



湖州师范学院

2023 届毕业设计(论文)

课 题 名 称: 基于计算机视觉的考教辅助管理监测系统

课 题 名 称 (英文): Examination and teaching assistant management monitoring system based on computer vision

学 生 姓 名: 王煜琦 学 号: 2019082501

专 业 名 称: 计算机科学与技术

指 导 教 师: 唐琦哲 职 称: 讲师

所 在 学 院: 信息工程学院

完 成 日 期: _____ 年 ____ 月 ____ 日

教务处制表

基于计算机视觉的考教辅助管理监测系统

摘要: 随着社会经济的发展,人们生活水平的提高,人们的生活用品逐渐智能化,监控技术也广泛应用于人类社会的各个领域。近年来,因为疫情的影响,大多考试都选择在线上进行,但监考形式依旧是监控记录考试状态,教师时刻关注监控,查看考生行为。监控的存在并未给监考老师减轻负担。如果能将实时监测考生考试情况,在监控的同时分析考生行为是否异常,一方面可以使监考老师不用时刻关注考试情况,提高工作效率;另一方面也可以时刻督促考生,减少考生的侥幸心理。本项目基于计算机视觉对考生考试情况进行监测,利用人脸识别技术进行考生身份确认,并利用人脸定位算法、图像分类和目标检测算法为监考老师打造适合各种考试场景的个性化监测内容,老师可以根据具体情况选择监测类型,包括口罩分类检测、闭眼检测、打哈欠检测、情绪分类检测、正脸检测和手机目标检测。并将监测到的异常情况进行统计存储,利用可视化技术为监考老师呈现考生考试情况,并评判考生违规作弊行为做依据。

关键字: 考试, 人脸识别, 人脸定位, 图像分类, 目标检测

Examination and teaching assistant management monitoring system based on computer vision

Abstract: With the development of social economy and the improvement of people's living standards, people's daily necessities are gradually becoming intelligent, and monitoring technology is also widely used in various fields of human society. In recent years, due to the impact of the epidemic, most exams have been conducted online, but the form of invigilation is still to monitor and record the exam status. Teachers always pay attention to monitoring and check the behavior of candidates. The existence of monitoring has not reduced the burden on invigilators. If you can monitor the examination situation of candidates in real time and analyze whether the behavior of candidates is abnormal while monitoring, on the one hand, it can make the invigilator not always pay attention to the examination situation and improve work efficiency; On the other hand, it can also constantly urge candidates to reduce their fluke mentality. This project monitors the exam situation of candidates based on computer vision, uses face recognition technology to confirm the identity of candidates, and uses face positioning algorithms, image classification, and target detection algorithms to create personalized monitoring content suitable for various exam scenarios for invigilators. Teachers can select monitoring types based on specific circumstances, including mask classification detection, eye closure detection, yawning detection, and mood classification detection Face detection and mobile phone target detection. The monitored anomalies will be statistically stored, and visualization technology will be used to present the invigilator with the examinee's exam situation, as well as to judge the examinee's illegal and cheating behavior as a basis.

Keywords: examination, face recognition, face localization, image classification, target detection

目 录

第一章 绪论.....	8
1.1 选题意义.....	8
1.2 计算机视觉技术的研究现状及发展趋势.....	8
1.2.1 研究现状.....	8
1.2.2 发展趋势.....	9
1.3 本章总结.....	10
第二章 相关技术.....	11
2.1 人脸识别技术.....	11
2.2 图像分类技术.....	11
2.3 目标检测技术.....	12
2.4 本章总结.....	13
第三章 基于计算机视觉的考教辅助监测算法.....	14
3.1 总体思路.....	14
3.2 基于 Retinaface+Facenet 的人脸识别方案.....	15
3.2.1 模型选择.....	15
3.2.2 模型构建.....	15
3.3 基于 Dlib 的人脸定位检测方案.....	16
3.3.1 模型选择.....	16
3.3.2 正脸检测.....	17
3.3.3 闭眼检测.....	17
3.3.4 打哈欠检测.....	18
3.4 基于 Mobilenet V1 的图像分类检测方案.....	19
3.4.1 模型选择.....	19
3.4.2 口罩检测数据集介绍.....	21
3.4.3 情绪检测数据集介绍.....	21
3.4.4 实验环境和参数设置.....	22
3.4.5 实验结果分析.....	22
3.5 利用相关异常数据进行专注度计算.....	22
3.6 基于 YOLOv5 的目标检测方案.....	22
3.6.1 模型选择.....	23
3.6.2 数据集介绍.....	24
3.6.3 实验环境和参数设置.....	24
3.6.4 实验结果分析.....	24

3.7 本章总结.....	24
第四章 系统设计与实现.....	25
4.1 系统目标.....	25
4.2 系统功能模块.....	25
4.3 系统实现流程.....	25
4.4 系统运行界面.....	27
4.4.1 考生子系统——GUI 界面.....	27
4.4.2 监考老师子系统——web 端.....	28
4.5 本章总结.....	31
第五章 总结与展望.....	32
5.1 总结.....	32
5.2 展望.....	32
参考文献.....	33

图目录

图 2-1 人脸识别过程	11
图 3-1 总体思路图	14
图 3-2 Facenet 结构.....	15
图 3-3 人脸识别模型构建	16
图 3-4 人脸关键点定位图	17
图 3-5 正脸检测流程图	17
图 3-6 眼睛定位坐标图	17
图 3-7 闭眼检测流程	18
图 3-8 嘴巴定位坐标图	18
图 3-9 打哈欠检测流程	19
图 3-10 深度可分离卷积块	21
图 3-11 Yolov5 结构.....	23
图 4-1 系统角色分类图	25
图 4-2 系统实现流程图	26
图 4-3 考生登录页面	27
图 4-4 考生监测界面	27
图 4-5 首页	28
图 4-6 考试任务设置页面	28
图 4-7 考试信息详情页面	29
图 4-8 专注度异常页面	29
图 4-9 专注度异常详情页面	30
图 4-10 行为异常页面	30
图 4-11 行为异常详情页面.....	31

表目录

表 3-1 Mobilenet V1 结构.....	19
表 3-2 口罩数据集	21
表 3-3 情绪数据集	22
表 3-4 Mobilenet V1 训练结果	22
表 3-5 手机数据集划分	24
表 3-6 YOLOv5 训练结果	24

第一章 绪论

1.1 选题意义

随着社会的发展，人们生活水平的提高，监控技术广泛应用于人类社会的各个领域，教室已普遍安装监测系统，成为日常学生安全和考试监控的重要工具。近年来，因为疫情的影响，越来越多的考试在线上进行。然而，无论是线上还是线下，考场上的监测系统主要起到录像记录的作用，有监控的同时，仍需人力查看监控，老师进行监考，利用效率较低，不能做到自动监测考生考试情况，给予老师评判依据。现在，计算机视觉日益发展，在很多领域发挥着重要作用，渴望用智能设备来辅助人们，实现更好的监测和信息的统计与管理。

本项目经过调研发现，不当的考试行为主要有替考、查看手机与纸质资料，在线上考试一人一个摄像头的情况下，则需要更多的监考人员来查看监控，所以需要专门的管理监测系统来识别学生身份，监测学生的考试行为，统计分析异常行为，以维护线上考试的考场秩序，提高减轻监考人员的工作负担，提高效率。

综上，本文将利用人脸识别、人脸定位、图像分类和目标检测技术，实现基于计算机视觉的考教辅助管理监测系统，监测的同时统计存储考生异常图像数据，并以可视化的形式展现给监考老师，作为评判考生考试行为的依据。

1.2 计算机视觉技术的研究现状及发展趋势

1.2.1 研究现状

计算机视觉技术是指利用计算机技术和算法来处理、分析和理解图像和视频数据的方法和技术。其主要目的是模拟和实现人类视觉系统的功能。随着人工智能技术和深度学习算法的发展，计算机视觉技术在图像识别、目标检测、虚拟现实、智能安防等领域都得到了广泛应用。下面将从几个方面介绍计算机视觉技术的研究现状。

(1).图像识别技术

图像识别技术是计算机视觉的基础性技术，主要包括特征提取、分类器训练、特征匹配等步骤。传统的基于手工特征的图像识别算法，在训练过程中需要手动耗费大量时间设计和提取特征。随着深度学习算法的发展，卷积神经网络(CNN)成为图像识别领域的主流算法，具有高精度和卓越的性能。

(2).目标检测和跟踪技术

目标检测和跟踪技术是计算机视觉中的热点研究领域。基于深度学习的目标检测算法（如 Faster R-CNN、YOLO、SSD 等）具有较高的检测精度和较快的检测速度^[1]。目标跟踪技术可以通过先前帧中目标物体的位置信息来预测目标的行进方向与速度等参数，并进行目标的追踪和分析。

3.3D 视觉技术

3D 视觉技术是将 2D 图像转换为 3D 模型及其表征的技术。它们是构建虚拟现实、进行三维建模和机器人视觉等面向现实世界的任务中的关键技术。常见的 3D 视觉技术包括三维重建、立体匹配、多视点计算机视觉等。

4. 场景理解技术

场景理解技术是指通过计算机视觉方法对场景数据进行特征提取、识别、分类、推理和动态管理等处理过程，从而对场景进行全面理解的技术。场景理解技术已经成功地应用于智能交通、智能家居、智能工厂等方面。

总之，计算机视觉技术的研究正在不断深入和发展，未来将涌现更多更高效的计算机视觉技术，为人类创造更好的智能化生活和工作环境。

1.2.2 发展趋势

计算机视觉技术是近年来发展迅速的一项技术，其广泛的应用前景和不断创新的潜力吸引了越来越多的研究和投资。随着人工智能技术和深度学习算法的发展，计算机视觉技术将会迎来以下趋势：

1. 大数据驱动

大数据是计算机视觉发展的重要基础，越来越多的图像和视频数据将被应用于模型训练和算法优化。利用大数据可以让计算机视觉系统更加准确、可靠和高效。

2. 深度学习算法

随着深度学习算法的发展，基于深度学习的计算机视觉系统将会越来越普遍。深度学习算法可以自动提取图像特征，并且在图像识别、目标检测和语义分割等领域达到极高的准确率和效率。

3. 立体感知和虚拟现实

随着虚拟现实和增强现实技术的发展，计算机视觉系统将会应用于立体感知和 3D 模拟等方面。在虚拟现实应用中，计算机视觉技术可以实现精确的头部追踪和手势识别等功能，从而带来更加真实的用户体验。

4. 智能交通

计算机视觉技术在智能交通领域的应用也将得到进一步发展，包括信号灯控制、车辆跟踪和自动驾驶等方面。计算机视觉可以对道路环境和车辆行驶状态进行实时感知和分析，从而提高交通效率和安全性。

5. 机器人和自动化生产线

计算机视觉技术可以应用于机器人和自动化生产线上，实现工业生产的智能化管理。机器人视觉系统可以进行目标检测、轨迹规划和控制等，从而实现更加灵活和高效的工业生产。

总之，计算机视觉技术的发展是一个不断创新的过程，新的技术和算法将不断诞生，同时应用场景也将越来越广泛。计算机视觉技术将通过大规模数据、人工智能和自动化技术等手段，逐步实现智能化管理和决策，为人类的生产和生活带来更多的创新和便利。

1.3 本章总结

本章主要介绍了计算机视觉技术的研究现状和发展趋势,通过分析表明利用计算机视觉技术实现考生考试的监测有着重大的意义,因而开发一个基于计算机视觉的考教辅助管理监测系统可以减轻监考老师的工作负担,提高工作效率。

第二章 相关技术

2.1 人脸识别技术

人脸识别技术是指利用计算机视觉技术,通过对人脸图像进行分析、比对、识别等操作,实现对不同人员身份进行判断、识别和辨认的技术。人脸识别技术是基于人脸特征的。对于输入的人脸图像或视频流,它首先确定是否存在人脸。如果存在面部,则它进一步提供每张脸的位置、大小和每个主要面部器官的位置信息。基于这些信息,进一步提取每个人脸中所蕴涵的身份特征,并将其与已知人脸进行比较,从而识别每个人脸的身份。

目前广泛使用的身份鉴定方法包括:指纹,视网膜,虹膜,声音,基因,人脸等。相对于其它身份鉴定技术来说,采用人脸作为身份鉴定的手段具有方便,友好,直接的优点。正因为具有这些优点,人脸识别技术的研究成为近年来计算机视觉、模式识别、人工智能等学科的研究热点^[2]。

一般来说,人脸识别的研究可以分解为如下过程:

- 1.人脸检测;
- 2.人脸面部特征提取;
- 3.人脸识别。

具体识别过程如图 2-1 所示:



图 2-1 人脸识别过程

2.2 图像分类技术

图像分类是根据图像的语义信息对不同类别图像进行区分,是计算机视觉的核心,是物体检测、图像分割、物体跟踪、行为分析、人脸识别等其他高层次视觉任务的基础^[3]。图像分类技术是指利用计算机视觉算法对图像进行分析和处理,以识别和分类不同的图像内容和特征的技术。与传统的图像处理不同,图像分类技术可以自动检测图像中的目标物体、人物、场景等内容,并按照各自的类别进行分类和识别。图像分类在许多领域都有着广泛的应用,例如在医疗领域中,可以用于图像诊断和疾病检测;在自动驾驶领域中,可以用于实现交通信号灯识别和行人检测等。

在实现图像分类的过程中,通常采用卷积神经网络(Convolutional Neural Network, CNN)的人工神经网络结构。CNN 是一种深度学习方法,可以自动地学习图像中的各种特征,从而实现图像分类和识别的目标。CNN 通常由多个卷积层和池化层构成,卷积层用来进行特

征检测和提取，池化层则用于对特征进行抽象和简化。通过不断的卷积和池化，CNN 可以从图像中提取出越来越多的高级特征，最终得到一个可以进行分类的向量表示。

图像分类的目标就是将不同图片划分到不同类别，实现最小的分类误差。在这个过程中存在诸多挑战，但机遇与挑战并存。近年来随着深度学习技术的兴起，图像分类得到飞速发展，并延伸出一系列的研究方向，如：多类别图像分类，细粒度图像分类，多标签图像分类，无/半监督图像分类，零样本图像分类等。

图像分类技术的应用前景广阔，但也存在一定的挑战和困难。例如，在处理大规模图像数据时，需要消耗大量的计算资源和存储空间；同时，在处理复杂图像时，误差率也较高，需要通过不断优化算法和进行数据预处理等手段来提高分类的准确率。

2.3 目标检测技术

目标检测技术是利用计算机视觉技术和机器学习算法，通过图像和视频数据对目标物体进行识别、定位和分类的技术。目标检测技术的应用非常广泛，包括智能交通、安防监控、军事侦察、无人机控制、医疗影像等等领域。本文将重点介绍目标检测技术的分类、算法原理、应用和未来发展方向等。

目标检测技术有多种分类方法，其中最常用的是按照目标检测算法的类型进行分类。根据算法的类型，可以将目标检测技术分为以下几种：

1.基于特征的方法：通过人工设计和选择一系列图像特征来实现目标检测和识别。代表算法有 SIFT、SURF、HOG 等。

2.基于生成模型的方法：使用数据驱动方法建立起具有一定统计特性的统计模型进行检测。代表算法有 Mixtures of Gaussians、HDP、LDA 等。

3.基于判别模型的方法：选取代表性准确率高、分布标签均衡的训练样本训练分类器。代表算法有 SVM、Adaboost、Random Forest 等。

4.基于深度学习的方法：利用深度学习模型，自动提取图像特征，并进行目标检测和分类。代表算法有 Faster R-CNN、YOLO、SSD 等。

目标检测技术的算法原理是实现目标检测和识别的关键。随着深度学习算法的发展和进步，现代目标检测技术主要借助于深度卷积神经网络（Deep Convolutional Neural Network, DCNN）实现对图像的特征提取和分类。DCNN 通常由多个卷积层和池化层组成，用于提取图像中的特征和信息。常见的目标检测系统通常包括以下步骤：

1.检测网络：对图像进行卷积操作，提取特征图。

2.功能层：通过特征图，确定图像中可能存在目标物体的候选区域。

3.分类网络：对每个候选区域，使用深度神经网络分类器进行分类，并输出分类结果和概率值。

4.回归器：对于每个候选区域，通过回归的方法获取最终的目标检测框。

目标检测技术在许多应用领域中得到了广泛应用，包括智能交通、安防监控、军事侦察、

无人机控制、医疗影像等等领域。未来，随着智慧城市建设和人工智能技术的发展，目标检测技术将会越来越重要和应用广泛。主要发展趋势包括以下几个方向：

- 1.实时目标检测：实现对大量数据进行快速处理和实时检测。
- 2.对复杂场景下目标的检测：解决雾天、夜间、多样性等复杂场景下的目标检测问题。
- 3.人性化管理：构建越来越贴近人性化的算法和界面。
- 4.隐私和安全保护：在数据分享和交互过程中，加强对隐私和安全保护。

总之，目标检测技术是计算机视觉领域的重要技术之一，具有广泛应用前景和不断创新应用的潜力。

2.4 本章总结

本章主要介绍了计算机视觉的相关技术。首先介绍了人脸识别技术的主要流程，然后介绍了图像分类技术和目标检测技术的概念和应用，为下文做了理论铺垫。

第三章 基于计算机视觉的考教辅助监测算法

3.1 总体思路

本项目利用摄像头采集图像旨在监测考生的异常行为,为监考老师提供实时的考生考试时的异常情况。总体思路如图 3-1 所示:

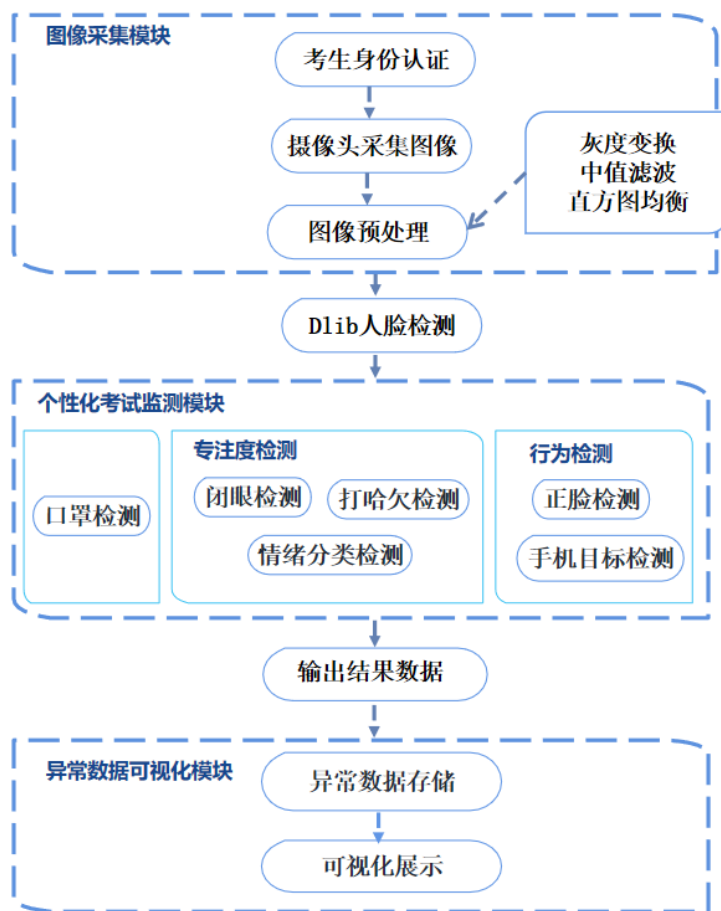


图 3-1 总体思路图

1.考生打开系统后,首先进行人脸识别考生身份认证,认证成功后,考生开始考试,开启摄像头并进行图像采集和处理,通过 OpenCV 调用电脑摄像头采集图像,并进行灰度变换、中值滤波、直方图均衡操作,对图像进行处理。

2.根据监考老师可选择的监测内容,系统主要分为一下三类监测,口罩检测、专注度检测和行为检测,其中专注度检测包括闭眼检测、打哈欠检测和情绪分类检测,行为检测包括正脸检测和手机目标检测。

3.口罩分类检测。根据处理后的图像进行分类检测,判断考生是否佩戴口罩。

4.闭眼检测和打哈欠检测。使用 Dlib 框架定位人脸,检测面部 68 个定位点,定位人眼和嘴巴区域,进行闭合度计算,判断是否长时间闭眼、打哈欠。

5.情绪分类检测。提取面部情绪特征,进行情绪分类检测,判断情绪是否异常。

6.若监考老师同时选择闭眼检测、打哈欠检测和情绪检测，系统则会自动统计闭眼、打哈欠和情绪异常次数。之后，通过三个数值计算专注度结果。

4.正脸检测和手机目标检测。通过 Dlib 人脸定位进行正脸检测和手机目标检测，并判断考生行为是否异常。

5.将异常数据统计存储到数据库，并可视化展示给监考老师。

3.2 基于 Retinaface+Facenet 的人脸识别方案

考生在使用本项目的系统时，首先需要进行考生身份确认，以确保是本人参加考试，身份认证成功后自动开始考试。

3.2.1 模型选择

Retinaface 在实际训练中使用 Mobilenet V1-0.25 和 Resnet 作为主干特征提取网络，使用 Resnet 可以实现更高的精度，使用 Mobilenet V1-0.25 可以实现更高的速度。

Facenet 是谷歌的人脸识别算法，利用相同人脸在不同角度等姿态的照片下有高内聚性，不同人脸有低耦合性，提出使用 CNN +triplet mining 方法，在 LFW 数据集上准确率达到 99.63%。Facenet 结构如图 3-2 所示。



图 3-2 Facenet 结构

Facenet 的训练步骤如下：

- 1.输入一张图片
- 2.通过深度卷积网络 Mobilenet V1 提取特征，根据初步特征获得长度为 128 的特征向量，特征浓缩。
- 3.进行 L2 标准化，使得不同人脸的特征向量可以属于同一数量级，方便比较。
- 4.构建分类器辅助训练，用于辅助 Triplet Loss 的收敛。

3.2.2 模型构建

在该人脸识别方案中，本项目主要使用 Retinaface 进行人脸提取，提取完成后的人脸首先进行人脸对齐，然后借助 Facenet 进行人脸特征的提取与比对，判定两张人脸照片是否属于相同的个体。

如图 3-3 所示，为本项目的基于 Retinaface+Facenet 的人脸识别模型构建流程。人脸识别模型首先使用 Retinaface 网络输出人脸预测框，将人脸图像输入至 Facenet 网络进行人脸特征提取，并与人脸数据库进行比对，比对完成输出识别结果。

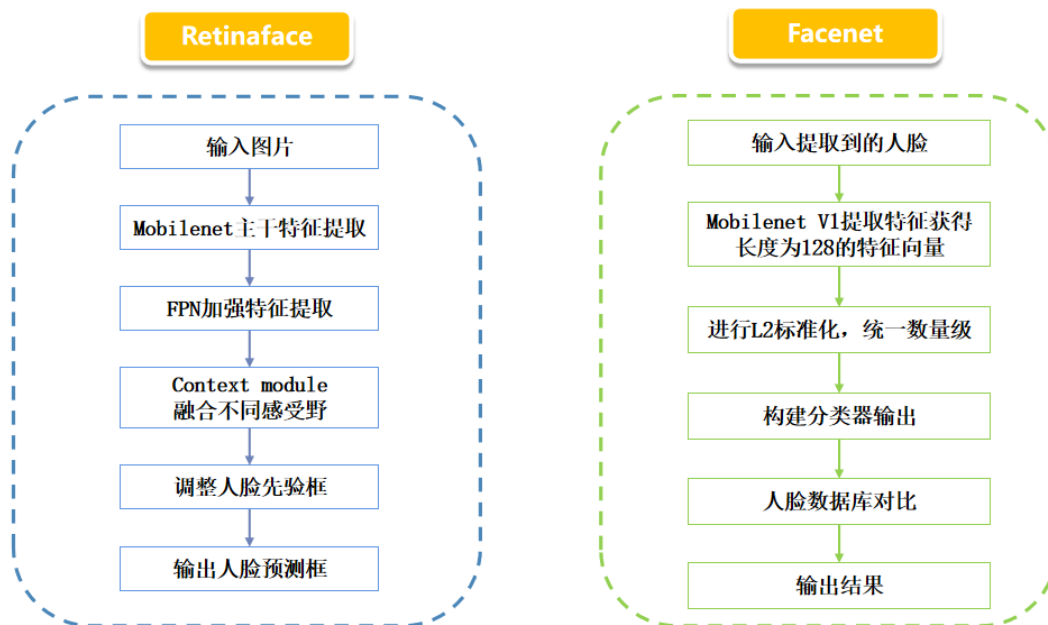


图 3-3 人脸识别模型构建

考生点击身份确认后，系统自动截取一张正脸图片，传入模型进行识别，并返回姓名，若与考生一致，则认证成功；若不一致，则认证失败，重新截取正脸图片进行认证。

3.3 基于 Dlib 的人脸定位检测方案

考生考试时容易出现：长时间闭眼、打哈欠、东张西望等情况。本项目从人脸获取情况和眼睛及嘴巴的坐标位置、开合度、频率等数据入手，并通过这些数据，实时分析考生的考试情况。

3.3.1 模型选择

本项目使用 Dlib 库进行人脸提取。Dlib 提供了人脸关键 68 点训练工具，并标出代表人脸关键特征的 68 个点^[4]。其中，shape_predictor_68_face_landmarks.dat 是一个用于人脸 68 个关键点检测的 dat 模型库^[5]，便于进行人脸对齐、人脸状态估计等应用。人脸关键点如图 3-4 所示：

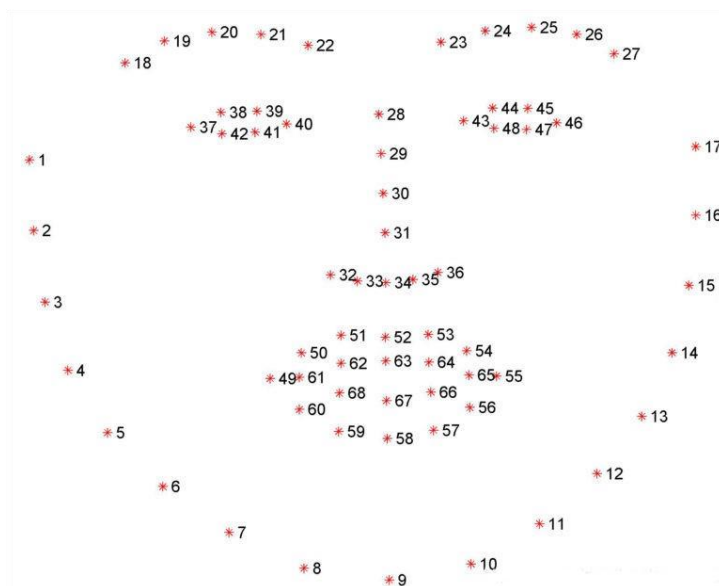


图 3-4 人脸关键点定位图

3.3.2 正脸检测

图像经过预处理后，用 Dlib 人脸检测算法检测人脸区域的界限，若获取不到正脸，则判定考生行为异常，若获取到正脸，则进行之后的检测。具体流程如图 3-5 所示：

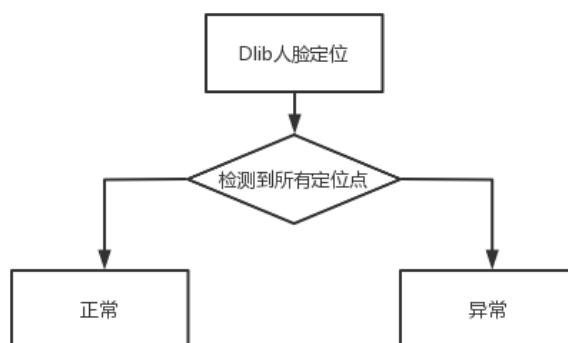


图 3-5 正脸检测流程图

3.3.3 闭眼检测

首先在人脸 68 个关键点中，提取眼部定位点 37~42，定位坐标如图 3-6 所示：

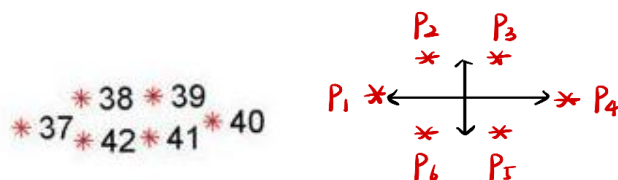


图 3-6 眼睛定位坐标图

将定位点 37~42 记为 P1~P6，如图 3-6，得到眼睛的长宽比 EAR，即闭合度为

$$EAR = \frac{||P2 - P6|| + ||P3 - P5||}{2||P1 - P4||}$$

当人眼睁开时，EAR 在某个值上下波动，当人眼闭合时，EAR 迅速下降，理论上会接近于 0^[5]。本项目设定某一阈值，当 EAR 低于该阈值时，眼睛处于闭合状态。然后通过判断连续几帧或几秒眼睛是否处理闭合状态，从而确定是否长时间闭眼。具体构建流程如图 3-7 所示：



图 3-7 闭眼检测流程

3.3.4 打哈欠检测

首先在人脸 68 个关键点中，提取嘴部定位点 49~68，定位坐标如图 3-8 所示：

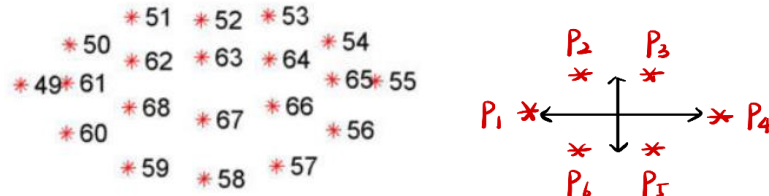


图 3-8 嘴巴定位坐标图

将定位点 49、51、53、55、57、59 记为 P1~P6，得到嘴巴的长宽比 MAR，即张合度为

$$MAR = \frac{||P2 - P6|| + ||P3 - P5||}{2||P1 - P4||}$$

当嘴巴张开时，张合度在某个值上下波动，当嘴巴闭合时，张合度迅速下降，理论上会接近于 0^[5]。我们设定某一阈值，当张合度低于该阈值时，嘴巴处于闭合状态。然后通过判断连续几帧或几秒嘴巴是否处理闭合状态，从而确定是否打哈欠。具体构建流程如图 3-9 所示：



图 3-9 打哈欠检测流程

3.4 基于 Mobilenet V1 的图像分类检测方案

在疫情环境下，为了保证考生的身体健康，考生集中考试需要佩戴口罩。但是考生在考试中的情绪变化往往会影响考生的考试专注度，基于此，本项目针对采集到的图像数据对考生进行口罩、情绪分类检测。

3.4.1 模型选择

本项目的分类模型采用基于 Mobilenet V1 的图像分类模型。

Mobilenet V1 模型是 Google 针对手机等嵌入式设备提出的一种轻量级的深层神经网络，其使用的核心思想便是深度可分离卷积块（depthwise separable convolution）。

Mobilenet V1 相比于 Resnet，具有以下优点：

①参数量较小：Mobilenet V1 模型相对于 Resnet 模型参数量较小，因此可以在计算和存储开销较小的情况下实现高效的模型训练和推理。

②计算速度快：相对于 Resnet 模型，Mobilenet V1 需要更少的计算资源来进行网络的前向计算，这使得 Mobilenet V1 的推理速度更快。

③便于在移动设备上迁移：Mobilenet V1 适用于在低功耗的移动设备上执行计算任务，比如智能手机、平板电脑等，这也是 Mobilenet V1 的主要应用场景。

考虑到本项目的实时性，选择了 Mobilenet V1 模型进行图像分类检测。

Mobilenet V1 网络结构如表 3-1 所示：

表 3-1 Mobilenet V1 结构

Type/Stride	Filter Shape	Input Size
-------------	--------------	------------

Conv / s2	3*3*3*32	224*224*3
Conv dw / s1	3*3*32dw	112*112*32
Conv / s1	1*1*32*64	112*112*32
Conv dw / s2	3*3*64dw	112*112*64
Conv / s1	1*1*64*128	56*56*64
Conv dw / s1	3*3*128dw	56*56*128
Conv / s1	1*1*128*128	56*56*128
Conv dw / s2	3*3*128dw	56*56*128
Conv / s1	1*1*128*256	28*28*128
Conv dw / s1	3*3*256dw	28*28*256
Conv / s1	1*1*256*256	28*28*256
Conv dw / s2	3*3*256dw	28*28*256
Conv / s1	1*1*256*512	14*14*256
5 * Conv dw / s1	3*3*512dw	14*14*512
5 * Conv / s1	1*1*512*512	14*14*512
Conv dw / s2	3*3*512dw	14*14*512
Conv / s1	1*1*512*1024	7*7*512
Conv dw / s2	3*3*1024dw	7*7*1024
Conv / s1	1*1*1024*1024	7*7*1024

Avg Pool / s1	Pool 7*7	7*7*1024
FC / s1	1024*1000	1*1*1024
Softmax / s1	Classifier	1*1*1000

其中，深度可分离卷积块是由两个部分组成，分别是深度可分离卷积和 1*1 普通卷积，深度可分离卷积的卷积核大小一般是 3*3 的，通常用于特征提取，而 1*1 的普通卷积可以完成通道数的调整。深度可分离卷积块的目的是使用更少的参数来代替普通的 3*3 卷积。深度可分离卷积块如图 3-10 所示：

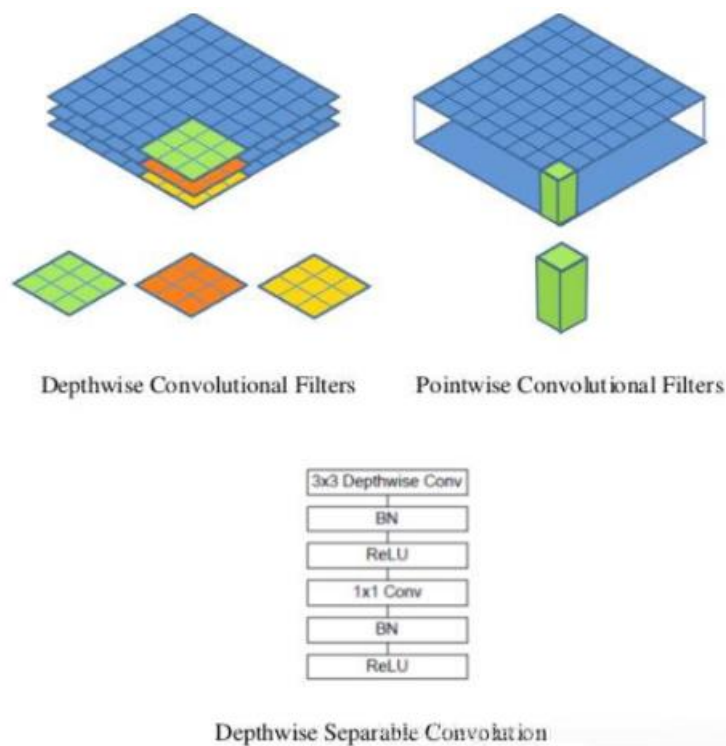


图 3-10 深度可分离卷积块

3.4.2 口罩检测数据集介绍

本项目下载了 4248 张实验数据集，分为佩戴口罩和未佩戴口罩图像数据。具体数量如表 3-2 所示：

表 3-2 口罩数据集

佩戴口罩	未佩戴口罩
3822	426

3.4.3 情绪检测数据集介绍

本项目采用 ExpW_image_align_filtrate 作为实验数据集。这是一个大规模的人脸表情数

数据集，包括大约 9 万张人脸表情图像。数据集主要包括 7 类情绪，具体类型和数量如表 3-3 所示：

表 3-3 情绪数据集

angry	disgust	fear	happy	neutral	sad	surprise
3585	3861	1053	29243	32642	10039	6882

3.4.4 实验环境和参数设置

本项目实验采用的 GPU 为 TITAN Xp(12GB)，Cuda 11.2。使用的深度学习框架为 TensorFlow 2.5.0，编译语言为 Python 3.8。本项目所使用的损失函数为交叉熵损失函数，优化器为 Adam。训练开始时，学习率初始化为 0.001。模型训练分为冻结阶段和解冻阶段，迭代次数各为 50 次，每个批次的数据量 BatchSize 为 32。^[3]

3.4.5 实验结果分析

在设置好上述参数后，经过多次反复训练，最后使得模型有了比较好的效果，最后得到的训练结果如表 3-4 所示：

表 3-4 Mobilenet V1 训练结果

Type	Method	Params	Alpha	Val_Loss
mask	MobilenetV1	4,253,864	1	0.304
emotion	MobilenetV1	4,253,864	1	0.413

3.5 利用相关异常数据进行专注度计算

考生可能出现的表情、动作等行为反应了考生在考试时的专注度。本项目根据考生长时间闭眼、打哈欠和情绪异常，记录异常产生次数，得到专注度值，并设定某一阈值，当专注度达到给定阈值时，判定为专注度异常。

将长时间闭眼、打哈欠、情绪异常的次数分别记为 c_1 , c_2 , c_3 ，权值记为 w_1 , w_2 , w_3 ，则专注度值的计算公式为：

$$Concentration = \sum_{i=1}^3 c_i * w_i$$

3.6 基于 Yolov5 的目标检测方案

考生在考试的过程中可能存在查看手机的情况，这种行为严重违反考试的规章制度，属于作弊情况，需要严肃对待。本项目针对采集到的考试图像数据进行手机的目标检测。

3.6.1 模型选择

本项目的目标检测模型采用基于 YOLOv5 的图像目标检测模型。

YOLOv5 采用了新的网络架构，平均训练和推理速度是 YOLOv4 的 4 倍和 3 倍，相对于 YOLOv3 更快，这种高效非常适合在实时应用中进行目标检测和跟踪。并且，YOLOv5 相对于其它目标检测模型的精度非常高，具有更好的可扩展性，从小型模型到大型模型都有不同的选择。故本项目选择了 YOLOv5 模型进行图像目标检测。

YOLOv5 可以分为三个部分，分别是 CSP Darknet，FPN 以及 YOLO Head。YOLOv5 的网络结构如图 3-11 所示：

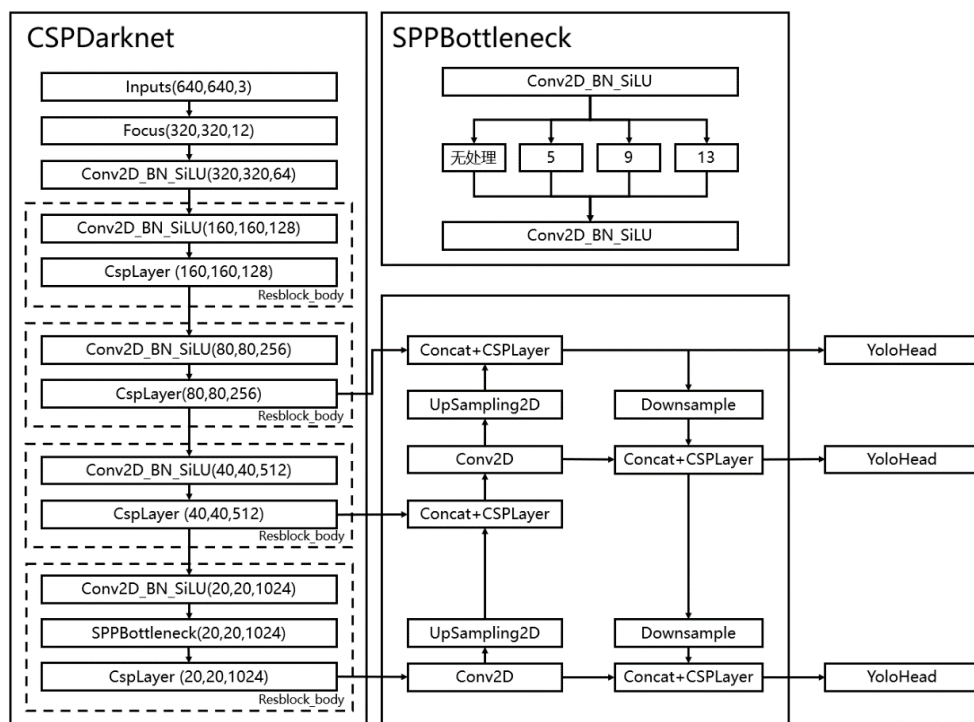


图 3-11 YOLOv5 结构

CSP Darknet 是 YOLOv5 的主干特征提取网络，输入的图片首先会在 CSP Darknet 里面进行特征提取，提取到的特征可以被称作特征层，是输入图片的特征集合。在主干部分，获取了三个特征层进行下一步网络的构建，这三个特征层称为有效特征层^[6]。

FPN 是 YOLOv5 的加强特征提取网络，在主干部分获得的三个有效特征层会在这一部分进行特征融合，特征融合的目的是结合不同尺度的特征信息。在 FPN 部分，已经获得的有效特征层被用于继续提取特征。在 YOLOv5 里依然使用到了 Panet 的结构，不仅会对特征进行上采样实现特征融合，还会对特征再次进行下采样实现特征融合。

YOLO Head 是 YOLOv5 的分类器与回归器，通过 CSP Darknet 和 FPN，已经获得三个加强过的有效特征层。每一个特征层都有宽、高和通道数，此时可以将特征图看作一个又一个特征点的集合，每一个特征点都有通道数个特征。YOLO Head 实际上所做的工作就是对特征点进行判断，判断特征点是否有物体与其对应。与以前版本的 YOLO 一样，YOLOv5 所用的

解耦头是一起的，也就是分类和回归在一个 1×1 卷积里实现。

3.6.2 数据集介绍

本项目下载了 301 张手机图片，通过 LabelImg 将图片中的手机进行框选，构建了 VOC 数据集。为了客观的衡量 YOLOv5 模型的性能，将数据集划分为训练集、测试集和验证集，数量如表 3-5 所示：

表 3-5 手机数据集划分

train	test	val
243	30	27

3.6.3 实验环境和参数设置

本项目实验采用的 GPU 为 TITAN Xp(12GB)，Cuda 11.2。使用的深度学习框架为 TensorFlow 2.5.0，编译语言为 Python 3.8。本项目所使用的损失函数为交叉熵损失函数，优化器为 SGD。训练开始时，学习率初始化为 0.01。模型训练分为冻结阶段和解冻阶段，迭代次数各为 50 次，冻结阶段每个批次的数据量 BatchSize 为 16，解冻阶段为 8。

3.6.4 实验结果分析

在设置好上述参数后，经过多次反复训练，最后使得模型有了比较好的效果，最后得到的训练结果如表 3-6 所示：

表 3-6 YOLOv5 训练结果

Method	Params	F1	mAP	Val_Loss
YOLOv5	7.2M	0.36	73.83%	0.991

3.7 本章总结

本章主要对监测算法的实现思路进行了介绍，具体包括模型选择工作、数据集收集工作、训练前实验环境和参数设置工作和训练后实验结果分析工作。

第四章 系统设计与实现

4.1 系统目标

本项目基于市场上缺少监测与管理相结合系统的情况，旨在设计一套完备的考教辅助管理监测系统，为监考老师减轻负担。本项目拟搭建一个 GUI 用户界面，用于考生身份认证和考试监测，和一个 web 平台，用于可视化展示异常行为及图像，直观的为监考老师展示考生情况，提高监考老师的工作效率。

4.2 系统功能模块

根据系统目标，可以将本项目的角色分类，如图 4-1 所示：

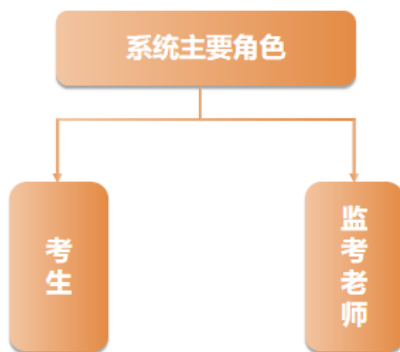


图 4-1 系统角色分类图

考生：是本系统的主要使用人员之一，本系统为监测考生考试佩戴口罩情况、考试专注度和行为而设计。本系统的监测数据来源皆来自于考生的考试数据。考生可以使用 GUI 用户界面进行人脸识别考生身份认证，监测系统的开启与关闭。

监考老师：是本系统的主要使用人员之一。监考老师可以在 web 端选择考试类型、考试时间和监测内容等信息并发布考试任务，在考生进行考试时，监考老师也可以查看考试情况、考生的异常信息，为判定考试是否违规违纪作依据。

4.3 系统实现流程

本项目为了更好为监考老师服务，设计了个性化的监测方式，监考老师可以根据考试类型选择监测内容，系统的具体实现流程如图 4-2 所示：

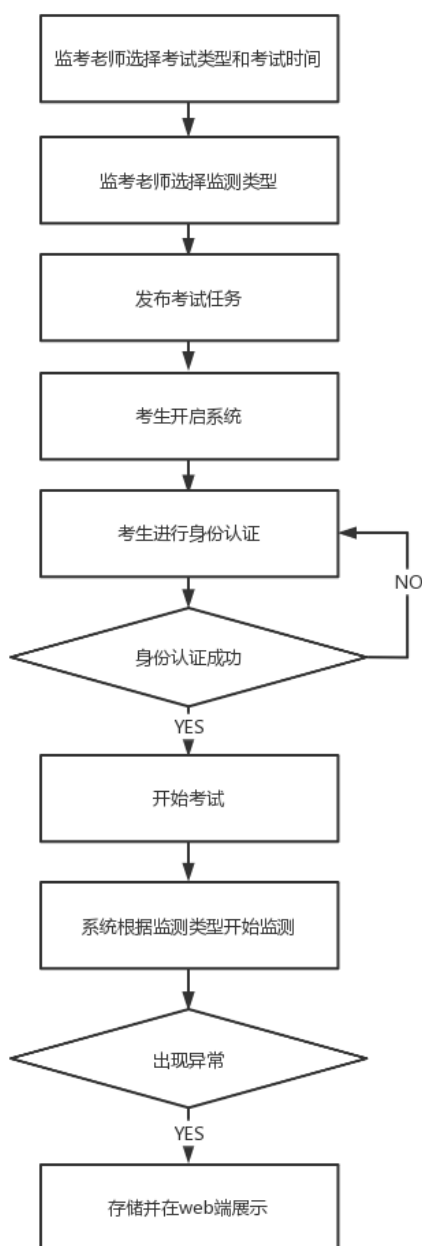


图 4-2 系统实现流程图

首先，监考老师在 web 端选择本场考试类型和考试开始时间、结束时间，并根据考试类型个性化选择监测内容，如：口罩检测、闭眼检测、打哈欠检测、情绪检测、正脸检测、手机检测等。监考老师发布考试任务后，考生可以打开考试监测系统。考生首先进行人脸识别考生身份确认，确认成功后自动开始考试，系统根据监考老师选择的监测内容开始监测，若监测出现异常，则进行统计存储，并在监考老师 web 端进行可视化呈现。

4.4 系统运行界面

4.4.1 考生子系统——GUI 界面

首先考生进入系统后进行身份确认，系统自动截取图片进行人脸识别，如图 4-3 所示：

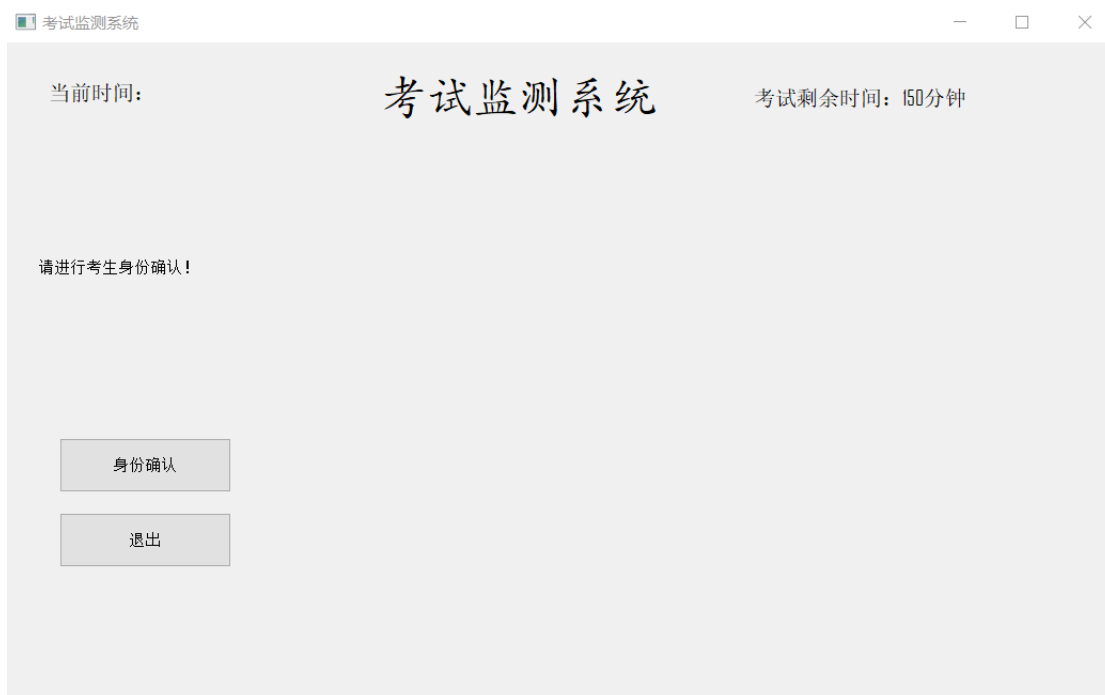


图 4-3 考生登录页面

考生身份确认成功后，在监测界面中自动开始考试，监测系统开始运行，若考试时间结束，监测系统关闭，考生停止考试，考生也可以通过点击结束考试按钮，提前停止考试，如图 4-4 所示：



图 4-4 考生监测界面

4.4.2 监考老师子系统——web 端

监考老师在首页点击加号按钮，可以新建考试监测任务，如图 4-5 所示。监考老师在考试任务设置页面，可以选择本场考试名称和考试开始时间、结束时间，并根据考试类型个性化选择监测内容，包括口罩检测、闭眼检测、打哈欠检测、情绪检测、正脸检测、手机检测，因进行口罩检测会影响面部检测效果，为呈现更多的检测内容，本次测试以选择闭眼检测、打哈欠检测、情绪检测、正脸检测和手机检测为例。监考老师点击发布按钮发布考试任务，如图 4-6 所示：

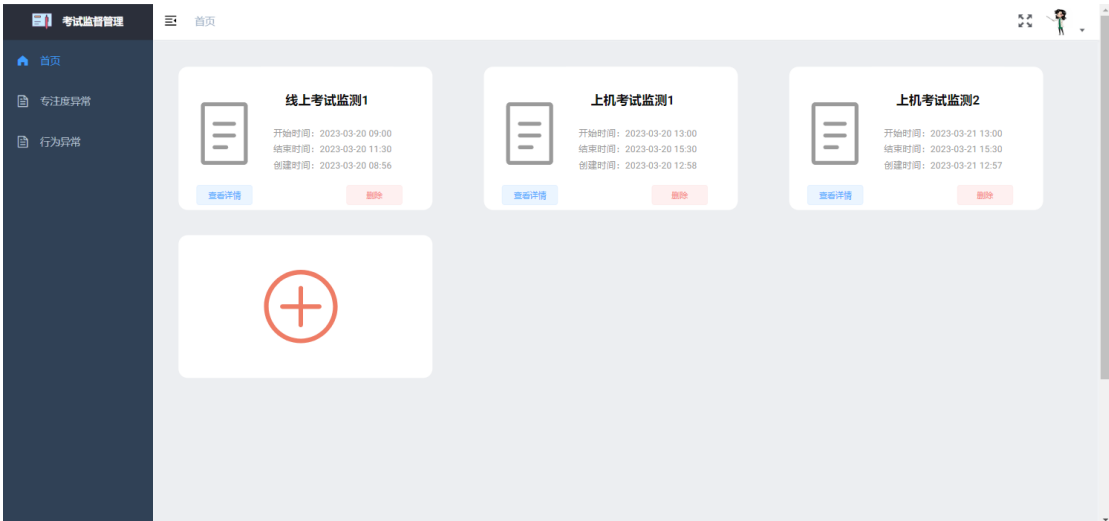


图 4-5 首页

考试任务设置

考试名

请输入考试名

开始时间

2023-03-22 13:00

结束时间

2023-03-22 15:30

持续时间

150

监测内容

☐ 口罩检测

☒ 闭眼检测

☒ 打哈欠检测

☒ 情绪检测

☒ 正脸检测

☒ 手机检测

发布

图 4-6 考试任务设置页面

点击查看详情，监考老师可以看到本场考试的情况，包括当前日期、考生人数、专注度异常次数、行为异常次数等数据，以及异常情况的可视化图表，包括本场考试的总异常情况对比、专注度异常情况对比、行为异常情况对比和根据考试时间的进行考试疲劳情况、情绪异常情况、行为异常情况，如图 4-7 所示：



图 4-7 考试信息详情页面

在专注度异常页面可以看到各个考生的长时间闭眼次数、打哈欠次数、情绪异常次数，以及判定为专注度异常的时间，如图 4-8 所示：

	姓名	闭眼次数	打哈欠次数	情绪异常次数	时间	操作
0	考生1	3	2	4	2023-03-22 14:45:32	查看
1	考生1	2	3	1	2023-03-22 14:44:32	查看

图 4-8 专注度异常页面

点击查看按钮，可以看到该考生的专注度异常图像，包括闭眼、打哈欠和情绪异常，如图 4-9 所示：

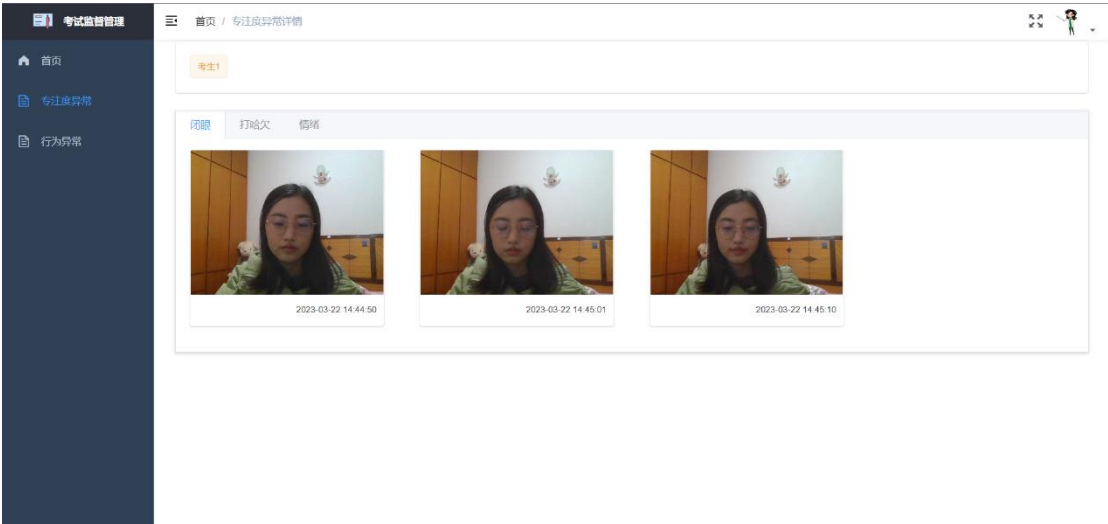


图 4-9 专注度异常详情页面

在行为异常页面可以看到各个考生出现异常的类型，以及异常出现的时间，如图 4-10 所示：

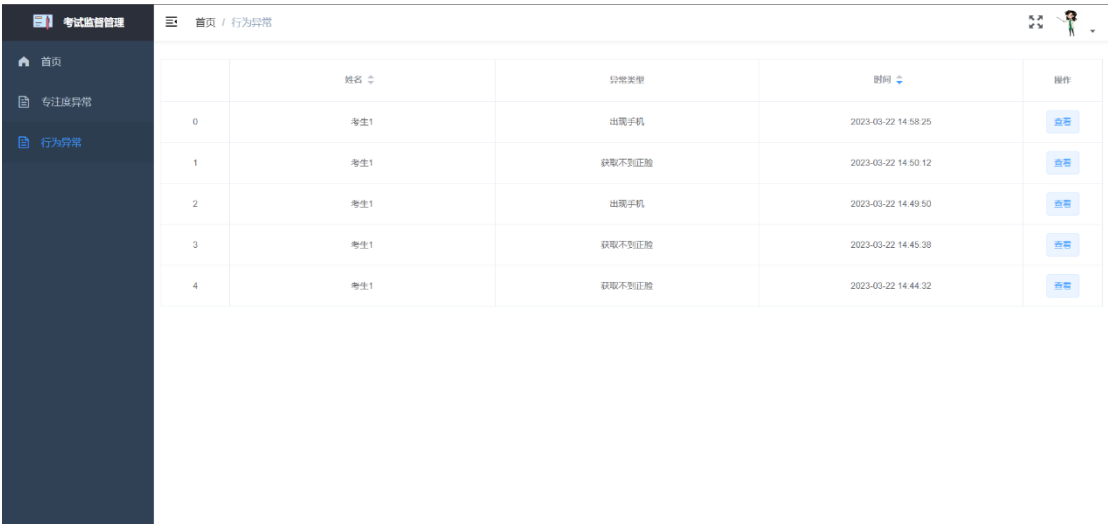


图 4-10 行为异常页面

点击查看按钮，可以看到该考生的行为异常图像，包括获取不到正脸、出现手机的行为，如图 4-11 所示：

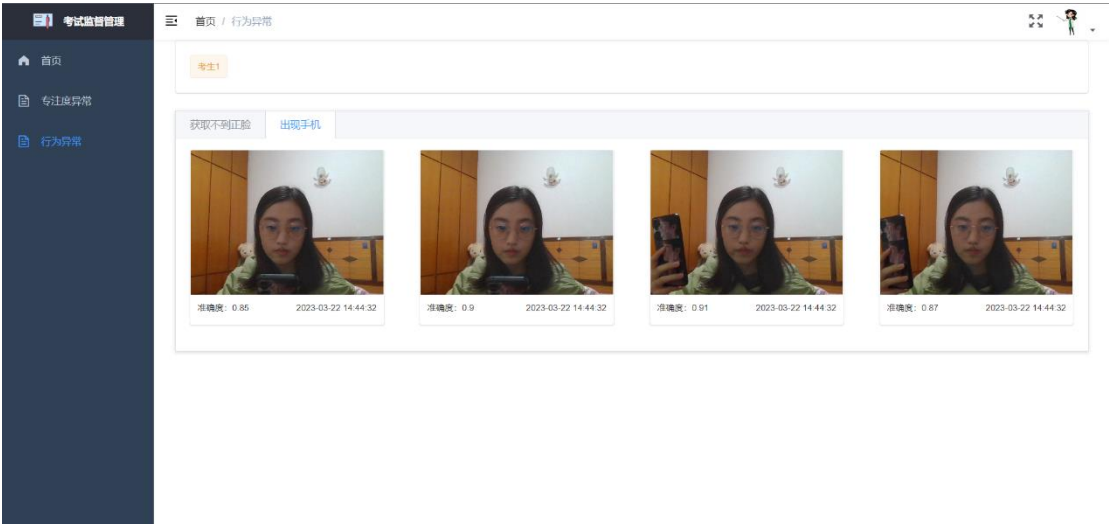


图 4-11 行为异常详情页面

4.5 本章总结

本章通过对现有问题进行分析,提出来本项目的系统目标并进行角色分类和系统流程设计。根据角色和系统流程划分为考生 GUI 界面和监考老师 web 端,最终完成了系统的搭建。

第五章 总结与展望

5.1 总结

随着经济的不断发展,智能化设备的增加,信息革命的前进,人们的生活水平日益提高,逐渐解放双手,利用智能化工具提高工作效率。在监控普及,又有疫情影响下的教学活动,设计一款基于计算机视觉的考教辅助管理监测系统,对考生考试情况进行监测管理十分重要。

为此,本项目查阅大量相关文献,对计算机视觉的相关技术进行学习和思考,从而最终确定了本项目的建设方案,具体步骤如下:

①利用人脸识别技术,进行考生身份确认。

②利用人脸定位、图像分类和目标检测技术,对考生考试情况进行监测,并将异常数据进行统计存储并可视化呈现给监考老师。

③利用 Vue、Flask 和 GUI 技术,搭建考生子系统和监考老师子系统,实现与用户的交互功能,为用户提供了便利。

5.2 展望

本项目设计的以计算机视觉为核心的管理监测系统,虽然能够对考生行为进行监测,但是仍然存在着些许弊端,有待提高:

①因目标检测算法需要自行制作 VOC 数据集,对检测内容进行框选,本项目采集的训练数据并不是很大,影响了算法的检测效果,后续可以采集更多的数据集进行训练。

②本项目的算法准确度还不够高,在后续的开发中,可以不断优化算法,提高算法的准确率,为用户提供更好的体验。

参考文献

- [1] 余佳.基于计算棒推理特性的目标检测技术研究[D].杭州师范大学.
- [2] 肖秀春.人脸检测与面部特征提取技术研究[D].湖南:湖南大学,2004. DOI:10.7666/d.y667246.
- [3] 张文华.场景自适应深度学习跟踪与识别[D].陕西:西安电子科技大学,2021.
- [4] 黄家才,旷文腾,毛宽诚.基于人脸关键点的疲劳驾驶检测研究[J].南京工程学院学报(自然科学版),2017,15(4):8-13. DOI:10.13960/j.issn.1672-2558.2017.04.002.
- [5] 侍新兰,杜友福,王同喜. 基于 OpenCV 的活体检测[J].电脑知识与技术,2020,16(15):211-213.
- [6] 陈志琳.基于面部特征的疲劳驾驶检测系统设计与实现[D].陕西:西安工业大学,2022.
- [7] 徐小康.基于深度学习的表情识别研究与应用[D].上海:东华大学,2022.
- [8] 周飞燕,金林鹏,董军.卷积神经网络研究综述[J].计算机学报,2017(6).
- [9] C.Stauffer and W.E.L.Grimson,“Adaptive backround mixture models for real—time tracking”,Proc.IEEE CVPR 1999,vol.2,pp.246.252(1999).
- [10] Fang H S, Xie S, Tai Y W, et al. Rmpe: Regional multi-person pose estimation [C]. Proceedings of IEEE International Conference on Computer Vision. 2017:2334-2343.
- [11] 陈胜勇,王其超,沃波海,et al. 一种基于滑动窗口局部匹配窗口的动作识别方法[J]. 计算机辅助设计与图形学学报.
- [12] 李珊,邓伟洪.深度人脸表情识别研究进展[J].中国图象图形学报.2020,(11).
- [13] 蔡玉宝,李德峰,连海根,等.基于表情识别的智慧教室授课评估系统[J].数字技术与应用.2020,(10).DOI:10.19695/j.cnki.cn12-1369.2020.10.49.
- [14] 韩丽,李洋,周子佳,等.课堂环境中基于面部表情的教学效果分析[J].现代远程教育研究.2017,(4).DOI:10.3969/j.issn.1009-5195.2017.04.011.
- [15] 张焕,张庆,于纪言.卷积神经网络中激活函数的性质分析与改进[J].计算机仿真.2022,39(4).DOI:10.3969/j.issn.1006-9348.2022.04.064.
- [16] 张珂,冯晓晗,郭玉荣,等.图像分类的深度卷积神经网络模型综述[J].中国图象图形学报.2021,(10).
- [17] Gao, Shang-Hua,Cheng, Ming-Ming,Zhao, Kai,等.Res2Net: A New Multi-Scale Backbone Architecture[J].IEEE Transactions on Pattern Analysis and Machine Intelligence.2021,43(2).652-662.DOI:10.1109/TPAMI.2019.2938758.
- [18] LeCun, Yann,Bengio, Yoshua,Hinton, Geoffrey.Deep learning[J].Nature.2015,521(May 28 TN.7553).
- [19] 郭波.基于 YOLOv5 网络的织物疵点检测算法研究[D].浙江:浙江理工大学,2021.
- [20] 武琦.基于视频压缩域的目标实时识别及其应用研究[D].河南:郑州大学,2022.