

文章编号: 2095-2163(2023)03-0169-07

中图分类号: TP399

文献标志码: A

# 基于YOLOv5的自习室人数监控与查询系统

郭遇强, 赵豪翔, 陈思源, 张 斌

(郑州大学 地球科学与技术学院, 郑州 450001)

**摘要:** 当前, 校园中自习室的空间资源利用存在不合理的问题, 为了减少学生寻找自习室的时间, 提高时间利用效率, 通过监控视频数据实时获取校园内各个教室的人数, 基于YOLOv5目标检测技术设计并开发了自习室人数监控与查询系统, 该系统使用OpenCV-Python模块的VideoCapture子模块, 调用YOLOv5算法处理视频帧, 将得到的人数数据储存在后端数据库中, 为后续功能应用提供数据支撑。

**关键词:** YOLOv5; 目标检测; OpenCV-Python; 人数监控

## Study room number monitoring and inquiry system based on YOLOv5

GUO Yuqiang, ZHAO Haoxiang, CHEN Siyuan, ZHANG Bin

(School of Geo-science and Technology, Zhengzhou University, Zhengzhou 450001, China)

**[Abstract]** At present, there are unreasonable problems in the utilization of space resources of study rooms on the campus. In order to reduce the time for students to search for study rooms and improve the efficiency of time utilization, a monitoring and query system for the number of study rooms is designed and developed based on YOLOv5 target detection technology. The system uses the VideoCapture submodule of OpenCV-Python module to call YOLOv5 algorithm to process video frames, and stores the number data obtained in the back-end database to provide data support for subsequent functional applications.

**[Key words]** YOLOv5; target detection; OpenCV-Python; population monitoring

## 0 引言

由于时间和空间等多种因素影响, 校园许多教室的空间没有得到合理的利用。例如, 学校一些教室人满为患, 而一些教室却只有寥寥数人; 有些时间段教室人数比较少, 而有些时间段却比较拥挤, 以至于没有位置提供给学生自习。但在很多情况下, 学生在前往自习室之前却无法得知自习室的使用情况, 不利于选择合适的自习室。这就对学生学习造成了些许不便, 同时降低了校园内的教室空间资源利用率。考虑到教室内都配有摄像头, 能够提供实时的摄像机图像, 如此一来则为改善该问题提供了思路和必要的数据基础。

研究可知, 如果需要准确获取自习室的人数, 通过摄像头视频数据进行人的识别是很重要的一项工

作, 也是后续各项研究的基础, 其运行速度、准确率等指标都会影响分析应用的效果<sup>[1]</sup>。2014年, R-CNN将PASCAL VOC 2012<sup>[2]</sup>数据集上的验证指标提升了30%, 成功将深度学习应用到了目标检测领域<sup>[3]</sup>, 在该领域内识别图片中的人逐渐成为热门的研究内容。最近几年, 研究者们提出了许多不同的、适用于各种场景的模型, 在识别人的准确度和速度中取得了可观的进展。

## 1 设计思路

当前, 基于深度学习目标检测算法模型主要分为基于候选框生成与分类的two-stage<sup>[4-5]</sup>方法和基于回归的one-stage方法<sup>[6]</sup>。其中, two-stage方法是在产生一系列候选框的基础上进行回归和分类, 代表性成果是最早将深度学习应用在目标检测领域

**基金项目:** 郑州大学全国大学生创新创业训练计划资助项目(2021cxcy119); 国家自然科学基金项目(41571394); 郑州大学高层次人才科研启动项目(237-32310319)。

**作者简介:** 郭遇强(2001-), 男, 本科生, 主要研究方向: 地理信息科学; 赵豪翔(2001-), 男, 本科生, 主要研究方向: 地理信息科学; 陈思源(2001-), 男, 本科生, 主要研究方向: 地理信息科学; 张 斌(1979-), 男, 讲师, 主要研究方向: 地理信息工程方面的教学和科研工作。

**通讯作者:** 张 斌 Email: 24852161@qq.com

**收稿日期:** 2022-05-03

的 R-CNN<sup>[7]</sup> 算法,使用了卷积神经网络对感兴趣的区域进行特征提取,但是其选择性过程无法通过 GPU 等并行计算设备进行加速,会分别提取每个区域的特征,造成重复计算,训练耗时费力。随后,研究人员提出了 SPP-Net<sup>[8]</sup>、Fast R-CNN<sup>[9]</sup> 等算法,加强了部分之间的联系,提升了检测的速度。而 one-stage 方法略过了候选框这一步骤,采用卷积神经网络对图像数据进行卷积操作,直接进行特征提

取并进行分类,虽然精确度比 two-stage 方法略低,但大大提升了目标识别速度,代表性成果为 YOLO 算法<sup>[10-11]</sup>,这是将对象检测视为一个回归问题、而不是一个分类问题,使用单个神经网络来完成检测对象的所有阶段,大大提高了检测速度,除此以外,还可以被训练来检测不同的对象。目前,主流算法的优缺点及使用场景见表 1。

表 1 目标检测算法优缺点及适用场景<sup>[12]</sup>

Tab. 1 Advantages, disadvantages and applicable scenarios of target detection algorithm<sup>[12]</sup>

算法	模型	优点	缺点	适用场景
R-CNN	two-stage	CNN 与候选框方法结合	耗时耗存储空间	适用于通用目标检测,效果不佳
SPP-Net	two-stage	图像卷积运算,实现多尺度卷积运算	耗时间	适用于通用目标检测场景,效果不佳
Fast R-CNN	two-stage	节省时间和特征存储空间	框架复杂,量化粗略导致定位不精确	适用于通用目标检测场景,效果不佳
YOLOv5	one-stage	网络较为简单,检测速度快	定位较差,小目标多目标检测较差	适用于检测速度要求较高的场景

通过分析各种算法模型特点可知, YOLOv5 算法网络结构简单、检测速度快,能较好地满足室内区域对人体目标的识别。本文设计采用 YOLOv5 的预训练模型作为目标检测算法模型,使用 OpenCV-Python 模块的 VideoCapture 子模块获取待处理图像。系统后端采用图像处理方法,应用 YOLOv5 网络模型处理教室视频,计算其中的人数,并允许学生对教室和人数进行查询,从而方便学生根据自身需要,灵活地选择自己想要前往的教室。系统前端利用百度地图 JavaScript API<sup>[13]</sup> 构建交互地图,使用

HTML、CSS、JavaScript<sup>[14]</sup> 搭建页面。

### 1.1 算法设计

自从 2016 年 Redmon 等学者<sup>[15]</sup> 提出目标识别领域的 YOLOv1 模型开始, YOLO 系列便不断推陈出新,直到 2020 年推出 YOLOv5 模型,模型性能最佳,体积更小,版本已相当成熟,更加适合实际工程的应用,而且可以通过 PyTorch 轻松完成部署与实现。YOLOv5 的网络结构主要由输入端、Backbone、Neck 以及 Prediction 四部分组成,如图 1 所示。

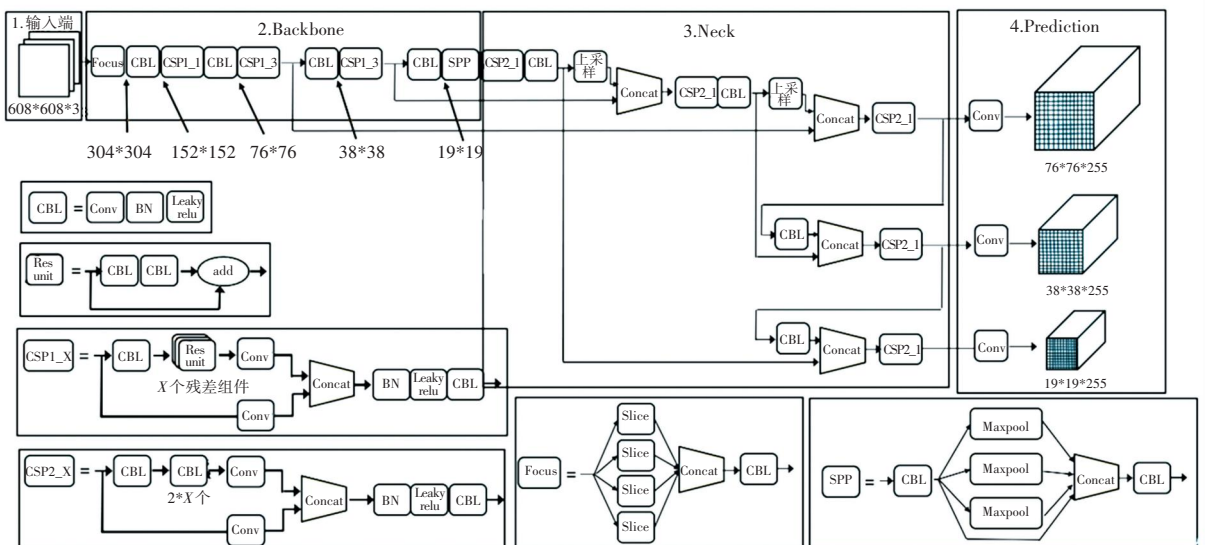


图 1 YOLOv5 网络结构<sup>[16]</sup>

Fig. 1 Network structure of YOLOv5<sup>[16]</sup>

由图 1 可知,输入端表示输入的图片或视频帧,为了方便后续的图像处理,该阶段会对输入的图片进行预处理,将图像缩放到网络的输入大小后再做归一化处理。Backbone 部分主要用于从图片中提取信息和特征,以供后续的网络使用。YOLOv5 使用 CSPDarknet 作为 Backbone, CSPNet 解决了其他大型卷积神经网络框架 Backbone 中网络优化的梯度信息重复问题,将梯度的变化从头到尾地集成到特征图中,因此减少了模型的参数量和 *FLOPS* 数值,既保证了推理速度和准确率,又减小了模型尺寸。Neck 网络部分通常位于 Backbone 网络部分和 Prediction 部分的中间位置,可用来进一步提升特征的多样性。Prediction 部分用来完成目标检测结果的输出。针对不同的检测算法,输出端的分支个数不尽相同,通常包含一个分类分支和一个回归分支。

相较于前期成果,YOLOv5 在如下方面进行了提高和改善:

(1)输入端。通过使用 Mosaic 数据增强的方式来丰富数据集和提高训练速度,该方式是在 CutMix 数据增强方法的基础上改进而来,相较于 CutMix 数据增强方法的 2 张图片,Mosaic 数据增强方法采用了 4 张图片,按照随机缩放、随机裁剪和随机排布的

方式进行拼接而成,适用面更广,同时通过自适应锚框计算和自适应图片缩放来提高算法的适用度和推算速度。

(2)Backbone 部分。采用 Focus 结构和 CSP 结构增强了 CNN 的学习能力,在确保准确率的情况下减少了计算量,并降低了内存成本。

(3) Neck 部分。YOLOv5 的 Neck 和 YOLOv4 中一样,都采用 FPN + PAN 的结构。YOLOv4 的 Neck 结构中,采用的都是普通的卷积操作。而 YOLOv5 的 Neck 结构中,采用借鉴 CSPNet 设计的 CSP2 结构,加强网络特征融合的能力。

(4) Prediction 部分。采用 *GIoU\_Loss*<sup>[17]</sup> 作为 Bounding box 的损失函数。该方法不仅简单,还具有尺度不变性。

目前,YOLOv5 的目标检测网络主要有 YOLOv5s、YOLOv5m、YOLOv5l、YOLOv5x 等 4 个主要模型,各模型计算速度和检测精度如图 2 所示。本文采用 YOLOv5s 网络模型,该模型是 YOLOv5 中各类模型的基础,具有深度最小、计算速度最快的优点,能很好地满足系统对运算速度的要求,其它的网络模型都是在该模型的基础上对网络的加深与加宽。

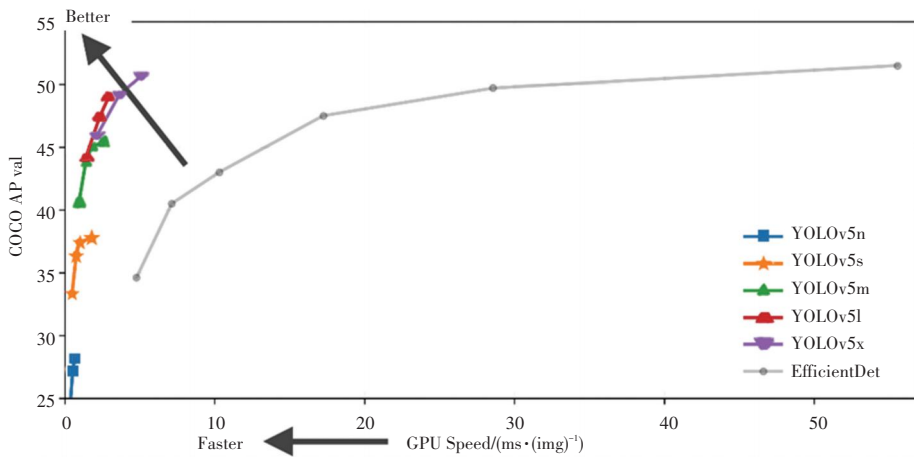


图 2 YOLOv5 的主要模型及其性能(基于 COCO 数据集)<sup>[18]</sup>

Fig. 2 Main models and performance of YOLOv5 (based on COCO data set)<sup>[18]</sup>

### 1.2 系统整体框架

按照“高内聚、低耦合”的设计思想,将系统分为表现层、业务层、数据层三层架构,分工明确,每层内功能集中,层与层之间相互独立,便于维护。具体架构如图 3 所示。这里对此将给出如下阐释论述如下。

(1)数据层,即系统所需的数据库。包含教室位置信息、教室人数、教室可承载的最大人数、教学

楼的课表。教室位置与教室最大承载人数需开发人员提前录入。

(2)业务层。通过连接数据库,获取教室的信息与课表。对课表进行查询与修改操作,对教室的信息进行查询与更新。

(3)表现层。为展示给用户的界面,用户可以通过查询功能获知目的地教学楼当前时段有哪些较为空闲的教室,并获知教室当前人数与最大承载数,

从而决定前往哪间教室。管理员可通过地图交互,对特定教学楼内的摄像头进行浏览,并可在 Web 页

面根据不同专业的课表信息及时修改教学楼课表,保证信息的实时性、准确性。

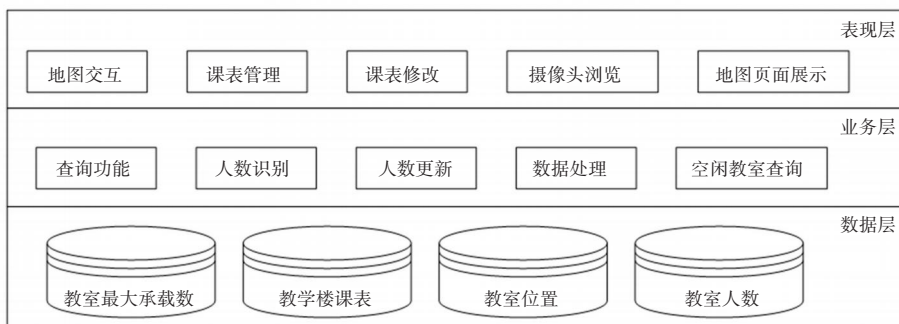


图 3 系统架构图

Fig. 3 System architecture

### 1.3 数据库结构设计

数据库的功能是对空闲教室查询提供数据支撑,实现对数据的存储、维护和查询等功能。根据用户查询的需要构建系统数据库,设置教学楼课表与教室信息表。

教学楼课表采用教学楼名称作为表名,也就是一栋楼即为数据库中的一张表,以星期作为该表一条记录的关键字,按照每天 10 节小课分为 10 个时间段。该表需要集合不同专业的上课需求,统计出每天哪个时段的哪间教室没有上课需求,按照“教室楼/教室编号”的格式提前录入,即表示当天该时段该教学楼内的所有空闲教室。若某一时间段有上课需求,则表中以空值“NULL”填入。无课表的结构见表 2。

表 2 教学楼课表数据结构图

Tab. 2 Class schedule data structure of teaching buildings

名称	类型	约束条件	说明
weekend	int	无重复	星期标识,主键
one	nchar(20)	初始值为 NULL	第一节无课的教室
two	nchar(20)	初始值为 NULL	第二节无课的教室
three	nchar(20)	初始值为 NULL	第三节无课的教室
four	nchar(20)	初始值为 NULL	第四节无课的教室
five	nchar(20)	初始值为 NULL	第五节无课的教室
six	nchar(20)	初始值为 NULL	第六节无课的教室
seven	nchar(20)	初始值为 NULL	第七节无课的教室
eight	nchar(20)	初始值为 NULL	第八节无课的教室
nine	nchar(20)	初始值为 NULL	第九节无课的教室
ten	nchar(20)	初始值为 NULL	第十节无课的教室

教室位置、教室人数、教室最大承载人数集中显示在同一张表中,其中教室位置按照“教学楼/教室

编号”的格式进行录入,作为一条记录的关键字,为了避免调用数据时由于不同编码导致的中文乱码问题,因此在录入时统一以英文方式录入。教室信息表结构见表 3。

表 3 教室信息数据结构图

Tab. 3 Classroom information data structure

名称	类型	约束条件	说明
place	nchar(20)	无重复	教室位置标识,主键
number	int	初始值为 0	教室当前人数
max	int	初始值为 0	教室所能承载的最大人数

### 1.4 教室人数信息的获取

OpenCV-Python 模块的 VideoCapture 子模块对 MP4 格式有着良好的支持,可以直接获取指定帧数并保存。利用该模块,将视频按帧循环读取,将当前视频帧以 .jpeg 格式进行存储,对存储路径进行设置,使路径中包含图片所对应的教室位置。具体流程如图 4 所示。

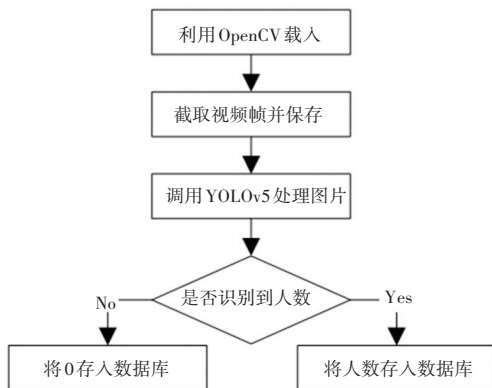


图 4 获取人数流程图

Fig. 4 Flow chart of obtaining the number of people

调用 YOLOv5 对保存的图片进行处理,即可得

到截取图片时刻的人数信息。通过设置后缀为 .pt 的模型包文件、识别类型等参数,为了能在框架运行时同步调用控制台,利用 Python 的 OS 模块的 popen

子模块,该模块允许开发者在 Python 内直接运行控制台命令。根据保存图片路径,构建命令行提示符(cmd)命令,处理结果如图 5 所示。

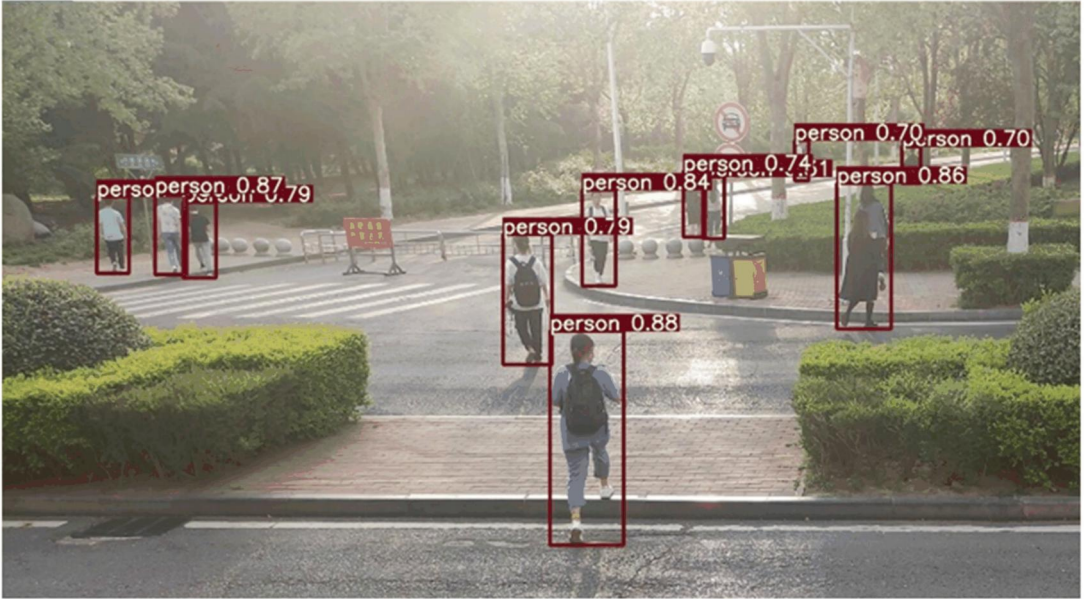


图 5 测试图片处理结果

Fig. 5 Test image processing results

图 5 中,将所有符合条件的人全部框出,标识框上方为置信度,该数字代表标识框内区域中存在人的概率,概率越大,准确性越高。遍历所有标识框,设置参数,过滤置信度较低的标识框,统计符合目标检测结果的数目,即为当前画面中的人数。

获取人数信息后,根据图片路径中所包含的位置信息,构建 SQL 语句<sup>[19]</sup>更新数据库中的人数信息。当没有人在图片中时,该模型会返回空值,此时对空值进行替换,将整型 0 代替空值存入数据库中。

## 2 自习室人数监控与查询试验

### 2.1 数据集介绍

本次实验的运行环境: CPU 为 Intel core i5 6300HQ,内存为 8 G,操作系统为 Windows 10,开发语言为 Python,Pytorch 的 CPU 版框架。

YOLOv5 自身模型文件的构建基于 COCO 数据集<sup>[20]</sup>,COCO 数据集是微软公司于 2014 年出资标注的 Microsoft COCO 数据集,该数据集是一个大型的物体检测数据集。本文通过手机在自习室录制了学生自习时的视频,按帧截取视频,获得总计 300 张教室图片作为测试数据集,每张图像的分辨率大小为 1920×1080。

### 2.2 测试结果与分析

#### 2.2.1 图片处理结果分析

单张图片处理结果如图 6 所示。从图 6 中可以看出,大部分人员可以准确识别并标出,但部分人员无法识别出来。产生此误差的原因是 YOLOv5 自身的识别误差。

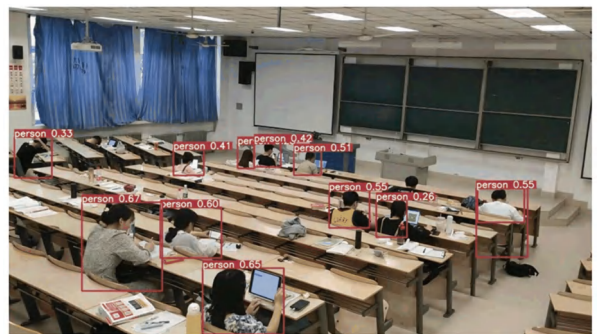


图 6 单张图片处理结果

Fig. 6 Single picture processing result

多次重复实验,得到每一次处理的准确率,制作图表如图 7 所示。从图 7 中可以看出,每次处理前后结果并不相同,但总体准确率稳定在 80% ~ 100%之间,可以满足日常生活中自习室检测人数的需求。

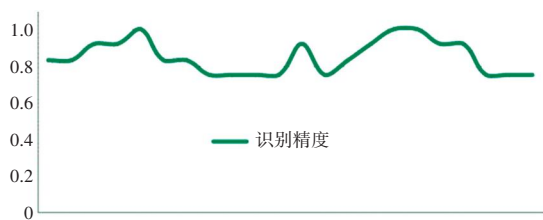


图7 图片处理准确率

Fig. 7 Processing accuracy of the image

每次处理图片所需时间如图8所示,由于Python为动态类型语言,在相同硬件条件下,每次处理同一段代码所需时间不一致。每次处理时间均在4s内,表示该框架处理图片所需时间短,处理速度快,可以满足实际应用的时间效率需求。

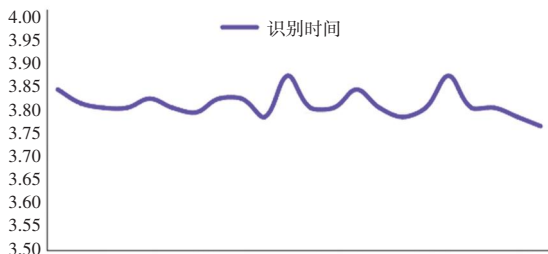


图8 图片处理时间

Fig. 8 Processing time of the picture

### 2.2.2 管理功能测试

管理员在浏览界面能够在学校地图上选择任意教学楼,且可以查看当前教室人数概况以及该摄像头下的监控画面中的人数,监控画面如图9所示。测试得到的浏览界面中,使用进度条概括显示当前人数占教室可以承载的总人数的比例,图9显示本系统相应功能实现完整,操作能够顺利准确运行。



图9 摄像头浏览界面

Fig. 9 Camera browsing interface

### 2.2.3 查询功能测试

用户选择目的地与自习时间,系统将根据用户的查询要求,结合当前各个自习室的人数情况和各个教室的排课情况,得出符合用户需求的自习室列表,以供用户选择。其具体查询界面如图10所示。图10表明,本系统能够正确完成相应的查询操作,

并且,将教室当前人数少于最大人数的一半的教室置为绿色,大于最大人数的一半置为红色,帮助用户可以更快地确定自己的目的地。



图10 查询结果

Fig. 10 Query results

## 3 结束语

本文从校园自习室空间利用的角度出发,通过视频摄像头获取影像数据,引入深度学习的计算方法,介绍了基于YOLOv5的自习室人数监控与查询系统的设计思路、具体实现及应用试验等内容。系统使用OpenCV-Python模块的VideoCapture子模块,调用YOLOv5算法处理视频帧,上传至数据库,同时通过调用数据库,将结果以网页端的形式展现出来,整个流程从视频帧处理到数据库数据更新,最后到网页端刷新显示仅需数分钟,对于短时间内人流量变动较小的公共空间,诸如自习室、图书馆等,有着较好的应用前景,能够加强对自习室的管理,给学生寻找自习室提供帮助,提高自习室的使用效率。后续将进一步对系统加以改进完善,在获取实时监控影像基础上优化数据流程及代码,缩短整个过程所需要的时间,以利于应用到更多场景中。

## 参考文献

[1] LIU Wei, ANGUELOV D, ERHAN D, et al. SSD: Single shot multitox detector [C]//Proceedings of European Conference on Computer Vision. Cham; Springer, 2016; 21-37.

[2] EVERINGHAM M, van GOOL L, WWILIAMS C K L, et al. The Pascal visual object classes (VOC) challenge [J]. International Journal of Computer Vision, 2010, 88(2): 303-338.

[3] 李柯泉, 陈燕, 刘佳晨, 等. 基于深度学习的目标检测算法综述 [J]. 计算机工程, 2022, 48(07): 1-12.

[4] GIRSHICK R, DONHAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]// Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus; IEEE, 2014; 580-587.