

基于深度学习的学生课堂注意力评价^①



孙绍涵¹, 张运楚^{1,2}, 王超¹, 张汉元^{1,2}

¹(山东建筑大学 信息与电气工程学院, 济南 250101)

²(山东省智能建筑技术重点实验室, 济南 250101)

通信作者: 张运楚, E-mail: yczhang@sdjzu.edu.cn

摘要: 针对学生课堂行为状态识别准确率较低的问题, 提出一种基于 YOLOv4 的改进模型. 通过建立学生课堂行为状态数据集, 调整 YOLOv4 算法训练模型的参数, 修改卷积块激活函数为 ELU 函数以优化模型, 同时提出将 DIoU-Soft-NMS 作为非极大值抑制机制, 识别分析教室中学生课堂行为状态; 根据各状态持续时长及状态变化频率计算学生听课有效时长, 并参考山东高考赋分原则, 建立学生课堂注意力量化评价准则, 同时建立教师课堂授课效果量化评价标准. 实验结果表明, 以同一评价指标衡量模型, 该模型在学生课堂行为检测速率不变的情况下, 平均精度均值 (mAP) 达到 98.8%, 比原 YOLOv4 模型提升了 3.53%, 学生服课堂注意力量化评价准则, 有较高的契合度.

关键词: 课堂行为识别; YOLOv4; 注意力评价; 课堂教学效果评价; 目标检测

引用格式: 孙绍涵, 张运楚, 王超, 张汉元. 基于深度学习的学生课堂注意力评价. 计算机系统应用, 2022, 31(6): 307-314. <http://www.c-s-a.org.cn/1003-3254/8553.html>

Evaluation of Students' Classroom Behavioral State Based on Deep Learning

SUN Shao-Han¹, ZHANG Yun-Chu^{1,2}, WANG Chao¹, ZHANG Han-Yuan^{1,2}

¹(School of Information and Electrical Engineering, Shandong Jianzhu University, Jinan 250101, China)

²(Shandong Provincial Key Laboratory of Intelligent Building Technology, Jinan 250101, China)

Abstract: Aiming at the low accuracy of students' classroom behavioral state recognition, an improved model based on YOLOv4 is proposed. We establish a dataset for students' classroom behavioral states, adjust the YOLOv4 algorithm to train the parameters of the model, and modify the activation function of the convolutional block as an ELU function to optimize the model. Meanwhile, DIoU-soft-NMS is proposed as a non-maximum suppression mechanism to identify and analyze students' classroom behavioral states. According to the duration of each state and the frequency of state changes, the effective length of students' attention on lectures is calculated, and by referring to the scoring principle of Shandong college entrance examinations, the quantitative evaluation criteria of students' classroom attention and the quantitative evaluation criteria of teachers' classroom teaching effects are established. The experimental results show that when the same evaluation index is used to measure the model, the mean average precision (mAP) of the model reaches 98.8% assumed that the detection rate of students' classroom behaviors remains unchanged, which is 3.53% higher than that of the original YOLOv4 model. Students approve the quantitative evaluation criteria of classroom attention, and it has a high degree of agreement.

Key words: classroom behavior recognition; YOLOv4; attention evaluation; classroom teaching effect evaluation; object detection

① 基金项目: 国家自然科学基金青年项目 (62003191)

收稿时间: 2021-09-18; 修改时间: 2021-10-25; 采用时间: 2021-11-07; csa 在线出版时间: 2022-05-26

我国已进入教育信息化 2.0 时代,通过实时识别学生课堂行为,跟踪学生状态变化,统计学生课堂不同行为占比,构建基于实证的课堂教学评价体系,可以有效掌握学生的课堂注意力程度,为学生制定个性化的课堂行为矫正策略。同时,通过分析学生课堂行为状态分布规律,为教师课堂教学质量评价和教学法改进提供一种数据驱动机制。

传统的考勤方式及听课状态评价识别方法大多为人工观察监控视频^[1]、问卷调查法^[2]等。由于课堂学生数量多,人工识别方法会出现效率低、错误率高、易疲劳等问题。借助机器学习技术,研究人员提出了基于人体骨骼关键点检测、光流特征、全局运动方向特征等学生课堂行为识别方法。Stanley^[3]使用 Kinect 设备采集人体骨骼关节点,通过计算学生脸部姿态的转向和身体的倾斜信息来分析其注意力状态,可以预测超过 60% 的学生情绪状态变化;党冬利^[4]提出基于运动历史图的 Zernike 矩特征和朴素贝叶斯分类器分类,再通过 Lucas-Kanade 光流特征和全局运动方向特征判断运动方向,从而判断学生举手、站立和坐下 3 种行为。

上述基于传统机器学习的学生课堂行为识别方法,多依赖人工特征选择,适应性和泛化能力差、准确率不高。近年来,随着深度学习理论和技术的发展,卷积神经网络得到了广泛应用,研究人员逐渐将深度学习技术应用到教育领域中。廖鹏等^[5]使用背景差分法提取感兴趣区域输入 VGG 网络训练,构建出识别学生课堂异常行为检测分析系统,有效识别了玩手机、上课睡觉等异常行为;Simonyan 等^[6]提出了时空双流网络结构 (two stream network) 并进行了基于视频的行为检测,该网络融合了两个相同的神经网络分别训练视频图像与光流,最后完成类别分类。

基于深度学习的目标检测主要解决图像的位置和分类两个问题,分为两个思路:一种是单步 (one-stage) 回归型目标检测算法,这类算法对输入网络检测的图像输出类别信息和定位回归性目标检测,例如 YOLO 系列^[7-10]和 SSD (single shot multi box detector)^[11];另一种是以 Faster R-CNN 系列^[12-14]为代表的两步 (two-stage) 算法,这类算法的思路是首先提取目标候选框,再在此基础上训练检测目标。

本文提出一种基于 YOLOv4 的改进模型,通过建立学生课堂行为状态数据集,调整 YOLOv4 算法训练模型的参数,修改卷积块激活函数及非极大值抑制机

制优化模型,识别分析教室中学生多种课堂行为状态,并根据各状态持续时长及状态变化频率,对学生课堂注意力进行量化评级,并以多种可视化方法展示学生课堂行为状态整体态势规律。同时建立教师课堂授课效果量化评价标准,以便教师通过学生整体听课状态和质量时空分布规律,了解授课效果和质量,做出改进。实验结果表明,改进模型在检测速率不变的情况下,准确率达到 98.8%,比原 YOLOv4 模型提升了 3.53%。提出的学生课堂注意力量化评级规则,与实际表现也有较好的一致性。

1 学生课堂注意力评价模型

本文将学生课堂行为状态分为“rise (抬头听讲)”“mobilephone (玩手机)”“side (交头接耳)”“book (低头看书)”和“bow (低头遮挡)”等 5 类状态,采用基于 YOLOv4 的改进模型实时识别这 5 种行为,统计每个学生各类行为发生的时刻及持续时间,在此基础上建立学生课堂注意力评价及教师课堂授课效果量化评价模型。

1.1 YOLOv4 目标检测算法原理

YOLOv4 算法是在 YOLOv3 基础上从主干特征提取网络、Mish 激活函数^[15]及损失函数等方面增加优化训练技巧而来,可以更快速有效地实现目标检测。其算法原理是首先将待检测图像分割成 76×76 、 38×38 与 19×19 的网格,分别对应 8 倍、16 倍、32 倍下采样层,若待检测学生落在其中某个网格中,则由该网格识别该学生课堂行为状态。其网络结构主要包括主干特征提取网络 (CSPDarknet53)、空间金字塔池化结构 (SPP)、路径聚合网络 (PANet) 和头部 (YOLO-head),网络结构图如图 1 所示。

YOLOv4 的主干提取网络结合了 CSPnet 与 Darknet53 残差网络,该模块通过优化网络结构减少计算量,降低了内存成本,并提取深层次网络特征保证了检测的正确率。同时,YOLOv4 在 CSPDarknet53 结构中采用 Mish 激活函数,Mish 函数在 x 负半轴时有较小的负梯度流入,能够保证信息的连续,从而有更好的梯度下降效果。Mish 激活函数的定义由式 (1) 给出,其曲线如图 2 所示。

$$f(x) = x \cdot \tan(\ln(1 + e^x)) \quad (1)$$

在特征金字塔部分,YOLOv4 使用了 SPP 结构:在

对 CSPDarknet53 的最后一个特征层进行 3 次卷积后, 分别利用 13×13、9×9、5×5、1×1 四个不同尺度的最大池化, 极大地增加了特征层的感受野, 分离出最显著

的上下文特征. 同时, YOLOv4 在 CSPDarknet53 的 3 个有效特征层上使用 PANet 结构, 实现实例分割算法, 对待识别图像反复提取特征.

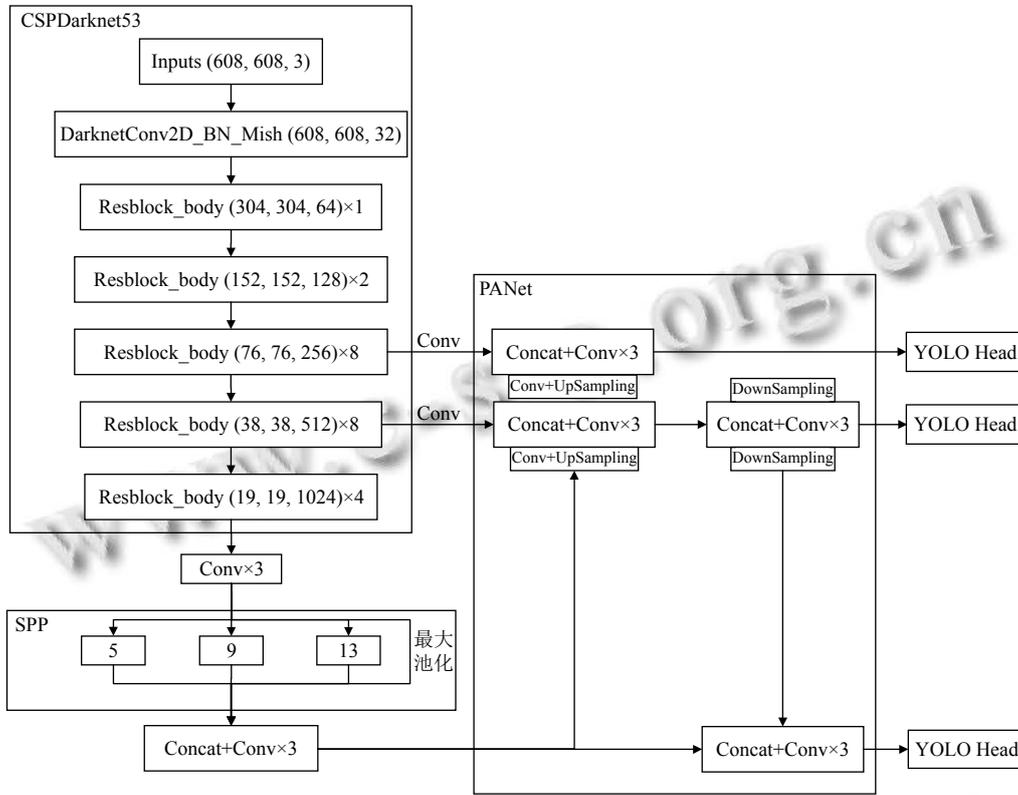


图 1 YOLOv4 整体网络结构

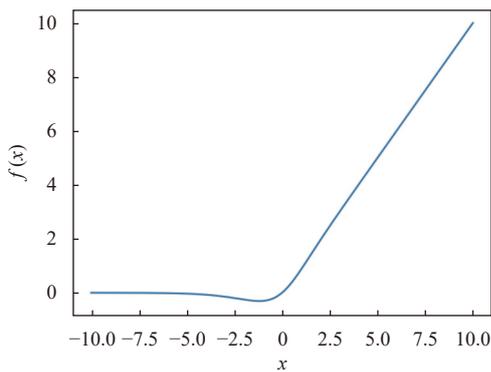


图 2 Mish 激活函数曲线

1.2 改进的 YOLOv4 目标检测算法

1.2.1 ELU 激活函数

激活函数是神经网络重要的组成部分, 对神经网络的训练及网络识别结果有直接影响的作用. ELU (exponential linear units) 激活函数^[16]是针对 ReLU

函数的一个改进型, 在神经网络中因其左侧具有软饱和性, 对噪声或变化的输入具有更好的鲁棒性; 其次 ELU 激活函数右侧无饱和性, 使其能够缓解梯度消失的问题; 最后 ELU 函数的输出平均值接近于零, 使梯度更接近自然梯度, 从而加快网络的学习速度. ELU 激活函数的定义由式 (2) 给出, 其曲线如图 3 所示.

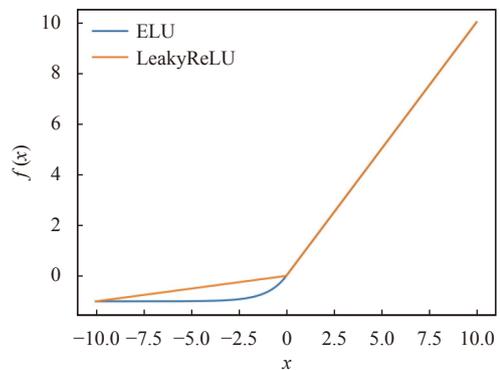


图 3 ELU 激活函数曲线

$$f(x) = \begin{cases} x, & \text{if } x > 0 \\ \alpha(\exp(x) - 1), & \text{if } x \leq 0 \end{cases} \quad (2)$$

其中, $\alpha > 0$.

YOLOV4 原卷积块 DarknetConv2D_BN_Leaky 为 DarknetConv2D + BatchNormalization + LeakyReLU, 将 LeakyReLU 激活函数改为 ELU 激活函数, 更改后的卷积块 DarknetConv2D_BN_ELU 具有更好的鲁棒性.

1.2.2 DIoU-Soft-NMS

在目标检测中, 首先产生一些候选区域, 其次通过分类网络得到类别得分, 与此同时通过回归网络得到更加精确的位置参数, 通过非极大值抑制去除冗余框, 得出最后的检测结果. 非极大值抑制 (non-maximum suppression, NMS)^[17] 是一种获取局部最大值并抑制非极大值的算法, 在计算机视觉中有着广泛的应用. 其核心思想是一个迭代-遍历-消除的过程: 首先, 将输出网络的预测框依据置信度大小排序, 筛选出置信度最大的预测框作为最大得分框, 计算其与集合中其余预测框的面积交并比 IoU , 删除超过预设值的预测框, 一轮筛选过后, 继续筛选出集合中剩余置信度最大的预测框, 重复以上步骤直至筛选结束. 传统的 NMS 处理方式可用式 (3) 表示:

$$s_i = \begin{cases} s_i, & IoU(M, B_i) < \varepsilon \\ 0, & IoU(M, B_i) \geq \varepsilon \end{cases} \quad (3)$$

其中, s_i 为候选框的原始得分, M 为得分最高的候选框, B_i 为预测框, $IoU(M, B_i)$ 为待预测框 B_i 和 M 的交并比, ε 为 NMS 预设阈值.

但在传统 NMS 中, IoU 指标对于两个距离相近的目标经常产生错误抑制. 因此, YOLOv4 应用 DIoU-NMS^[18] 作为 NMS 的抑制准则, DIoU 同时考虑到真实框与待测框的重叠区域面积以及中心坐标的欧氏距离. 因此 DIoU-NMS 的 s_i 更新公式定义如下:

$$s_i = \begin{cases} s_i, & IoU - R_{DIoU}(M, B_i) < \varepsilon \\ 0, & IoU - R_{DIoU}(M, B_i) \geq \varepsilon \end{cases} \quad (4)$$

$$R_{DIoU} = \frac{\rho^2(b, b^{gt})}{c^2} \quad (5)$$

其中, $\rho^2(b, b^{gt})$ 代表真实框与待测框的欧氏距离, c 代表真实框与待测框最小闭包矩形的对角线长度.

在识别学生课堂行为状态时, 后排同学相对密集, 常常出现相互遮挡的情况, 当两个候选框之间重叠面积较大, 其中分数较低的框会因为与得分最高的框交

并比超过设置阈值被过滤, 从而出现漏检, DIoU-NMS 在筛选候选框时有一定的局限性. 因此, 本文使用 Soft-DIoU-NMS 算法对 YOLOv4 网络进行优化. 在该算法中, 不直接删除超过设定阈值的预测框, Soft-DIoU-NMS 应用“权值惩罚”机制, 在筛选出最大得分框后, 计算与集合剩余预测框的 DIoU, 再将超过预设值的框输入连续的高斯线性曲线中, 给予一个惩罚因数来衰减预测框得分. 最终的改进如式 (6) 所示:

$$s_i = s_i e^{-\frac{(IoU - R_{DIoU}(M, B_i))^2}{\sigma}}, \forall B_i \notin D \quad (6)$$

这样的软性非极大值抑制当两个框重合度越高, s_i 取值越小, 相应预测框的得分越小, 在一定程度上解决了 DIoU-NMS 的局限性, 能更大程度的保留需要的框, 剔除不需要的框.

1.3 学生位置及课堂行为状态跟踪统计

由于教师座椅位置固定, 即使学生上课活动幅度大, 其位置坐标也会固定在一定范围之内, 因此只要确定好每个位置的初始坐标, 并对后面视频中每秒取滑动平均即可完成目标对应. 具体做法为: 首先判断每个位置范围内的横纵坐标平均值作为初始位置, 再对后续帧中每个 bounding box 的中心点与初始位置计算欧氏距离, 距离最近的点所在的位置即为框的位置; 同时对 5 帧之内取平均值, 5 帧之后进行滑动平均更新坐标.

考虑到学生每秒钟之内状态不会发生变化, 以秒为单位计算每类状态数量最大作为这一秒钟的状态标签, 并据此统计每个学生课堂行为状态持续时间分布饼图及时序图.

1.4 学生课堂注意力量化评价

对每位同学的课堂行为状态时序图进行分析, 统计同学各类课堂行为发生的时刻及持续时间, 据此探索构建学生课堂注意力量化评价标准, 对每节课每位学生的听课质量打分, 并以课堂行为状态时序图、分布饼图等可视化方法, 使学生了解自身听课质量. 具体方法如下:

步骤 1: 得到抬头、看书、侧脸及玩手机的真正有效时长.

首先, 考虑到动作持续频率与持续时间, 设立“抬头听讲”动作有效的最短时长, 从而将抬头听讲与其他动作区分: 如果抬头动作持续时间大于 30 s, 则认为抬头有效, 该同学再在真听讲; 如果持续时间小于 30 s:

若后续动作为看书,则认为抬头有效,该同学为抬头听讲并学习课本内容;若后续动作为交头接耳或玩手机,则认为抬头无效,因此将该抬头时间计入后续动作中,据此对学生课堂行为状态时序图进行修正。

其次,针对遮挡情况,如果“低头遮挡”时长小于等于5 min,且整节课“抬头听讲”时间超过20 min,则将“低头遮挡”时长计算为“看书”时长,否则计算为“玩手机”时长。

步骤2:赋予各状态权重,得到初始课堂注意力得分。

由于课堂以“听讲”为主,“低头看书”次之,两者均为与课堂相关的动作,则赋予两者权重分别为0.6、0.3,而“交头接耳”可能与同学探讨问题或说与课堂无关话题,则赋予权重0.1,而“玩手机”为与课堂无关动作,赋予权重0,据此得到学生课堂注意力初步量化公式:

$$s_0 = \frac{0.6 \times t_{rise} + 0.3 \times t_{book} + 0.1 \times t_{side}}{t_{rise} + t_{book} + t_{side} + t_{mobilephone}} \times 100 \quad (7)$$

步骤3:为了让全班同学成绩连续并符合高斯规律,按照山东高考分级赋分原则^[19]计算学生课堂注意力得分。

首先,将所有成绩按比例分为8个等级,每个等级对应一个分数区间,如表1所示。

表1 山东等级方案分换算对应表

等级	A	B+	B	C+	C	D+	D	E
比例(%)	3	7	16	24	24	16	7	3
赋分区间	91-100	81-90	71-80	61-70	51-60	41-50	31-40	21-30

接着,根据每个等级所对应的原始分区间与相应的等级赋分区间,按等比例转换法则计算每个考生等级分,公式为:

$$\frac{s_2 - s_0}{s_0 - s_1} = \frac{t_2 - t_0}{t_0 - t_1} \quad (8)$$

其中, s_1 、 s_2 分别表示某个等级所对应学生课堂注意力得分区间的下限和上限; t_1 、 t_2 分别表示相应等级的等级分赋分区间的下限和上限; s_0 表示相应等级内某学生课堂注意力得分; t_0 表示相应等级内此学生的最终等分。

最后将 t_0 与表2对照得到学生本节课课堂注意力量化等级。

表2 学生课堂注意力量化等级

注意力量化得分	85-100	75-85	60-75	0-60
评价等级	优	良	中	差

1.5 教师课堂授课效果量化评价

学生听课抬头率是一个评价教师授课效果的重要指标。由于学生抬头低头变化频率不高,因此此量化标准每分钟计算一次班级抬头率,课程结束后计算所有抬头率的平均值。教师授课效果等级由表3所示。

表3 教师授课效果量化表

抬头率	75%-100%	50%-75%	25%-50%	0-25%
评价等级	优	良	中	差

2 实验及结果分析

2.1 实验平台

本实验摄像设备为海康威视 PTZ,实验硬件为 PC,操作系统是 Windows 10 64 位系统,PC 处理器型号为 Intel(R) Core(TM) i7-7700,显卡型号为 NVIDIA GTX 1050Ti,深度学习框架 TensorFlow-CPU 1.13.1、Keras 2.1.5,编程语言为 Python, GPU 加速库为 CUDA 10.2 和 CUDNN 7.0。

2.2 实验数据集的采集和标注

本实验数据集是使用山东建筑大学信息楼 417 教室监控摄像头,拍摄学生自然上课状态,截取了 3 200 张照片。考虑到教室中人数较多,学生之间存在遮挡情况,无法区分学生低头看书或玩手机,故增加一类“低头遮挡”。本研究使用 Labelimg 标注工具标注矩形框,注意力状态为“rise(听讲)”“mobilephone(玩手机)”“side(交头接耳)”“book(低头看书)”和“bow(低头遮挡)”5 类状态,并以 PASCAL VOC 格式保存 XML 文件。

本实验对同一教室内不同场景进行测试,实验图片如图4所示。本研究共建立 3 200 张照片,其中包含不同光照、不同角度的丰富数据集,提高了实验的可靠性。

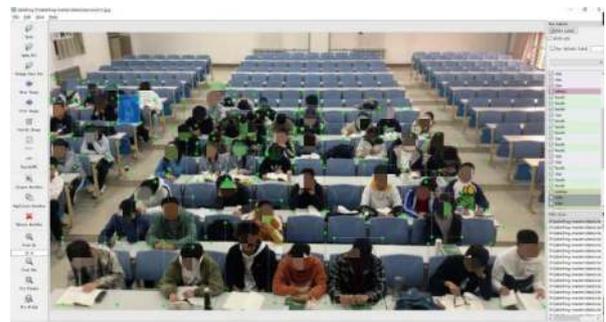


图4 标注后的视频帧

2.3 模型的训练

在模型训练之前需要设置 YOLOv4 相关参数。本

实验采用学习率机制,并冻结训练加快训练速度,共训练 150 个训练周期 (Epoch),前 75 个训练周期 batch_size 为 2,初始学习率为 $1e^{-3}$,后 75 个训练周期 batch_size 为 1,初始学习率为 $1e^{-4}$.为防止数据集过拟

合,应用 callback 实现早停法,当学习率到 $1e^{-6}$ 时停止学习.将数据集分按照 9:1 的比例划分为训练集与测试集,分别为 2880、320 张,训练集与测试集中各类型标签数量如表 4 所示.

表 4 样本类型与数量

数据类型	训练集				测试集					
	rise	book	bow	mobile phone	side	rise	book	bow	mobile phone	side
数量	57 903	20 852	24 632	11 276	6 530	5 964	2 243	2 587	1 331	788

2.4 结果分析

用一段课堂视频,通过识别学生课堂行为状态,按照上述模型进行计算评价.

2.4.1 学生课堂行为状态识别结果分析

本研究使用 *mAP* 与 *F1-score* 作为模型的评价指标.

(1) *mAP*

精准率 *P* 代表被正确识别的样本占所有被识别样本的比例,召回率 *R* 代表正确识别样本占所有应该被识别样本的比例,计算公式如下:

$$P = \frac{TP}{TP+FP} \times 100\% \quad (9)$$

$$R = \frac{TP}{TP+FN} \times 100\% \quad (10)$$

其中, *TP* 表示正确识别的学生课堂行为状态的数量; *FP* 表示错误识别的学生课堂行为状态数量; *FN* 表示漏识别的学生课堂行为状态数量.

根据式 (9) 可以绘制出 *P-R* (precision-recall) 曲线,单类别分类准确率 *AP* (average precision) 是对 *P-R* 曲线做平滑处理,积分计算曲线下的面积. *mAP* (mean average precision) 是各个分类准确率 *AP* 的平均值, *mAP* 的有关计算公式如下:

$$AP_i = \int_0^1 p(r)dr, i = 1, 2, \dots, n \quad (11)$$

$$mAP = \frac{\sum AP_i}{n}, i = 1, 2, \dots, n \quad (12)$$

其中, *p* 为精准率, *r* 为召回率, *n* 为类别数.

(2) *F1-score*

F1-score 又称 *F1* 分数,是分类问题中精准率 *P* 和召回率 *R* 的调和平均数,常用作多分类问题的最终指标.多类问题的 *F1* 为所有类别的平均值, *F1* 的计算公

式如下所示:

$$F1_i = 2 \times \frac{R_i \times P_i}{R_i + P_i}, i = 1, 2, \dots, n \quad (13)$$

$$F1 = \left(\frac{1}{n} \sum F1_i \right)^2, i = 1, 2, \dots, n \quad (14)$$

针对上述模型评价指标,本研究以相同的实验数据集分别使用 YOLOv3、YOLOv4、改进 YOLOv4 和本文算法网络模型进行训练,在相同的情况下进行训练识别对比.结果如下所示.

从表 5 可以看出,在使用相同数据集情况下, YOLOv4 算法比 YOLOv3 算法 *mAP* 提高 3.24%, *F1* 值提高 0.015;将 YOLOv4 卷积块中的激活函数更改为 ELU 函数后, *mAP* 提高 3.29%, *F1* 值提高 0.036;使用 Soft-DIoU-NMS 进行非极大值抑制改进后, *mAP* 提高 0.24%, *F1* 值提高 0.01, 本文的改进算法相比较于原始的 YOLOv4 算法 *mAP* 提高了 3.53%, *F1* 值提高了 0.046. 本文使用的改进算法识别教室中学生自然状态下上课图片效果如图 5 所示,对于后排遮挡的同学依然有很好的检测效果,注意力状态标签统计结果如表 2 所示.

表 5 不同网络的模型评估结果

模型	激活函数		非极大值抑制		<i>mAP</i> (%)	<i>F1</i>
	LeakyReLU	ELU	DIoU-NMS	Soft-DIoU-NMS		
YOLOv3	√	—	√	—	92.03	0.907
YOLOv4	√	—	√	—	95.27	0.922
改进YOLOv4	—	√	√	—	98.56	0.958
本文算法	—	√	—	√	98.80	0.968

由表 6 可知,本研究改进的算法对学生课堂行为状态的 5 个状态都具有较好的识别效果.



图5 本文算法识别结果

表6 各类标签识别效果

注意力行为	TP	FP	AP (%)	F1
bow	3953	1410	97.63	0.96
side	1028	739	98.38	0.94
book	3832	1748	99.33	0.97
mobilephone	1724	221	98.85	0.98
rise	11309	1605	99.82	0.99

2.4.2 学生课堂注意力量化评价结果

选取山东建筑大学信息楼 417 教室内学生课堂监控视频 45 min, 分析其中一名学生的课堂行为状态, 本研究算法对学生课堂行为状态识别统计结果如图 6(a) 所示: 其中, 该生本节课“抬头听讲”类别共 1084 s, 小于 20 min, 根据第 1.4 节制定的量化规则, 该生“低头遮挡”的时间应划分至“玩手机”的类别中; 同时, 2000 s 之前的“抬头听讲”时常均大于 30 s, 但第 7 段“抬头听讲”时长小于 30 s, 紧接着该生“玩手机”, 因此, 将第 7 段“抬头听讲”划分至“玩手机”类别中, 修正后的时序图如图 6(b) 所示, 饼图如图 7(b) 所示。

根据统计结果, 该生本节课“抬头听讲”“看书”“交头接耳”和“玩手机”4 类状态的有效时长分别为 1052 s、198 s、61 s 和 1389 s, 按照式 (7) 可得, 初始得分 $s_0=25.80$ 分, 是该同学所在班级本节课成绩的前 82%, 查询表 1 可知该生为 D+ 等级, $s_1=20.6$ 、 $s_2=41.9$ 、 $t_1=41$ 、 $t_2=50$, 根据式 (8), 得到该学生课堂注意力量化评价最终得分为 43.19, 注意力评价为差。

同时使用人工观察法统计该生本节课的行为状态, 结果对比如表 7 所示, 人工观察法所示各类状态时长计算初始分数 $s_0=25.75$, 课堂注意力量化评价最终得分为 43.56, 注意力评价为差。与本研究算法结果相同, 因此本研究算法具有良好的可靠性。

为了验证本研究提出的学生课堂注意力量化准则

的合理性, 将该学生的本节课课堂监控视频交由 5 名无关同学, 使用“优”“良”“中”“差”4 个等级评价其课堂注意力, 得出结果分别为“差”“中”“差”“差”“差”, 总体结论为该学生本节课课堂注意力表现为“差”, 与本研究提出的量化准则评价结果一致, 证明本研究提出的量化准则可以正确评价学生课堂注意力表现。

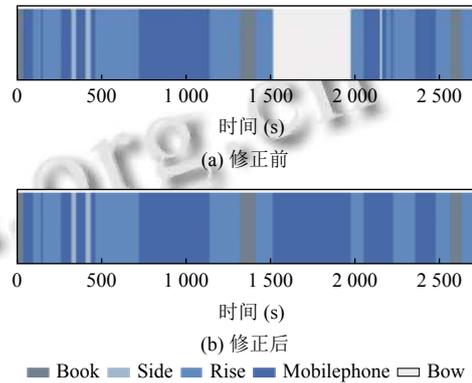


图6 修正前后的该学生课堂行为状态时序图对比

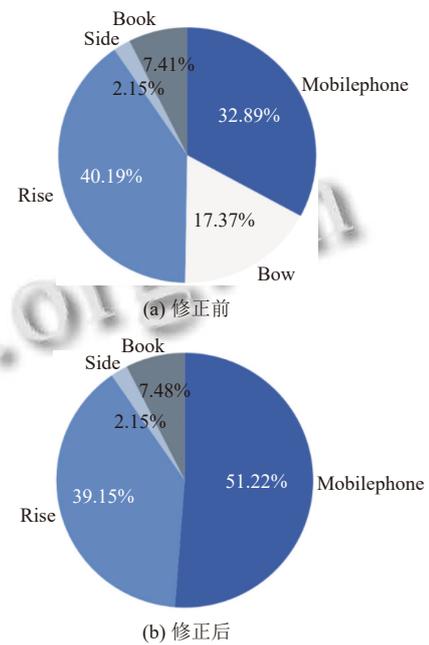


图7 修正前后的该学生课堂行为状态饼图对比

表7 两种方法的行为状态对比

方法	课堂行为状态				量化得分	等级评价
	抬头听讲	看书	交头接耳	玩手机		
本研究算法	1052	198	61	1389	43.19	差
人工观察法	1044	212	53	1391	43.56	差

2.4.3 教师课堂授课效果量化评价结果

对山东建筑大学信息楼417教室某节课45 min中学生抬头状态进行识别,每分钟统计一次抬头学生数量,根据统计结果计算本节课抬头率为51.4%,人工观察法计算得到抬头率为51.3%,两者结果相差不大,据表2的量化标准量化等级均为“良”,两种方法计算的抬头率折线对比图如图8所示。

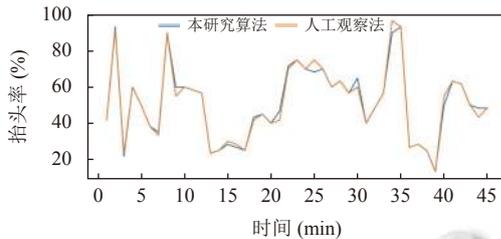


图8 两种方法的学生课堂抬头率统计折线对比图

3 结论

随着深度学习的不断发展,传统的教室授课模式呈现出很多弊端,已经无法适应新的教学环境与观念的转变.使用深度学习的方法识别学生上课注意力状态,从而判断老师授课效果与学生的课堂听课态度,可以端正学生听课态度,提高教师的授课效率,有效地提升课堂教学效果。

本文使用改进的YOLOv4目标检测算法检测教室中学生上课状态,统计各状态时长及变换频率,据此判断学生听课及教师授课效果.识别结果准确率较高,能够准确地反映课堂状况.对于遮挡程度过高的情况,仍存在无法准确识别的问题,今后可考虑使用双目摄像头,多角度拍摄学生上课视频,提升遮挡情况下识别的准确率。

参考文献

- 魏艳涛,秦道影,胡佳敏,等.基于深度学习的学生课堂行为识别.现代教育技术,2019,29(7):87-91.
- 袁笑文.情感教学法在大学英语课堂教学中的应用研究[硕士学位论文].长春:吉林大学,2014.
- Stanley D. Measuring attention using microsoft kinect [Master's thesis]. Rochester: Rochester Institute of Technology, 2013.
- 党冬利.人体行为识别及在教育录播系统中的应用[硕士学位论文].西安:西安科技大学,2017.
- 廖鹏,刘宸铭,苏航,等.基于深度学习的学生课堂异常行为检测与分析系统.电子世界,2018,(8):97-98.
- Simonyan K, Zisserman A. Two-stream convolutional networks for action recognition in videos. arXiv: 1406.2199, 2014.
- Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 779-788.
- Redmon J, Farhadi A. YOLO9000: Better, faster, stronger. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 6517-6525.
- Redmon J, Farhadi A. YOLOv3: An incremental improvement. arXiv: 1804.02767, 2018.
- Bochkovskiy A, Wang CY, Liao HYM. YOLOv4: Optimal speed and accuracy of object detection. arXiv: 2004.10934, 2020.
- Liu W, Anguelov D, Erhan D, et al. SSD: Single shot MultiBox detector. arXiv: 1512.02325, 2016.
- Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014. 580-587.
- Girshick R. Fast R-CNN. Proceedings of the 2015 IEEE International Conference on Computer Vision. Santiago: IEEE, 2015. 1440-1448.
- Ren SQ, He KM, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks. Advances in Neural Information Processing Systems 28. Montreal: Curran Associates, Inc., 2015. 91-99.
- Misra D. Mish: A self regularized non-monotonic neural activation function. arXiv: 1908.08681, 2020.
- Clevert DA, Unterthiner T, Hochreiter S. Fast and accurate deep network learning by exponential linear units (ELUs). arXiv: 1511.07289, 2016.
- Neubeck A, van Gool L. Efficient non-maximum suppression. 18th International Conference on Pattern Recognition (ICPR'06). Hong Kong: IEEE, 2006. 850-855.
- Zheng ZH, Wang P, Liu W, et al. Distance-IoU loss: Faster and better learning for bounding box regression. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12993-13000. [doi: 10.1609/aaai.v34i07.6999]
- 郑焕,邹樱.新高考选考科目等级赋分方式的衍变.教育导刊,2020,(5):51-57.