

# 探索教室场景中的学生课堂行为检测难点问题

陈 发

西安翻译学院 信息工程学院 陕西 西安 710105

**【摘 要】：**近年来，基于深度学习的学生课堂行为检测成为教育领域关注的热点，然而并取得显著进展。现有研究直接将通用目标检测技术直接应用于教室场景，未对教室场景学生行为检测的难点进行系统性分析和研究。因此，本文探索了教室场景学生行为检测难点问题。首先，将学生课堂行为总结为抬头、低头、扭头、玩手机、看书、举手、趴下7类，建立真实教室场景学生行为数据集。其次，通过观察学生行为数据，发现可能存在遮挡、小目标和多尺度等影响检测。最后，在Faster R-CNN模型、Faster R-CNN+DPS模型、Faster R-CNN+CABM模型、YOLOv1模型、SSD模型和YOLOv5模型上进行对照实验，实验结果表明多尺度问题影响最为显著。研究为选择合适的学生行为检测方法提供依据，有助于解决学生课堂行为检测问题。

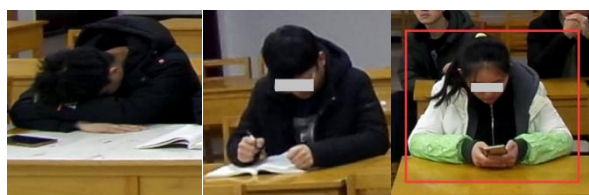
**【关键词】：**学生课堂行为检测；检测难点问题；多尺度问题

DOI:10.12417/2705-1358.25.19.014

## 1 教室场景中学生课堂行为的研究现状

传统人工方法从视频中提取学生课堂行为信息工作量大且时效性差，研究者们逐渐转向采用计算机视觉技术检测学生课堂行为信息<sup>[3]</sup>。主要包含两种方法：一是基于图像分类的识别方法，该方法利用卷积神经网络对图像

进行处理和特征提取，从而准确地识别出学生的行为类型，相关示例如图1(a、b)；二是基于目标检测的检测方法，该方法可针对学生在图像中的动作行为进行定位和识别，不仅输出行为类别，还可准确的标注相关行为的空间位置，具体展现如图1(c)所示：



a 睡觉

b 书写

c 玩手机

图1 核心方法示例

基于图像分类的方法识别学生课堂行为。徐家臻等<sup>[4]</sup>采用基于人体骨架模型结合XGBoost网络模型，成功识别出听讲、看书、站立、举手及写字等五种常见课堂学习行为，平均准确率达到了85.49%；黄勇康等<sup>[17]</sup>结合实时目标检测和跟踪，获取每个学生的实时图片流，通过深度时空残差卷积神经网络学

习每个学生行为的时空特征，验证了提出的课堂教学视频中实时多人学生课堂行为识别模型，实现了88.5%的准确率；蒋沁沂等<sup>[18]</sup>运用残差网络在其构建的课堂学习行为数据集上，成功识别出上课、玩手机、记笔记、东张西望、看书及睡觉等六种常见课堂行为，其平均识别率达到了91.91%；而魏艳涛等<sup>[12]</sup>对VGG16网络进行迁移学习，利用其构建的课堂学习行为数据集，成功识别出听课、看书、书写、站立、举手、睡觉、左顾右盼七种行为，其准确率为93.33%。基于图像分类的学生课堂行为识别方法，虽能有效识别学生课堂行为，但不能提供学生在教室的空间位置信息且缺乏实时性，限制了实际应用。因此，研究者使用目标检测方法进行检测。郑锐等<sup>[15]</sup>改进经典目标检测框架Faster R-CNN，提出多尺度感知探测头、特征融合策略、并采用OHEM(Online Hard Example Mining)，实现了对举手、站立和睡觉等行为的检测，平均精度为57.6%，比原方法提高了3.4%；唐林等<sup>[19]</sup>结合了合并的感兴趣区域(ROI)池化和局部保持学习的方法改进Faster R-CNN框架，实现对站立、坐着及趴下等课堂行为的检测，该模型的平均精度可达79.89%；吕文超等<sup>[20]</sup>改进SSD算法，引入ResNet模块，并添加FPN模块构建RF-SSD，在自制的听课、记笔记、玩手机和睡觉等4种学生行为状态数据集上进行训练和测试，在输入图像大小为300×300像素和512×512像素时，平均精度分别为62.3%和70.6%，与原始SSD模型相比，平均检测精度分别提高了3.6%和3.7%；而王泽杰等<sup>[6]</sup>用OpenPose算法提取的人体姿态全局特征，并融合YOLOv3算法提取的交互物体局部特征，对学生正坐、侧身、低头及举手行为进行识别，平均精

度达到 95.45%；此外，李丽娜等<sup>[21]</sup>提出 ET-YOLOv5s，即改进 YOLOv5s，结合 ESRGAN（Enhanced Super-Resolution Generative Adversarial Network）和一个微小目标检测模块，能够检测教室中 11 种学生的行为，包括玩手机、上课、手托脸、向右转身、向左转身、鞠躬、睡觉等，平均精度达到了 96.8%。进一步地，一些研究者尝试直接从视频中进行行为分析，周鹏霄等<sup>[22]</sup>的工作实现了对课堂教学视频中 S-T 行为智能识别算法，能够对较为标准的教学视频分析得到较高准确率。然而该研究课题较为复杂，视频中存在许多不清晰的特征，且整体技术还不够成熟，因此直接采用视频进行的研究相对较少。

综上所述，研究者们虽在学生课堂行为检测上达到了一定的准确率，但这些成果都直接将通用场景的检测识别技术应用在教育场景中，忽略了教室场景的特殊性，缺乏教室场景学生行为检测难点问题的讨论。因此，本研究针探索了教室场景学生课堂行为检测难点问题。

## 2 实验设计

### 2.1 实验原理

图 2 是数据集图片，从图中观察到教室场景学生行为检测中可能存在影响检测结果的遮挡问题、小目标问题。及多尺度问题。为了研究不同问题的研究程度，进行实验设计（表 2）。



图 2 检测难点问题示例：遮挡问题（蓝色框）、小目标问题（黄色框）、多尺度问题（两红色框对比）

表 2 实验设计

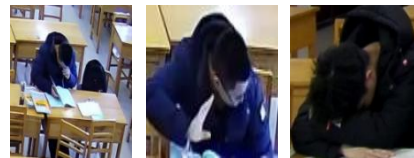
| 挑战  | 基线模型         | 对比模型                |
|-----|--------------|---------------------|
| 遮挡  | Faster R-CNN | Faster R-CNN+DPS    |
| 小目标 | Faster R-CNN | Faster R-CNN+CBAM   |
| 多尺度 | Faster R-CNN | YOLOv1、SSD 及 YOLOv5 |

### 2.2 学生课堂行为数据集的构建

本实验收集了约 2000 分钟的大学生课堂视频，总结了 7 种具有代表性的学生课堂行为（图 4），包括抬头、低头、扭头、玩手机、看书、举手以及趴下。



抬头 低头 扭头 玩手机



看书 举手 趴下

图 4 种典型的课堂动作行为示例图

## 3 实验结果与分析

### 3.1 遮挡问题对照实验

遮挡问题实验结果如表 4，观察 7 种动作的检测准确率，Faster RCNN+DPS 相比于 Faster RCNN 显著提高。由上述对照实验结果可得，Faster RCNN 算法在引入 DPS 模块后，检测的准确率明显提升，DPS 模块是针对遮挡问题提出的，说明在教育场景中，遮挡问题存在且会对学生课堂行为检测产生影响，同时也表明加入 DPS 模块可以在一定程度上解决由遮挡问题引起的检测不准确问题。

表 4 遮挡检测准确率

| 模型 mAP | Faster R-CNN | Faster R-CNN+DPS |
|--------|--------------|------------------|
| 抬头     | 80%          | 85.45%           |
| 低头     | 79.4%        | 81.44%           |
| 扭头     | 68%          | 70.22%           |
| 玩手机    | 69%          | 69.1%            |
| 看书     | 60.3%        | 79.41%           |
| 举手     | 89%          | 92.38%           |
| 趴下     | 78%          | 80.68%           |

### 3.2 验证教室场景中小目标问题的实验结果

由表 5 可以观察到 7 种动作的检测准确率，Faster R-CNN+CBAM 相比于 Faster R-CNN 有明显的提升。结合上述表格和图片实验结果，加入 CBAM 模块的 Faster R-CNN 检测准确率有所提高。这说明小目标问题存在于学生行为检测中，影响检测结果。

表 5 小目标检测准确率

| 模型 mAP | Faster R-CNN | Faster R-CNN+ CBAM |
|--------|--------------|--------------------|
| 抬头     | 80%          | 89%                |
| 低头     | 79.4%        | 73%                |
| 扭头     | 68%          | 73%                |
| 玩手机    | 69%          | 71%                |
| 看书     | 60.3%        | 46%                |
| 举手     | 89%          | 89%                |
| 趴下     | 78%          | 82%                |

3.3 验证教室场景中的多尺度问题实验结果

由表 7 纵向数据对比,可观察到四种不同算法模型在特定尺度区间内对学生动作检测准确率的表现存在差异。

表 6 四种算法模型的平均准确

| 模型 mAP       | 全部     | 小于 1658 | 1658-2000 | 2000-5000 | 大于 5000 |
|--------------|--------|---------|-----------|-----------|---------|
| SSD          | 62.60% | 5.09%   | 10.28%    | 24.83%    | 69.45%  |
| YOLOv1       | 77.0%  | 3.59%   | 4.61%     | 28.63%    | 79.4%   |
| Faster R-CNN | 80.11% | 4.33%   | 4.04%     | 30.33%    | 81.55%  |
| YOLOv5       | 81.64% | 6.22%   | 11.83%    | 31.27%    | 83.68%  |

4 结语

本研究创建了真实教室场景的学生课堂行为数据集并进行了一系列对照实验。统性地探索了教室场景中学生课堂行为识别难点问题,发现遮挡、小目标、多尺度问题影响了教室场景学生行为检测且多尺度问题最为显著。

参考文献:

[1] 祝智庭,胡姣.教育数字化转型的本质探析与研究展望[J].中国电化教育,2022,(04):1-8.

[2] 刘献君.智慧教育背景下高等教育管理变革探究[J].高校教育管理,2024,18(01):24-32.

[3] 何秀玲,杨凡,陈增照等.基于人体骨架和深度学习的学生课堂行为识别[J].现代教育技术,2020,30(11):105-112.

[4] 徐家臻,邓伟,魏艳涛.基于人体骨架信息提取的学生课堂行为自动识别[J].现代教育技术,2020,30(05):108-113.

[5] Abedi A, Khan S. "Improving state-of-the-art in detecting student engagement with resnet and TCN hybrid network,"2021 18th Conference on Robots and Vision (CRV), Burnaby, BC, Canada, 2021, pp. 151-157,

[6] 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,(02):55-66.