

# 基于改进 YOLOv5s 算法的危险区域入侵报警<sup>①</sup>



沈 杰, 黄晓华

(南京理工大学 机械工程学院, 南京 210094)  
通信作者: 沈 杰, E-mail: alpha298@163.com

**摘 要:** 工厂环境复杂多变, 存在很多危险区域, 违规进入会给工人的生命健康带来严重的危害. 针对传统的检测方法操作复杂、识别效果差, 提出了一种基于改进 YOLOv5s 模型的危险区域工人入侵警报系统. 首先将基于 SGBM 算法双目测距技术融合进 YOLOv5s 目标检测中, 增加空间距离这一触发条件, 使得工人只有走近摄像头一定范围内才会触发声光报警. 进一步地, 在 YOLOv5s 中引入注意力机制, 通过对比实验证明了 CA 模块的引入对模型的平均准确率  $mAP@0.5$  提升最明显为 1.86%. 结果显示此方法能够较为准确的识别出工人是否进入危险区域, 并进行声光报警, 提醒工人注意, 为工厂安全管理提供了新的手段.

**关键词:** YOLOv5s; 危险区域; 注意力机制; 双目测距; 目标检测

引用格式: 沈杰, 黄晓华. 基于改进 YOLOv5s 算法的危险区域入侵报警. 计算机系统应用, 2023, 32(3): 157-162. <http://www.c-s-a.org.cn/1003-3254/9019.html>

## Intrusion Alarm of Dangerous Area Based on Improved YOLOv5s Algorithm

SHEN Jie, HUANG Xiao-Hua

(School of Mechanical Engineering, Nanjing University of Science and Technology, Nanjing 210094, China)

**Abstract:** The factory environment is complex and changeable, with many dangerous areas, and illegal entry can bring serious harm to the life and health of workers. Considering the complex operation and poor recognition effect of traditional detection methods, this study proposes an alarm system for workers' intrusion in dangerous areas on the basis of the improved YOLOv5s model. Firstly, the binocular ranging technology based on the SGBM algorithm is integrated into YOLOv5s object detection, and the trigger condition of spatial distance is added. Hence, the sound and light alarm will be triggered only when workers approach the camera within a certain range. Furthermore, the attention mechanism is introduced into YOLOv5s. Comparative experiments prove that the introduction of the CA module improves the average accuracy of  $mAP@0.5$  by 1.86%. The results show that this method can accurately identify the intrusion of a worker in dangerous areas and gives a sound and light alarm to remind the worker, which provides a new means for factory safety management.

**Key words:** YOLOv5s; dangerous area; attention mechanism; binocular ranging; object detection

近些年来, 深度学习技术快速发展, 在各个领域都有广泛的应用, 极大地赋能了“智能工厂”的建设. 在工厂中存在很多危险区域, 如机械设备工作半径内、高温、高电压等场所. 一旦人员进入, 就会对工人的生命健康造成不可想象的损害. 因此, 如何保障工厂内人员

的安全已经成为管理者们越来越重视的问题. 要预防这些重大安全事故的发生, 就得对工厂的危险区域进行管控, 不仅仅需要限制人员的随意进出, 而且要对进入的人员及时的声光提醒.

传统的工人危险区域入侵检测一般是通过 UWB

<sup>①</sup> 收稿时间: 2022-08-16; 修改时间: 2022-09-15; 采用时间: 2022-10-14; csa 在线出版时间: 2022-12-16  
CNKI 网络首发时间: 2022-12-19

(ultra-wideband) 和 GPS (global positioning system) 定位技术将采集到的施工现场环境和人员坐标位置数据信息整合进 BIM (building information modeling) 建筑信息模型中<sup>[1]</sup>, 通过服务器的判断、分析、处理、传输, 从而对现场工作人员的不安全行为进行监控. 但这种方式实际操作过程比较复杂, 而且干扰工人的正常操作, 也给管理工作增加负担. 近两年随着深度学习技术的逐渐兴起, 基于机器视觉的目标检测技术逐渐运用到这一领域, 如任磊等<sup>[2]</sup> 针对不同的危险区域的范围设计了两种施工场地人员图像数据集制作方式, 并且通过 YOLOv3 神经网络训练权重的方式, 用实验验证了该方法有较高的可行性和准确率. 但是这样, 工人只要出现在画面图像内, 识别到就会直接报警, 必须保证画面内全都是危险区域. 这对于危险区域的划分提出了更加细致的要求, 而且相机的安装、调试比较困难.

因此本研究在 YOLOv5s 目标检测的基础上, 嵌入了基于 SGBM 算法的双目测距技术, 增加了空间距离这一触发条件, 工人只有走近距离摄像头一定范围内, 并且识别到才会触发声光报警. 同时针对光照不足检测图像模糊的问题, 引入注意力机制强化特征通道, 增强重要特征的权重, 减少背景的特征提取, 提高目标识别的精度.

## 1 入侵报警系统总体设计

本文设计的工人危险区域入侵警报系统具体流程图如图 1.

1) 将双目测距代码嵌入到改进后 YOLOv5s 主程序中, 双目相机把实时捕获的左、右图像沿着长度方向合并为一张图片输入进程序. 因此在双目测距程序开始前需要将捕捉到的一张照片分成左、右两张图片, 通过 SGBM 立体匹配算法经过读取相机内外参、立体校正、消除畸变、立体匹配等步骤, 计算以左图像为基准的视差图, 并经过重投影矩阵转换得到相机坐标系下三维点的坐标  $(x, y, z)$ .

2) 在 YOLOv5s 目标检测中的 process detections 模块只将左图作为检测图像输入到模型中进行目标检测, 输出预测框位置 (框左上点和右下点的纵横坐标).

3) 双目测距和目标检测两个模块都以左图为准. 因此, 可以按上述公式计算目标检测每一个预测框的中心点与左相机坐标系下原点的距离, 考虑到因视差图成像质量问题可能导致的中心点坐标丢失的情况,

需设置循环语句, 以预测框中心点为中心, 沿图像坐标系横纵两个方向在预测框范围内寻找距离在设定合理范围内的点, 以此点代替中心点坐标输出<sup>[3]</sup>.

4) 将检测点 3 轴方向的距离以及计算得到的总距离值通过 cv.putText 函数输出到预测框一侧, 类别的置信度显示在另一侧. 如图 2 所示.

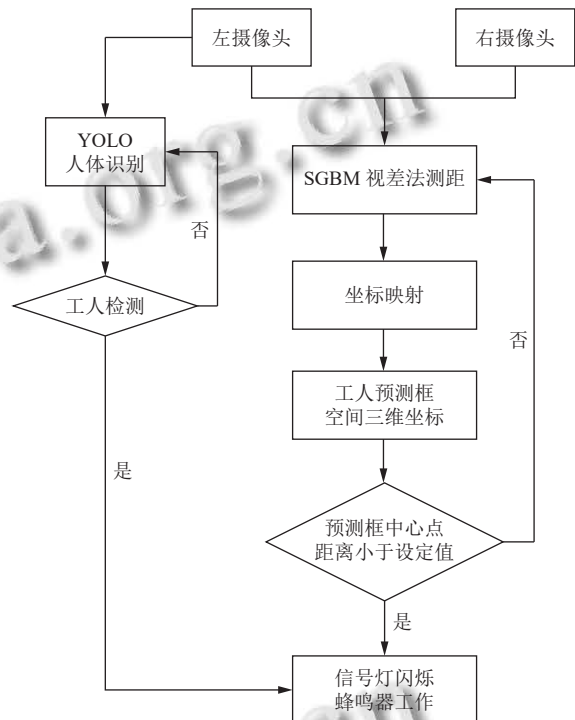


图 1 报警系统总体设计流程图

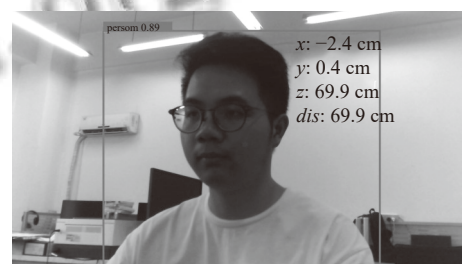


图 2 实验效果

5) 将程序部署在基于 Linux 系统的 Jetson NX 的工控机上, 分别将警报灯和蜂鸣器连接在工控机的 30 pin 的多功能插针中 GPIO\_1, GPIO\_2, 配置端口文件, 设定输出的电压值与方向, 如下代码.

```

os.system('echo 421 > /sys/class/gpio/export')
os.system('echo out > /sys/class/gpio/gpio421/direction')
os.system('echo 1 > /sys/class/gpio/gpio421/value')
  
```

6) 在程序运行时,若检测图像画面中有到工人,且工人预测框中心点的距离值在设定值内,则将 GPIO\_1, GPIO\_2 口的 value 值置 1 通电,实现灯光、蜂鸣报警.

7) 将工控机及摄像头部署在工厂危险区域中,实际测试效果如图 3.



图 3 工厂实测效果图

程序运行稳定,且能够很好地识别出工人及距离,并且发出声光报警,提醒工人注意.

以下分别介绍双目测距模块和改进的 YOLOv5s 目标检测模块.

## 2 双目测距模块

双目测距是模仿人类的双眼通过视差的大小来感受物体的远近,其主要步骤有相机标定、图像校正、立体匹配、三维重构 4 大部分.

### 2.1 相机标定

相机标定主要有两大目的:首先是为了求解真实世界坐标系到相机坐标系的变换矩阵,其次是为了求解畸变系数,进行图像矫正.其基于张正友平面模板法原理,以棋盘格为标定参照物,使用双目摄像头拍摄的不同角度的图片,各 14 张,如图 4 所示.

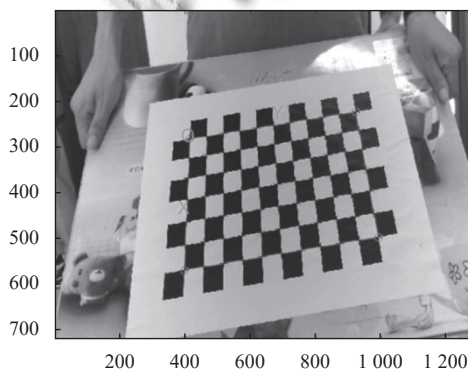


图 4 棋盘格标定

利用 Matlab 中 camera calibration 工具箱将拍摄的样张图片通过提取角点、预估畸变误差等操作对左右两相机分别进行单目标定,得到每个相机的内部参数、畸变系数.再利用上述标定的图片利用 stereo camera calibrator 工具箱将误差较大图像移除之后重新计算,对外部参数进行标定,得到双目相机的参数如表 1.

表 1 双目标定参数

参数	表达式
内部参数	$MatrixLeft = \begin{bmatrix} 1135.660 & 1.342 & 749.440 \\ 0 & 1145.088 & 377.729 \\ 0 & 0 & 1 \end{bmatrix}$
	$MatrixRight = \begin{bmatrix} 1135.660 & 1.342 & 749.440 \\ 0 & 1145.088 & 377.729 \\ 0 & 0 & 1 \end{bmatrix}$
畸变参数	$DL = \begin{bmatrix} 0.02813 & 0.24039 & -0.0010 & 0.00049 & -1.32553 \end{bmatrix}$
畸变参数	$DR = \begin{bmatrix} 0.02646 & 0.11626 & -0.00057 & -0.00118 & -0.43902 \end{bmatrix}$
外部参数	$R = \begin{bmatrix} 0.99997 & 0.00014 & 0.00755 \\ -0.00009 & 0.99997 & -0.00710 \\ -0.00755 & 0.00712 & 0.99995 \end{bmatrix}$
	$T = \begin{bmatrix} -59.75972 & -0.21550 & -0.04529 \end{bmatrix}$

### 2.2 图像校正

利用 OpenCV 中 Undistort Points() 和 StereoRectify() 函数分别依据上述的标定好的参数进行畸变及立体校正,同时得到重投影矩阵.最终得到一个消除左右原始图像的径向、切向畸变误差,经过旋转和重新投影映射的行对准的双目图像,如图 5 所示.



图 5 行对齐图像

### 2.3 立体匹配

立体匹配算法是依据匹配策略,在左图基准图像上提取匹配单元对应的图像信息,根据极线约束准则<sup>[4]</sup>在右图参考图像的交线上寻找最相似的匹配点,最后由左右最佳匹配点的位置提取视差值计算深度距离.视差值计算深度值公式如下:

$$depth = \frac{f \times b}{d + (c_{xr} - c_{xl})} \quad (1)$$

其中,  $f$  为焦距长度,  $b$  为基线长度,  $d$  为视差,  $c_{xr}$  和  $c_{xl}$  是两个相机主点的列坐标.

因此本文采用以 SGM 算法为核心的 SGBM 算法.其匹配代价由两部分组成:在矩形窗口内,将经过 Sobel



算子预处理后的图像梯度图计算得到的 BT 代价与原始图片计算得到的 BT 代价简单相加,减少由图像采样离散化造成的误差。

以上代价函数计算的每一个数值储存在代价空间  $C(x,y,d)$  内,通过多个方向的一维路径上平等的进行代价聚合<sup>[5]</sup>,如图 6 所示,代价聚合的路径可以是 8 个或者 16 个,这样就可以覆盖全部的图像,通过这种方式来近似替代二维最小化全局能量函数。

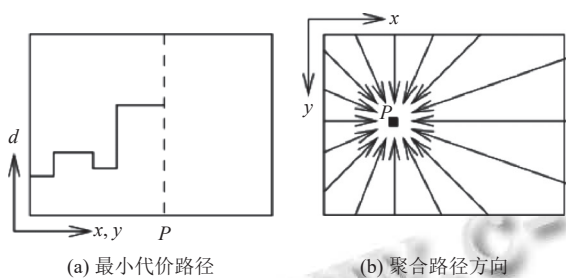


图 6 代价聚合路径

SGM 算法巧妙的使用一维聚合代替了二维聚合这一非确定性问题。首先选择某一路径,在该方向上计算有效范围内的匹配代价,接着对所有的路径进行代价计算,将这些匹配代价相加求和,选取有效视差范围内的最小值作为此像素点聚合后的匹配代价。

以下为任意像素  $p$ , 选择视差为  $d$ , 方向为  $r$  上的代价计算公式:

$$L'_r = C(p,d) + \min \left\{ \begin{array}{l} L'_r(p-r,d) \\ L'_r(p-r,d-1) + P_1 \\ L'_r(p-r,d+1) + P_1 \\ \min_i L'_r(p-r,i) + P_2 \end{array} \right\} - \min_k L_r(p-r,k) \quad (2)$$

式 (2) 中第 1 项是当前代价  $C$ , 第 2 项平滑项为视差不变时的代价, 视差差值为 1 时的代价, 视差差值大于 1 的代价这三者的最小值, 其中  $P_1, P_2$  惩罚项。第 3 项是防止因总聚合值随着路径不断叠超出上限, 必须减去常数项  $\min_k L_r(p-r,k)$ , 像素  $p$  的任意视差值都对应一个常数, 因此不会对结果产生影响<sup>[6]</sup>。但最终代价聚合的上限因此会降为:

$$L \leq C_{\max} + P_2 \quad (3)$$

最终聚合的代价是所有路径上的和:

$$S(p,d) = \sum_r L_r(p,d) \quad (4)$$

经过代价聚合后按照“Winer Takes All”原则求解视差:

$$d^* = \arg \min S(p,d) \quad (5)$$

原始的视差图需要进一步的优化处理, 通过唯一性检测、亚像素插值、连通域噪声滤波等操作剔除错误的匹配点, 提高视差精度。最终视差图如图 7 所示。



图 7 视差图

## 2.4 三维重构

为测算空间物体到摄像头的距离, 还需将视差图经过 `cv2.reprojectImageTo3D()` 函数变换, 得到与视差图大小相同的映射图像, 其中每一个像素的 3 个通道分别存储着在相机坐标系下对应点的三维坐标  $(x,y,z)$ , 因此每个点空间距离为:

$$distance = \sqrt{x^2 + y^2 + z^2} \quad (6)$$

## 3 改进 YOLOv5s 模型

YOLO 模型一直以检测速率快而闻名。随着新的优化技术如 Focal Loss, FPN 等的不断加入, 到如今 YOLOv5 检测精度已经有了极大的提升, 且其模型的权重文件相比上代缩小近 90%, 使得 YOLOv5 模型非常适合部署在嵌入式设备实现对目标的实时检测<sup>[7]</sup>。因此本模型以 YOLOv5s 模型为基础架构, 通过引入注意力机制增强重要特征, 提高检测精度。

### 3.1 引入注意力机制

人体视觉的注意力机制可以使得我们在大量无效信息干扰下, 将注意力集中到需要关注的区域, 极大地提高了大脑处理信息的效率。通过借鉴人体视觉的注意力机制, 在 YOLOv5s 的网络结构中的 C3 模块中 Bottleneck 模块加入 CA (coordinate attention) 注意力机制模块, 将传递通道赋予不同的权重, 使得在梯度计算过程中, 权重较大的通道占主导地位, 最终呈现在原始图像中网络能更明显的检测重点关注部分<sup>[8]</sup>。

### 3.2 CA 模块

相比之前发布的 SE、ECA、CBAM 模块仅考虑

建模通道以及特征图空间之间的关系来构建注意力特征图, CA 模块 (图 8) 还考虑了特征空间的位置信息. 将原始输入特征分解为两个一维特征的编码过程, 分别沿着长宽两个空间方向聚合特征进行全局池化, 在一个空间方向建立依赖关系的同时, 在另一方向捕获精准的位置特征<sup>[9]</sup>, 二者相辅相成, 共同应用于特征图, 以增强关注对象的表示. 再将生成的 attention map 进行 concat 操作, 使通道叠加, 在经过卷积及非线性函数变换, 为空间信息各自在长宽方向进行编码的中间特征映射, 将得到的新特征按照  $H$  和  $W$  的值拆分, 在经过激活函数与最原始的输入相乘, 得到 CA 模块最终的注意力特征图.

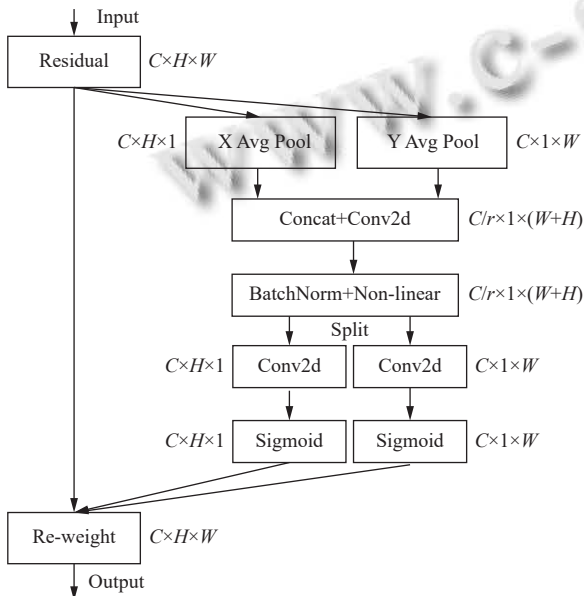


图 8 CA 模块结构

### 3.3 对比实验

将上述 4 个注意力机制模块分别插入到 YOLOv5s 的 C3 块的 Bottleneck 结构中, 如图 9 所示的 attention module 位置, 分别组成新的 YOLOv5s-SE, YOLOv5s-ECA, YOLOv5s-CBAM, YOLOv5s-CA 四个网络模型, 将原始的 YOLOv5s 模型作为空白对照组进行对比实验.

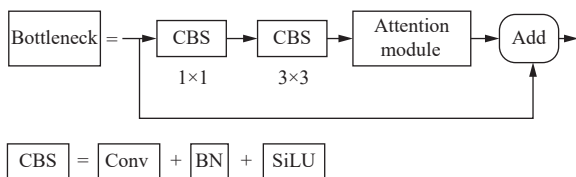


图 9 注意力模块插入位置

### 3.4 数据集

实验采用 COCO 数据集中 10 822 张“人”类别图片, 其中 8 129 张图片作为训练集, 共计 33 199 个目标; 2 693 张图片作为测试集, 共计 11 004 个目标.

### 3.5 实验参数

本实验基于 Windows 10 平台下 PyTorch 1.10.2 深度学习架构. 其中 CPU 型号为 AMD Ryzen 75800 8 核处理器@3.4 GHz, 系统运行内存 16 GB, GPU 为 8 GB NVIDIA GeForce RTX 3070Ti, 学习率(lr)设为 0.01, 动量因子 (momentum) 设为 0.937, 权重衰减 (weight\_decay) 设为 0.000 5, 批处理单元 (batch\_size) 设为 32, 训练 300 轮, 从头开始训练, 当验证损失 (validation loss) 平稳不下降时, 即视为模型收敛<sup>[10]</sup>.

### 3.6 评价指标

本实验通过目标检测模型的通用评价指标来定量分析上述 4 个改进网络模型对人体识别的检测效果的影响, 分别为准确率 (precision,  $P$ ), 查全率 (recall,  $R$ ), 通过绘制  $P$ - $R$  曲线, 将曲线做积分即为检测精度 (average precision,  $AP$ ), 将 IoU0.5 设定为阈值, 大于阈值的预测框被判为正例, 反之则为负例<sup>[11]</sup>, 得到平均值  $mAP@0.5$  (mean average precision) 为最终评价指标.

### 3.7 数据分析

为了验证注意力模块 CA 模块的引入对整体模型性能的提升效果, 本实验以原始的 YOLOv5s 模型为对照组, 设计一系列对比实验, 实验结果以  $mAP@0.5$  为参考标准, 如图 10 和表 2 所示, 原始的 YOLOv5s 模型在训练轮次 240 轮后模型发生轻微的模型退化现象, 导致检测精度下降. CBAM 注意力机制的引入, 虽然其在 SE 模块的基础上增加了空间注意力模块, 但在给 YOLOv5s 模型增加计算量的同时, 并没有得到很好的效果. 而 ECA 模块和 SE 模块对整体准确率的提升效果差不多, 但是 ECA 模块使用  $1 \times 1$  Conv1d 代替 SE 模块中的全连接层, 将模型的总层数从 342 layers 较少到 319 layers, 从 7 092 342 parameters 减少到 7 063 578 parameters, 减少了模型的计算量.

表 2 各模型实验结果对比

序号	模型	$mAP@0.5$ (%)
1	YOLOv5s	68.85
2	YOLOv5s-SE	70.34
3	YOLOv5s-CBAM	70.06
4	YOLOv5s-ECA	70.39
5	YOLOv5s-CA	<b>70.71</b>

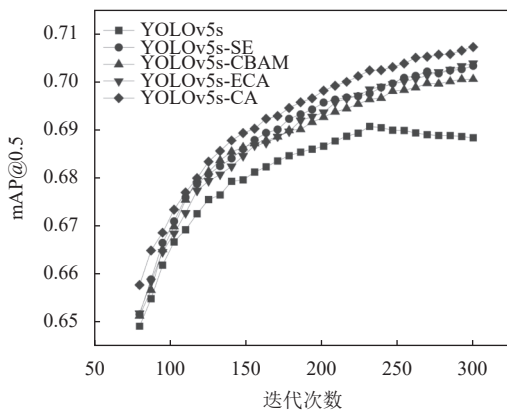


图 10 各个模型 mAP@0.5 曲线

而上述提到的 CA 模块相比于其他 3 种模型考虑了特征空间的位置信息,提高了预测框定位的准确率,因此对整体模型平均正确率的提升是最明显的,将 YOLOv5s 模型的 mAP@0.5 提升了 1.86%。如图 11,对密集小目标的检测效果也较为理想。因此将 YOLOv5s-CA 模型作为下述 YOLOv5 的危险区域工人入侵报警系统的基础模型。

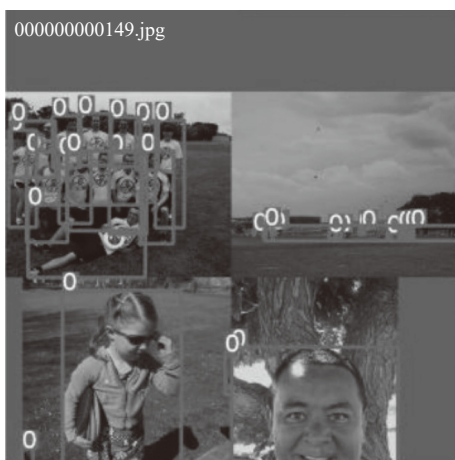


图 11 改进模型预测效果

## 4 结论

本文提出了一种基于改进 YOLOv5s 模型的检测方法用于工厂危险区域工人入侵的检测。一方面,本文通过将基于 SGBM 算法的双目测距算法嵌入 YOLOv5s 模型中,增加空间距离这一触发条件,提高了人员侵入报警的有效性 with 可靠性。另一方面,在 YOLOv5s 中引

入了 CA 注意力机制,通过训练实验验证了相比于其他模型 CA 模块更能提高整体模型的精度,提升了 1.86%。最后将程序部署在工控机上,在发现危险区域人员入侵时能够及时的进行声光报警。

## 参考文献

- 于用庆,邹树琪,张兴军.基于 BIM 技术与定位技术的地铁施工人员安全预警.城市轨道交通研究,2021,24(6):129-132,136.
- 任磊,苗作华,李自强,等.基于 YOLOv3 算法的危险区域工人识别.土木工程信息技术,2022,14(2):10-17.
- 杨荣坚,王芳,秦浩.基于双目图像的行人检测与定位系统研究.计算机应用研究,2018,35(5):1591-1595,1600. [doi:10.3969/j.issn.1001-3695.2018.05.068]
- 彭泽龙,李铨.基于双目测距和 YOLOv5s 的红外人体测温精度补偿系统.信息与电脑(理论版),2022,34(7):80-83.
- Hirschmuller H. Accurate and efficient stereo processing by semi-global matching and mutual information. Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR2005). San Diego: IEEE, 2005. 807-814.
- 高聪聪.基于 SGM 的立体匹配算法研究[硕士学位论文].西安:西安电子科技大学,2021.
- 闫彬,樊攀,王美茸,等.基于改进 YOLOv5m 的采摘机器人苹果采摘方式实时识别.农业机械学报,2022,53(9):28-38,59. [doi:10.6041/j.issn.1000-1298.2022.09.003]
- 徐诚极,王晓峰,杨亚东. Attention-YOLO: 引入注意力机制的 YOLO 检测算法.计算机工程与应用,2019,55(6):13-23. [doi:10.3778/j.issn.1002-8331.1812-0010]
- Hou QB, Zhou DQ, Feng JS. Coordinate attention for efficient mobile network design. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE 2021. 13713-13722.
- 顾星,詹伟达,崔紫薇,等.基于注意力机制的红外目标检测方法.激光与光电子学进展,2022:1-14. <http://kns.cnki.net/kcms/detail/31.1690.TN.20220713.1942.591.html>. (2022-07-18)[2022-10-18].
- 张秀再,邱野,张晨.改进 YOLOv5s 算法的地铁场景行人目标检测.激光与光电子学进展,2022:1-20. <http://kns.cnki.net/kcms/detail/31.1690.TN.20220713.1944.609.html>. (2022-07-16)[2022-10-18].

(校对责编:孙君艳)