

基于机器视觉的工业巡检过程监控分析系统^①



贾金明, 宋焕生, 梁浩翔, 云旭, 戴喆

(长安大学 信息工程学院, 西安 710064)

通讯作者: 宋焕生, E-mail: 772972603@qq.com

摘要: 工业巡检中人员的行为与生产安全息息相关, 有关巡检监测方法的设计成为了研究热点. 针对目前巡检监控分析依赖于人工判断且精度低的问题, 本文提出了一种基于机器视觉的工业现场巡检过程监控分析系统. 首先利用 YOLOv3 网络对将视频流中的人员进行检测, 根据检测结果, 使用人员行为分析方法, 剔除场景内干扰并获取巡检人员真实的行为信息, 最后根据人员行为信息对巡检过程进行评估, 将评估结果存储至数据库同时发布至网页. 本文使用多个监控视角的视频进行实验, 实验结果表明, 本文所提系统在复杂环境下, 能够准确检测巡检人员并分析其行为, 同时满足实时处理的需求. 本文可以为工业巡检的智能化监测提供参考.

关键词: 巡检; 深度学习; 图像处理; 行为分析; 数据库存储

引用格式: 贾金明, 宋焕生, 梁浩翔, 云旭, 戴喆. 基于机器视觉的工业巡检过程监控分析系统. 计算机系统应用, 2021, 30(2): 70-76. <http://www.c-s-a.org.cn/1003-3254/7731.html>

Monitoring and Analysis System of Industrial Inspection Based on Machine Vision

JIA Jin-Ming, SONG Huan-Sheng, LIANG Hao-Xiang, YUN Xu, DAI Zhe

(School of Information Engineering, Chang'an University, Xi'an 710064, China)

Abstract: People's behavior in industrial inspections is closely bound up with safe production, and the design of inspection monitoring methods has become a hot research area. Aiming at the problem that current monitoring and analysis of inspection depend on manual judgment with low accuracy, this study proposes a monitoring and analysis system for industrial inspection based on machine vision. Firstly, the people in the video stream are detected by the YOLOv3 network. According to the detection results, in-scene interferences are removed by behavior analysis to obtain real behavior of inspectors. Finally, the inspection process is evaluated based on the behavior, and then results are stored in the database and posted to the web page. Videos with multiple monitoring perspectives are used for experiments. Results demonstrate that the system proposed in this study can accurately detect inspectors and analyze their behavior in complex environments, while achieving real-time processing. This result can serve as a reference for the intelligent monitoring of industrial inspection.

Key words: inspection; deep learning; image processing; behavior analysis; database storage

目前工业生产设备缺乏巡检人员定期维护、排查异常现象和隐患, 导致设备频繁损坏, 造成巨大的损失.

因此完整的巡检过程监控分析在巡检人员对设备维护管理中起着重要的作用. 工业现场巡检过程监控分析

① 基金项目: 国家自然科学基金 (61572083); 教育部联合基金 (6141A02022610); 陕西省重点研发计划重点项目 (2018ZDXM-GY-047); 中央高校团队培育项目 (300102248402)

Foundation item: National Natural Science Foundation of China (61572083); Joint Fund of Ministry of Education (6141A02022610); Major Project of Key Research and Development Program of Shaanxi Province (2018ZDXM-GY-047); Team Incubation Project of the Central Universities of China (300102248402)

收稿时间: 2020-05-20; 修改时间: 2020-06-16; 采用时间: 2020-06-28; csa 在线出版时间: 2021-01-27

是未来人工智能的一个重要研究领域,也是计算机视觉中的一个热门研究课题.工业现场巡检过程监控分析是对工业生产场景巡检过程视频中巡检人员的行为进行监控与分析,判断其行为是否符合规范与标准.现有的行为监控分析方法主要分为基于视频和图像两种方式.其中基于视频的行为监控分析方法使用视频序列之间的时空、姿态等信息做处理.根据实现方式不同,主要分为基于马尔可夫链状态转移图^[1],基于模板匹配^[2]与基于深度学习的方法.而基于图像的行为监控分析方法则分别判断每一帧图像中人员的行为状态信息,此方法并未结合视频的时空信息进行综合分析.本文根据先前的研究,设计了基于计算机视觉的工业现场巡检过程监控分析系统.

巡检过程监控分析系统包含对场景下人员的检测,随着深度学习技术^[3]的飞速发展,卷积神经网络在计算机视觉领域的应用更加广泛,目标检测作为计算机视觉领域的研究热点之一,已经应用于军用、民用各个领域,包括智能交通^[4,5]、智能监控^[6-8]、图像分割^[9,10]等,其中智能监控下的人员检测^[11,12]取得了巨大的成功.首先提出的目标检测^[13]方法是两级目标检测算法,其中 R-CNN^[10]由 Girshick 等在 2014 年提出,该算法首次将卷积神经网络应用于目标检测.同年 SPP-Net 被提出^[12],其可对整张图像仅进行一次卷积操作,避免了 R-CNN 中对所有候选区域分别提取特征的重复操作.接着 Fast R-CNN^[14]仿照 SPP-Net 在 R-CNN 的基础上进行了改进,提高了检测精度和速度.2016 年, Ren

等提出了 Faster R-CNN^[15],真正实现了网络的端到端训练.随后出现的 R-FCN 沿用 Faster R-CNN 的框架,使用全卷积神经网络,但是计算量仍较大,针对此情况出现了 Mask R-CNN^[16]网络,实现了对实例的分割.2016 年,Redmon 等深入分析目标检测算法速度差的原因,提出了一级目标检测算法 YOLOv1^[17].YOLOv1 使用网格预测的形式代替生成候选区域网络,提高了网络的检测速度.YOLOv2 在 YOLOv1 的基础上进行改进,引入批量归一化操作.YOLOv3^[18]继续改进,增加了多尺度预测方法,可以得到较高的检测精度.

1 系统总体结构

巡检人员在进入本文所述巡检系统前,首先需经过一套人脸识别系统,以保证进入巡检系统的人员唯一性.由于该系统并不属于本文重点研究内容,因此文章内部不进行讨论.如图 1 所示,本系统将多相机下的实时视频流数据通过 RTSP 协议传输至服务器,读取巡检路线中每个相机视角下的单帧图像进行巡检预设置.巡检预设置结束后,将视频流数据输入深度神经网络中进行人员检测.对巡检人员行为进行分析,得到人员实时位置信息并记录,分析记录结果进行干扰的剔除以及评估结果的获取,获取的评估结果包括实际巡检路线,巡检路线完成度,每个巡检区域对应的实际巡检时长,巡检区域完成度与巡检总完成度.最后将评估结果与分析结果得出的巡检事件存储并在网页端发布与显示.

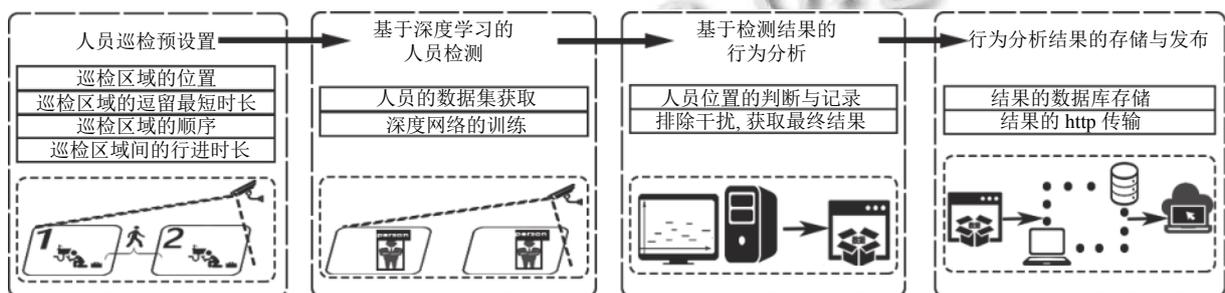


图 1 系统结构

本系统的特色及优势有以下几个方面:

- (1) 系统使用 YOLOv3 网络进行人员的检测,检测精度高且速度快,对复杂多变的实际环境检测能力强.
- (2) 系统设计了一种基于人员检测结果的巡检人员行为分析算法,该算法可以同时绑定多相机下的多个巡检区域,算法精度高、速度快且有效地剔除了

干扰.

- (3) 系统采用数据库存储模式对数据进行存储,并将图像压缩编码成 Base64 的形式快速通信,减少了网络传输量.

- (4) 系统包含一套完整的人机交互界面与可以长期稳定运行的系统,操作方便快捷.

2 系统算法

2.1 预设置

如图2所示, n 为巡检路线所包含的相机总数, 为巡检路线下的巡检区域总数. 本系统首先要进行相机名称的绑定, 绑定结果如图2中 Camera 1, Camera 2, ..., Camera n 所示, 并对应相机进行巡检预设置. 系统调用相机读取当前待配置相机下的图像作为巡检预设置背景, 通过本系统设计的人机交互界面进行巡检区域的绘制, 将巡检区域位置信息保存至数据库, 区域绘制结

果如图2中 Area 1, Area 2, ..., Area ($m+1$) 所示. 随后对应巡检区域设置最小逗留时长, 并保存至数据库中, 结果如图2中 Stay time 1, Stay time 2, ..., Stay time ($m+1$) 所示. 再配置巡检路线下的相机顺序, 根据配置的相机顺序以及绘制各个相机下的巡检区域顺序进行整体巡检区域顺序的自动更新, 将结果保存至数据库. 最后按照巡检区域顺序进行相邻巡检区域之间行进时长的设置, 设置结果保存至数据库中, 结果如图2中 Travel time 1, Travel time 2, ..., Travel time ($m+1$) 所示.

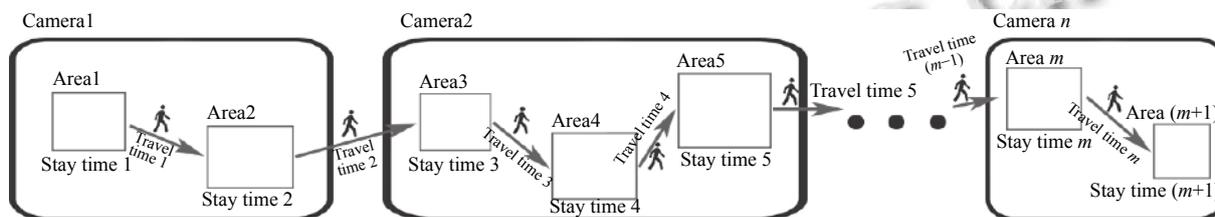


图2 巡检预设置说明图

2.2 目标检测算法

本系统中对于人员的检测, 选择采用 COCO 数据集训练 YOLOv3 网络进行人员检测, 以实现快速准确地判断出人员的位置信息. COCO 数据集中采用海量数据和标签的训练结果作为主干的卷积神经网络, 部分人员样本如图3所示.



图3 COCO 数据集中的人员样本

本系统采用的端到端目标检测算法 YOLOv3^[19] 将人员检测问题转化为一个回归性问题. 给定一张输入图像, 它能够直接在多个位置上回归出人员的边界框. 本目标检测算法不仅能保证实时性, 也能保证系统的准确性.

2.3 基于人员检测结果的行为分析算法

系统首先将视频流数据通过多线程的方式输入深度学习网络中得到人员检测结果, 将此结果进行行为分析. 首先当系统并未开启巡检监测时, 当相机场景下出现人员则系统报该巡检路线有人员闯入事件产生; 当系统开启巡检监测时, 首先判断人员是否在巡检区域内以及所在巡检区域编号. 判断示意图如图4所示.

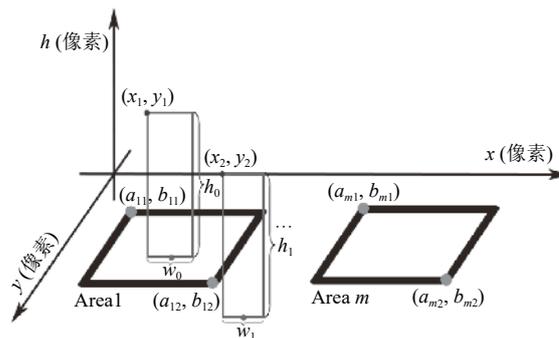


图4 人员位置判断示意图

图4中 a_{m1} 为巡检区域 m 左上角 x 坐标, b_{m1} 为巡检区域 m 左上角 y 坐标, a_{m2} 为巡检区域 m 右下角 x 坐标, b_{m2} 为巡检区域 m 右下角 y 坐标. x_k 为人员检测框 k 左上角 x 坐标, y_k 为深度学习人员检测框 k 左上角 y 坐标, w_k 为深度学习人员检测框 k 宽度, h_k 为深度学习人员检测框 k 高度.

人员位置判断方法如式 (1)、式 (2) 所示, 式中 k 为人员检测框序号, 检测框与人员一一对应, m 为巡检区域序号. 当同时满足式 (1) 和式 (2) 的时候, 人员 k 处于巡检区域 m 内.

$$a_{m1} \leq x_k + \frac{1}{2}w_k \leq a_{m2} \tag{1}$$

$$b_{m1} \leq y_k + h_k \leq b_{m2} \tag{2}$$

将满足式 (1)、式 (2) 的人员检测结果输入行为分析算法中进行干扰排除以及最终巡检行为分析结果的

获取. 假设有一条巡检路线包含5个巡检区域, 经深度学习人员检测并分析人员位置结果, 结果信息送入巡检人员行为分析算法进行处理, 输入信息如图5所示. 图5中每个点代表某时刻在某位置检测到人员. 当同一时刻有两个或者两个以上区域同时出现信息点的时候, 代表同一时刻多个区域同时有人的出现, 则先保存此刻数据, 假设信息连续次数小于某一给定阈值 $v=4$ 时则认为此处为干扰, 进行筛选排除, 如图中三角形点所示, 三角形点为干扰信息. 并且由于从一个区域行至下一区域至少需要某一时间, 所以当从某一区域离开再出现至下一区域时小于此时间, 则认为是外来干扰, 需要排除, 如图中正方形点所示, 正方形点也为干扰信息.

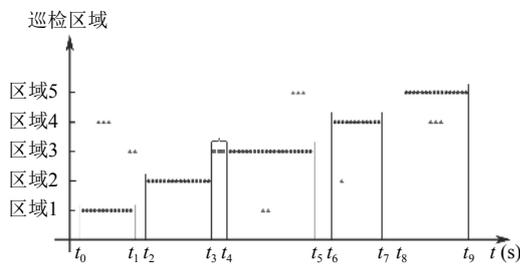


图5 行为分析算法示意图

排除干扰信息后得到正确信息, 如图5所示圆点为正确巡检信息, 并且由图可得知巡检正确路线及每个巡检区域真实逗留时长. 假设区域1到区域5最短逗留时长依次为 T_1, T_2, T_3, T_4, T_5 . 如图所示区域1到区域5实际逗留时长依次为 $t_1 - t_0, t_3 - t_2, t_5 - t_4, t_7 - t_6, t_9 - t_8$. 则各巡检区域完成度及巡检区域总完成度如式(3)、式(4)所示.

$$c_m = \begin{cases} 1, & \text{if}(t_m - t_{m-1} \geq T_m) \\ (t_m - t_{m-1})/T_m, & \text{if}(t_m - t_{m-1} < T_m) \end{cases} \quad (3)$$

$$C = \frac{\sum_{i=1}^{S_m} c_i}{s_m} \quad (4)$$

式(3)中, c_m 代表第 m 个巡检区域的巡检区域完成度, 式(4)中 C 代表巡检区域总完成度, T_m 代表巡检区域 m 最短逗留时长, S_m 代表巡检区域总数, m 为1, 2, 3, 4, 5.

如图5假设正确巡检路线为1至5, 实际巡检路线为 r_1, r_2, r_3, r_4, r_5 , 则采用式(5)、式(6)进行巡检路线总完成度 P 的计算.

$$P_m = \begin{cases} 1/m, & \text{if}(r_m == m) \\ 0, & \text{if}(r_m \neq m) \end{cases} \quad (5)$$

$$P = \sum_{i=1}^{S_m} P_i \quad (6)$$

假设巡检总完成度 S 中巡检区域总完成度所占评估权重为 l_1 , 巡检路线总完成度所占权重为 l_2 . 则该巡检总完成度计算方法如式(7)所示.

$$S = l_1 C + l_2 P \quad (7)$$

最后根据计算结果进行巡检事件判断, 判断算法如下所示. 当满足 $c_m < 1$ 时, 第 r_m 巡检区域不合格, 生成区域逗留不合格事件; 当满足 $P < 1$ 时候, 该巡检路线不合格, 生成巡检路线不合格事件. 根据 S 的值可以评判本次巡检情况, 本文所使用评价指标为, 当满足 $S < 0.85$ 时, 本次巡检情况差; 当满足 $0.85 \leq S < 1$ 时, 本次巡检情况良好; 当满足 $S = 1$ 时, 本次巡检情况优异, 即巡检合格完成.

2.4 数据存储、传输与结果发布

2.4.1 数据存储

数据的存储过程采用一组可以完成特定功能的SQL^[20]语句集, 经编译后存储进数据库, 最终用户可以通过指定存储过程的名称来进行调用. 这个过程可以看作是对编程中面向对象方法的模拟. 本系统数据库中所存储信息如图6所示. 图6(a)所示为巡检区域预设置存储信息. 图6(b)所示巡检人员行为分析结果存储信息.

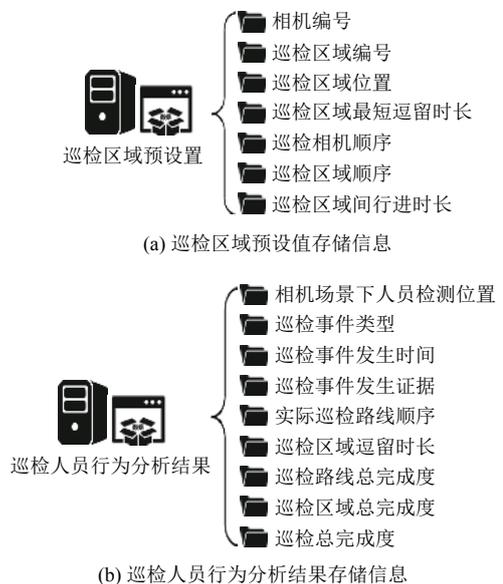


图6 数据库存储信息示意图

2.4.2 数据的传输与发布

本系统采用HTTP协议将数据库存储如图6(b)所示结果进行传输发布. 信息传输及发布过程展示如图7所示, 在进行数据的传输时, 客户向服务器请求服务只

需请求方法和路径, 通信简单快速, 且允许传输任意类型的数据对象. 客户机和服务器之间通信不需要建立持久的连接, 节省了传输时间. 如图 7 所示网页端可以向服务器端传输相机与算法的开关信号, 服务端接收信号后进行相应的操作处理. 服务端向网页端传输巡检行为分析结果及巡检事件信息, 巡检事件信息包含巡检过程中所产生巡检事件的类型、巡检事件产生时

间及巡检事件证据信息, 借此可以获知该巡检过程具体行为信息. 其中巡检事件中的证据图片在进行 HTTP 传输之前, 先进行 Base64 编码压缩, 将压缩信息传输, 这样可以节约网络资源, 网页端接收到信息后进行 Base64 解码, 还原出证据图片显示. 巡检行为分析结果信息可以直观评估该巡检过程. 数据传输结果在服务器网页端进行发布.

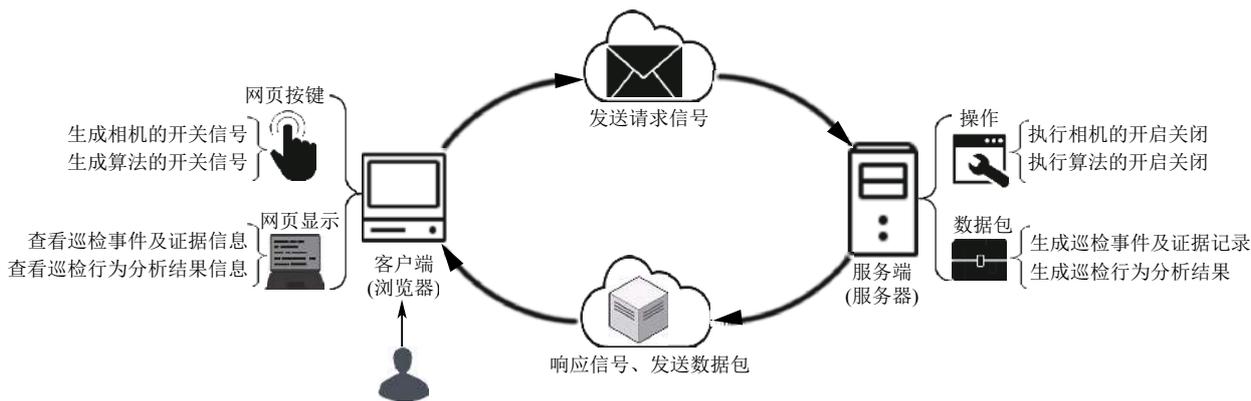


图 7 信息传输及发布过程示意图

3 实验结果分析

3.1 实验条件

本实验所采用的服务器硬件 CPU 为 Intel(R) Core (TM) i7-8700K, 显卡使用 NVIDIA GTX1080Ti, 摄像头使用 TL-IPC42A-4, 分辨率为 1080 P, 采集频率 15 帧/秒, 交换机使用华为 S1724G-AC 24 口千兆交换机, 网线使用千兆 CAT6 类网线. 本系统软件采用 QT、OpenCV、Microsoft SQL Server Management Studio 等相关软件环境. 本实验在 A、B 两个场景中进行测试, 每个场景下都有 5 段相互独立的巡检视频. 场景 B 较场景 A 实验环境更为复杂, 尤其在光照条件, 遮挡情况有明显的区分, 场景具体信息如表 1 所示. 下述实验中所采用的视频数据皆为相同数据源文件.

3.2 实验结果及分析

3.2.1 人员检测实验

如表 2 所示, 本实验通过获取在 A、B 两个场景下采用 YOLOv3 目标检测算法进行人员检测的结果与实际人员总数进行对比得出算法检测准确率进行评估. 实验得出场景 A 的平均检测准确率为 93.16%, 场景 B 的平均检测准确率为 90.94%. 由于场景 B 环境较场景 A 环境更为复杂, 因此在场景 B 下 YOLOv3 平均检测准确率低于场景 A 下的平均检测准确率. 尽管如此,

实验中场景 B 仍然拥有 90.94% 的平均检测准确率, 因此本文所提出的人员检测方法可以满足实际场景中人员检测的应用.

表 1 实验场景信息

场景	巡检点个数	正常巡检顺序	光照	遮挡	每个点的逗留时长(s)	区域间行进时长(s)	视频名称	视频帧数
A	5	1-2-3-4-5	较好	较少	5-5-5-5-5	3-3-6-4	test1	4300
					12-12-8-8-10	4-2-6-2	test2	4900
					6-8-10-8-4	8-4-6-3	test3	5700
					10-8-6-10-10	2-6-4-2	test4	6500
					12-10-8-6-8	3-9-6-2	test5	8000
B	4	1-2-3-4	较差	较多	8-8-8-8	4-5-8	test6	4300
					10-6-8-4	6-4-6	test7	4900
					12-15-8-8	10-5-7	test8	5700
					10-12-10-10	8-4-7	test9	6500
					8-6-9-8	9-10-12	test10	8000

3.2.2 巡检检测实验

如表 3 所示, 本实验将在场景 A 和场景 B 下的 YOLOv3 人员检测结果的基础上对本系统提出的巡检人员行为分析算法进行测试, 将分析结果与对应实际结果进行对比. 实验假设在巡检完成度中, 巡检区域总完成度所占权重为 0.7, 巡检路线总完成度所占权重为 0.3. 通过式 (7) 的计算得知场景 A 中的巡检路线监测平均准确率为 92%, 巡检逗留时长监测平均准确率为 98.24%. 场景 B 中的巡检路线监测平均准确率为 85%,

巡检逗留时长监测平均准确率为 95.96%。场景 B 由于人员遮挡情况更多使得巡检路线监测平均准确率和巡检逗留时长监测平均准确率略低于场景 A, 但其依然具有较高的监测准确率, 所以本系统所提出的巡检人员行为分析算法有效地剔除了干扰, 可以满足在实际场景下对巡检路线和巡检逗留时长的监测。

表 2 人员检测实验结果

场景	视频名称	视频帧数	实际人员数	检测人员数	准确率(%)
A	test1	4300	10800	10066	93.2
	test2	4900	12500	11588	92.7
	test3	5700	15100	14134	93.6
	test4	6500	19000	17461	91.9
	test5	8000	24000	22656	94.4
平均准确率					93.16
B	test6	4300	10240	9349	91.3
	test7	4900	13600	12362	90.9
	test8	5700	15220	13850	91.0
	test9	6500	18700	16793	89.8
	test10	8000	25300	23200	91.7
平均准确率					90.94

表 3 巡检监测实验结果

场 景	视频 名称	巡检路线		巡检时长			
		实际	监测	准确率(%)	实际(s)	监测(s)	准确率(%)
A	test1	1-2-3-4-5	1-2-3-4-5	100	38	38	100
	test2	5-4-3-2-1	5-4-3-1	60	49	48	98.0
	test3	1-2-4-3-5	1-2-4-3-5	100	48	46	95.8
	test4	2-4-3-5-1	2-4-3-5-1	100	45	45	100
	test5	2-3-1-4-5	2-3-1-4-5	100	39	38	97.4
平均准确率				92			98.24
B	test6	1-2-3-4	1-2-3-4	100	40	38	95.0
	test7	4-2-3-1	4-2-1	50	29	27	93.1
	test8	1-3-2-4	1-3-2-4	100	45	44	97.8
	test9	4-1-3-2	4-1-3-2	100	49	46	93.9
	test10	1-2-4-3	1-2-4	75	28	28	100
平均准确率				85			95.96

3.2.3 系统效果实验

如表 4 所示, 本实验将在场景 A 和场景 B 下基于 YOLOv3 人员检测结果的行为分析结果得出的巡检结果分别进行评估, 评估结果的准确率与由本文 3.2.2 章节巡检监测实验所得出的巡检路线及巡检逗留时长准确率相关。同时本实验采用实时率计算方法对输入系统视频流的总时长与经过系统处理后输出的视频流总时长作对比, 其计算结果为二者的比值, 得到场景 A 下系统平均实时率为 98.88%, 场景 B 下系统平均实时率为 98.38%。通过对比场景 A、B 获取的平均实时率结果, 表明系统实时率仍受场景中环境因素的干扰。但是经过对场景 A、B 实时率的分析说明本系统近似满足实时性要求, 可以进行实际场景的应用。

表 4 系统效果实验结果

场 景	视频 名称	巡检路线总 完成度(%)	巡检区域总 完成度(%)	巡检总完 成度(%)	视频总 时长(s)	系统处理 时间(s)	实时 率(%)
A	test1	100	100	100	287	291	98.6
	test2	20	73.3	57.3	327	329	99.4
	test3	60	92.0	82.4	380	386	98.4
	test4	20	88	67.6	433	437	99.1
	test5	40	79.7	67.8	533	539	98.9
平均实时率							98.88
B	test6	100	100	100	287	292	98.3
	test7	25	75	60	327	333	98.2
	test8	50	85.4	74.8	380	385	98.7
	test9	25	100	77.5	433	438	98.9
	test10	50	68.8	63.2	533	545	97.8
平均实时率							98.38

文献 [21] 也进行了巡检路径的检测, 文章采用基于 GPS 的大圈以及小圆两种距离计算方法进行巡检路线的监测, 成本高且精度最高为 81.3%, 低于本文提出的方法, 且未谈及有关处理速度的研究。此方法与本文提出的方法相比, 本文方法仅使用视频流信息, 节约成本, 且考虑实际场景信息更为详尽, 能够排除场景内的大多数干扰, 获取准确的人员巡检信息。本文提出的方法在普通场景下巡检监测系统精度为 95.12%, 在复杂场景下巡检监测系统精度为 90.48%, 且速度可以满足实时处理的需求, 均优于文献 [21]。因此本文所提供的算法以及巡检监测系统, 是能够实际应用于工业场景的智能化巡检监测。

4 结语

为了快速、准确地分析工业现场巡检过程中巡检人员的巡检行为是否符合规范和标准, 本文提出了一种基于机器视觉的工业现场巡检过程监控分析系统。系统对每个相机场景进行巡检时长、巡检路线等参数的预设置, 设置完成后使用 YOLOv3 网络监控多相机下的人员位置信息。系统使用的检测算法检测精度高、速度快, 并适用于多种复杂的检测环境。接着使用本文设计的巡检人员行为分析方法分析人员位置信息, 并评估巡检过程。该方法能够并行处理多个相机下的多个巡检区域的视频流数据, 有效排除了干扰。最后将评估结果存储、传输与发布。根据网页发布结果, 可以查看巡检过程中巡检人员具体行为信息。系统采用数据库存储模式对数据进行存储, 保证系统数据信息的安全。在数据传输时, 系统使用 Base64 对图像数据进行编码及解码, 减少了网络通信量。本系统包含一套完整成熟的人机交互界面与可以长期稳定运行的系统, 且设备安装简单, 操

作方便,能够满足实时数据处理的需求。

参考文献

- 1 Noorit N, Suvonvorn N. Human activity recognition from basic actions using finite state machine. In: Herawan T, Deris MM, Abawajy J, eds. Proceedings of the First International Conference on Advanced Data and Information Engineering (DaEng-2013). Singapore, Singapore. 2014. 379–386.
- 2 Bobick A, Davis J. Real-time recognition of activity using temporal templates. Proceedings of the 3rd IEEE Workshop on Applications of Computer Vision. Sarasota, FL, USA. 1996. 39–42.
- 3 Lecun Y, Bengio Y, Hinton G. Deep learning. Nature, 2015, 521(7553): 436–444. [doi: [10.1038/nature14539](https://doi.org/10.1038/nature14539)]
- 4 Jin XY, Davis CH. Vehicle detection from high-resolution satellite imagery using morphological shared-weight neural networks. Image and Vision Computing, 2007, 25(9): 1422–1431. [doi: [10.1016/j.imavis.2006.12.011](https://doi.org/10.1016/j.imavis.2006.12.011)]
- 5 Wang L, Lu Y, Wang H, *et al.* Evolving boxes for fast vehicle detection. Proceedings of 2017 IEEE International Conference on Multimedia and Expo. Hong Kong, China. 2017. 1135–1140.
- 6 Al-Nawashi M, Al-Hazaimeh OM, Saraee M. A novel framework for intelligent surveillance system based on abnormal human activity detection in academic environments. Neural Computing and Applications, 2017, 28(1): 565–572.
- 7 Nguyen-Meidine LT, Granger E, Kiran M, *et al.* A comparison of CNN-based face and head detectors for real-time video surveillance applications. Proceedings of the 2017 7th International Conference on Image Processing Theory, Tools and Applications. Montreal, QC, Canada. 2017. 1–7.
- 8 Yu RC, Wang HC, Davis LS. Remotenet: Efficient relevant motion event detection for large-scale home surveillance videos. Proceedings of 2018 IEEE Winter Conference on Applications of Computer Vision. Lake Tahoe, NV, USA. 2018. 1642–1651.
- 9 van de Sande KEA, Uijlings JRR, Gevers T, *et al.* Segmentation as selective search for object recognition. Proceedings of 2011 International Conference on Computer Vision. Barcelona, Spain. 2011. 1879–1886.
- 10 Girshick R, Donahue J, Darrell T, *et al.* Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA. 2014. 580–587.
- 11 Tian YL, Luo P, Wang XG, *et al.* Pedestrian detection aided by deep learning semantic tasks. Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA. 2015. 5079–5087.
- 12 Zhou XY, Gong W, Fu WL, *et al.* Application of deep learning in object detection. Proceedings of the 2017 IEEE/ACIS 16th International Conference on Computer and Information Science. Wuhan, China. 2017. 631–634.
- 13 Ouyang WL, Wang XG. Joint deep learning for pedestrian detection. Proceedings of 2013 IEEE International Conference on Computer Vision. Sydney, Australia. 2013. 2056–2063.
- 14 Girshick R. Fast R-CNN. Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile. 2015. 1440–1448.
- 15 Ren SQ, He KM, Girshick R, *et al.* Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137–1149. [doi: [10.1109/TPAMI.2016.2577031](https://doi.org/10.1109/TPAMI.2016.2577031)]
- 16 He KM, Gkioxari G, Dollár P, *et al.* Mask R-CNN. Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy. 2017. 2980–2988.
- 17 Redmon J, Divvala S, Girshick R, *et al.* You only look once: Unified, real-time object detection. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA. 2016. 779–788.
- 18 Ju MR, Luo HB, Wang ZB, *et al.* The application of improved YOLO V3 in multi-scale target detection. Applied Sciences, 2019, 9(18): 3775. [doi: [10.3390/app9183775](https://doi.org/10.3390/app9183775)]
- 19 Lin TY, Dollár P, Girshick R, *et al.* Feature pyramid networks for object detection. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA. 2017. 936–944.
- 20 Kagalwala RA, Thompson JP. Database schema for Structured Query Language (SQL) server: US, 7653652. 2010-01-26.
- 21 孙华. GPS智能化设备巡检 [硕士学位论文]. 济南: 山东大学, 2005.