Computer Science and Application 计算机科学与应用, 2026, 16(2), 40-49
Published Online February 2026 in Hans. https://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2026.162037

Hans汉斯

基于YOLOv11与空间特征增强的学生课堂行为检测方法研究

康雯婷[1]，平　源[2,3*]，林　琳[1]，李乐俭[2]

[1]吉林化工大学信息与控制工程学院，吉林　吉林
[2]许昌学院信息工程学院，河南　许昌
[3]河南省大数据安全与应用工程技术研究中心，河南　许昌

收稿日期：2025年12月25日；录用日期：2026年1月22日；发布日期：2026年1月29日

摘　要

针对课堂场景中学生行为目标尺度较小、姿态变化幅度有限及不同行为在外观特征上有较高相似性等问题，本文构建了一种新的YOLOv11-RMS模型。该模型以YOLOv11为基础，在主干网络中引入基于Restormer的特征提取模块，通过建模长程依赖关系增强特征的全局建模能力；其次，在主干网络末端融合多级通道注意力机制(MLCA)，以强化关键语义特征并抑制冗余信息；最后，在检测头阶段引入基于空间自适应特征调制的SAFMP模块，对上采样后的浅层特征进行重构与增强，提升模型在复杂课堂环境下检测稳定性。实验结果表明，YOLOv11-RMS模型在SCB数据集和自建学生课堂行为数据集CLASS上的mAP50分别达到72.6%和84.9%，较主流模型取得了明显提升。

关键词

课堂行为，目标检测，YOLOv11，注意力机制，特征增强

Research on Student Classroom Behavior Detection Method Based on YOLOv11 with Spatial Feature Enhancement

Wenting Kang[1], Yuan Ping[2,3*], Lin Lin[1], Lejian Li[2]

[1]College of Information and Control Engineering, Jilin University of Chemical Technology, Jilin Jilin
[2]School of Information Engineering, Xuchang University, Xuchang Henan
[3]Henan Province Engineering Technology Research Center of Big Data Security and Applications, Xuchang Henan

Received: December 25, 2025; accepted: January 22, 2026; published: January 29, 2026

*通讯作者。

文章引用: 康雯婷, 平源, 林琳, 李乐俭. 基于 YOLOv11 与空间特征增强的学生课堂行为检测方法研究[J]. 计算机科学与应用, 2026, 16(2): 40-49. DOI: 10.12677/csa.2026.162037

Abstract

To address the challenges of small target scales, limited posture variations, and high visual similarity among different student behaviors in classroom scenarios, a novel YOLOv11-RMS model is proposed. Based on the YOLOv11 framework, a Restormer-based feature extraction module is incorporated into the backbone to enhance global feature modeling through long-range dependency learning. Subsequently, a multi-level channel attention mechanism (MLCA) is integrated at the end of the backbone to strengthen discriminative semantic features while suppressing redundant information. Furthermore, a spatially adaptive feature modulation module (SAFMP) is introduced in the detection head to reconstruct and enhance upsampled shallow features, thereby improving detection robustness in complex classroom environments. Experimental results demonstrate that the proposed YOLOv11-RMS model achieves mAP50 scores of 72.6% and 84.9% on the SCB dataset and the self-collected classroom behavior dataset (CLASS), respectively, outperforming existing mainstream methods.

Keywords

Classroom Behavior, Object Detection, YOLOv11, Attention Mechanism, Feature Enhancement

1. 引言

随着智慧教育的不断推进，利用计算机视觉技术对课堂教学过程进行自动化分析，逐渐成为教育技术与人工智能交叉领域的重要研究方向。学生课堂行为作为反映课堂教学状态与学生学习表现的重要外在特征，其自动检测与识别对于课堂管理、教学行为分析以及智慧课堂系统的构建具有重要的应用价值[1]。

目标检测任务的现有方法可分为双阶段检测和单阶段检测两类。双阶段目标检测方法先生成候选区域再进行分类和定位优化，其在精度上具有一定优势，但其计算开销较大，如 R-CNN [2]、Faster R-CNN [3]等方法。单阶段方法将定位与分类统一为一次回归，能够实现更高的推理速度和实时性，如 SSD [4]、YOLO [5]系列。在学生课堂行为检测领域，研究者们通过改进目标检测模型的结构，以应对小尺度目标和复杂背景等带来的挑战。例如，Chen 等人[6]在 YOLO-v4 中引入连接模块和 Repulsion Loss，以缓解遮挡导致的误检和漏检。Jia 等人[7]将 YOLOv5 与坐标注意力结合，并配合 OpenPose 姿态估计提升检测效果。Peng 等人[8]提出基于 YOLOv10s 的 YOLO-CBD 模型，通过注意力机制、特征聚合和自适应损失优化多尺度目标检测表现。Sheng 等人[9]提出基于 YOLO 架构，通过多尺度大卷积核模块与特征优化模块加强不同尺度信息的建模能力。尽管上述方法在一定程度上提升了检测性能，但在远距离学生目标尺度较小、不同课堂行为在外观特征上高度相似的复杂场景下，仍然存在较为显著的误检和漏检现象。

针对上述问题，本文以 YOLOv11 [10]为基础架构，通过对特征提取、语义建模与空间信息表达进行协同增强，改善模型在复杂课堂场景下对关键行为区域特征的表达能力，进而构建了一种新的 YOLOv11-RMS 模型，为学生课堂行为检测提供了一种可行的技术方案。

2. YOLOv11-RMS 模型

2.1. YOLOv11

YOLOv11 整体延续了 YOLO 系列端到端、单阶段检测的设计思想。模型采用了 C3k2 模块以增强特

征提取与融合能力。C2PSA 结构通过空间注意力机制增强特征图在空间维度上的响应能力，使网络能够更加关注具有判别意义的空间区域。颈部网络采用 PAN 结构，通过自顶向下与自底向上的双向特征路径，使模型能够在融合高层语义信息的同时有效保留浅层空间细节，从而提升不同尺度特征之间的交互质量。在检测头方面，对分类与回归任务进行解耦建模，有助于缓解不同任务之间的优化冲突，提高边界框回归精度和类别预测的稳定性。

2.2. YOLOv11-RMS

2.2.1. 整体框架

YOLOv11-RMS 模型整体架构如图 1 所示。首先，模型在特征提取阶段引入基于 Restormer [11]的特征建模单元，以增强对长程依赖关系的刻画能力。其次，在特征融合与增强阶段，将 C2PSA 结构与混合局部–全局通道注意力机制 MLCA [12] (Mixed Local Channel Attention)相结合，利用局部与全局通道统计信息的协同建模。同时，引入基于空间自适应调制思想的特征增强模块 SAFM [13] (Spatially-Adaptive Feature Modulation)，对特征进行进一步重构与增强，以提高模型在复杂课堂场景中的表达稳定性与鲁棒性。整体结构在保证计算效率的同时，实现了对关键行为区域信息的更有效表征。

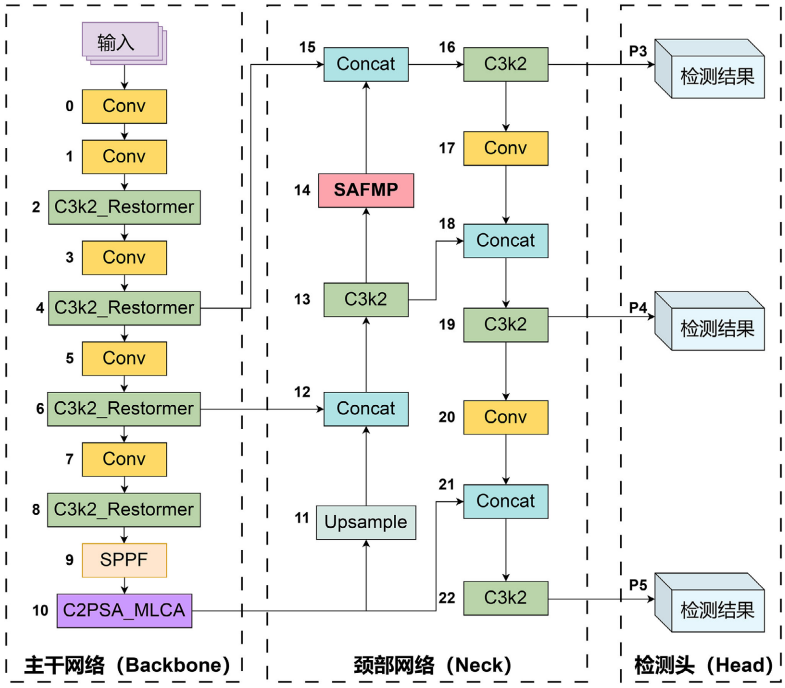

Figure 1. Framework of the YOLOv11-RMS architecture
图 1. YOLOv11-RMS 结构图

2.2.2. C3k2_Restormer

Restormer 核心思想在于通过重构自注意力计算方式，在保持全局建模能力的同时显著降低计算复杂度。此外，Restormer 通过引入深度可分离卷积结构，使注意力计算同时具备局部感受野建模能力，从而在效率与表达能力之间取得良好平衡。在课堂行为检测任务中，传统卷积网络在建模长距离依赖与全局上下文方面存在一定局限。为此，本文在主干网络中引入了 C3k2_Restormer 模块，其整体结构如图 2 所示。该模块首先通过卷积层对输入特征进行通道调整与基础特征提取，随后引入由多尺度动态注意力(Multi-Dconv Head Transposed Attention, MDTA)和门控深度前馈网络(Gated-Dconv Feed-Forward Network,

GDFN)组成的特征增强单元。其中，MDTA 模块通过在通道维度上构建全局相关性，实现对长距离依赖关系的有效建模，从而弥补卷积操作在建模全局语义信息方面的不足；GDFN 模块则通过门控机制对特征进行非线性变换与筛选，在增强判别性特征表达的同时抑制冗余信息；MDTA 与 GDFN 通过残差连接嵌入到网络中，有效缓解深层网络训练过程中的梯度消失问题，并增强特征传递的稳定性。通过将该结构引入主干网络(图 1 的 Backbone 部分)，模型不仅能够充分利用局部细节信息，还能够捕获更具判别力的全局上下文特征，从而显著提升在复杂场景下的特征表达能力与检测性能。

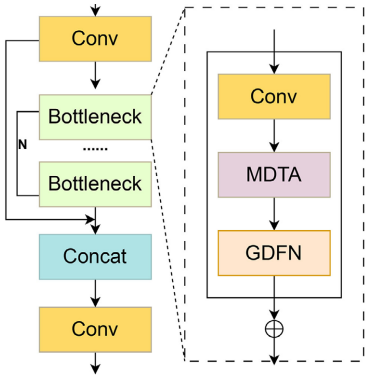

Figure 2. Architecture of the C3k2_Restormer module
图 2. C3k2_Restormer 模块结构图

2.2.3. C2PSA_MLCA

C2PSA_MLCA 模块是在原有 C2PSA 结构基础上引入混合局部–全局的通道注意力 MLCA，旨在进一步提升复杂场景下的特征建模能力与表达可靠性，其结构如图 3 所示。MLCA 通过对输入特征分别进行局部池化与全局池化，提取不同尺度下的通道响应信息，并通过一维卷积对通道间相关性进行建模，进而生成融合局部与全局信息的通道权重。该权重用于对特征进行重标定，从而突出具有判别性的语义特征并抑制冗余信息。将 MLCA 嵌入至 C2PSA 的特征变换分支中，并在 PSABlock 内作为前馈增强单元参与特征更新，使其与原有的多分支特征融合机制协同工作。一方面，C2PSA 通过并行结构实现多尺度信息交互与特征聚合，为后续检测提供更丰富的候选表征；另一方面，MLCA 对聚合后的特征进行通道级重标定，使网络在特征融合过程中能够更有效地保留有用信息并抑制背景干扰。二者的结合在保证计算效率的同时，有效增强了特征表达的判别性与稳定性，为后续目标检测任务提供更加可靠的特征基础。

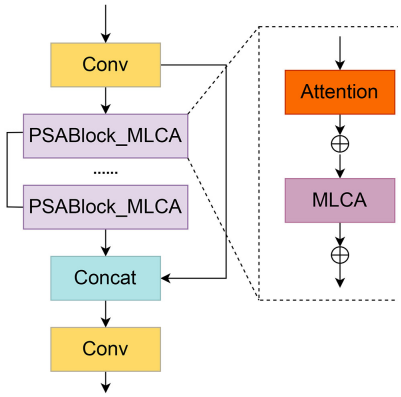

Figure 3. Architecture of the C2PSA_MLCA module
图 3. C2PSA_MLCA 模块结构图

2.2.4. SAFMP

基于 SAFM 的空间自适应特征建模思想，本文在其基础上引入了一种更加轻量化的特征增强模块 SAFMP (Spatially-Adaptive Feature Modulation Package)，如图 4 所示。该模块首先通过卷积映射对输入特征进行通道维度统一，为后续特征建模提供一致的表示空间。随后，引入空间自适应特征调制单元，对特征进行重加权。该过程通过结合局部区域的响应信息与由下采样操作所引入的全局上下文信息，实现对不同空间位置特征的重要性建模，从而增强模型对关键区域的感知能力并抑制冗余背景响应。在完成空间维度上的特征增强后，特征进一步输入通道混合模块(Convolutional Channel Mixing, CCM)，以建模通道之间的相关性。该模块通过卷积形式实现跨通道信息交互，相较于传统基于全连接结构的通道建模方式，在保证表达能力的同时显著降低了计算开销，并增强了局部特征的建模能力。最后，模块采用基于像素重排(Pixel Shuffle)的方式对特征进行空间重构，从而恢复高分辨率特征表示，使增强后的特征更有利于后续检测头进行判别与定位。

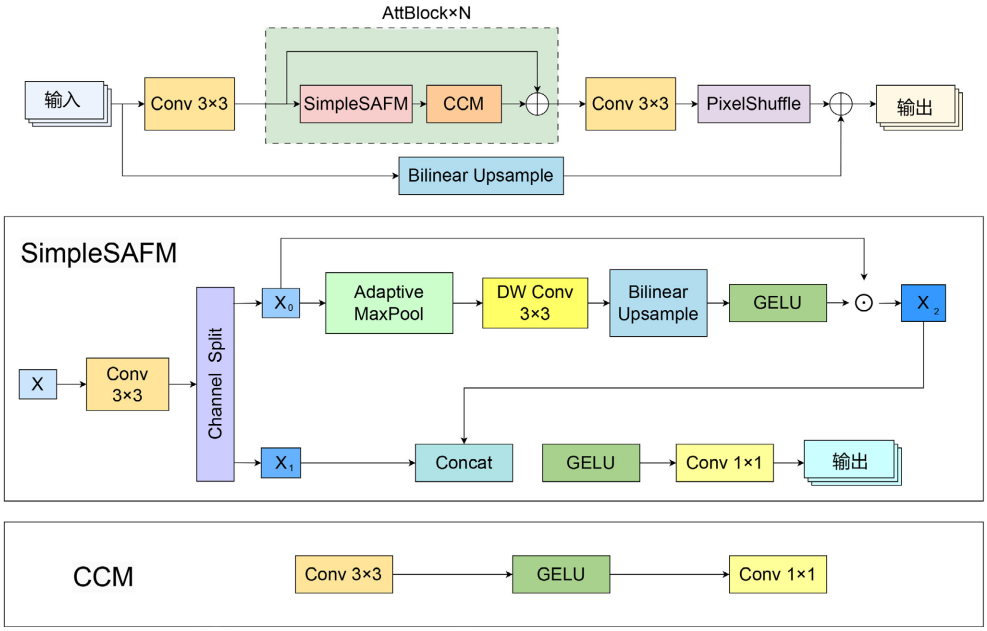

Figure 4. Architecture of the SAFMP module
图 4. SAFMP 模块结构图

3. 实验与分析

3.1. 数据集及实验环境

3.1.1. 数据集

本文实验采用公开数据集 SCB [14]和自建学生课堂行为数据集 CLASS 进行验证。自建数据集 CLASS 共包含 12 类检测目标，分别为使用手机(using_phone)、端坐(upright)、举手(hand-raising)、阅读(reading)、身体弯曲(bend)、低头(bow_head)、转头(turn_head)、书写(writing)、抬头(raise_head)、睡觉(sleep)等学生课堂行为及手机(phone)、书本(book)等与行为高度相关的物品类别。本文对 CLASS 数据集中各类别样本数量进行了统计，结果如表 1 所示。所有数据在采集与使用过程中均进行了匿名化与去标识化处理，仅用于学术研究目的。代表样例如图 5 所示。为保护学生隐私安全，样例中涉及的人脸区域均进行脱敏处理。

Table 1. Number of samples for each class in the CLASS dataset
表 1. CLASS 数据集各类别样本数量

类别	手机	睡觉	低头	使用手机	书本	书写
数量	6879	2239	54,284	19,031	24,372	10,298
类别	转头	阅读	弯曲身体	端坐	抬头	举手
数量	4635	33,963	16,861	61,945	15,891	27,231

Figure 5. Representative samples of each category
图 5. 类别代表样例

3.1.2. 实验环境及实验方案设置

相关实验在 NVIDIA GeForce RTX 4060 Laptop GPU 平台上完成。模型实现基于 Python 版本为 3.9.21，采用 PyTorch 2.2.0 深度学习框架，并通过 CUDA 12.1 进行加速。训练阶段统一将输入图像分辨率调整为 640 × 640，优化器采用 SGD，初始学习率为 0.01，动量因子设置为 0.937，权重衰减系数为 0.0005，批大小设置为 16。

3.2. 评价指标

为衡量模型在学生课堂行为检测任务中的检测准确性与整体性能，本文采用精确率(Precision, P)、召回率(Recall, R)、平均精度(Average Precision, AP)、平均精度均值(mAP)作为模型性能评价指标，计算方法如式(1)~(4)所示。其中，TP 表示模型正确地将正类预测为正类的数量。FP 表示模型错误地将负类预测为正类的数量。FN 表示模型错误地将正类预测为负类的数量。

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{1}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

$$\text{AP} = \int_0^1 \text{P}(\text{R})\,\text{dR} \tag{3}$$

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{4}$$

本文同时引入参数量(Params)、计算复杂度(GFLOPs)以及推理速度(Frames Per Second, FPS)作为效率评价指标。其中，Params 用于衡量模型的规模与存储开销。GFLOPs 表示模型单次前向推理所需的浮点运算量。FPS 用于衡量模型在实际推理阶段的实时处理能力。

3.3. 实验结果及分析

3.3.1. 训练过程与收敛性分析

为分析所提出模型在训练过程中的优化特性与收敛行为，本文对模型训练阶段的损失变化情况进行了系统分析。如图 6 所示，YOLOv11-RMS 在自建数据集 CLASS 上的训练与验证过程中表现出良好的收敛特性。随着训练轮次的增加，训练阶段的 box_loss、cls_loss 和 dfl_loss 均呈现持续下降趋势，且两者之间保持较小差距，这表明模型在学习过程中具有较好的稳定性与泛化能力。边界框回归损失与分类损失均在前期快速下降，随后逐渐趋于平稳，说明模型能够有效学习目标的位置信息与类别特征。精确率、召回率以及 mAP50 和 mAP50-95 均随训练轮次增加而逐步提高，并在训练后期趋于收敛。整体来看，改进模型在自建数据集上的训练过程稳定且能够有效学习特征，为后续性能对比实验提供了可靠基础。

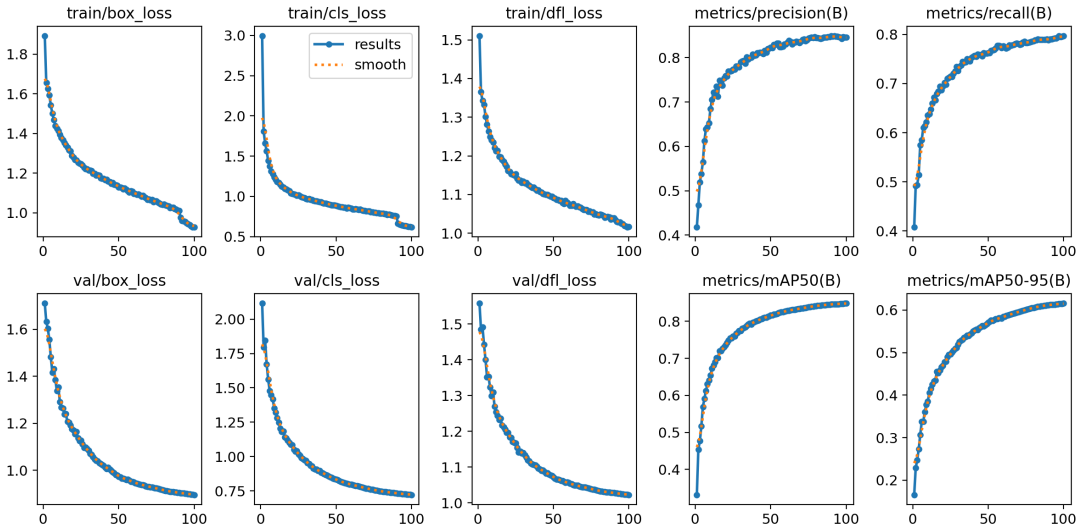

Figure 6. Changes of training loss over epochs
图 6. 训练损失随轮次的变化

3.3.2. CLASS 数据集上的性能对比

为评估在自建数据集 CLASS 上的检测性能，表 2 给出了 YOLOv11-RMS 在不同行为类别上的检测结果。其中，mAP 指标反映了模型在整体类别层面的综合检测性能，各行为类别的平均精度值用于衡量模型对不同课堂行为的检测性能。可以看出，改进模型在端坐、低头、抬头等姿态特征明显的行为类别上取得了较高的平均精度，说明模型能够较好地捕获此类行为的关键空间特征。手机目标尺度较小且易被手部等物体遮挡，导致边界框回归结果不稳定并伴随预测置信度偏低。“使用手机”与“阅读”“书写”等类别高度相似，因而易产生误检现象。尽管引入的 C3k2_Restormer 与 C2PSA_MLCA 提升了特征的全局依赖建模能力和通道语义选择效果，但在极小目标与弱可见性条件下仍受到输入特征信息不足的制约。后续工作可在现有模型基础上引入更高分辨率特征层，以保留更多低层空间细节信息，从而进一

步提升小目标的定位精度与检测稳定性。

Table 2. Detection performance of YOLOv11-RMS on various classroom behavior
表 2. YOLOv11-RMS 在各类课堂行为上的检测性能

模型	mAP50	AP			
YOLOv11-RMS	84.9	手机	睡觉	低头	使用手机
		59	91.2	95.6	77.6
		书本	书写	转头	阅读
		89.4	80.6	75.8	81.3
		弯曲身体	端坐	抬头	举手
		91.7	97.6	95.5	83

表 3 对比了 YOLOv11-RMS 与 YOLOv5s [15]、YOLOv8n [16]等主流目标检测模型在 CLASS 数据集上的检测结果。实验结果表明，YOLOv11 在该数据集上的整体性能优于 YOLOv5s 和 YOLOv8n，说明更强的特征建模能力在复杂课堂场景中具有一定优势。改进后的模型在各项评价指标上均取得提升，其中召回率由 78.6%提升至 79.7%，mAP50 由 84.1%提升至 84.9%，mAP50-95 由 60.4%提升至 61.6%。以上结果表明，模型通过 C3k2_Restormer 模块，在特征提取阶段更充分地建模全局语义信息，结合 C2PSA_MLCA 模块提升了高层特征的判别能力，SAFMP 模块进一步增强了对局部细节与空间结构的刻画能力。

Table 3. Performance comparison of mainstream models on the CLASS dataset
表 3. 主流模型在 CLASS 数据集上性能对比

模型	Precision (%)	Recall (%)	mAP50 (%)	mAP50-95 (%)
YOLOv5s	82.3	76.7	81.3	59.8
YOLOv8n	83.9	78.2	82.4	60.1
YOLOv11n	84.4	78.6	84.1	60.4
YOLOv11-RMS	**84.6**	**79.7**	**84.9**	**61.6**

3.3.3. SCB 数据集上的性能对比

为进一步验证 YOLOv11-RMS 在不同数据分布条件下的泛化能力，本文在公开数据集 SCB 上开展了对比实验，实验结果如表 4 所示。与 YOLOv11 原始模型相比，YOLOv11-RMS 在精确率、召回率以及 mAP50 指标上均取得了一定提升，其中精确率由 70.9%提升至 71.5%，召回率由 63.8%提升至 65.2%，mAP50 由 71.1%提升至 72.6%。在 mAP50-95 指标上，改进模型同样表现出小幅提升。YOLOv11-RMS 在判别目标与背景时更加严格，从而在一定程度上抑制了噪声区域和低置信度预测。但在部分边界模糊或尺度较小的目标上，可能导致少量真阳性样本被过滤，从而使召回率略低于 YOLOv8。

Table 4. Performance comparison of mainstream models on the SCB dataset
表 4. 主流模型在 SCB 数据集上性能对比

模型	Precision (%)	Recall (%)	mAP50 (%)	mAP50-95 (%)
YOLOv5s	61.8	67.8	67.9	47.5
YOLOv8n	63.9	**69.6**	72.2	**52.1**
YOLOv11n	70.9	63.8	71.1	51.4
YOLOv11-RMS	**71.5**	65.2	**72.6**	51.8

3.3.4. 消融实验

为验证引入的各模块对模型性能的影响，以 YOLOv11 为基准模型，逐项加入 C3k2_Restormer、C2PSA_MLCA 模块、SAFMP 模块来进行消融实验，实验结果如表 5 所示。从表中可以看出，基准模型在未引入任何改进模块时，mAP50 和 mAP50-95 分别为 84.1%和 60.4%，同时具备较高的推理速度和较低的模型复杂度。引入 C3k2_Restormer 模块后，尽管模型参数量和计算量有所增加，但 mAP50 和 mAP50-95 分别提升至 84.4%和 60.9%，表明基于 Restormer 的特征建模能够增强特征的全局表达能力。在此基础上进一步引入 C2PSA_MLCA 模块，模型检测性能继续提升，多层通道注意力机制对关键语义特征具有增强作用。当三种模块同时引入时，模型取得最优检测性能，mAP50 和 mAP50-95 分别达到 84.9%和 61.6%。综合来看，多模块协同引入后，模型在检测精度与模型复杂度之间取得了较为合理的平衡，验证了本文改进策略的有效性。

Table 5. Ablation experiment results
表 5. 消融实验结果

C3k2_Restormer	C2PSA_MLCA	SAFMP	Params	GFLOPs	FPS	mAP50 (%)	mAP50-95 (%)
×	×	×	2.59	6.5	833.3	84.1	60.4
√	×	×	6.35	29.1	588.2	84.4	60.9
√	√	×	6.28	29.1	588	84.7	61.3
√	√	√	9.48	39.3	500	**84.9**	**61.6**

4. 结论

本文围绕课堂场景下学生行为检测任务，对目标检测模型的结构进行了系统性改进，提出了基于 YOLOv11 的改进模型 YOLOv11-RMS。通过在主干网络中引入基于 Restormer 的特征增强模块，并在特征融合阶段融合多层通道注意力机制，有效提升了模型对复杂场景中多尺度目标的表征能力。实验结果表明，所提出的方法在保持较高检测精度的同时，进一步增强了模型在复杂背景下的鲁棒性与稳定性。未来工作可进一步从模型轻量化等方面展开深入研究。

致　谢

谨向所有给予支持与帮助的人员致以诚挚的感谢。感谢在数据采集与整理过程中给予帮助的相关成员。正是他们在技术支持和资源保障方面的付出，使得实验工作得以高效、有序地开展。团队成员之间良好的合作氛围和严谨的学术态度，为本研究的顺利推进创造了良好的条件。

基金项目

河南省科技攻关计划(242102211065、252102211075)，河南省创新型科技人才队伍建设工程(CXTD2017099)，河南省研究生教学质量工程项目(YJS2025GZZ36, YJS2024AL112, YJS2024JD38)。

参考文献

[1] Liu, Q., Jiang, X. and Jiang, R. (2025) Classroom Behavior Recognition Using Computer Vision: A Systematic Review. *Sensors*, **25**, Article 373. https://doi.org/10.3390/s25020373

[2] Girshick, R., Donahue, J., Darrell, T. and Malik, J. (2014) Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. 2014 *IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, 23-28 June 2014, 580-587. https://doi.org/10.1109/cvpr.2014.81

[3] Ren, S., He, K., Girshick, R. and Sun, J. (2017) Faster R-CNN: Towards Real-Time Object Detection with Region

Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 1137-1149. https://doi.org/10.1109/tpami.2016.2577031

[4] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C., *et al*. (2016) SSD: Single Shot MultiBox Detector. In: *Lecture Notes in Computer Science*, Springer, 21-37. https://doi.org/10.1007/978-3-319-46448-0_2

[5] Redmon, J., Divvala, S., Girshick, R. and Farhadi, A. (2016) You Only Look Once: Unified, Real-Time Object Detection. 2016 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Las Vegas, 27-30 June 2016, 779-788. https://doi.org/10.1109/cvpr.2016.91

[6] Chen, H. and Guan, J. (2022) Teacher-Student Behavior Recognition in Classroom Teaching Based on Improved YOLO-V4 and Internet of Things Technology. *Electronics*, **11**, Article 3998. https://doi.org/10.3390/electronics11233998

[7] Jia, Q. and He, J. (2024) Student Behavior Recognition in Classroom Based on Deep Learning. *Applied Sciences*, **14**, Article 7981. https://doi.org/10.3390/app14177981

[8] Peng, S., Zhang, X., Zhou, L. and Wang, P. (2025) YOLO-CBD: Classroom Behavior Detection Method Based on Behavior Feature Extraction and Aggregation. *Sensors*, **25**, Article 3073. https://doi.org/10.3390/s25103073

[9] Sheng, X., Li, S. and Chan, S. (2025) Real-Time Classroom Student Behavior Detection Based on Improved YOLOv8s. *Scientific Reports*, **15**, Article No. 14470. https://doi.org/10.1038/s41598-025-99243-x

[10] Khanam, R. and Hussain, M. (2024) YOLOv11: An Overview of the Key Architectural Enhancements. https://www.arxiv.org/pdf/2410.17725

[11] Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S. and Yang, M. (2022) Restormer: Efficient Transformer for High-Resolution Image Restoration. 2022 *IEEE*/*CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), New Orleans, 18-24 June 2022, 5718-5729. https://doi.org/10.1109/cvpr52688.2022.00564

[12] Wan, D., Lu, R., Shen, S., Xu, T., Lang, X. and Ren, Z. (2023) Mixed Local Channel Attention for Object Detection. *Engineering Applications of Artificial Intelligence*, **123**, Article 106442. https://doi.org/10.1016/j.engappai.2023.106442

[13] Sun, L., Dong, J., Tang, J. and Pan, J. (2023) Spatially-Adaptive Feature Modulation for Efficient Image Super-Resolution. 2023 *IEEE*/*CVF International Conference on Computer Vision* (*ICCV*), Paris, 2-6 October 2023, 13144-13153. https://doi.org/10.1109/iccv51070.2023.01213

[14] Yang, F. (2023) SCB-Dataset: A Dataset for Detecting Student and Teacher Classroom Behavior. https://doi.org/10.48550/arXiv.2304.02488

[15] Khanam, R. and Hussain, M. (2024) What Is YOLOv5: A Deep Look into the Internal Features of the Popular Object Detector. https://doi.org/10.48550/arXiv.2407.20892

[16] Varghese, R. and Sambath, M. (2024) YOLOv8: A Novel Object Detection Algorithm with Enhanced Performance and Robustness. 2024 *International Conference on Advances in Data Engineering and Intelligent Computing Systems* (*ADICS*), Chennai, 28-29 March 2024, 1-6. https://doi.org/10.1109/adics58448.2024.10533619