

基于 Yolov5 和 Dlib 的在线违规行为监控识别算法*

封彦^{1**}, 鲁磊²

(1. 西安工程大学, 西安 710048; 2. 西安交通大学信通学院, 西安 710049)

摘要: 使用双摄像头监控视频, 设计一种考试作弊异常行为检测算法, 检测考生替考、考试现场出现考试无关的物品及人员、考生在考试过程中说话、考生注意力漂移等四种常见的考试异常行为。首先使用基于注意力 YOLOv5s 算法检测考试中的禁用物品; 然后用基于 Dlib 的人脸识别确认考生的身份信息和有无替考行为; 最后利用唇动检测和头部姿态估计进行考生的异常行为检测。双摄像头的使用减少了视野死角, 防止考生在摄像头的视野死角区域放置作弊工具。实验表明所提的算法具有较好的检测精度和实时性。

关键词: 图像分类; 人脸识别; 异常行为检测; 注意力机制; 卷积神经网络

中图分类号: TP391.41 **文献标志码:** A **文章编号:** 1005-488X(2023)03-0276-07

A Violation Behavior Detection Algorithm Based on Yolov5 and Dilb for Online Video Surveillance

FENG Yan, LU Lei

(1. Xi'an Polytechnic University, Xi'an 710048, CHN; 2. School of Information and Communication Engineering, Xi'an Jiaotong University, Xi'an 710049, CHN)

Abstract: A dual-camera monitor was proposed and a detection algorithm was designed for abnormal cheating behavior in exams, which detected four common exam abnormal behaviors of candidates, such as: carrying prohibited items, surrogate exam-taker, speaking during the exam, and drifting of candidates' attention. Firstly, the attention-based YOLOv5s algorithm was used to detect prohibited items in exams. Then, Dlib-based facial recognition was applied to confirm the candidates' identity information. Finally, lip movement detection and head pose estimation were utilized to detect abnormal behavior of candidates during the exam. The use of dual cameras could reduce blind spots in the field of view to prevent candidates from using cheating tools. The experiment results showed that the proposed algorithm had good detection accuracy and real-time performance.

Key words: image classification; face recognition; abnormal behaviors detection; attention mechanism; convolutional neural network

收稿日期: 2023-07-15

* 基金项目: 中国博士后科学基金面上资助项目(2022M712555)

作者简介: 封彦(1984—), 女, 硕士, 主要研究方向为图像处理、深度学习、信号分析; (E-mail: 303660626@qq.com)

鲁磊(1988—), 男, 博士, 讲师, 主要研究方向为图像处理、深度学习、信号分析。

** 通讯作者

引言

受 COVID-19 疫情影响,2020 年全球有 180 多个国家近 13 亿在校学生无法正常到校上课。2020 年度春季,我国 1400 多所高校开出了 700 多万门次在线课程。除此之外,像 MOOC 这样的在线课程也越来越受学生欢迎。传统的高等教育向网络的转型为学生提供了额外的学习途径,促使更多的学生学习到更加丰富的知识。与此同时,学生接受了在线课堂教育则需经过线上的方式完成结课考试以获取学分或结课证书。

然而,少数学生绞尽脑汁的想着各种作弊措施,而视频会议只具有简单的图像视频录制、存储和传输的功能,教师采用分屏视频显示监考超过 30 人的在线考试时纷纷表示看不过来。市面上的主流在线考试平台则从监控考生多媒体设备的角度防止作弊行为,提供身份识别、视频监控、切屏强制交卷、随机题目随机选项等功能。随着人工智能在视频异常行为检测领域的发展^[1],如何利用深度学习和机器视觉技术开发一个在线考试异常行为识别算法,减少因为人工疲劳造成的漏检现象,提高审查效率,保证线上考试的公平性是一个关键问题。

考试作弊异常行为检测算法则是要完成对异常行为的自动检测,其中包括行为建模和行为分类^[2]。传统的考场作弊行为识别研究通常采用对异常行为进行建模的方式,根据异常行为的特点提取具有代表性和易识别性的特征,然后使用模式匹配计算匹配度从而判断该行为是否为异常行为。如果使用机器学习分类器对异常和正常行为的特征进行分类则需要带标签的正常和异常样本训练模型。常用的特征包括头发和皮肤的颜色、头部姿态及视线等。

李慧^[3]使用 RGB 色度分布和 YCgCr 肤色分割技术分别提取桌面区域和手部区域后,提出了根据桌面区域手部数量的异常行为判断准则锁定作弊考生。雷宇辉^[4]在 HOG-LBP+SVM 人脸检测的基础上,结合考生肤色、发色特征,提出一种新的基于人脸、肤色发色的考生目标检测方法和发色-混合高斯模型的考生异常行为识别方法,通过圈定考生的头部圈活动范围内的发色变化判断考生的身体位姿变换对教室中考场作弊行为进行识别并预警。李永恒^[5]首先在肤色分割的基础上利用 LBP 和 Ada-

Boost 进行人脸检测,然后用 LK 光流实现 ORB 特征点运动估计并用 Kmeans 对运动矢量进行分类,最后将与前一帧相同区域的直方图相似度作为作弊概率。除了使用颜色特征外,李永恒还采用了基于平面图像的头部运动姿态估计计算欧拉角 yaw 参数识别异常行为。Cote 等人^[6]同样提出基于头部姿态估计的考生异常行为检测,通过建立隐马尔可夫模型检测考生的头部姿态异常行为。

在视线追踪方面,Javed 等人^[7]提出了一种基于卡尔曼滤波的眼动跟踪在线作弊检测方法。当检测并跟踪到的考生瞳孔区域超过设定的阈值时,检测方法就会判定考生具有作弊行为。Martynow 等人^[8]首先使用 Haar+AdaBoost 算法检测人脸和眼睛区域,然后用 MeanShift 算法跟踪眼睛眼部区域,最后使用 SVM 分类器判断被跟踪眼睛区域的注视区域是在屏幕外还是屏幕内,从而判断受试者是否有异常行为。Zou 等人^[9]开发了一个只需要少量样本的基于视线检测的辅助作弊检测系统原型。

针对传统算法特征难提取和识别率低的问题,田佳莉^[10]首先建立了包含传递材料、转身扭头、低头看资料、前倾站立和弯腰拣资料 5 种考场异常行为的数据集,然后使用基于冻结网络优化的 YOLOv3 和 YOLOv4 模型实现异常行为检测算法。付康^[11]则按照东张西望、趴在桌上、手在桌下和传递物品这四种考试异常行为构建数据集,用于训练 YOLOv3 与 Mobilnet-SSD 模型进行考试异常行为检测。Fang 等人^[12]在 YOLOv3 算法的基础上,通过使用 K-Means 算法、GIoUloss、焦点丢失和 Darknet32 进行改进,提高检测精度和检测速度。洪钟^[13]首先使用改进的 YOLO-D 网络进行目标检测,然后使用运动历史图像确定考生正常活动范围用于检测其是否存在左右偷看、传递纸条等异常行为。作者在使用肤色检测手部区域后,根据手部区域离开桌面的时间判断考生是否存在偷看手机的异常行为。于明学等人^[14]首先利用改进的 3D 卷积神经网络分别提取双路监控视频中的行为特征,然后将两路特征向量进行融合,最后使用 LibSVM 分类器判断视频中是否存在异常行为,提高识别准确性。Alairaji 等人^[15]将头部运动检测、虹膜运动检测和手部运动检测结合在一起,以识别同一学生的手和脸之间以及不同学生之间的接触。

在线考试中的异常行为监控方面,Hu 等人^[16]通过网络摄像头获取考生的头部姿势和口腔状态信

息,并对在线考试中考生的异常行为进行判别。Atabay 等人^[17]则利用 BeatGAN 算法根据考生的骨骼运动信息,检测远程电子考试期间考生的异常行为。

基于 CNN 的目标检测主要分为两种:一阶段检测方法和两阶段检测方法。两阶段检测法需要对候选区域做二次修正,检测精度较高但检测速度较慢。近年来,随着一阶段算法的检测精度不断提高,YOLO 系列^[18-20]、RetinaNet^[21]、EfficientDet^[22]等方法被广泛的应用在各种工业领域视觉检测任务中。虽然 YOLO 系列算法中最新的 YOLOv6^[19]和 YOLOv7^[20]对 Darknet-53 主干网络、标签分配策略、预测头和损失函数都做了进一步的优化,但 YOLOv5^[18]由于其性能(实时性和准确性)的平衡仍具有广泛的应用。由于注意力机制复杂,在大数据集上取得了优异性能的 DETR^[21]、Swin Transformer^[22]在工业场景中的部署效率较低。为了工业部署方便,SqueezeNet^[23]、MobileNet 系列^[24]、ShuffleNet 系列^[25]算法主要研究轻量级模型。

相比于教室考场环境作弊异常行为识别算法研究,在线考试的作弊异常行为识别算法研究较少。考虑到使用一个摄像头的方式并不能很好监控学生的考试状态,提出了一种基于双摄像头监控视频的人工智能在线考试作弊行为识别算法,将 Dlib 算法和基于注意力与 ShuffleNetv2 改进的 YOLOv5 模型结合完成检测考生替考、考试现场出现考试无关的物品及人员、考生在考试过程中说话、考生注意力漂移四种常见的作弊行为。该算法除了可以为疫情期间的在线考试提供初步筛选,提醒监考教师重点关注考生外,还可以为在线的证书类学习课程(如 MOOCs)的考试提供监控评估,提高考试公信力。

1 算法设计

1.1 算法整体框架

文章提出的考试作弊异常行为检测算法主要研究以下四种典型的潜在作弊行为:

(1) 考试现场出现与考试无关的物品和人员

在考试的过程中,如果摄像头采集到的图像中出现纸条、书本、手机或 pad 等其他电子设备、纸质工具以及除考生外的其它人员,则判断为潜在作弊,需要对考生进行警告并对监考教师进行预警。

(2) 考生替考

在考试开始前,首先需要对考生的身份进行验证,并采集考生端坐面向屏幕的照片作为模板或正样本。在考试的过程中,实时对屏幕前的考生是否为考生本人进行验证。

(3) 考生在考试过程中说话

如果视频中考生嘴唇的运动超过预设的范围及时间,则判断考生有与人交流等潜在作弊行为,需要对考生进行警告及对监考老师进行预警。

(4) 考生注意力漂移

如果在考试过程中,考生的视线不集中在屏幕区域超过阈值时间,则判断考生有查看资料等潜在作弊行为,需要对考生进行警告及对监考老师进行预警。

算法的具体流程框图如图 1 所示,首先利用改进的 YOLOv5^[18]算法对考试人员以及相关物品进行检测,若检测到考试作弊物体,则做出作弊预警;在检测出人脸后将人脸图像输入到基于 Dlib 的人脸识别模块中,通过人脸识别判断考生在考试过程中有无替考行为;每 3 s 利用人脸特征点对唇部与头部姿态进行建模,检测有无说话行为和检测考生的头部转向来判断考生注意力是否漂移。

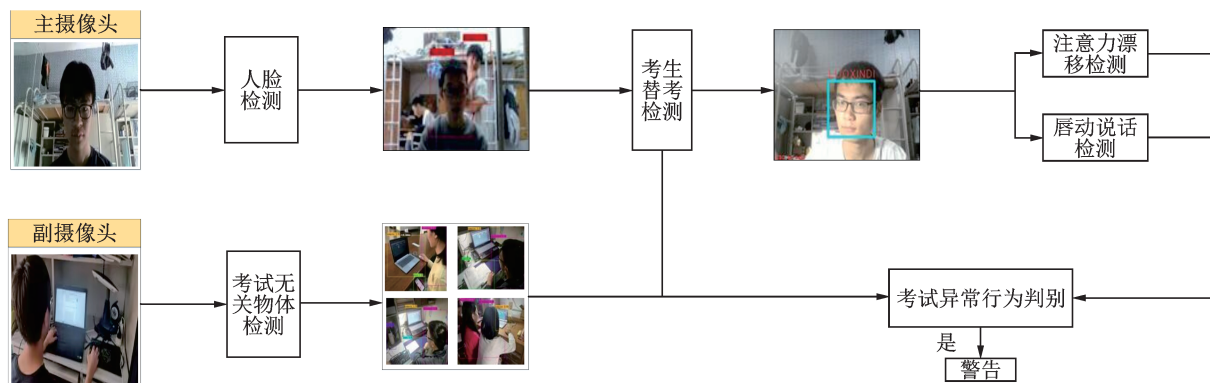


图 1 算法流程框图

Fig.1 Algorithm flow diagram

1.2 考试无关的物品和人员检测算法

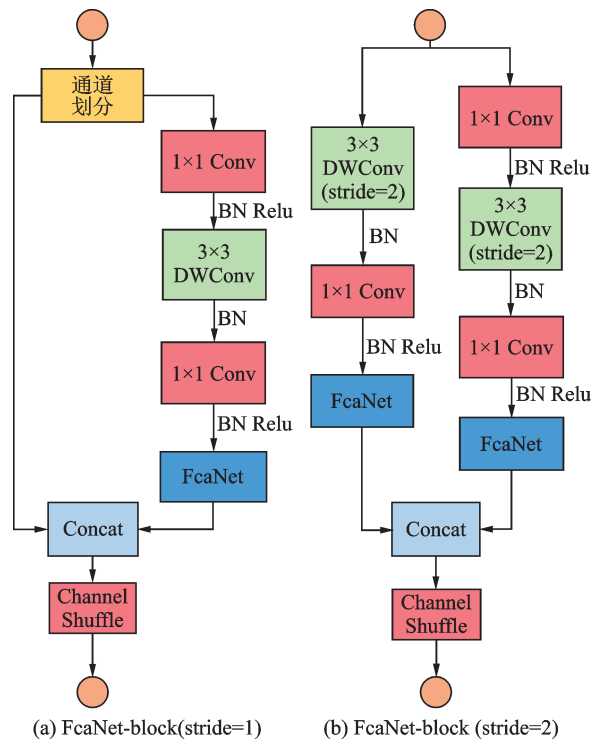
在考试中,考生可能会使用手机、书本、纸张、ipad等物体完成作弊行为,也可能通过他人完成考试作弊。考虑到需要同时对多路视频进行考试无关物品和人员的检测,文章采用了高性能轻量级的ShuffleNetV2^[25]网络替代YOLOv5的特征提取主干网络,并用注意力机制对ShuffleNetV2网络中的两种基本单元(block)进行了改进,实现高效和精准的目标和人员检测。

图2(a)为加入了注意力机制的stride=1的基本单元,首先在输入端利用通道划分(channel split)将输入特征图分为两个分支,在原本右侧主分支的1×1卷积、3×3深度可分卷积和1×1卷积后加入了FcaNet注意力模块^[26],左侧分支则为恒等映射。图3(b)为加入了注意力机制的stride=2的基本单元。为了加倍输出特征图,该基本单元在输入端没有进行通道划分就将输入特征图分为两个分支。在保持右侧分支与stride=1的改动相同的基础上,文中在左侧3×3深度可分卷积和1×1卷积后同样加入了FcaNet注意力模块。FcaNet是一种根据不同通道的DCT变换后频率分量的重要性选择对高性能的频率分量所对应的通道进行加权的通道注意力模块。

1.3 考生替考检测算法

传统考试在进入考场前教师会使用打印好的考生表人为的对每一个考生进行人脸识别,检测考生是否存在替考行为。在线考试同样需要在考试前对考生进行人脸识别从而确定考生的身份,替考行为检测流程如图3所示。

在主摄像头获得考生的正面人脸图像后,文章使用了Dlib库进行人脸和关键点检测。首先对图像进行预处理,包括归一化、人脸光线补偿、灰度处理、直方图均衡化和图像滤波;其次使用基于梯度直方图(HOG)算法得到人脸特征点;然后使用预训练好的深度残差网络模型产生每副输入人脸图像的128维特征向量;最后与本地数据库进行对比,来识别考生的身份。本算法对考生每隔3s检测一



(a) FcaNet-block (stride=1) (b) FcaNet-block (stride=2)

图2 改进的ShuffleNetV2两种基本单元

Fig.2 Improved ShuffleNetV2 with two blocks

次,从而判断有无替考行为。

1.4 考生在考试过程中说话检测

在获得68个人脸特征点后,分别取左内嘴角61号点分别与62和68号点的夹角 α ,上下内嘴角63和67号点之间的距离 h_1 ,左右内嘴角61和65号点之间的距离 w ,对唇部进行建模,如图4所示。当 α 和 h_1/w 大于阈值时,则判断考生有张嘴现象。

1.5 考生注意力漂移检测

在正常考试中,考生的活动范围有限,其正常答题状态应是大幅度运动。当考生在考试中使用书本手机等作弊工具时,会发生头部转向行为。文章使用的头部姿态估计依赖于考生的人脸检测中检测到的68个面部特征点中的8个特征点(如图5所示),分别对应眼睛、鼻尖、嘴角和下巴。在使用张正友棋盘格算法进行摄像机标定后,在不考虑摄像机径向畸变的情况下,可将2D人脸特征点转换

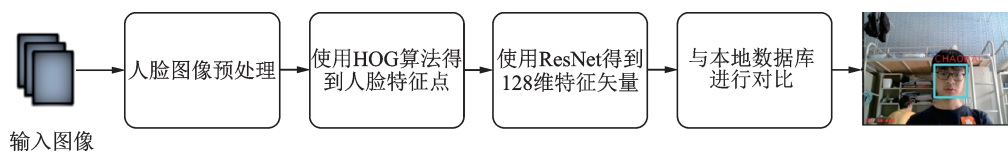


图3 替考行为检测流程框图

Fig.3 Flow diagram of proxy test behavior detection

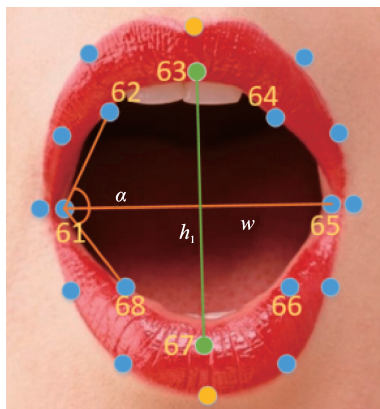


图 4 嘴唇区域建模示意图

Fig.4 Schematic of lip area modeling

到 3D 坐标系下。文章首先使用了 OpenCV 提供的 SolvePnP Ransac 姿态估计 API 完成考生的头部姿态估计,然后用面部的横向角度和纵向角度即可描述考生当前的视线方向。当考生的视线方向偏转大于阈值时,则判断考生有注意力漂移现象。

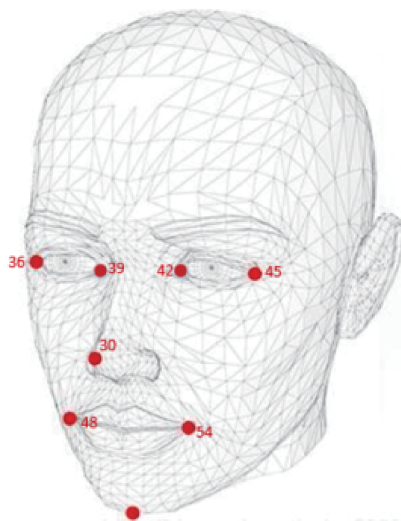


图 5 头部姿态建模示意图

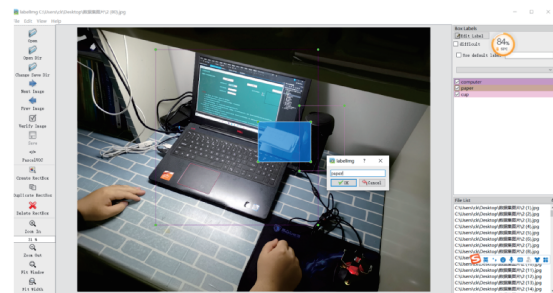
Fig.5 Schematic of head pose modeling

2 实验结果及分析

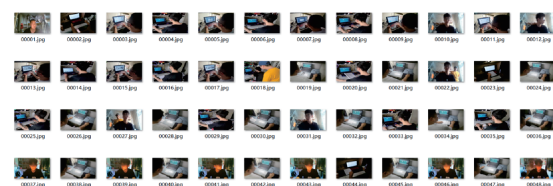
2.1 数据集

实验中拍摄 10 000 张图像,在 python3.8.1 中利用 labeling 库对 10 000 张 640×640 分辨率图片中的人和书本、纸张、手机、平板和杯子等物体,以及考生的姓名、是否有注意力偏移和说话现象作为标签,建立数据集 TCS2021(Test Cheat System)。标注过程和部分图片如图 6 所示。由于采用了双视角,TCS2021 分为正视角和副视角两部分。正视角

主要采集考生的正面图像,包括考生头部转动和开口等异常行为,一共 3 000 张。副视角主要包括视野中出现书本、纸张、手机等 7 类作弊物品,一共 7 000 张。然后将数据集按照 7:2:1 的方式划分为训练集、验证集和测试集。



(a) 标注过程



(b) 部分图片

图 6 标注过程和部分图片

Fig.6 Annotation process and partial images

2.2 考试无关的物品和人员检测实验结果与分析

提出的基于 FcaNet 和 SufferNetv2 改进的 YOLOv5 算法(FS-YOLOv5)首先使用 COCO 数据集进行预训练,得到对笔记本电脑、手机、书籍等常见目标和人的检测的预训练模型,然后使用 TCS2021 进行微调并加入了适当的数据增强。测试使用了一张 RTX 2080 显卡,初始学习率设置为 0.01,并且以每 10 个 epoch 5% 的速率衰减,训练过程共 30 个 epoch,SGD 动量为 0.937,优化器权重衰减为 0.000 5,批归一化大小为 16,置信区间设置为 95%,IOU 交并比为 0.5。部分检测结果如图 7 所示。

为了验证提出的基于注意力和 SufferNetv2 改进的 YOLOv5 算法的性能,研究使用有类别的平均



图 7 考试无关人员和物品检测结果

Fig.7 Examination results for unrelated personnel and items

精度(mean average precision, mAP)和检测速度作为性能指标,与工业上主流使用的目标检测算法YOLOv4,Faster R-CNN,SSD进行对比实验,实验结果如表1所示。

表1 算法性能对比

Tab.1 Performance comparison of algorithms

方法	mAP@0.5	FPS(帧/s)
YOLOv4	64.1	22.8
YOLOv5	66.3	37.6
Faster R-CNN	67.2	13.1
SSD	55.7	15.3
FS-YOLOv5(ours)	69.6	36.3

2.3 考生替考检测算法实验结果与分析

本次实验录用正常2个教学班61位同学的人脸信息,分别进行20次测试。总体的识别准确率可以达到93.6%。部分测试结果如图8所示。

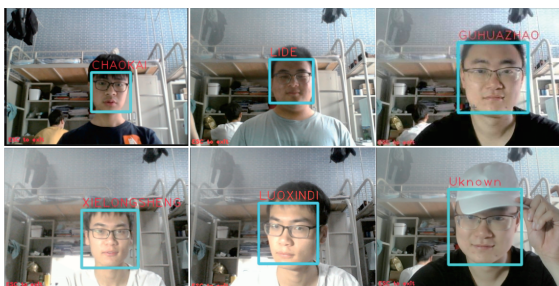


图8 考生替考检测结果

Fig.8 Test results of candidate substitution

从图8可以看出,在灯光较暗的环境下以及考生面部被部分遮挡时(佩戴口罩,帽子等物品),算法检测正确率会降低。

2.4 考试过程中说话检测算法实验结果与分析

考生的说话检测每隔3s运行一次。当考生在考试过程中说话时,嘴唇运动的建模参数 $\alpha > 15^\circ$ 或 $h_1/\tau_w > 0.5$ 时会做出预警。部分检测结果如图9所示。本部分算法的识别准确率为93.1%。

2.5 考生注意力漂移检测算法实验结果与分析

考生注意力漂移检测每隔3s运行一次,若考生头部发生转动的转向角度大于 40° 时且时间长于2s就会发生预警。部分检测结果如图10所示。本部分算法的识别准确率为87.6%。当考生的头部转向幅度过大时无法准确检测结果。

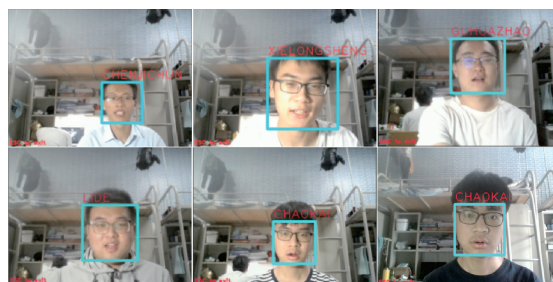


图9 考生说话检测结果

Fig.9 Test results of candidates' speaking detection

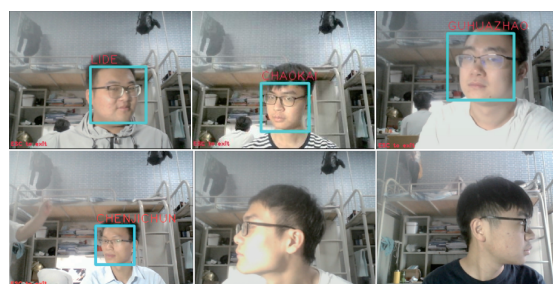


图10 考生注意力漂移检测结果

Fig.10 Test results of examinees' attention drift detection

3 结束语

针对在线考试监考困难的问题,文章采用双摄像头监控视频设计了一种基于注意力YOLOv5和Dlib的在线考试作弊异常行为识别算法。首先对采集到的图像和视频使用基于注意力YOLOv5算法进行目标检测,检测考试中严禁使用的物品;然后用基于Dlib的人脸识别进行考生的面部识别,及时确认考生的身份信息和有无替考行为;最后利用抖动检测和头部姿态估计进行考生的异常行为检测。双摄像头的使用减少了视野死角,防止考生在摄像头的视野死角区域放置作弊工具。所提的算法具有较好的检测精度和实时性。

参 考 文 献

[1] Mu H, Sun R, Yuan G, et al. Abnormal human behavior detection in videos: A review[J]. Information Technology and Control, 2021, 50(3): 522-545.

[2] Balderas A, Caballero-Hernández J A. Analysis of learning records to detect student cheating on online exams: Case study during COVID-19 pandemic [C]. Eighth International Conference on Technological Ecosystems for Enhancing Multiculturality, Association for Computing Machinery, New York, USA, 2020: 752-757.

[3] 李 慧. 基于视频的考场异常行为识别研究[D]. 天津:天津财经大学, 2012.

[4] 雷宇辉. 基于视频监控的考场异常行为识别方法及实现[D].

- 成都:成都理工大学, 2019.
- [5] 李永恒. 基于头部运动分析的考场内异常行为检测[D]. 长春: 长春理工大学, 2018.
- [6] Cote M, Jean F, Albu A B, et al. Video summarization for remote invigilation of online exams[C]. 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Placid, NY, USA: IEEE, 2016: 1-9.
- [7] Javed A, Aslam Z. An intelligent alarm based visual eye tracking algorithm for cheating free examination system[J]. International Journal of Intelligent Systems and Applications, 2013, 5(10): 86.
- [8] Martynow M, Zielińska A, Marzejon M, et al. Pupil detection supported by Haar feature based cascade classifier for two-photon vision examinations[C]. International Symposium on Image and Signal Processing and Analysis (ISPA), Dubrovnik, Croatia, IEEE, 2019: 54-59.
- [9] Zou Y, Shi C, Wang L. Real-time cheating detection system based on sight detection[C]. International Conference on Computer, Artificial Intelligence, and Control Engineering, Zhuhai, China, SPIE, 2022, 12288: 235-241.
- [10] 田佳莉. 基于多特征的考场异常行为识别研究[D]. 沈阳: 沈阳工业大学, 2021.
- [11] 付 康. 基于深度学习的考试异常行为识别研究及应用[D]. 荆州: 长江大学, 2020.
- [12] Fang M, Chen Z, Przystupa K, et al. Examination of abnormal behavior detection based on improved YOLOv3[J]. Electronics, 2021, 10(2): 197.
- [13] 洪 钟. 基于视频监控的考场异常行为检测方法研究与实现[D]. 北京: 北京邮电大学, 2019.
- [14] 于明学, 金 鑫, 李晓东, 等. 基于3D卷积双路神经网络的考场行为异常识别[J]. 北京电子科技学院学报, 2018, 26(4): 60-71.
- [15] Alairaji R M, Aljazaery I A, ALRikabi H T H S. Abnormal behavior detection of students in the examination hall from surveillance videos[C]. Proceedings of ICACCP, Singapore, Springer, 2022: 113-125.
- [16] Hu S, Jia X, Fu Y. Research on abnormal behavior detection of online examination based on image information[C]. International Conference on Intelligent Human-Machine Systems and Cybernetics, Hangzhou, China, IEEE, 2018, 2: 88-91.
- [17] Atabay H A, Hassanpour H. Abnormal behavior detection in electronic-exam videos using BeatGAN[C]. Proceedings of ICSPIS, Behshahr, Islamic Republic of Iran, IEEE, 2022: 1-5.
- [18] Zhu X, Lyu S, Wang X, et al. TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios[C]. Proceedings of the ICCV, Virtual: IEEE, 2021: 2778-2788.
- [19] LI C, LI L, JIANG H, et al. YOLOv6: A single-stage object detection framework for industrial applications [EB/OL]. 2022.9. arXiv: 2209.02976.
- [20] Wang C Y, Bochkovskiy A, Liao H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, Canada, IEEE, 2023: 7464-7475.
- [21] Categorization D, Koonce B. Efficient Net [J]. Convolutional Neural Networks with Swift for Tensorflow: Image Recognition and Dataset Categorization, 2021: 109-123.
- [22] Liu Z, Lin Y, Cao Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, Virtual: IEEE, USA, 2021: 10012-10022.
- [23] Iandola F N, Han S, Moskewicz M W, et al. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size [J]. arXiv preprint arXiv: 1602.07360, 2016: 1-17.
- [24] Howard A G, Zhu M, Chen B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications[J]. arXiv preprint arXiv: 1704.04861, 2017.
- [25] Ma N, Zhang X, Zheng H T, et al. Shufflenet v2: Practical guidelines for efficient cnn architecture design[C]. Proceedings of the ECCV, Munich, Germany: IEEE, 2018: 116-131.
- [26] Qin Z, Zhang P, Wu F, et al. Fcanet: Frequency channel attention networks [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, Virtual: IEEE, USA, 2021: 783-792.