表情检测算法

Facial expression recognition based on improved Faster RCNN

智能系统学报. 2021, 16(2): 210-217 <https://dx.doi.org/10.11992/tis.201910020>

多感知兴趣区域特征融合的图像识别方法

Image recognition method based on multi-perceptual interest region feature fusion

智能系统学报. 2021, 16(2): 263-270 <https://dx.doi.org/10.11992/tis.201906032>

基于注意力机制的显著性目标检测方法

Salient object detection method based on the attention mechanism

智能系统学报. 2020, 15(5): 956-963 <https://dx.doi.org/10.11992/tis.201903001>

多标记学习自编码网络无监督维数约简

Unsupervised dimensionality reduction of multi-label learning via autoencoder networks

智能系统学报. 2018, 13(5): 808-817 <https://dx.doi.org/10.11992/tis.201804051>

高斯核函数卷积神经网络跟踪算法

Convolutional neural network tracking algorithm accelerated by Gaussian kernel function

智能系统学报. 2018, 13(3): 388-394 <https://dx.doi.org/10.11992/tis.201612040>

行人重识别研究综述

Survey on pedestrian re-identification research

智能系统学报. 2017, 12(6): 770-780 <https://dx.doi.org/10.11992/tis.201706084>

DOI: 10.11992/tis.202510035

网络出版地址: <https://link.cnki.net/urlid/23.1538.TP.20260302.1410.002>

课堂密集场景类内感知半监督学生行为检测

张睿¹, 张延春¹, 张佳琪¹, 白峭峰², 樊光瑞¹, 寇旭鹏¹

(1. 太原科技大学 计算机科学与技术学院, 山西 太原 030024; 2. 太原科技大学 机械工程学院, 山西 太原 030024)

摘要: 针对课堂学生行为检测中座位密集、个体像素占比低且差异大, 类间分布不均衡导致的漏检、误检问题以及高质量标签数据获取困难的问题, 提出课堂密集场景类内感知半监督学生行为检测方法。构建半监督学生行为检测模型, 设计双流交互融合模块和颗粒化多层次感知模块, 通过挖掘行为个体特征及交互特征, 缓解行为交互信息淹没问题; 提出动态类间分布损失与检测头, 优化因高频标签类别导致的网络偏向性; 针对伪标签置信度分布不一致问题, 构建类内分布特征感知的半监督学习策略, 通过动态划分伪标签筛选阈值, 抑制噪声样本。通过自建数据集实验, 在 10% 的标注条件下, 学生行为检测模型的 mAP50 相较 YOLOv11m 提升 3.2%; 在引入半监督策略后, 进一步提升 5.0%; 并进一步通过公共数据集验证了方法的有效性与泛化能力。结论可为智慧课堂场景下的学生行为识别、课堂学情智能分析提供有力的技术支撑。

关键词: 深度学习; 计算机视觉; 半监督学习; 目标检测; 神经网络; 特征提取; 伪标签; 动态阈值

中图分类号: TP391.4 **文献标志码:** A **文章编号:** 1673-4785(2026)04-1013-17

中文引用格式: 张睿, 张延春, 张佳琪, 等. 课堂密集场景类内感知半监督学生行为检测 [J]. 智能系统学报, 2026, 21(4): 1013-1029.

英文引用格式: ZHANG Rui, ZHANG Yanchun, ZHANG Jiaqi, et al. Class-wise distribution-aware semi-supervised student behavior detection in crowded classrooms[J]. CAAI transactions on intelligent systems, 2026, 21(4): 1013-1029.

Class-wise distribution-aware semi-supervised student behavior detection in crowded classrooms

ZHANG Rui¹, ZHANG Yanchun¹, ZHANG Jiaqi¹, BAI Qiaofeng², FAN Guangrui¹, KOU Xupeng¹

(1. College of Computer Science and Technology, Taiyuan University of Science and Technology, Taiyuan 030024, China; 2. School of Mechanical Engineering, Taiyuan University of Science and Technology, Taiyuan 030024, China)

Abstract: A semi-supervised student behavior detection model for dense classroom environments is proposed to address missed and false detections caused by high seat density, low individual pixel proportion, large intra-class variation, imbalanced class distribution, and limited availability of high-quality labeled data. The proposed model integrates a dual-stream interactive fusion module and a granular multi-level perception module to extract individual and interactive behavioral features and mitigate the suppression of behavioral interaction information. A dynamic interclass distribution loss and detection head are introduced to reduce the bias of the network's bias toward frequent categories. To address inconsistencies in pseudo-label confidence distributions, a class-wise distribution-aware semi-supervised learning strategy is employed, using dynamic thresholding for pseudolabel selection to suppress noisy samples. Experiments on a self-constructed classroom dataset demonstrated that, under 10% labeled data conditions, the proposed model improves mAP50 by 3.2% over YOLOv11m, and the semi-supervised strategy further increases performance by 5.0%. Validation on public datasets confirms the effectiveness and generalization capability of the model. These findings provide robust technical support for student behavior recognition and intelligent classroom learning status analysis in smart classroom scenarios.

Keywords: deep learning; computer vision; semisupervised learning; object detection; neural networks; feature extraction; pseudolabel; dynamic threshold

收稿日期: 2025-10-29. 网络出版日期: 2026-03-02.

基金项目: 教育部人文社会科学研究项目 (23YJCZH299); 山西省高等学校教学改革创新项目 (J20250145); 山西省研究生教育教学改革项目 (2025JG138, 2024JG165).

通信作者: 张睿. E-mail: zhangrui@tyust.edu.cn.

2025 年 1 月 19 日, 中共中央、国务院印发了《教育强国建设规划纲要 (2024—2035 年) 》, 将“建设学习型社会, 以教育数字化^[1]开辟发展新赛

道、塑造发展新优势”作为重点任务之一^[2]。课堂密集学生行为检测作为教育数字化的关键技术,其重要性愈发凸显。

在智能化教育评估^[3]中,学生行为的精准检测是实现客观、高效的教育质量评估的关键技术,其准确性对评估指标有着直接的影响^[4]。随着计算机视觉的发展,利用目标检测技术^[5]来实现学生课堂行为的精准检测已成为一种切实可行的方案^[6]。Tan等^[7]在CA-YOLOv9(coordinate attention YOLOv9)网络的基础上加入了坐标注意力模块,并将训练好的网络应用于全景多尺度课堂行为检测,证明该网络可以在不降低推理速度的情况下提高检测精度,初步验证了网络在智能课堂中实时应用的可行性;张浩鹏等^[8]针对真实课堂环境中存在的学生行为多样、遮挡频繁及背景干扰严重等问题,提出一种局部跨阶段视觉图卷积模型,旨在提升复杂课堂环境下的学生行为识别精度;胡锦涛等^[9]提出了一种融合全局注意力机制和时空图卷积网络的人体骨架行为检测网络,并通过迁移学习实现学生在线课堂行为检测功能,实验结果表明,提出的方法是一种有效的学生在线课堂行为检测方法;周珍玉等^[10]提出了一种基于人物交互的学生课堂行为检测研究方法,通过分析教室监控视频,检测出学生和物品目标,并基于人与物的交互关系来检测其课堂行为。

然而,以上检测技术大多聚焦于近景或单体目标较清晰的场景,对远景下个体密集^[11]、目标尺度存在差异^[12]的课堂行为检测关注不足。且在此类密集课堂场景环境中,仅依赖人工构建的大规模高质量标注数据集不仅成本高昂,而且难以覆盖多样化行为。为此,人们开始探索利用未标注数据辅助训练的半监督学习方法^[13]。Xu等^[14]提出了高效教师框架,用于可扩展且有效的基于锚点的单阶段半监督目标检测训练,这是第一次尝试将半监督对象检测应用于YOLOv5;Mi等^[15]从数据初始化的角度研究了师生学习,并提出了一种名为Active Teacher的新算法,该算法可以在提高伪标签质量的同时,最大限度地发挥有限标签信息的效果;Li等^[16]提出了一个简单而有效的基于梯度的采样框架,从两种类型的确认偏差的角度解决了类不平衡问题,并通过实验证明其优于当前类别的不平衡目标检测器作为类不平衡未来研究的基线;Liu等^[17]提出了UnbiasedTeacher-v2框架,它展示了半监督目标检测方法对无锚点检测器的一般化,并引入了Listen2Student机制用

于半监督回归损失,该方法既适用于无锚点方法,也适用于基于锚点的方法,并且始终如一地表现出色。

由以上研究可知,基于目标检测技术的学生行为检测虽取得一定进展,但仍存在明显局限。以YOLO系列^[18]为代表的单阶段检测器兼具高效性与精度,但对高质量标注数据依赖较强;双阶段检测器^[19]生成的伪标签质量较高,但推理速度较慢,检测效果有限。此外,现有方法多聚焦于近景或单体目标清晰的场景,对远景下个体密集、目标尺度差异大的课堂行为检测关注不足。在此类密集场景中,仅依靠人工构建大规模高质量标注数据集,不仅成本高昂,也难以覆盖多样化行为。

综上所述,尽管现有检测技术已经取得了一定的成效,但面对远景课堂场景行为检测任务时,仍存在个体密集且目标尺度差异大的难点,为此,本文针对数据间分布及人物交互特征设计全监督模型,并对网络中各模块及类平衡损失进行详细介绍;此外,鉴于学生人数众多且行为多样,致使高质量标签的制作成本高昂,从半监督策略入手,挖掘标签类内分布特征,使模型自主选择伪标签阈值,合理利用无标注数据;最后,通过自建数据集与公共数据集实验,验证所提方法的有效性,并展望未来研究方向。

1 半监督目标检测相关工作

1.1 YOLOv11

YOLOv11^[20]于2024年发布,在继承前代架构的基础上,通过改进结构与训练流程,降低了参数量并提升了检测精度^[21]。该系列具备良好的部署灵活性,可适配边缘设备^[22]、云平台^[23]及NVIDIA GPU系统。为满足不同场景在资源与精度间的权衡需求,YOLOv11提供了多尺度网络(包括n、s、m、l和x),覆盖了参数量、计算复杂度及检测精度^[24]的合理区间。在学生行为检测任务中,精确识别细微行为对课堂管理与学习评估具有重要意义^[25]。在大规模课堂环境下,实现学生多样化行为的实时检测,既需高精度网络,也对推理速度有较高要求。基于此,选用精度与效率平衡较好的YOLOv11s作为基准网络。YOLOv11延续了YOLO系列的核心架构,包含Backbone(主干网络)、Neck(颈部网络)和Head(检测头)3个阶段。其整体结构如图1所示。

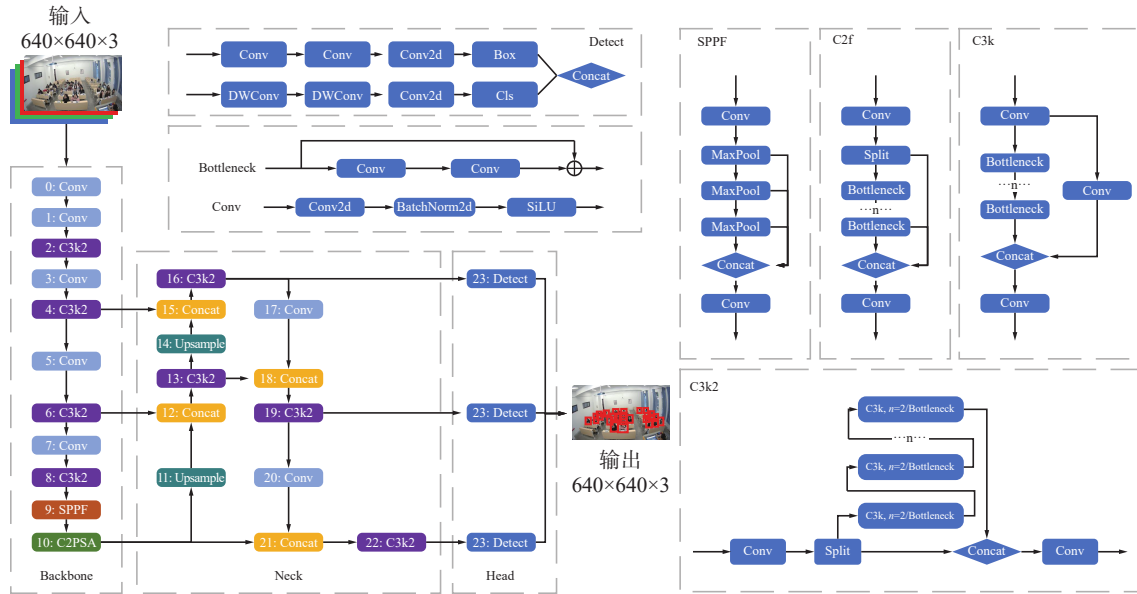


图1 YOLOv11 结构示意图

Fig. 1 YOLOv11 structure diagram

1.2 EfficientTeacher 半监督网络框架

Efficient Teacher^[14]采用师生架构,针对单阶段基于锚点的半监督目标检测问题。鉴于YOLO系列在性能上优于早期半监督所用的双阶段模型,为延续其单阶段检测优势,参照Efficient Teacher框架对CSB-YOLO进行进一步改进。

该框架通过Dense Detector、Pseudo Label Assigner和Epoch Adaptor这3个模块实现了可扩展且高效的训练流程。其中,Dense Detector基于RetinaNet并融入YOLOv5的密集采样技术^[26],提升了伪标签质量与推理效率;Pseudo Label Assigner将伪标签分为可靠与不确定两类,通过精细化分配策略缓解了伪标签不一致问题;Epoch Adaptor则利用域适应和分布适应技术,实现了稳定高效的端到端^[27]训练。

2 课堂学生行为检测模型构建

针对课堂密集场景下,学生个体像素占比低、尺度差异大及类间分布不均衡等问题,本文以YOLOv11s为基准网络,提出课堂学生行为检测模型CSB-YOLO(classroom student behavior YOLO),其结构如图2所示。该模型通过双流交互融合模块(dual-stream interaction fusion block, DSIFBlock)与颗粒化多层次感知模块(granular hierarchical perception block, GHPBlock)协同作用,增强对微小目标的多维度感知与精细特征提取。DSIFBlock通过空间流与语义流的双向特征交互机制,将局部细粒度信息与全局上下文语义进行动态融合,从而在复杂课堂场景中强化学生微小

动作(如手势、视线、姿态变化等)的表达能力;GHPBlock则通过分层卷积结构与多尺度聚合机制,对不同尺度的学生个体进行颗粒化表征,扩大了小目标的感受野并保留边缘与细节特征,从特征层面增强模型对微小目标的分辨能力。同时,引入动态类间分布检测头(adaptive class rebalancing detect, ACRDetect)与动态类间分布损失(adaptive class rebalancing loss, ACRLoss),通过对类别分布进行自适应加权,平衡了小样本行为类别在训练中的影响力,避免了模型对高频大目标类别的偏置。综合以上设计,CSB-YOLO在密集课堂场景下能够更准确地捕获微小目标的空间细节与交互特征,提升了学生行为识别的鲁棒性与精度。

2.1 双流交互融合模块

课堂场景下学生座位紧密、人数众多、学生行为多变、目标像素占比低且差异较大,这些因素使得学生行为检测变得更加困难。为此,设计了DSIFBlock,通过空间、通道、细节及全局等多维度特征增强与协同计算,赋予模型更精细的特征表达能力,并在结构上突出“双流”与“交互”的优势。

DSIFBlock首先通过通道压缩层利用1×1卷积与批归一化,将高维输入特征压缩为低维中间表示,以减少计算复杂度并保留主要语义信息。随后在空间流中引入多尺度深度可分离卷积(卷积核大小分别为1、3、5),各分支在不同感受野下提取局部细节与小目标特征,经过拼接后通过空间融合层的1×1卷积与Sigmoid激活生成空间注意力图,对输入特征进行空间加权,使模型聚焦

于显著位置区域。与此同时,通道流分支采用自适应平均池化提取全局通道信息,经过两层 1×1 卷积实现通道间的非线性映射与权重生成,动态调节不同语义通道的响应强度,以突出与个体判别相关的特征。为进一步刻画学生之间的行为关系,设计交互特征建模层,利用逐通道卷积与批归一化提取结构化语义信息,并通过计算特征图的自相似矩阵实现位置间的依赖建模。该关系矩阵经 softmax 归一化后,用于加权原始特征,

实现显式的全局交互增强。最终,模块将空间流与通道流的增强结果与交互建模特征拼接,经特征细化层的 1×1 卷积与 BatchNorm 实现融合压缩,再通过可学习参数 γ 与残差连接进行输出调节,实现稳定优化与多维特征的互补融合。整体而言,DSIFBlock 通过通道与空间维度的差异化特征聚焦,并结合交互关系矩阵强化个体间关联,使模型在密集课堂环境下更精确地感知学生行为细节与群体互动,从而提升检测性能。

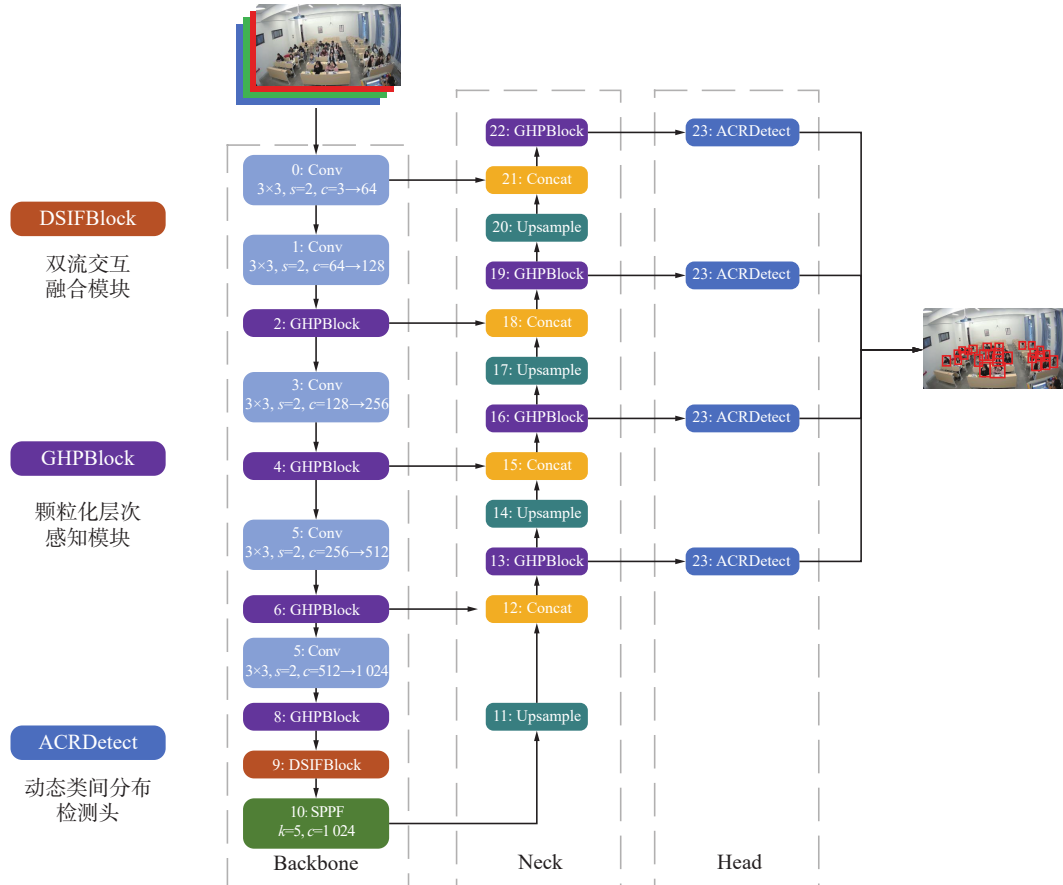


图 2 CSB-YOLO 结构示意图
Fig. 2 CSB-YOLO structure diagram

在通道流中,模块采用自适应通道注意力机制,为不同通道动态分配权重,使网络能够聚焦于与个体判别相关的关键特征,减少冗余信息干扰;而在空间流中,利用多尺度卷积核和深度可分离卷积有效捕捉局部与小目标的细节特征,从而提升对复杂动作和细粒度行为的感知能力。同时,模块嵌入了交互特征建模,通过特征相似性矩阵刻画目标间依赖关系,显式建模学生个体行为互动,强化群体场景下的全局关联表达。最后,将空间流、通道流的增强结果与交互建模特征融合,经细化与残差连接实现多维特征互补与稳定优化。整体上,双流交互融合模块不仅在局部细节与全局关系之间建立了紧密联系,还显著

提升了模型对人物细节和交互行为的感知与判别能力,从而有效应对课堂场景中复杂多变的行为检测任务。其结构如图 3 所示。

模块处理过程如下:

对于输入的特征,其中 B 是批次大小, C 是通道数, H 和 W 分别是特征图的高度和宽度。首先,特征缩放器(feature scaler, FS)通过全局特征来为每个通道分配不同的重要性权重。

$$F_1 = \text{ReLU}(\text{Conv2d}(\text{AvgPool}(x)))$$

$$A_{FS} = \text{Sigmoid}(\text{Conv2d}(F_1))$$

式中: x 为特征图张量;AvgPool为平均池化操作;Conv2d为二维卷积层;Sigmoid和ReLU是一种广泛使用的激活函数,用来引入非线性特性。然

后,输入特征图通过维度变换器(dimensional transformer, DT)进行增强,使用大小为1、3、5的卷积核进行空间卷积,并将所有尺度的特征图拼接在一起,得到一个多维度的特征图 F_{concat} ,随后使用一个 1×1 的卷积层进行融合得到权重。

$$F_k = \text{Conv}_{k \times k}(\text{Conv}_1(\mathbf{x}))$$

$$F_{\text{concat}} = \text{Concat}(F_1, F_3, F_5)$$

$$A_{\text{DT}} = \text{Sigmoid}(\text{Conv2d}(F_{\text{concat}}))$$

式中: $\text{Conv}_{k \times k}$ 为卷积核大小为 k 的卷积, Conv_1 为图3中DT卷积操作, Concat 为张量拼接操作。交互建模层用于捕捉图中不同位置之间的相似度关系,通过计算特征图中不同空间位置之间的内积,从而衡量了它们在特征空间中的相似程度,交互增强特征的表达式为

$$\mathbf{x}_{\text{interaction_enhanced}} = \text{Softmax}(\mathbf{x}^T \cdot \mathbf{x}) \cdot \mathbf{x}$$

式中: Softmax 为一种激活函数; $\mathbf{x}^T$ 为特征图 $\mathbf{x}$ 的转置,用来将输入值转换为概率分布。

在得到空间、通道和交互信息权重之后,将其进行拼接以及特征细化操作以提高特征的表达能力。

$$\mathbf{F}_2 = \mathbf{x} \cdot \mathbf{A}_{\text{DT}} + \mathbf{x} \cdot \mathbf{A}_{\text{FS}}$$

$$\mathbf{F}_{\text{refined}} = \text{Conv}_4(\text{Concat}(\mathbf{F}_2, \mathbf{x}_{\text{interaction_enhanced}}))$$

式中 Conv_4 为一种复合卷积模块,由 CBR(Conv2d-BatchNorm2d-ReLU)操作序列构成。

最后,为了保持信息的流动和稳定性,将细化后的特征图与原始输入特征图进行结合,并通过一个可学习的调节参数 γ 对输出进行调节。

$$\mathbf{y} = \mathbf{x} + \gamma \cdot \mathbf{F}_{\text{refined}}$$

式中: γ 是一个可以训练的参数,最终输出 $\mathbf{y}$ 即为经过 DSIFBlock 增强后的特征图。

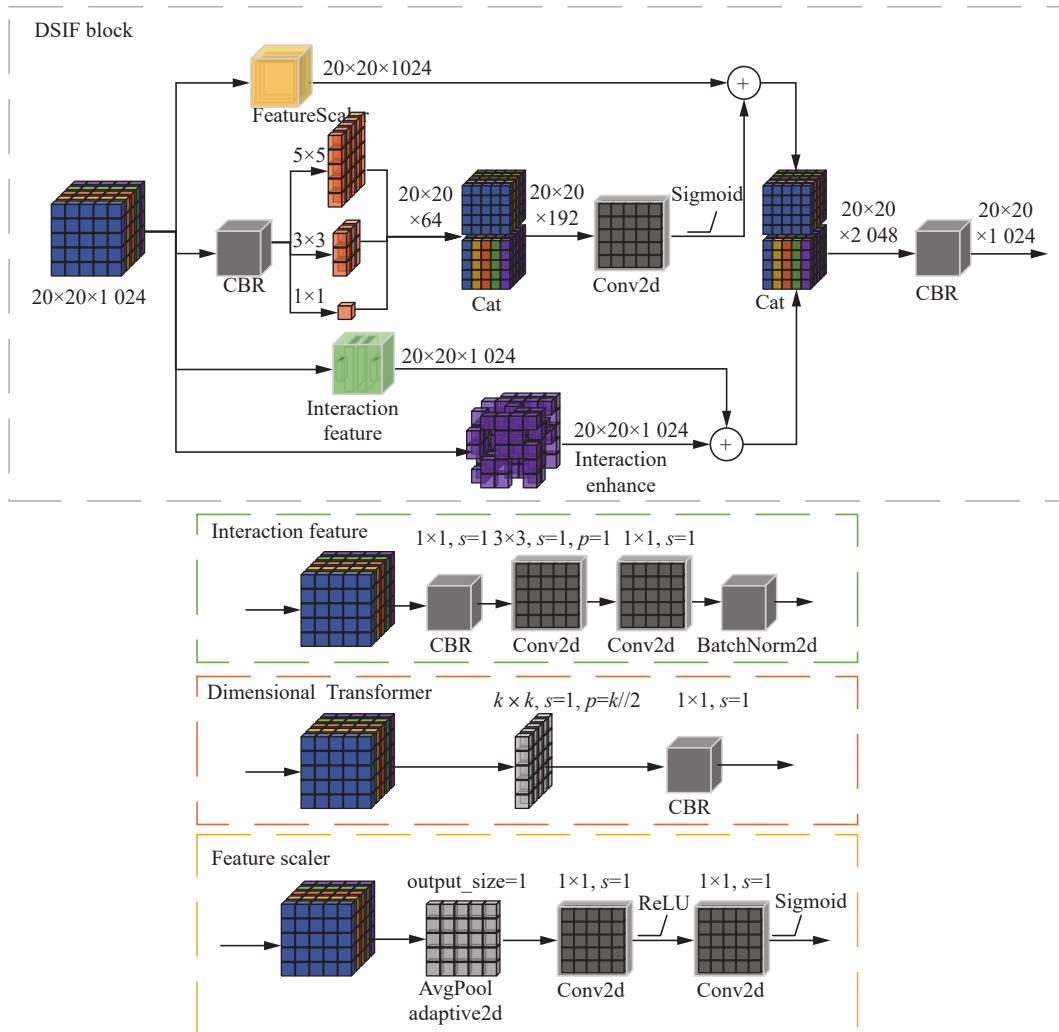


图3 DSIFBlock 结构示意图

Fig. 3 DSIFBlock structure diagram

2.2 颗粒化多层次感知模块

由于教学场景中,远景学生行为活动幅度较

小(如学生微小的头部动作)以及前后排学生像素占比差异大,致使网络无法将其准确捕捉。为

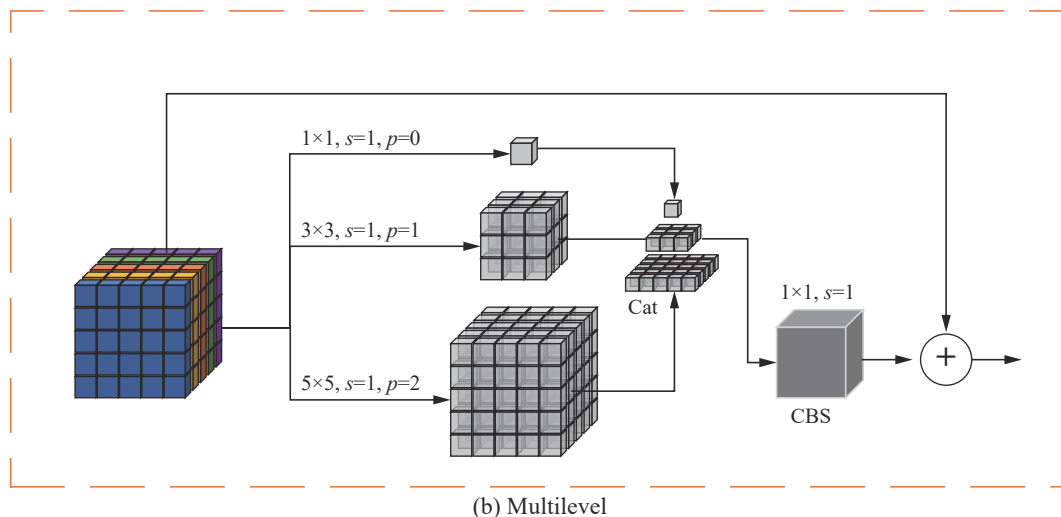
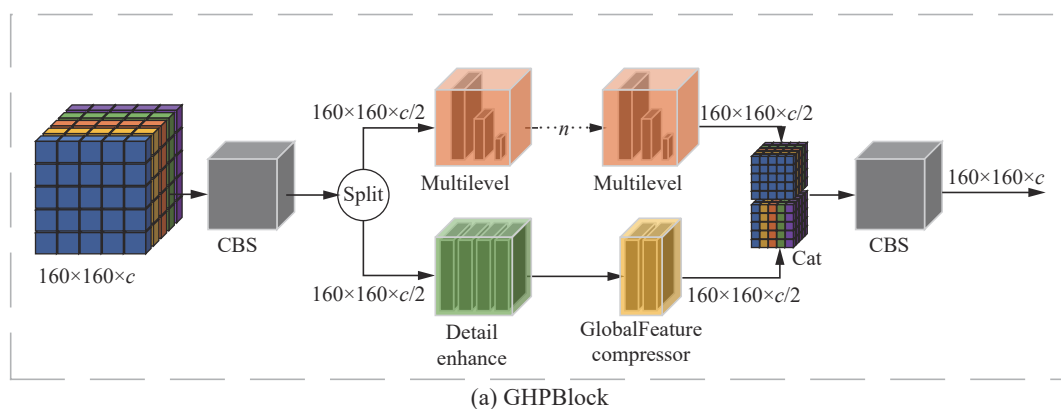
解决以上问题,设计 GHPBlock,通过细致化的特征处理流程,采用逐层细化的方式在多级层次上处理学生行为检测任务中的微小目标。其核心创新体现在特征的“颗粒化”操作中,模块首先将输入特征进行通道上的分解,通过 1×1 卷积将输入特征的通道维度从输入通道数扩展至输入通道数对应的两倍扩展维度,随后通过多个并行的卷积分支对不同“颗粒”特征进行处理。通过这种逐层精化和分解的方式,模块能够自适应地增强课堂场景中微小、局部化的目标特征。而后,使用信息融合方式将各颗粒特征层次整合,确保细节和全局信息的有效结合,从而提升模型对学生行为微小变化的敏感度和检测精度,缓解个体特征模糊和行为交互信息淹没的问题。

GHPBlock 首先在浅层特征提取阶段利用 1×1 卷积与 BatchNorm 进行通道扩展,保留输入特征的空间分辨率,并通过特征分块的方式为后续多分支处理提供更高的特征维度支撑。在此基础上,引入多尺度特征提取结构,通过 1×1 、 3×3 与 5×5 共 3 种卷积核的并行操作,实现不同感受野下的特征表达,以充分捕获学生头部、手

部及上身等不同比例目标的语义信息。多尺度特征在拼接后通过 1×1 卷积与 BatchNorm 进行通道融合,从而在特征层面实现全局语义与局部细节的统一表达。

在中层特征建模阶段,细节增强分支通过主干路径的两层卷积提取局部细节,辅助路径利用边缘卷积提取边界响应,两者经拼接与 1×1 卷积融合,在保持上下文一致性的同时增强学生个体边缘、姿态及动作区域的可辨性。该过程补偿了下采样导致的小目标细节损失,提升密集区域中相邻学生行为的区分能力。

在高层特征聚合阶段,全局特征压缩器被设计以实现行为相关区域的自适应权重分配。该模块通过通道维度的最大池化与平均池化生成双通道注意力图,形成空间权重分布,对增强特征进行加权,使网络聚焦于动作显著区域并抑制背景响应。经多尺度语义融合、细节增强与注意力加权的特征输入检测头,实现复杂课堂场景下微小目标的精准检测与动作识别。整体方法在保持较高检测速度的同时,提升了小目标场景下的召回率与特征表达能力。其模块结构如图 4 所示。



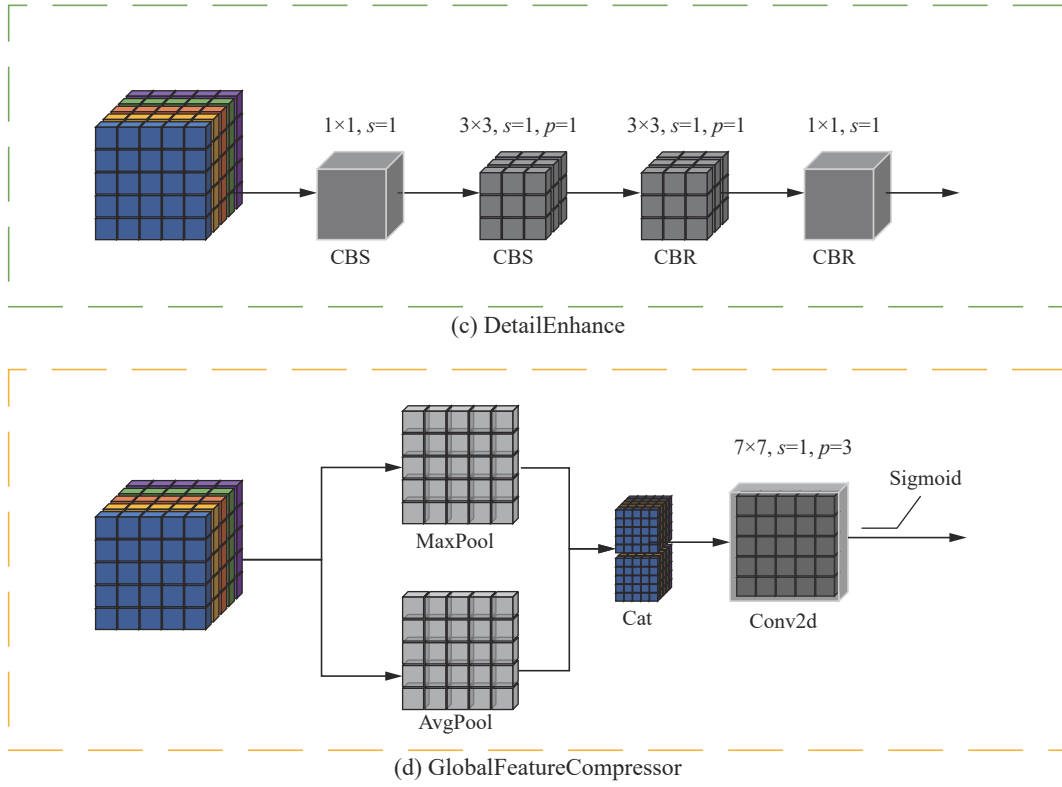


图4 GHPBlock 结构示意图

Fig. 4 GHPBlock structure diagram

处理过程如下:

首先,将输入的特征 $\mathbf{X} \in \mathbf{R}^{B \times C \times H \times W}$ 通过一个 1×1 的卷积进行通道扩展,使特征通道数扩展为 $2C_1$,并将扩展后的特征传入多个并行的多层次特征提取模块(Multilevel),对输入特征分别进行最细颗粒度(1×1 卷积)、中等颗粒度(3×3 卷积)和较大颗粒(5×5 卷积)的信息提取,并进行融合,其数学表达为

$$\begin{cases} \mathbf{y}_1 = \text{Conv}_{1 \times 1}(\mathbf{x}) & \mathbf{y}_1 \in \mathbf{R}^{B \times 2C_1 \times H \times W} \\ \mathbf{y}_2 = \text{Conv}_{1 \times 1}(\mathbf{y}_1) \\ \mathbf{y}_3 = \text{Conv}_{3 \times 3}(\mathbf{y}_1) \\ \mathbf{y}_4 = \text{Conv}_{5 \times 5}(\mathbf{y}_1) \end{cases}$$

$$\mathbf{y}_5 = \text{Concat}(\mathbf{y}_2, \mathbf{y}_3, \mathbf{y}_4) \quad \mathbf{y}_5 \in \mathbf{R}^{B \times 3C_1 \times H \times W}$$

细节增强分支则先后经过细节增强模块和空间增强模块进行细节增强处理,增强边缘及细节特征表达为

$$\mathbf{y}_6 = \text{Conv}_{3 \times 3}(\text{Conv}_{1 \times 1}(\mathbf{y}_1)) \quad \mathbf{y}_6 \in \mathbf{R}^{B \times C_2 \times H \times W}$$

$$\mathbf{F}_1 = \text{ReLU}(\text{BatchNorm2d}(\text{Conv2d}(\mathbf{x})))$$

$$\mathbf{y}_7 = \text{Concat}(\mathbf{y}_6, \mathbf{F}_1) \quad \mathbf{y}_7 \in \mathbf{R}^{B \times C_2 \times H \times W}$$

$$\mathbf{F}_2 = \text{ReLU}(\text{BatchNorm2d}(\text{Conv2d}(\mathbf{y}_7)))$$

$$\mathbf{y}_8 = \mathbf{A}_{\text{GFCCompressor}}(\mathbf{F}_2) \quad \mathbf{y}_8 \in \mathbf{R}^{B \times C_2 \times H \times W}$$

最后,将原始特征与得到的多层次颗粒化信息、细节增强信息融合,得到最终的特征图 $\mathbf{y}_{\text{out}}$

$$\mathbf{y}_{\text{out}} = \text{Conv}_{1 \times 1}(\text{Concat}(\mathbf{y}_1, \mathbf{y}_5, \mathbf{y}_8))$$

最终输出 $\mathbf{y}_{\text{out}}$ 即为经过 GHPBlock 后得到的特征图。

2.3 动态类间分布损失函数与检测头

课堂环境中,受时间、情景和教师引导等因素影响,学生的某些行为通常在特定的课堂活动中才会出现,致使课堂行为标签存在长尾分布的现象。针对这一现象,提出了带有自适应类间分布损失机制的 ACRDetect 与 ACRLoss,通过在检测头与损失函数两侧同时引入分布感知模块,构建了一个相互作用、动态调整类间平衡闭环。

在检测头部分,模型在原有分支之外增设轻量化类间平衡分支,该分支利用深度可分离卷积与 1×1 卷积建模类别权重,并通过 Sigmoid 归一化实现动态重加权,使得权重不仅取决于全局统计,更能随图像内容与上下文灵活调整。与此同时,在损失函数部分引入指数滑动平均(exponential moving average, EMA)^[28] 对类别分布进行动态追踪,以平滑批次间的波动并获得更稳定的类别统计信息。在此基础上,采用类别频率的倒数生成目标权重,即对于高频类别赋予较低权重,而对低频类别赋予较高权重,从而提升稀有类别在损失计算中的相对重要性,避免其在训练过程中被多数类掩盖。最后,再通过均方误差(mean squared error, MSE)对齐损失约束预测结果与目

标分布一致,从而实现“数据分布-权重分配-梯度更新”的自适应优化,形成检测头预测与损失函

数约束协同作用的闭环机制。ACRDetect 结构图如图 5 所示,其损失计算过程如下:

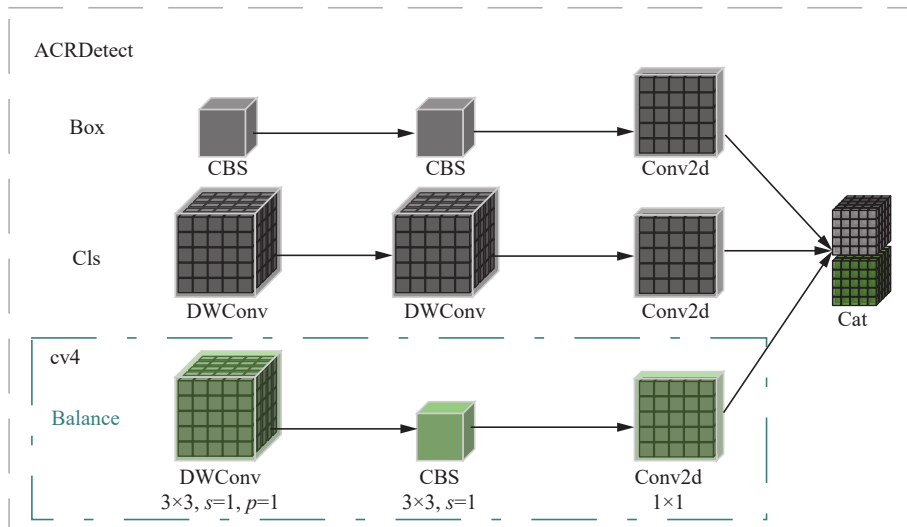


图 5 ACRDetect 结构示意图

Fig. 5 ACRDetect structure diagram

首先,模型通过 ACRDetect 中对应的卷积得到每个预测框的标签平衡权重 $\text{predicted}[i]$,并根据实际类别频率计算得到目标平衡权重 $\text{target}[i]$,用于反应在训练数据中不同类别的不平衡情况。

$$\begin{aligned} \text{predicted}[i] &= \text{cv4}[i](x[i]) \\ w[i] &= \frac{1}{\text{class_frequency}[c_i] + \varepsilon} \\ \text{target}[i] &= \frac{w[i]}{\sum_{i=1}^N w[i]} \end{aligned}$$

式中: cv4 为一种复合卷积模块,由检测头新添加的 Balance 分支使用; $x[i]$ 为第 i 个实例的预测类别; c_i 为目标框的类别标签; $\text{class_frequency}[c_i]$ 为类别 c_i 出现的频率; ε 是一个极小常数,避免除零。

最后,利用得到的每个预测框的标签平衡权重和目标平衡权重计算 L_{acr} 。

$$L_{\text{acr}} = \frac{1}{N} \sum_{i=1}^N (\text{predicted}[i] - \text{target}[i])^2$$

得到总损失表达为

$$L = L_{\text{box}} + L_{\text{cls}} + L_{\text{dfl}} + \lambda \cdot L_{\text{acr}}$$

式中: L_{box} 、 L_{cls} 和 L_{dfl} 分别为原 YOLOv11 使用的边界框损失、类别损失和分布焦点损失, L_{acr} 为新增动态类间分布损失, λ 取值为 4.0 (经 4.4.1 实验对比后取值)。

相较于传统的静态重加权或重采样方法,该闭环机制可以有效缓解类间分布不平衡问题。其一,权重能够随训练批次和课堂情境自适应变化,因此在不同课堂环节行为先验发生改变时,模型依旧能保持对低频类别的关注;其二,权重

预测直接依赖特征,使模型能够感知局部密集目标、小目标以及遮挡等复杂上下文,避免将同一类别在不同难度或姿态下进行统一处理;其三, MSE 对齐约束将预测权重逐渐引导至逆频率目标,既增强了低频类别的有效梯度,又因其为软性约束而避免对高频类别的过度惩罚,从而保持整体平衡与训练稳定。

综上所述,ACRDetect 与 ACRLoss 能够应对课堂行为长尾分布情况,改善课堂场景中学生行为类间分布不平衡导致的检测精度问题,有效提升网络对学生行为检测的精度。

3 类内分布特征感知的半监督策略

针对课堂场景下学生行为类间特征相似度高、密集遮挡导致同类行为内部置信度分布差异较大的问题,传统半监督方法采用固定阈值筛选伪标签,易丢失有效难分样本并引入噪声。为此,提出类内分布特征感知的半监督目标检测策略 (intra-class distribution-aware semi-supervised method, IDSM) 并构建类内分布感知半监督学生行为检测模型 (intra-class awareness classroom student behavior YOLO, ICSB-YOLO)。具体而言,教师模型利用无标注数据,通过弱图像增强生成伪标签供学生网络使用,并通过强图像增强实现师生模型间的一致性平衡。训练分为预热与半监督两个阶段:预热阶段仅使用标注数据训练教师模型;半监督阶段,教师模型对弱增强后的无标注数据生成伪标签,用于指导学生模型优化。整体框架如图 6 所示。

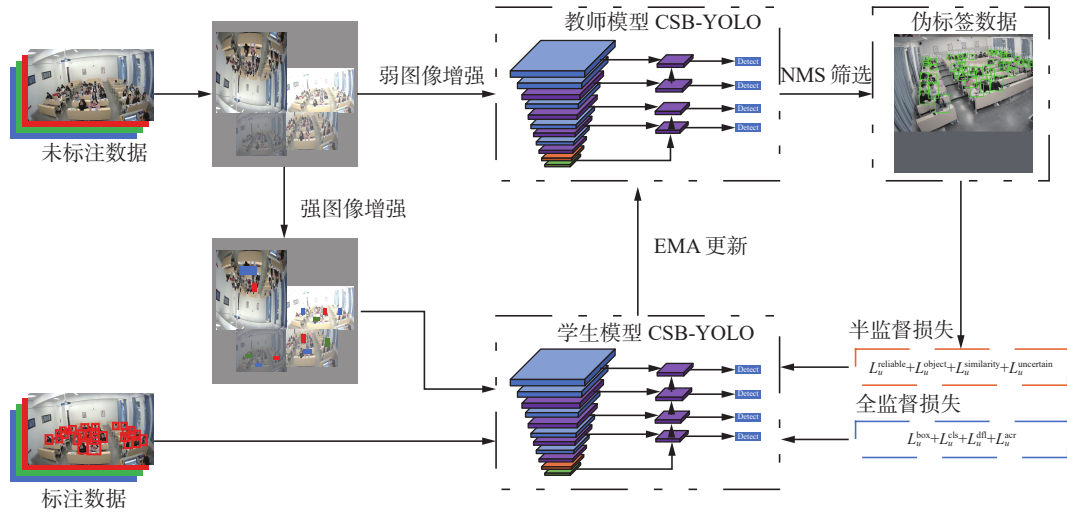


图6 ICSB-YOLO 结构示意图

Fig. 6 ICSB-YOLO structure diagram

在课堂行为检测场景中,个体尺度小、遮挡频繁等因素导致不同行为类别间特征分离度较低,同类样本的预测置信度呈现显著离散现象,形成类内异质性。固定阈值在此情况下易产生结构性偏置:高离散类别中的低置信有效样本被过度过滤,而类间边界处的高置信误检样本被保留,加剧类别混淆。针对上述问题,本文将形式化为类内置信分布不一致性问题,并提出如下假设:同一类别伪标签置信度的统计分布能够刻画该类别的学习状态,其均值与离散程度可用于构建类别自适应的决策边界。由此,设计了基于类别级均值与方差的动态双阈值机制,对伪标签进行可信度分级,实现类别自适应的样本筛选。该策略通过对类内置信分布建模并依此调整阈值,缓解伪标签噪声对梯度更新的干扰,提升半监督训练的稳定性与泛化能力。

教师网络根据优化后的学生模型通过 EMA 来更新更新自身参数,从而生成更高质量的伪标签,推动半监督训练的持续优化。教师模型的更新公式为

$$\theta_t^k = \alpha \theta_t^{k-1} + (1 - \alpha) \theta_s^k$$

式中: θ_t 和 θ_s 分别为教师模型和学生模型的参数; α 是更新速率,文中使用值为 0.999。

3.1 类内分布感知伪标签分配策略

课堂场景下学生像素占比低、行为差异小、类间特征相似度高,使得模型易受伪标签噪声影响,干扰对特征关联的学习,加剧判断偏差。为此,设计了一种基于类内关系的伪标签分配策略,通过类内数据的期望值和离散程度划分高低阈值,以减少类间关系的混淆。首先利用非极大值抑制^[29]筛选冗余的检测框;随后,在经过筛选

的检测框中设置了两个阈值 η_1 和 η_2 , 进一步将伪标签划分为可信、不确定和不可信 3 种类型,其划分规则如下:

$$B = \begin{cases} \text{Reliable}, & \rho(B) \geq \eta_1 \\ \text{Uncertain}, & \eta_1 < \rho(B) < \eta_2 \\ \text{Background}, & \rho(B) \leq \eta_2 \end{cases}$$

为了确保可靠伪标签和不确定伪标签分布合理, η_1 和 η_2 的值依据迭代次数采用以下策略进行动态调整

$$\eta_1 = \text{Mean}(\mathbf{O}_c^k) + \text{VAR}(\mathbf{O}_c^k) \times 2.0$$

$$\eta_2 = \text{Mean}(\mathbf{O}_c^k) - \text{VAR}(\mathbf{O}_c^k) \times 2.0$$

式中: $\mathbf{O}_c^k$ 为第 k 个批次的 c 类伪标签分数列表, Mean 为得分列表中所有得分的期望值, VAR 为得分列表中所有得分的方差, η_1 和 η_2 分别为高置信度阈值和低置信度阈值。

均值代表了类内数据的中心位置,而方差则反应了数据的离散程度,通过这种计算,模型可以基于类内关系的分布情况来确定适合的阈值,减少了类间关系的混淆,使模型能够在训练过程中更好地适应不同的类分布,从而提高模型的稳定性与泛化能力。

3.2 一致性约束半监督损失函数

为确保数据在经过弱图像增强和强图像增强仍然能够保持相同的预测结果,设计了一致性损失函数。半监督学习结合了标注数据和未标注数据,通过利用未标注数据提升模型的泛化能力,其损失函数通常包括监督损失和半监督损失。监督损失基于有标签数据计算预测误差,而半监督损失通过伪标签生成或自监督方法在未标注数据上优化。最终损失为两者的加权和,通过调整超参数平衡监督与半监督部分,其数学表达式为

$$L_{\text{total}} = \lambda_1 \cdot L_s + \lambda_2 \cdot L_u$$

$$L_s = L_s^{\text{box}} + L_s^{\text{cls}} + L_s^{\text{dfl}} + 4.0 \cdot L_s^{\text{acr}}$$

$$L_u = L_u^{\text{reliable}} + L_u^{\text{object}} + L_u^{\text{similarity}} + L_u^{\text{uncertain}}$$

式中: L_s 为监督损失, 仍然使用优化后的损失函数; L_u 为半监督损失, 由可靠损失、置信度损失、一致性损失和不确定损失组成; λ_1 和 λ_2 为平衡超参数, 用于调整监督和半监督损失的贡献, 设置为 1.0 和 3.0。

$$L_u^{\text{reliable}} = \sum_{B \in \text{reliable}} (\text{CE}(X_{(h,w)}^{\text{cls}}, \hat{Y}_{(h,w)}^{\text{cls}}) + \text{CIoU}(X_{(h,w)}^{\text{reg}}, \hat{Y}_{(h,w)}^{\text{reg}}))$$

$$L_u^{\text{object}} = \sum_{B \in \text{Background}} \text{CE}(X_{(h,w)}^{\text{obj}}, 0)$$

$$L_u^{\text{similarity}} = T^2 \cdot \left(\theta_t \cdot \log \frac{\theta_t}{\theta_s} + (1 - \theta_t) \cdot \log \frac{1 - \theta_t}{1 - \theta_s} \right)$$

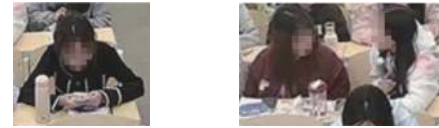
$$L_u^{\text{uncertain}} = \sum_{B \in \text{uncertain}} (\text{CE}(X_{(h,w)}^{\text{obj}}, \text{obj}_{(h,w)}))$$

式中: $X_{(h,w)}^{\text{cls}}$ 、 $X_{(h,w)}^{\text{reg}}$ 和 $X_{(h,w)}^{\text{obj}}$ 分别为学生模型在位置 (h, w) 处预测的类别得分、回归边框的回归得分和目标性预测值; $\hat{Y}_{(h,w)}^{\text{cls}}$ 和 $\hat{Y}_{(h,w)}^{\text{reg}}$ 分别为伪标签的类别得分和回归得分; $\text{obj}_{(h,w)}$ 为伪标签的目标性预测值; CE 为交叉熵损失函数; CIoU 为完全交并比损失函数; θ_t 和 θ_s 分别为教师和学生模型概率分布; T 为温度系数, 用来控制概率分布的平滑程度, 能够使模型更好的理解类别之间的关系, 使用值为 5.0。

4 半监督学生行为检测实验与分析

4.1 学生课堂行为数据集

为验证所提方法的有效性, 实验采用的数据集为某大学学生课堂行为数据集 SCAD, 该数据集由大学教室内安装的智能教学督导摄像头采集, 涵盖阶梯教室、平层教室和多功能教室的课堂影像, 并从中提取图像构建而成。共包含图像 2740 张, 其中标注数据 820 张。使用 LabelImg 工具对学生行为进行标注, 划分为抬头听课、低头学习、使用电子设备以及私下交流 4 类。数据集图像实例见图 7, 标注分布见表 1。



(f) 使用电子设备

(g) 私下交流

图 7 自建数据集 SCAD 部分图像
Fig. 7 Partial image samples from the SCAD

表 1 数据集分布情况
Table 1 Dataset distribution

标注情况	有标签	无标签	验证集	测试集
1% 已标注	28	2166	273	273
2% 已标注	55	2139	273	273
5% 已标注	137	2057	273	273
10% 已标注	274	1920	273	273

4.2 实验环境

实验平台搭载 NVIDIA GeForce RTX 4070Ti SUPER, 显存 32 GB, 深度学习框架为 Pytorch 2.4.1+cu124, 编程语言为 Python3.8.10, 优化器采用 AdamW, 输入网络的图像尺寸为 640, 每批次处理数量为 2, 实验设置训练总批次为 300, 评价指标使用模型参数量 (parameters) 和平均精度^[21] (mean average precision, mAP)。

4.3 评价指标

在目标检测任务中, mAP 是最常用且最具代表性的性能评价指标。mAP 的计算过程通常先针对单一类别得到平均精度 (average precision, AP), 即通过精度-召回曲线衡量模型在该类别上的检测效果, 再对所有类别的 AP 进行平均, 从而得到整体检测性能。与单一的准确率不同, mAP 同时考虑了检测框的位置精度和分类正确性, 能够更全面地反映模型的检测质量。其计算方式为

$$A_p = \int_0^1 P(R) dR$$

$$m_{AP} = \frac{1}{C} \sum_{i=1}^C A_{p_i}$$

式中: R 为召回率, 即正确检测到的目标占所有真实目标的比例; $P(R)$ 为在不同召回率 R 下的精度; C 为类别总数。

在实际评估中, 常用 mAP50 与 mAP50:95 两个指标。mAP50 表示在交并比 (intersection over union, IoU) 阈值为 0.5 时计算得到的 mAP, 衡量模型在较宽松定位要求下的性能; 而 mAP50:95 则是在 IoU 阈值从 0.50 到 0.95 (步长 0.05) 共 10 个阈值下分别计算 mAP 后再取平均, 更加严格、全

面地反映检测性能。

在学生课堂行为检测任务中,选择 mAP 作为核心评价指标主要基于以下两方面考虑:一方面,远景课堂场景下行为类别较多且分布不均衡,若仅采用准确率等指标,易受高频类别影响而掩盖低频行为的检测效果。mAP 对每类行为分别评估后取平均,能够客观反映模型在所有类别上的性能。另一方面,课堂行为检测需同时判断行为类别与定位目标位置,mAP 结合精度与召回率,能更全面地衡量模型在多目标、多类别场景下的综合表现。因此,实验采用 mAP 作为评价指标。

4.4 全监督部分

4.4.1 损失平衡超参数配比

为选取合适的动态类别平衡损失权重,使用 10% 标注数据进行训练,在保持原损失部分不变的基础上,调整平衡超参数 λ 的值,分别设置为 1.0、2.0、3.0、4.0 和 5.0,实验结果如图 8 所示。实验结果表明,模型检测精度在最初能够随着 λ 的增大而提升,但在超过 4.0 后出现下降趋势,因此综合考虑整体效果,最终确定 λ 的最佳取值为 4.0。

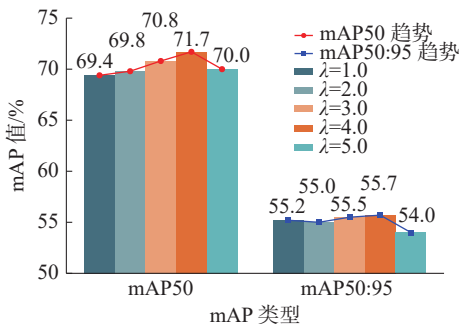


图 8 损失平衡超参数实验情况

Fig. 8 Evaluation of loss balance settings

4.4.2 CSB-YOLO 消融实验

为验证各模块及损失函数的有效性,使用自建数据集 SCAD 中 10% 标注数据开展实验,结果如表 2 所示。

消融实验系统地验证了各模块对模型性能的贡献。原始模型在 mAP50 和 mAP50:95 上分别为

65.2% 和 51.5%;当逐步引入 DSIFBlock 与 GHPBlock 时,性能得到持续提升,最终分别达到 70.2% 和 55.3%;在此基础上进一步结合 ACLoss 后,mAP50 提升至 71.7%,mAP50:95 提升至 55.7%。整体结果表明,随着模块的逐步叠加,检测精度呈现稳步增长趋势,相较于原始模型,最终配置在 mAP50 和 mAP50:95 上分别提升 6.5 % 和 4.2%,证明了所提方法在课堂学生行为检测任务中的有效性。

表 2 消融实验情况

Table 2 ablation study results

DSIF Block	GHP Block	ACLoss	Param/ 10^6	mAP50/%	mAP50:95/%
—	—	—	9.43	65.2	51.5
√	—	—	8.71	67.6	52.9
√	√	—	11.89	70.2	55.3
√	√	√	12.72	71.7	55.7

注:“√”表示模型使用该模块,“—”表示模型未使用该模块。

4.4.3 CSB-YOLO 对比实验

为验证 CSB-YOLO 的全监督性能,在 10% 标注数据条件下,与 YOLOv11、YOLOv12 等主流版本进行对比,结果如表 3 所示。在检测精度方面,CSB-YOLO 的 mAP50 达到 71.7%,优于所有对比模型。与 YOLOv11m 相比,mAP50 提升 3.2 个百分点,mAP50:95 提升 0.9 个百分点;与 YOLOv11L 相比,mAP50 提升 1.7 个百分点,mAP50:95 提升 0.1 个百分点;与 YOLOv12L 相比,mAP50 提升 6.1 个百分点,mAP50:95 提升 1.6 个百分点。在模型效率方面,CSB-YOLO 的参数量为 12.72×10^6 ,较 YOLOv11m 减少 36.7%,较 YOLOv11L 减少 51.3%;计算复杂度为 90.8×10^9 ,与同精度级别的 YOLOv11L 基本持平;推理速度为 25.39 f/s,虽低于 YOLOv11m,但高于课堂实时监测的常规需求阈值 (≥ 25 f/s)^[30]。综合而言,CSB-YOLO 在保持实时处理能力的前提下,以更低的模型复杂度实现了检测精度的有效提升,在密集课堂场景下表现出良好的精度与效率平衡。

表 3 对比实验情况

Table 3 Results of Comparison Experiments

Model	FLOPs/ 10^9	Param/ 10^6	mAP50/%	mAP50:95/%	Δ (mAP50)	FPS/(f/s)
YOLOv8m	79.1	25.86	67.5	54.4	4.2↑	67.96
YOLOv8L	165.3	43.63	68.6	55.1	3.1↑	44.13
YOLOv9m	77.6	20.16	69.7	54.4	2.0↑	51.59
YOLOv9c	103.7	25.53	67.0	54.6	4.7↑	53.48
YOLOv9e	192.5	58.20	67.1	54.8	4.6↑	24.79
YOLOv10m	64.0	16.49	67.7	54.3	4.0↑	50.27

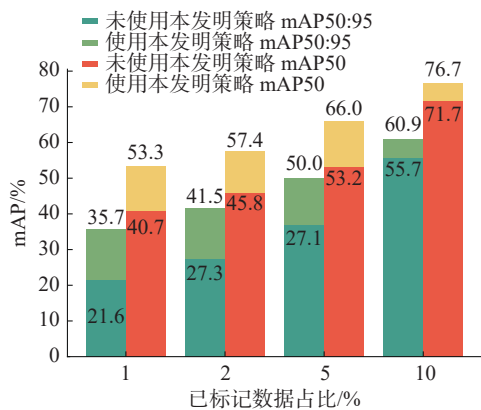
续表 3

Model	FLOPs/ 10^9	Param/ 10^6	mAP50/%	mAP50:95/%	Δ (mAP50)	FPS/(f/s)
YOLOv10b	98.7	20.46	69.9	55.1	1.8 $\uparrow$	43.80
YOLOv10L	127.2	25.78	67.8	55.2	3.9 $\uparrow$	37.43
YOLOv11m	68.2	20.10	68.5	54.8	3.2 $\uparrow$	69.17
YOLOv11L	90.2	26.10	70.0	55.6	1.7 $\uparrow$	56.12
YOLOv12m	67.8	20.10	66.4	53.5	5.3 $\uparrow$	57.49
YOLOv12L	89.4	26.40	65.6	54.1	6.1 $\uparrow$	39.03
CSB-YOLO	90.8	12.72	71.7	55.7	—	25.39

注：“ Δ ”表示“CSB-YOLO”相比其他网络的变化量，“ $\uparrow$ ”表示提高，加粗表示最优效果。

4.5 IDSM 策略有效性实验

1) 使用 IDSM 策略前后性能对比。将未使用 IDSM 策略的 CSB-YOLO 与使用 IDSM 策略的 ICSB-YOLO 分别在 1%、2%、5% 和 10% 标注数据集中进行比较, 实验结果如图 9 所示。实验结果表明, 在采用 IDSM 策略后, 模型在 1%、2%、5% 和 10% 占比的标注数据训练中, mAP50 分别提升了 12.6%、11.6%、12.8% 和 5.0%。这说明模型能够合理、高效地筛选利用无标注数据进行训练, 进一步证明了该策略的有效性。

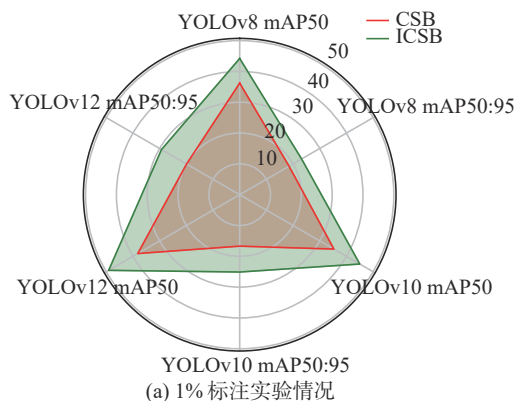


注: 左侧为 mAP50:95 指标, 右侧为 mAP50 指标。

图 9 ICSB-YOLO 消融实验直方图

Fig. 9 Histogram of ablation study for ICSB-YOLO

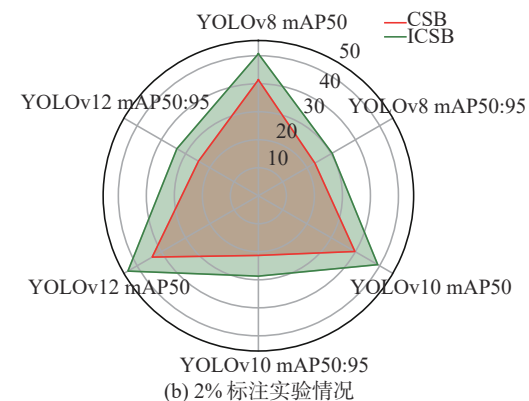
2) IDSM 半监督策略跨 YOLO 版本模型性能



验证。为进一步验证 IDSM 半监督策略的有效性与泛化性, 将该策略应用于不同版本 YOLO 模型并开展实验对比, 结果如图 10 所示。整体来看, IDSM 在 3 个版本上均带来了稳定增益。与未使用 IDSM 半监督策略的模型相比, 使用 IDSM 策略的 YOLOv8 的 mAP50 平均提升约 8 个百分点, mAP50:95 提升约 6 个百分点; YOLOv10 的 mAP50 平均提升接近 10 个百分点, mAP50:95 提升约 7 个百分点; YOLOv12 的 mAP50 提升约 10 个百分点, mAP50:95 提升约 8 个百分点。可以发现, IDSM 策略在不同版本的 YOLO 模型上均能带来一定程度的增益, 证明了该策略的通用性。

4.6 ICSB-YOLO 对比实验

为进一步验证 ICSB-YOLO 的检测性能, 选择 EfficientTeacher^[14]、Active Teacher^[15]、Unbiased Teacher^[31]、Unbiased Teacher-v2^[17] 和 4 种半监督方法进行对比。各方法均按其原始论文推荐的骨干网络与训练配置复现, EfficientTeacher 基于 YOLOv5, 其余基于 Faster R-CNN。需要说明的是, 半监督方法通常与其骨干网络存在一定耦合, 若统一更换骨干可能影响其原有性能的发挥。因此, 本文采用领域内对比研究的常见做法, 确保各方法处于其常规应用状态下进行比较。实验结果如表 4 所示。



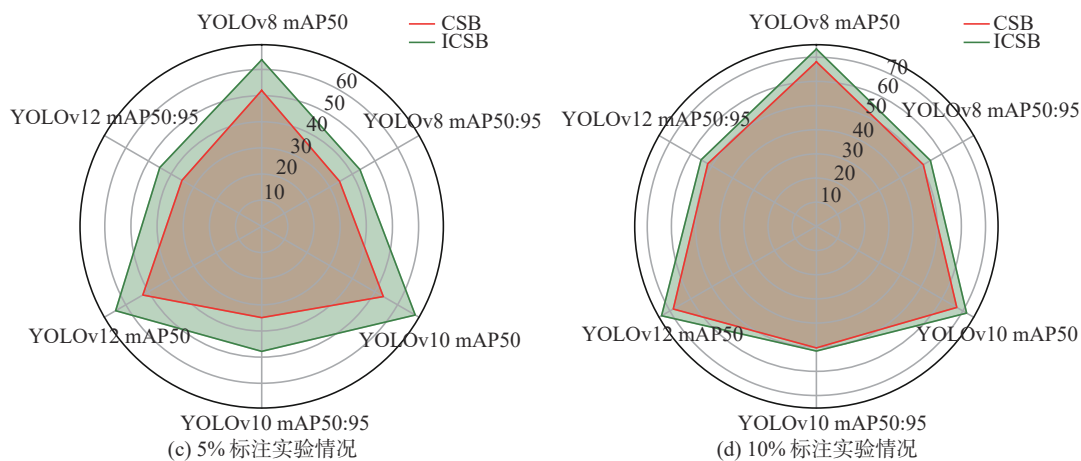


图 10 IDSM 效果展示
Fig. 10 IDSM Performance

表 4 半监督对比实验情况
Table 4 Semi-Supervised Comparison Results %

模型	mAP	已标注数据占比			
		1	2	5	10
Efficient Teacher	mAP50	35.8	45.1	54.3	58.1
	mAP50:95	22.7	28.6	37.5	41.7
Active Teacher	mAP50	38.5	41.1	38.9	65.1
	mAP50:95	18.6	20.7	26.7	47.2
Unbiased Teacher	mAP50	39.8	41.8	41.9	65.2
	mAP50:95	20.4	22.9	28.5	47.4
Unbiased Teacher-v2	mAP50	38.9	41.2	41.0	64.9
	mAP50:95	20.9	23.9	28.6	47.9
ICSB-YOLO	mAP50	53.3	57.4	66.0	76.7
	mAP50:95	35.7	41.5	50.0	60.9

注: 加粗表示最优效果。

从表 4 可以看出, 在不同标注比例下, ICSB-YOLO 的检测精度均优于对比方法。以 10% 标注为例, ICSB-YOLO 的 mAP50 达 76.7%, 较 EfficientTeacher(58.1%)提升 18.6%。这一差距在一定程度上反映了 ICSB-YOLO 核心策略(intra-class distribution-aware semi-supervised method, IDSM)在利用无标注数据方面的潜在优势。此外, 随着标注比例从 1% 增至 10%, ICSB-YOLO 的性能提升幅度也更为平稳, 表明其对标注数据的依赖性相对较低, 在半监督场景下具有较好的适用性。同时, 为评估各方法的实际部署效率, 补充其骨干网络的计算量(floating-point operations per second, FLOPs)与推理速度(frames per second, FPS)指

标。实测数据如下: EfficientTeacher 所基于的 YOLOv5, FLOPs 为 102.73×10^9 , FPS 为 36.8; Active Teacher、Unbiased Teacher 及 Unbiased Teacher-v2 所基于的 Faster R-CNN, FLOPs 为 145.43×10^9 , FPS 为 20.1。需要说明的是, 半监督策略仅作用于训练阶段, 不改变模型的推理架构, 因此上述 FLOPs 与 FPS 即为各方法部署时的实际效率指标。综合来看, ICSB-YOLO 在取得高检测精度的同时, 其推理速度(25.39 f/s)满足课堂实时监测的应用需求, 且计算复杂度(90.8×10^9)低于基于 Faster R-CNN 的方法, 有效保持了 YOLO 系列的高效特性。

4.7 泛化性验证

为更好地说明本文方法的泛化能力, 使用公共数据集 SCB-Dataset3-U^[32] 进行实验, 结果如图 11 所示。可以看到, CSB-YOLO 不仅在自建数据集上展现出较好的优势, 在公共数据集上同样体现出良好的鲁棒性与推广性。从结果来看, 其在仅有 12.72×10^6 参数量的情况下即可实现 95.2% 的 mAP50, 相比参数量为 25.86×10^6 的 YOLOv8m, 精度保持完全一致而模型规模减少一半。进一步对比 YOLOv8l, 虽然该模型的参数量达到 43.63×10^6 , 但其精度仅为 94.4%, 反而低于 CSB-YOLO, 而后者的参数规模仅为前者的 20% 左右。继续观察其他版本, 如 YOLOv9e 和 YOLOv10l, 二者的参数量分别高达 58.20×10^6 和 25.78×10^6 , 但其精度水平与 CSB-YOLO 基本处于同一范围。这些结果充分表明, CSB-YOLO 在降低模型复杂度的同时, 仍能保持甚至略优的检测精度, 凸显出其在轻量化、高效性与实际部署价值上的优势。

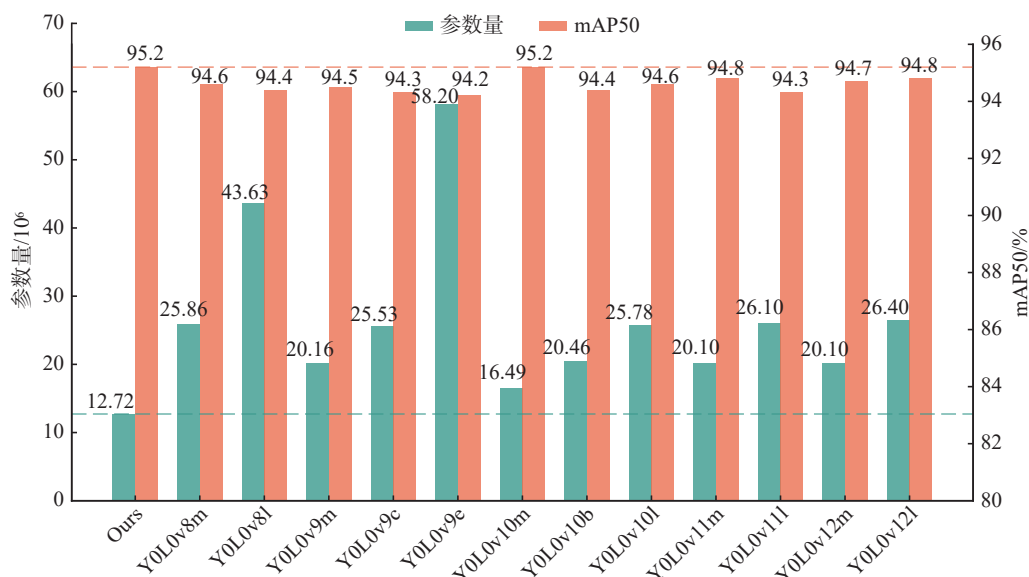


图 11 SCB-Dataset3-U 效果

Fig. 11 SCB-Dataset3-U performance

4.8 可视化效果

为直观展示网络在教学课堂场景中的检测效果,使用自建数据集 SCAD 进行检测并使用公共数据集 SCB-Dataset3^[32] 的大学课堂部分(图 12(e)、图 12(f)展示为 0.355k_university_yolo_Dataset 下训练图像 40030092.jpg)进行泛化性对比,如图 13 所示。展示了网络在多功能教室、平层教室和阶梯教室的表现以及 ICSB-YOLO 和 YOLOv11 在未训练场景中的预测情况,通过图 13(a)、图 13(b)的对比可以看出,ICSB-YOLO 能够更准确地检测出学生目标及学生行为,漏检和误检现象减少(YOLOv11 在图中左侧部分存在的漏检、误检现象得到改善),证明了其在未知场景下的泛化效果。

4.9 实验结果分析

从上述实验结果可以看出,本文提出的 CSB-YOLO 及其半监督学习模型 ICSB-YOLO 在检测精度、鲁棒性及模型效率方面均表现出一定的优势。在全监督实验(见表 3)中,CSB-YOLO 相较于 YOLOv11m 和 YOLOv11L 模型,在使用 10% 标注数据的条件下, mAP50 分别提升了 3.2% 和 1.7%,同时参数量减少约 36.7% 与 51.3%。这一结果表明,CSB-YOLO 在有效降低模型复杂度的同时,仍能保持甚至提升检测性能,体现了其良好的结构设计及高效的特征表达能力。通过消融实验(见表 2)进一步发现,随着 DSIFBlock 与 GHPBlock 模块的逐步引入,模型的 mAP50 由 65.2% 提升至 70.2%, mAP50:95 由 51.5% 提升至 55.3%;在此基础上结合 AC Loss 后,最终检测精度进一步提升至 71.7% 与 55.7%,充分验证了各模块设计的有效性。



图 12 数据集图像

Fig. 12 Dataset image visualization

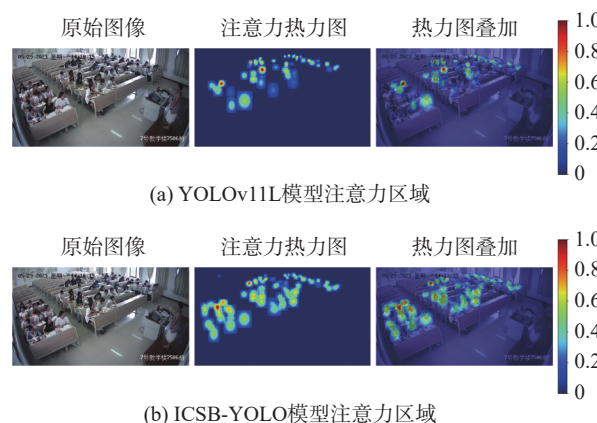


图 13 可视化效果

Fig. 13 Visualization results

半监督实验(见图9)中, IDSM 半监督学习策略增强了模型对无标注数据的利用能力。在1%、2%、5%和10%标注比例下, ICSB-YOLO相较于CSB-YOLO的mAP50分别提升了12.6%、11.6%、12.8%和5.0%,说明IDSM策略能够有效筛选高置信度伪标签并抑制噪声伪标注的干扰,从而实现稳定的性能提升。进一步在不同YOLO版本(v8、v10、v12)上的实验结果(见图10)显示, IDSM策略在各版本模型中均带来了6%~10%的mAP50提升,证明了所提出策略的通用性与鲁棒性。

在与主流半监督检测方法的对比实验(见表4)中, ICSB-YOLO在各标注比例下均取得最优性能。当标注比例仅为1%时, mAP50达到53.3%,较EfficientTeacher提升17.5%;当标注比例增加至10%时, mAP50进一步提升至76.7%,且优于Active Teacher、Unbiased Teacher和Unbiased Teacher-v2半监督检测方法。这一结果表明, ICSB-YOLO在低标注率条件下依然能够保持较高检测精度,验证了其在数据稀缺场景下的学习效率与泛化能力。此外,在公共数据集SCB-Dataset3-U上的泛化实验(见图11)中, CSB-YOLO以仅 12.72×10^6 的参数量实现了95.2%的mAP50,与参数量为 25.86×10^6 的YOLOv8m性能持平,甚至超过参数量高达 43.63×10^6 的YOLOv8L,显示出参数效率和轻量化优势。可视化结果进一步表明, ICSB-YOLO在复杂课堂场景中能有效减少漏检与误检现象,检测框定位更加准确,说明该模型在复杂环境下仍具较为可观的鲁棒性与稳定性。

综上所述,本文提出的CSB-YOLO及其半监督模型ICSB-YOLO在检测精度、模型复杂度、数据利用率以及跨场景泛化能力等方面均表现出较好性能,验证了所提方法在课堂密集场景下学生行为检测任务中的有效性与应用潜力。

5 结束语

针对课堂场景中学生密集、个体像素占比低、差异大及类间分布不平衡等问题,提出CSB-YOLO,通过DSIFBlock和GHPBlock增强特征提取能力,解决密集场景下的特征模糊问题;利用ACRDetect和ACLoss缓解类间不平衡,提升检测精度和泛化性。同时,设计IDSM半监督学习策略并构建ICSB-YOLO,动态调整伪标签筛选标准,有效挖掘未标注数据信息,减少噪声干扰。通过在自构建的SCAD数据集和SCB-Dataset3公共数据集实

验。结果表明, CSB-YOLO参数量为 12.72×10^6 ,相比其他YOLO版本平均参数量降低47.90%,且在mAP50指标上,相比YOLOv12L提升幅度可达6.1%;与YOLO系列其他版本相比, mAP50平均提升了3.72%。在此基础上构建半监督框架,与CSB-YOLO相比, ICSB-YOLO在10%标注情况下, mAP50提升了5.0%,且在不同数据标注量情况下, IDSM策略均能在对比实验中保持较好的效果。此外,经过可视化分析发现,在未经训练的场景中,通过半监督训练得到的模型仍能保持较强的检测能力,验证了模型的泛化能力。

尽管课堂密集场景类内感知半监督学生行为检测能够取得预期效果,但当前学生行为检测模型仍依赖单一视觉模态,难以全面捕捉学生在课堂中的细微状态与情绪,导致信息利用不足。下一步将聚焦于多模态信息融合驱动半监督行为检测研究,充分结合视觉、语音及姿态等多源数据,在提升检测精度的同时降低对人工特征设计的依赖,为课堂质量监督提供更具鲁棒性和智能化的支撑。

参考文献:

- [1] NWAKOBY C, ILOKA P. Regulatory challenges in implementing digital education in anambra state for sustainable development[J]. Unizik journal of educational laws and leadership studies, 2025, 1(1): 224–236.
- [2] 张炜, 刘怡. 中国高等学校分类的背景与思路——基于实施《教育强国建设规划纲要》的思考[J]. 高等工程教育研究, 2025(5): 1–8, 45.
ZHANG Wei, LIU Yi. Background and approach to classification of Chinese higher education institutions—based on the implementation of master plan on building China into a leading country in education[J]. Research in higher education of engineering, 2025(5): 1–8, 45.
- [3] SHI Yuxuan, HUANG Wen, SANG Yijing. A narrative review of utilizing generative artificial intelligence in classroom instructions[C]//2024 4th International Conference on Educational Technology. Wuhan: IEEE, 2024.
- [4] KUANG Yuanyuan, TANG Ying, XIE Tao. Objectives, methods, and challenges of applying intelligent assessment in education: a systematic review[C]//2024 International Symposium on Educational Technology. Macao: IEEE, 2024.
- [5] SHEHZADI T, HASHMI K A, LIWICKI M, et al. Object detection with transformers: a review[J]. Sensors, 2025, 25(19): 6025.
- [6] 广田薰, 周政, MERSHA Bemnet Wondimagegnehu, 等. 教室环境下的教学质量 AI 机器评估研究进展[J]. 北京

- 理工大学学报, 2024, 44(10): 1077–1088.
- HIROTA Kaoru1, ZHOU Zheng, MERSHA Bemnet Wondimagegnehu, et al. AI machine evaluation level on classroom teaching quality[J]. *Transactions of beijing institute of technology*, 2024, 44(10): 1077–1088.
- [7] 谭苏燕, 王祖焯, 何高大, 等. 基于 CA-YOLOv9 网络的实时全景多尺度课堂行为识别[J]. *现代教育技术*, 2024, 34(7): 123–130.
- TAN Suyan, WANG Zuxuan, HE Gaoda. Real-time panoramic multi-scale classroom behaviors recognition based on CA-YOLOv9 network[J]. *Modern educational technology*, 2024, 34(7): 123–130.
- [8] 张浩鹏, 施铮, 刘峰, 等. CPViG-Net: 基于局部跨阶段视觉图卷积的学生课堂行为识别[J]. *计算机科学*, 2026, 53(2): 57–66.
- ZHANG Haopeng, SHI Zheng, LIU Feng. CPViG-Net: students' classroom behavior recognition based on cross-stage visual graph convolution[J]. *Computer science*, 2026, 53(2): 57–66.
- [9] 胡锦林, 齐永锋, 王佳颖. 基于时空图卷积网络的学生在线课堂行为识别[J]. *光电子·激光*, 2022, 33(2): 149–156.
- HU Jinlin, QI Yongfeng, WANG Jiaying. Recognition of Students' Online Classroom Action Based on Spatio-temporal Graph Convolutional Network[J]. *Optoelectronics & laser*, 2022, 33(2): 149–156.
- [10] 周珍玉, 秦学, 蔡芳, 等. 基于人物交互的学生课堂行为识别研究[J]. *现代教育技术*, 2024, 34(2): 53–61.
- ZHOU Zhenyu, QIN Xue, CAI Fang, et al. Research on classroom behavior recognition of students based on human-object interaction[J]. *Modern educational technology*, 2024, 34(2): 53–61.
- [11] LU Sheng, GUO Yangming, LONG Jiang, et al. Dense small target detection algorithm for UAV aerial imagery [J]. *Image and vision computing*, 2025, 156: 105485.
- [12] CHEN Yuming, YUAN Xinbin, WANG Jiabao, et al. YOLO-MS: rethinking multi-scale representation learning for real-time object detection[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2025, 47(6): 4240–4252.
- [13] 刘建伟, 刘媛, 罗雄麟. 半监督学习方法[J]. *计算机学报*, 2015, 38(8): 1592–1617.
- LIU Jianwei, LIU Yuan, LUO Xionglin. Semi-supervised learning methods[J]. *Chinese journal of computers*, 2015, 38(8): 1592–1617.
- [14] XU Bowen, CHEN Mingtao, GUAN Wenlong, et al. Efficient teacher: semi-supervised object detection for YOLOv5[EB/OL]. (2023–02–15)[2025–05–20]. <https://arxiv.org/abs/2302.07577>.
- [15] MI Peng, LIN Jianghang, ZHOU Yiyi, et al. Active teacher for semi-supervised object detection[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022.
- [16] LI Jiaming, LIN Xiangru, ZHANG Wei, et al. Gradient-based sampling for class imbalanced semi-supervised object detection[C]//2023 IEEE/CVF International Conference on Computer Vision. Paris: IEEE, 2023.
- [17] LIU Yen-chen, MA Chih-yao, KIRA Z. Unbiased teacher v2: semi-supervised object detection for anchor-free and anchor-based detectors[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022.
- [18] 徐彦威, 李军, 董元方, 等. YOLO 系列目标检测算法综述[J]. *计算机科学与探索*, 2024, 18(9): 2221–2238.
- XU Yanwei, LI Jun, DONG Yuanfang, et al. Survey of development of YOLO object detection algorithms[J]. *Journal of frontiers of computer science and technology*, 2024, 18(9): 2221–2238.
- [19] ALHASHMI S, AL-AZAWI A. A review of the single-stage vs. two-stage detectors algorithm: comprehensive insights into object detection[J]. *International journal of environmental sciences*, 2025, 11(3s): 775–787.
- [20] KHANAM R, HUSSAIN M. YOLOv11: an overview of the key architectural enhancements[EB/OL]. (2024–10–23)[2025–08–20]. <https://arxiv.org/abs/2410.17725>.
- [21] 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.
- [22] SUN Kailai, WANG Xinwei, MIAO Xi, et al. A review of AI edge devices and lightweight CNN and LLM deployment[J]. *Neurocomputing*, 2025, 614: 128791.
- [23] ALONSO J, ORUE-ECHEVARRIA L, CASOLA V, et al. Understanding the challenges and novel architectural models of multi-cloud native applications—a systematic literature review[J]. *Journal of cloud computing*, 2023, 12(1): 6.
- [24] ABNAR S, SHAH H, BUSBRIDGE D, et al. Parameters vs FLOPs: scaling laws for optimal sparsity for mixture-of-experts language models[EB/OL]. (2025–01–21)[2025–08–20]. <https://arxiv.org/abs/2501.12370>.
- [25] MAHMOUD E, BAWANEH A K. Best practices of effective classroom management strategies supported by digital ICT in higher education[J]. *International journal of evaluation and research in education*, 2025, 14(3): 2337.
- [26] QU Feilong, ALI KHAN N, JHANJHI N Z, et al. Improved YOLOv5 lane line real time segmentation system integrating seg head network[C]//The 7th International Global Conference Series on ICT Integration in Technical Education & Smart Society. Aizuwakamatsu City: MDPI, 2025.

- [27] CAI Boyu, WANG Huibing, YAO Mingze, et al. Focus more on what? guiding multi-task training for end-to-end person search[J]. *IEEE transactions on circuits and systems for video technology*, 2025, 35(7): 7266–7278.
- [28] KLINKER F. Exponential moving average versus moving exponential average[J]. *Mathematische semesterberichte*, 2011, 58(1): 97–107.
- [29] DEVERNAY F. A non-maxima suppression method for edge detection with sub-pixel accuracy[D]. Rocquencourt: INRIA, 1995.
- [30] 国家标准化管理委员会. 中小学、幼儿园安全防范要求[S]. 北京: 中国标准出版社, 2022.
Standardization Administration of China. Safety requirements for primary, secondary schools and kindergartens [S]. Beijing: China standards press, 2022.
- [31] LIU Yen-cheng, MA Chih-yao, HE Zijian, et al. Unbiased teacher for semi-supervised object detection [EB/OL]. (2023–10–04)[2025–08–20]. <https://arxiv.org/abs/2102.09480>.
- [32] YANG Fan, WANG Tao. SCB-Dataset3: a benchmark for detecting student classroom behavior[EB/OL]. (2023–10–04)[2025–08–20]. <https://arxiv.org/abs/2310.02522>.

作者简介:



张睿,教授,CCF 高级会员,主要研究方向为计算机视觉、智能信息处理。先后主持教育部人文社科项目、山西省重点研发计划子课题等科研项目 20 余项,获授权国家发明专利 16 项,发表学术论文 40 余篇。E-mail: zhangrui@tyust.edu.cn。



张延春,硕士研究生,主要研究方向为计算机视觉。E-mail: Springer0415@163.com。



张佳琪,硕士研究生,主要研究方向为智能信息处理。E-mail: 1004366758@qq.com。

2026 年智能检测与运动控制技术国际会议征稿通知

2026 年 9 月 18—20 日,由南京理工大学、中国人工智能学会联合主办的“2026 年国际智能检测与运动控制技术国际会议”将在南京召开。

一、重要信息

会议官网: www.idmct.com

投稿网址: <https://www.ais.cn/attendees/index/JMUJZ2?invite=CPNTU>

主办方投稿邀请码: CPNTU(投稿减免 400 元/篇)

二、征稿主题

本会议面向广大学者征集稿件,主题包括但不限于:

主题一 智能检测技术:

同步定位与建图(SLAM)、视觉-惯性里程计(VIO)、激光雷达感知与点云处理、基于视觉的目标检测与跟踪、多传感器融合与标定、面向实时感知的边缘人工智能、物联网感知与无线传感网络、异常/故障检测与状态监测、语义地图构建与场景理解、多模态感知(视觉/激光/惯性/雷达)

主题二 运动控制技术:

模型预测控制(MPC)、强化学习驱动的控制、最优与轨迹规划、多智能体协同控制、分布式/网络化控制系统、容错与鲁棒控制、自主导航与避障控制、实时嵌入式控制系统、信息物理系统控制与集成、人机交互与共享控制

...

如不确定主题是否符合,请联系文末咨询通道的会议秘书林老师。

三、论文出版

本会议所有的投稿都必须经过 2~3 位组委会专家审稿,经过严格的审稿之后,最终所录用的论文将被递交 ACM(ISBN: 979-8-4007-3194-5)出版社,见刊后由出版社整理提交 EI Compendex、Scopus 检索。

四、联系我们

投稿咨询: 林老师

电话/微信: 13922153442(欢迎微信咨询)