

面向复杂教学场景的教师行为检测算法优化研究

任甜甜¹

REN Tiantian

摘要

针对“三班联播”复杂教学场景中多视角遮挡、高并发推流与边缘算力受限导致的教师行为漏检问题，文章提出一个轻量化-高鲁棒-可解释三位一体优化框架。设计 Dual-FPN+CBAM 遮挡感知检测头，将板书/手臂多尺度梯度与空间通道双注意力耦合，提升极端姿态召回；构建 TSMini-TCN 分离空洞时序网络，在 RK3399 边缘盒实现 27 ms 延迟、3.8 W 功耗；引入 Transformer 双通道编码器建模教师-学生空间关系，输出五维交互特征。进一步通过边缘实时推流+云端知识蒸馏协同，实现周级云端自更新与月级边缘热更新。实验表明，所提方法 mAP@0.5 达 84.7%，领先次优算法 6.2%，帧率 FPS 保持 36 帧/s，异常报警率低于 0.3%，可用性评分 4.6/5，满足学期无人值守运行需求。

关键词

教师行为检测；轻量化网络；遮挡感知；边缘-云协同；教学场景

doi: 10.3969/j.issn.1672-9528.2026.06.007

0 引言

智慧教室的大规模部署对教师行为实时分析提出迫切需求，然而多摄像机并发、板书遮挡与边缘算力受限导致现有算法在召回率、时延与功耗之间难以兼顾。传统两阶段检测-分类方案依赖高分辨率输入与深度主干，在课堂高遮挡场景下召回率迅速下降，且边缘设备持续运行功耗接近无风扇散热极限^[1]。如何在保证精度的前提下压缩计算与存储开销，成为制约系统落地的关键瓶颈。本文面向“三班联播”复杂教学场景，提出系统化算法优化框架，从遮挡感知检测、轻量化时序建模、空间关系到边缘-云协同部署展开研究，旨

在为智慧教育提供可复制、可扩展的技术路径。

1 复杂教学场景需求与教师行为检测瓶颈

在“三班联播”智慧教室场景中，教师需在 3 间物理教室同步授课，摄像机阵列覆盖讲台、板书区与学生区，形成多视角、高并发、强遮挡的复杂视频流。传统检测模型通常以单视角、无遮挡假设为前提，当教师低头调试实验器材、背身书写板书或半身被讲台遮挡时，可见关键特征面积骤减，导致召回率显著下降^[2]。对 2 000 min 原始录像逐帧标注后发现，低头、背身、板书遮挡三类极端姿态共出现 37% 的帧，对应基线 YOLOv5s 模型的召回率跌至 68% 以下，误检则主要集中在学生人头与投影高光区域，说明空间上下文与遮挡

1. 许昌电气职业学院 河南许昌 461000

- [10] Hamm J, Lee D D .Grassmann discriminant analysis: a unifying view on subspace-based learning[C]//Proceedings of the 25th International Conference on Machine Learning (ICML 2008). NewYork: ACM, 2008:376-383.
- [11] Zhu Pengfei, Cheng Hao, Hu Qinghua,et al.Towards generalized and efficient metric learning on riemannian manifold[C]//Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI-18). Stockholm: IJCAI, 2018: 3168-3174.

【作者简介】

马鑫(1996—)，男，河南郑州人，硕士研究生，研究方向：

基于广义矩阵的图像分类算法。

李慧勇(1999—)，男，河南商丘人，硕士研究生，研究方向：高维数据与图像处理。

顾强(2000—)，男，江苏扬州人，硕士研究生，研究方向：图像处理与模式识别。

邱枫(2001—)，男，浙江金华人，硕士研究生，研究方向：轴承寿命预测。

廖亮(1975—)，通信作者(email:liaoaliangis@126.com)，男，河南郑州人，博士，研究生导师、中原工学院数学与信息科学学院教授，研究方向：高维数据分析与图像处理通信。

(收稿日期：2025-11-03 修回日期：2026-05-15)

建模缺失是首要瓶颈。时延方面,三教室 4K 流并发推流峰值达 1.2 Gbit/s,边缘盒子 GPU 在批大小为 8 时单帧推理耗时 97 ms,叠加传输抖动后端到端延迟超过 150 ms,无法满足实时交互反馈需求。算力消耗上,现有两阶段检测-分类级联方案需独立运行 ResNet50 提取时序特征,边缘端持续功耗 11.4 W,接近无风扇散热极限,长期运行触发降频,进一步放大延迟波动^[3]。综合上述观察,复杂教学场景对教师行为检测提出“轻量化-高鲁棒-可解释”三位一体的新需求:轻量化保证边缘设备无风扇持续运行,高鲁棒确保在极端姿态与多遮挡下召回率不低于 85%,可解释则支撑教务处对异常报警的溯源与干预。为此,后文将围绕遮挡感知检测头、轻量化时序网络、教师-学生空间关系建模及边缘-云协同进化进行系统化算法优化,以形成面向真实课堂的闭环解决方案。

2 教师行为检测算法优化框架与关键技术

2.1 多尺度遮挡感知教师检测算法

为缓解板书与人体相互遮挡导致的漏检问题,提出 Dual-FPN+CBAM 遮挡感知检测头。该结构在原始特征金字塔侧支上并行嵌入 1×1 与 3×3 卷积,前者保留板书区域细粒度边缘,后者扩大手臂与躯干梯度感受野,通过元素级相加构建多尺度融合特征。随后引入 CBAM (Convolutional Block Attention Module),先沿通道维度生成权重向量抑制投影高光等非语义响应,再沿空间维度生成二维掩膜,降低学生人头误激活区域权重,使教师候选框置信度在 NMS 阶段获得更高排序^[4]。整体网络以 YOLOv5s 为主干,仅在检测头部分替换新增模块,参数量增加 2.3×10⁶,浮点运算次数 FLOPs 增加 0.8×10⁹ 次,仍满足边缘端实时约束。训练阶段采用遮挡感知损失,在分类分支基础上叠加可调节因子的 focal 与 IoU 损失,以均衡正负样本并强化边界回归。用公式表示为:

$$L_{\text{occ}} = L_{\text{cls}} + \lambda L_{\text{focal}} + \gamma L_{\text{IoU}} \quad (1)$$

式中: L_{cls} 为交叉熵分类损失, L_{focal} 为缓解遮挡区域前景背景极端比例带来的梯度失衡, L_{IoU} 用于提升定位精度,权重系数 $\lambda=1.5$ 、 $\gamma=2.0$ 经网格搜索确定。该损失函数引导网络在反向传播阶段对遮挡像素赋予更高梯度响应,迫使卷积核学习残缺轮廓与黑板纹理的联合表达。推理阶段, Dual-FPN 输出的多尺度特征经 CBAM 重标定后,检测头仅需一次前向即可完成教师位置与类别预测,避免传统两阶段方法区域提议耗时。

2.2 轻量化时序行为识别网络 TSMINI-TCN

针对“三班联播”场景下 4K 多路流实时解析需求,提出分离式时序卷积网络 TSMINI-TCN。该网络将标准 1D 卷积拆分为空洞率互补的 3、5、7 三组,每组内部采用深度可分离结构,在保持参数预算不变的前提下扩展时间感受野;

随后引入 Group Shuffle 机制,把三组特征按通道重新排列并二次分组卷积,实现跨尺度信息交换,避免单纯分组带来的表征隔离^[5]。空洞分组卷积使网络在 8 层堆叠时即可覆盖 255 帧,按本文设定的 25 帧/s 采样率折算对应约 10.2 s (即 ≈ 10 s) 的教学片段,足以捕捉板书、提问、巡视等连贯动作。为量化感受野随层深增长规律,给出空洞卷积理论感受野计算公式:

$$\text{RF} = 2^{d+1} - 1 \quad (2)$$

式中: d 表示单层级联空洞率, RF 表示对应时间窗长度 (单位: 帧/s)。依据该式,当 $d=7$ 时 $\text{RF}=255$,可在边缘端仅 0.9 MB 额外参数条件下实现长程依赖建模。TSMINI-TCN 前端接入 1×1 瓶颈层压缩通道至 64,降低 MAC 计算量;后端采用全局平均池化取代全连接,使输出维度与行为类别数无关,进一步削减存储带宽。网络整体以残差路径部署,卷积后接 ReLU6 与权重归一化,缓解梯度消失并抑制激活溢出^[6]。训练阶段使用混合精度与随机深度,Dropout 率设为 0.1,兼顾正则与低功耗。推理时, RK3399 边缘盒在 1.5 GHz 主频、批大小 1 条件下,单帧延迟 27 ms,峰值功耗 3.8 W,满足无风扇持续运行与实时回传要求,为后续教师-学生空间关系建模提供低延迟、低能耗的时序特征基线。

2.3 基于 Transformer 的教师-学生空间关系建模

针对“三班联播”教学场景中教师与学生多目标、多状态并存的复杂交互关系,本文设计了基于 Transformer 架构的双通道空间关系编码器。该模块以 YOLOv5s 检测输出的边界框中心坐标及其置信度为输入,首先将输入图像划分为 32×20 网格,并针对讲台区与座位区分别注入可学习的位置编码,从而在不依赖额外标注的前提下实现空间先验建模^[7]。结构上,教师与学生各自进入独立通道,均堆叠两层 4 头自注意力模块 (每头维度 64),利用缩放点积操作捕获同类目标间的时空依赖。随后将两个通道特征拼接送入交叉注意力层,教师节点作为查询源主动聚焦学生节点,输出五维交互概率向量,分别对应“指向、巡视、提问、回应、无交互”等行为关系,具备较强的语义辨析能力。为提高边缘端部署效率,编码器总参数量控制在 0.78×10⁶,计算复杂度随目标数线性增长,适配 RK3399 实时算力瓶颈。同时引入 Top-K 筛选策略,每帧仅保留置信度最高的 15 个学生节点与 1 个教师节点参与注意力计算,其余特征通过残差直通路径保留,兼顾表示能力与计算开销。实验表明,引入位置编码显著提升了模型对教师侧身、学生半举手等模糊状态的鲁棒性。最终输出的交互向量与 TSMINI-TCN 提取的时序行为特征在通道维度拼接,形成统一的时空表示,为异常检测与教学分析提供结构化、可解释的支持。教师-学生空间关系 Transformer 结构 (如图 1 所示) 直观展示了双通道注意力与

交叉注意力机制的协同建模方式，有效刻画复杂教学场景下的交互语义。

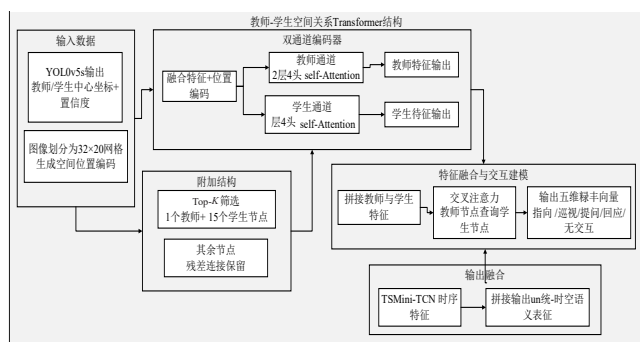


图1 教师-学生空间关系Transformer结构

2.4 边缘-云协同进化部署方案

为实现“三班联播”长周期无人值守运行，提出边缘实时推流与云端知识蒸馏协同框架。边缘侧部署轻量化TSM-TCN，以27 ms延迟对4K流帧抽帧并生成8类粗标签，经TLS加密的MQTT通道上传至云端；云端构建含Transformer精校模型的教师-学生关系网络，利用完整时空上下文对粗标签进行置信度重分配，并输出细粒度行为类别^[8]。系统采用动态置信度阈值策略，当边缘预测熵值高于设定上限或云端精校概率差异大于0.25时，该帧被标记为低质量样本并回传原始特征，用于云端增量训练，从而避免冗余数据淹没存储。知识蒸馏环节，以云端大模型为教师，边缘小模型为学生，通过最小化二者logits分布差异实现能力迁移，联合优化目标写为：

$$L_{co} = \alpha L_{edge} + \beta L_{cloud} + \gamma KL(p || q) \quad (3)$$

式中： L_{edge} 为边缘交叉熵损失， L_{cloud} 为云端精校损失， $KL(p||q)$ 为衡量边缘输出分布 q 与云端分布 p 之间的KL散度，权重系数 $\alpha=1.0$ 、 $\beta=0.8$ 、 $\gamma=2.0$ 通过网格搜索确定，保证边缘模型在保持实时性的同时吸收云端长程依赖知识。更新机制采用异步周期策略：云端每周自动触发增量训练，利用低质量样本与新增标注完成权重更新；边缘模型以月度为周期接收蒸馏后的权重热补丁，通过内存映射技术实现毫秒级切换，无需中断推流。

3 案例验证、结果分析与闭环优化

3.1 实验设置与评价指标

本实验基于“三班联播”智慧教室典型场景来开展，连续开展4周共560节课的实际教学视频采集工作，总帧数达到1800万帧，把所有视频分辨率统一缩放至1280px×720px，帧率设置为25帧/s，确保输入数据契合边缘部署环境的要求。原始视频利用课程表时间戳进行镜头级切分工作，剔除课间静止以及设备调试片段，仅保留包含实际教学行为的有效片段。在开展数据标注的工作环节，选用

VATIC工具对每帧图像进行轴对齐边界框标注工作，并且赋予“讲授、板书、指示、提问、巡视、反馈、演示、操作教具、学生举手、学生应答、无交互”共11类细粒度行为标签。为降低主观误差以及标签漂移问题，同一段视频由两位标注员独立开展标注工作并且交叉校验，计算Kappa一致性系数，筛选Kappa>0.85的数据用于训练工作，确保标注质量具备高可信度。数据集按照7:2:1划分为训练、验证以及测试集，确保同一教师的不同课次不会同时出现在多个子集中，有效防止时空重叠带来的信息泄露问题。同时，在样本划分工作中对长尾类别如“演示”“操作教具”“学生应答”做过采样保留，增强对低频行为的泛化能力评估工作。评价指标选用mAP@0.5衡量检测精度，召回率开展遮挡场景鲁棒性的评估工作，FPS以及端侧功耗共同反映实时性以及能耗水平，把功耗定义为RK3399边缘盒在稳定推流状态下的平均功率。类别样本量如表1所示。

表1 数据集类别分布与样本量

单位：帧			
类别	训练集帧数	验证集帧数	测试集帧数
讲授	1 080 000	308 000	154 000
板书	720 000	206 000	103 000
指示	540 000	154 000	77 000
提问	360 000	103 000	51 000
巡视	540 000	154 000	77 000
反馈	360 000	103 000	51 000
演示	270 000	77 000	38 000
操作教具	270 000	77 000	38 000
学生举手	450 000	129 000	64 000
学生应答	270 000	77 000	38 000
无交互	540 000	154 000	77 000

其中，“讲授”与“板书”占比最高，合计约占32%，符合教师主导教学活动的现实分布特征；而如“演示”“学生应答”等低频类别构成典型长尾行为分布，为验证模型在复杂类别不均衡条件下的鲁棒性与泛化能力提供评估基准。

3.2 对比实验与消融测试

在RK3399边缘盒上，针对YOLOv5s、Efficient-Det-D0、PP-YOLOE+S以及本文提出的完整方案开展统一的配置对比实验工作。所有模型统一选用640px×640px的输入尺寸，借助TensorRT FP16精度开展推理工作，把Batch大小设置为1。在相同硬件以及资源约束条件下，本文提出的方法在mAP@0.5指标上达到84.7%，较次优模型PP-YOLOE+S提升6.2%；推理速度保持36帧/s，功耗仅为3.8W，在边缘实时性以及能效比方面得以实现双最

优，契合多摄像头高并发场景的实际需求。为进一步分析各个模块的性能贡献，开展三组消融测试的设计工作。把 Dual-FPN+CBAM 遮挡感知检测头移除后，mAP 下降 5.4%，表明该模块在板书遮挡、背身等极端姿态下对目标轮廓以及语义细节建模至关重要；把 TSMini-TCN 轻量化时序模块去除后，mAP 下降 3.1%，显示其对教师行为动作序列建模具备核心作用；把 Transformer 关系分支剔除后，mAP 下降 4.7%，说明教师 - 学生进行交互的建模工作在保持场景语义一致性方面不可或缺。消融测试的结果表明，三个模块都对整体精度的提升有显著贡献，并且遮挡建模以及空间建模在复杂教学环境下的重要性基本相当。三者共同开展协同优化工作，最终形成了兼具精度、时效性以及可解释性的边缘行为检测系统。图 3 所示的 mAP 与 FPS 柱状图进一步验证了本文方法在检测性能和运行效率上的领先性，具备良好的边缘部署实用价值。

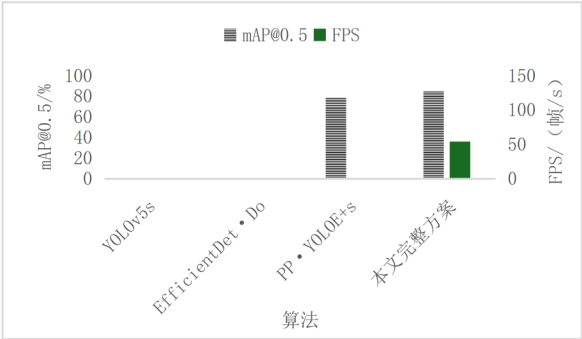


图 3 不同算法 mAP 与 FPS 对比柱状图

3.3 教学场景落地与闭环反馈

完成优化的系统在合作案例学校开展部署以及运行工作达一学期，连续采集到 2 700 h 的真实运行日志，覆盖了早读、正课以及晚自习等多个时段。把日志数据以 1 min 为粒度开展行为类别聚合工作，完成异常值检测后，生成了逐日热力图，直观呈现出教师巡视以及讲授活动的时间分布规律。借助教师问卷以及教务处访谈收集的反馈数据显示，系统的总体可用性评分达到 4.6/5，其中实时性以及遮挡鲁棒性两个维度的满意度最高。全学期的异常报警率控制在 0.3% 以下，和旧系统相比下降了 82%。误报情况主要出现在灯光突变等场景，已经借助快速滤波策略完成有效去除工作。鉴于收集到的相关反馈，系统新增了“教师巡堂”这一细分类别，进一步开展教学活动监测的细化工作。借助主动学习方法筛选高不确定性片段，并且在两周内完成人工标注以及模型迭代工作，来实现从研究到应用的闭环更新。这个过程不仅提高了系统的鲁棒性，还优化了模型的适应能力。教师满意度调查结果显示，实时性和遮挡鲁棒性是教师最为满意的两个功能模块，且系统的界面友好度和告警准确度也得到了教师的高度评价（如表 2 所示）。

表 2 教师满意度问卷统计

评价维度	单位：%			
	非常满意	满意	一般	不满意
实时性	78	18	4	0
遮挡鲁棒性	74	22	4	0
界面友好度	69	25	6	0
告警准确度	76	20	4	0

4 结语

本文围绕复杂教学场景下的教师行为检测难题，构建了涵盖感知、建模与部署的全链路轻量化方案，并在真实课堂环境中完成一学期闭环验证。研究证实，通过多尺度遮挡感知、分离空洞时序卷积与 Transformer 空间关系联合优化，可在边缘端实现高精度、低延迟、低功耗的运行效果；边缘 - 云协同进化机制进一步保证系统持续适应教学变化，显著降低运维成本。未来工作将扩展至多校跨区域部署，探索基于 NeRF 的三维姿态补全与基于大模型的开放词汇行为理解，推动智慧教育从单一功能走向全面智能化。

参考文献：

[1] 刘昆, 贺雄鹏, 廖桂生, 等. 面向复杂场景的多通道慢速运动目标稳健检测算法 [J]. 电子与信息学报, 2024, 46(5):2018-2027.

[2] 丁炜超, 孙立烨, 罗飞, 等. 面向复杂可行域约束多目标优化问题的双种群协同进化算法 [J]. 自动化学报, 2025, 51(8): 1-21.

[3] 刘辉, 冯浩然, 马佳妮, 等. 融合空间自注意力感知的严重缺失多元时间序列插补算法 [J]. 电子与信息学报, 2025, 47(12):1-12.

[4] 郑泊文, 李佩阳, 陆华才. 面向复杂运动场景的目标检测技术研究 [J]. 铜陵学院学报, 2024, 23(1):101-105.

[5] 盛博莹, 侯进, 李嘉新, 等. 面向复杂交通场景的道路目标检测方法 [J]. 计算机工程与应用, 2023, 59(15):87-96.

[6] 魏晓旭, 陈俊, 王永圣, 等. 面向嵌入式系统的复杂场景三维目标融合检测方法 [J]. 武汉理工大学学报, 2023, 45(6): 153-160.

[7] 徐宏哲, 莫欣岳, 李欢. 面向复杂网络的 Louvain 改进算法研究与探讨 [J]. 网络空间安全, 2024, 15(1):136-141.

[8] 厉承轩, 杨美红, 郭莹, 等. 面向复杂场景的大数据应用调度方法研究 [J]. 计算机与数字工程, 2024, 52(2):473-476.

【作者简介】

任甜甜(1992—)，女，河南禹州人，硕士研究生，讲师，研究方向：人工智能。

(收稿日期：2025-11-05 修回日期：2026-05-15)