

基于深度学习的学生课堂行为识别研究

王禹钧, 马致明

(新疆师范大学计算机科学技术学院, 新疆 乌鲁木齐 830054)

✉ 627885569@qq.com; 406287175@qq.com



摘要:智慧课堂面临学校硬件设备难以普适、课堂数据缺失等问题,给学生课堂行为识别研究带来了一定的难度和挑战。为了降低对课堂场景部署设备的要求,提高识别精度与速度,在已有识别模型基础上提出了一种基于深度学习的学生课堂行为识别模型。首先融合重影模块(Ghost)实现轻量化,其次加入坐标注意力机制(Coordinate Attention)提升检测精度。实验结果表明,改进后的模型识别精度(mAP)达到了 86.2%,与原模型相比提高了 3.5%,推理时间减少了 16.7%,参数量降低了 35.5%,速度与精度均有一定提升,符合智慧课堂的基本要求。

关键词:深度学习;学生课堂行为;行为识别;智慧课堂

中图分类号:TP3-05 **文献标识码:**A

Research on Classroom Behavior Recognition of Students Based on Deep Learning

WANG Yujun, MA Zhiming

(School of Computer Science and Technology, Xinjiang Normal University, Urumqi 830054, China)

✉ 627885569@qq.com; 406287175@qq.com

Abstract: Smart classrooms face problems such as difficulty in adapting school hardware equipment and missing classroom data, which poses certain difficulties and challenges to the research of student classroom behavior recognition. In order to reduce the requirements for deploying equipment in classroom scenarios and improve recognition accuracy and speed, this paper proposes a deep learning-based student classroom behavior recognition model based on existing recognition models. Firstly, the Ghost module is integrated to achieve lightweight, and secondly, a Coordinate Attention mechanism is added to improve detection accuracy. The experimental results show that the improved model has a recognition accuracy (mAP) of 86.2%, which is 3.5% higher than the original model. The inference time is reduced by 16.7%, and the parameter quantity is reduced by 35.5%. The speed and accuracy are both improved to a certain extent, which meets the basic requirements of smart classrooms.

Keywords: deep learning; student classroom behavior; behavior recognition; smart classroom

0 引言(Introduction)

近年来,深度学习技术在教育领域中的应用越来越深入。《中国教育现代化 2035》《关于加强新时代教育管理信息化工作的通知》等都不约而同地提及要充分且正确利用现代技术,

大力发展智慧课堂^[1]。在课堂中,学生是学习活动的主体,学生的行为是课堂教学质量好坏最直接的反映^[2]。通过对课堂中学生行为进行识别,能更好地分析学生的上课情况、了解学生的学习兴趣。因此,利用深度学习技术实时掌握学生课堂学

有研究显示:当加入少量运算复杂度时,加入注意力机制可以改善轻量化模型的性能^[10]。课堂场景下,尽管 YOLOv5

模型能提取出富有细节的特征图,但课堂情景具有复杂性,如使模型能够学习重要场景中的特征以增强关键特征提取的能力成为必须解决的问题。Coordinate Attention 获取通道之间的特征信息时,实现了空间方向上准确位置信息及感兴趣区域的获取^[11]。这种注意力机制的每个权重都包含通道间信息、横向与纵向空间信息,可以帮助网络更加准确地定位目标信息,提高识别能力。它不但得到通道之间的信息,与方向有关的位置信息也被考虑在内,能帮助模型对目标进行定位与识别,并且灵活轻量,能在网络核心结构上进行简单插入。Coordinate Attention 的具体构造如图 3 所示,依次实现坐标信息嵌入模块和坐标注意力生成模块两个功能。

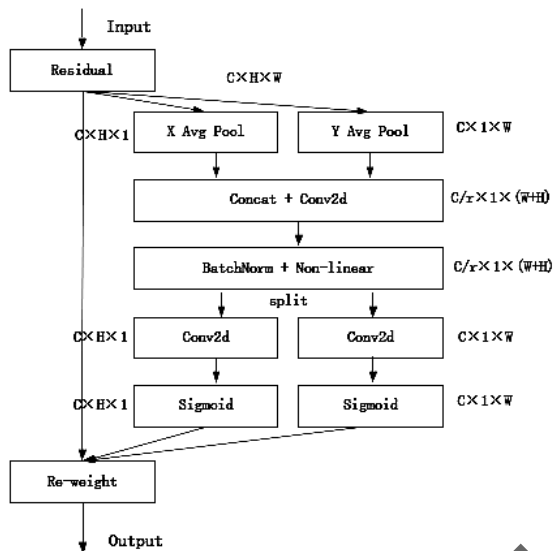


图 3 坐标注意力机制结构图

Fig. 3 Coordinate Attention structure diagram

YOLOv5 模型大部分由卷积神经网络组成,如何能更加高效地抽取特征信息是比较关键的,并且在实际课堂环境下学生被遮挡的问题尤为严重,存在小目标学生难以识别的问题。为此,本文提出加入 Coordinate Attention 改善模型的特征提取能力,使模型的关键特征提取能力有了很大的提升。YOLOv5 的主干网络的主要工作是提取输入处理后照片的特征,也就是说抽取特征的能力在很大程度上取决于主干网络。为使网络能自主学习关键特征并摒弃其他不重要的特征,所以在主干网络上增加了 Coordinate Attention。

3 实验与结果分析(Experimentation and analysis of results)

3.1 数据集构建

由于课堂类数据集场景具有独特性与保密性,因此学生课堂行为识别的数据集相对缺乏。目前,并不存在专用公开数据集,需要自行构建数据集完成模型训练。研究人员观察大量课堂视频后发现,课堂教学过程中学生的行为是多样的,不同行为能够体现学生不同的学习状况。结合上述文献的行为分类及观察课堂实录中的学生行为,本文确定书写、听课、举手、低头、左顾右盼、小组讨论 6 种行为类型用于学生课堂行为数据集的构建。数据集来自小学公开课视频分帧形成的图片,视频

来源网站为国家教育资源公共服务平台,共采集 50 节不同课堂的教学视频,每节课的时间平均为 40 min 左右,作为原始实验数据。考虑到视频识别模型会更加复杂,难以应用于智慧课堂,尽管本文收集到的原始数据为课堂视频,但仅是针对单帧图片的学生行为进行识别。数据处理的流程如下:使用 Python 进行脚本的编写,将间隔设置为 30 s,以均匀采样帧并按照一定帧数间隔将视频解码成图片。经过多轮删减筛选,筛选出清晰的课堂学生图片。处理完成后,收集并整理获得 3 002 张小学生的课堂图片数据,截取数据集图片如图 4 所示。



图 4 数据样例图片

Fig. 4 Sample data images

由于收集的学生课堂行为图像过少,为达到训练所需的样本量,对现有图像数据做了扩充处理。在智慧课堂实际应用中,存在各类不同的场景,例如光照的强弱、教室的环境不同等。可以通过对数据集进行增强以增加不同条件下的数据,从而提高训练模型的适应力。本文利用一些常见的数据增强方式对训练集进行扩展,包括图像旋转、将噪声随机加入原始图像中、随机变换图像色彩等方法。扩增数据集之后,重新过滤图像数据集,共得到 4 120 张学生课堂图片,行为数量共 71 016 个。

3.2 实验环境

实验硬件环境配置如下:处理器为 intel(R) Xeon(R) E5-2699v3,显卡为 RTX3060 12 GB 显存。软件环境配置如下:编程语言为 Python,深度学习框架为 PyTorch1.9.0,操作系统为 Windows 10。训练集、验证集、测试集的比例按照 7:2:1 进行划分。*epochs* 为 200,*Batch-size* 为 16。在训练期间,各训练轮次会根据训练情况调整学习率的取值,初始学习率设定为 0.001,模型采用余弦退火算法更新学习率的动态取值。输入端采用 Mosica 数据增强包括随机尺寸拼接等技术进行叠加处理等。

3.3 评价指标

mAP 为类别平均像素准确率,能衡量模型在所有类别的效果,它是评价算法性能的最终指标之一。除了评价精度的指标,还有检测速度的指标如 FLOPs(浮点运算数),处理一张图片所需的浮点运算量可以衡量模型的复杂度;*InferenceTime* 为推理时间,是指检测每张图片需要使用的的时间,用来衡量模型推理速度的快慢;*Params* 为参数量,指模型训练中需要训练的参数总数,模型参数量决定了模型的体积,也影响推理时间。

3.4 对比实验

为了验证 YOLOv5-GC 融合模型对学生课堂行为的识别效果,同时方便对比原始模型的提升程度,本文选取快速区域生成网络(Faster R-CNN)、YOLOv4、YOLOv5 几种经典模型进行对比实验,这些模型常用于目标检测任务。为了更好地比较改进模型和传统模型,需要将基础参数保持一致,以免因参数不一致而影响结果。

运用改进后的 YOLOv5 模型对 6 种行为的识别精度都明显上升,相对其他算法表现出了更高的精度,如表 1 所示。

表 1 不同模型识别学生行为结果

Tab. 1 Recognition results of student behavior using different models

课堂行为	YOLOv4	Faster R-CNN	YOLOv5s	改进后 YOLOv5
书写	0.79	0.81	0.8	0.82
听课	0.93	0.97	0.94	0.96
举手	0.83	0.86	0.84	0.89
低头	0.82	0.87	0.83	0.86
左顾右盼	0.59	0.79	0.69	0.73
小组讨论	0.86	0.89	0.87	0.91

如表 2 所示,分别从 4 个方面对 Faster R-CNN、YOLOv4、YOLOv5 等经典模型进行对比实验。本文所提方法的 mAP 高于除 Faster R-CNN 外的其他相似的算法。Faster R-CNN 是两阶段算法,它的精度较高,但是占用的内存也较大。根据智慧课堂的建设需求,YOLOv5-GC 模型更加适用于真实课堂中学生行为的识别。由于最终目标是将模型应用于课堂终端,因此模型的参数规模、运算符号数、推理耗时也是重要的衡量指标。

由表 2 可知, mAP 从高到低排序为 Faster R-CNN、YOLOv5-GC、YOLOv5、YOLOv4。YOLOv5-GC 的 mAP 为 86.2%,较原模型有了明显的提高,Faster R-CNN 是两阶段算法,有精度高的优势,它的 mAP 达 86.5%,虽然优于 YOLOv5-GC 融合模型,但是 Faster R-CNN 的参数数量巨大,不符合智慧课堂的应用要求,难以部署在真实的课堂场景中。本文所提模型使用的参数量最少,较原模型降低 2.7 MB,更加轻量且方便在教室环境下布置。YOLOv5-GC 模型比原模型推理时间减少 16.7%,耗时越短,模型实时检测能力越好。从浮点运算量这个指标来看,YOLOv5-GC 模型的复杂度大大降低,方便在智慧课堂部署。综上可知,本文提出的模型速度指标上表现最好,它具有精度高、参数量少、识别耗时短等优点,平衡检测速度与精度,满足实时检测的条件,符合智慧课堂设备的应用要求。

表2 对比实验结果

Tab. 2 Comparison of experimental results

模型	mAP /%	参数量/MB	推理时间/ms	$FLOPs$ /GB
YOLOv4	80.0	61.4	43	60.53
YOLOv5	82.7	7.6	24	16.7
Faster R-CNN	86.5	130.7	98	270.21
YOLOv5-GC	86.2	4.9	20	10.8

3.5 消融实验

为验证实验的延展性和适用性,本小节选取 YOLOv5、YOLOv5+Ghost、YOLOv5+Coordinate Attention、YOLOv5+Ghost+Coordinate Attention 进行研究,融合形成四种模型进行消融实验,消融实验结果如表 3 所示。

表3 消融实验结果

Tab. 3 Results of ablation experiments

模型	mAP /%	参数量/MB	推理时间/ms	$FLOPs$ /GB
YOLOv5	82.7	7.6	24	16.7
YOLOv5+Ghost	83.2	5.1	11	9.9
YOLOv5+Coordinate Attention	85.8	7.7	28	17.5
YOLOv5+Ghost+Coordinate Attention	86.2	4.9	20	10.8

由表 3 可知,融合 GhostNet+Ghost+Coordinate Attention 在各个指标上都表现最好,相较 YOLOv5 模型, mAP 上升 3.5%,参数量减少 2.7 MB,运算量、推理时间也显著减少,模型的复杂度降低,使得模型更加轻量化,在降低模型复杂度的同时提升了模型检测的精度。只加入 GhostNet 后, mAP 上升较小,但是参数量和推理时间显著减少,识别速度显著提升,由此可见,Ghost 在轻量化参数量方面具有优越性。轻量化网络替换后,可以达到实时检测的要求,对部署设备要求较低,易于课堂安装。只加入 Coordinate Attention 后, mAP 上升 3.1%,提取特征能力的提高,使能够关注到关键特征,但是参数量和速度略微增加。Coordinate Attention 是为轻量级网络设计的,它可以让轻量化后的网络在更大区域拥有注意力,同时弥补 GhostNet 轻量化但特征提取不足的缺点,达到了提升 mAP 的效果。所以,从几个指标来看,加入 GhostNet 和 Coordinate Attention 对模型轻量化、提升关键特征有很大的帮助。本文提出的 YOLOv5-GC 实现 YOLOv5 网络轻量化和识别精度的提升。通过改进 YOLOv5 模型,并将其应用于学生课堂行为的识别,同智慧课堂的建设需要紧密结合起来。

4 结论(Conclusion)

考虑到实时检测的速度与精度,本文根据真实课堂场景在 YOLOv5 的基础上改进学生课堂行为识别模型。在分析当前学生课堂行为识别方法存在的问题后,提出一种基于 YOLOv5 改进的 YOLOv5-GC 模型,并在轻量化模型、注意力机制添加等方面提出了改进策略,对效果进行验证。通过对比实验验证模型的泛化能力,使用消融实验验证各模块的有效性。本文提出的 YOLOv5-GC 模型参数规模显著减少,识别精度和速度也有一定的提升,便于在移动设备端部署,可应用于智慧课堂中对学生的行为进行精确识别,具有一定的应用价值。后续将开展教师行为研究,探究师生互动行为如何影响学生学习行为,为优化智慧课堂环境下师生互动效果提供一定依据,其理论意义和现实意义更值得期待。

(下转第 62 页)