

# 基于改进 YOLOv5 障碍物检测的智能导盲杖研究

李直鸿, 谢娜, 禹锦明, 谭腾飞

(湖南工程学院 计算机与通信学院, 湘潭 411104)

**摘要:** 针对目前视障人士居家出行安全的问题, 我们设计出一款基于深度学习目标检测算法的智能导盲杖, 该导盲杖创新在于以改进后的 YOLOV5 算法为核心, 同时结合超声波测距模块, 实现对障碍物的准确检测。然后并通过语音播报使用者, 并且搭配了精准的北斗定位等功能, 能够安全方便的帮助视障人士出行。所以我们设计的是更加方便用户使用的东西, 而不仅仅是个的拐杖, 它不仅能告诉你有一个物体挡住了你的去路, 还能告诉你那个物体是什么, 然后帮助你绕过它。此外在障碍检测方面, 实际的遍历调查发现, 小型障碍物以及多背景障碍物在公园小区等地方尤为常见。首先由于公园小区障碍物的种类大小不一, 容易出现一些小型或中小型的障碍物, 我们增加通道注意力机制 (ECANet), 同时结合使用上采样模块 (PAN) 进行自下而上的上采样用于捕捉小型障碍物的局部特征信息, 增加小型障碍物特征信息的提取能力。其次针对复杂环境下, 容易出现同类型障碍物多背景下的场景, 导致检测回归精度低、模型泛化能力低的问题。我们将采用改进后的检测头 (Decoupled Head) 模块, 将分类和回归任务单独进行, 并且结合新的 SIOU 边界框损失函数, 提高对多背景下障碍物的检测精度和模型的鲁棒性。经实验结果表明, 运用改进后的 YOLOv5 算法, 其障碍物检测结果 mAP 提高了 2.2%, Precision 提高了 0.03; Recall 提高了 0.025, 并在硬件层次的实际场景检测效果中也得到了显著改进。

**关键词:** 智能导盲杖、障碍物检测、特征信息、损失函数、YOLOv5

## Research on YOLOv5 Park Obstacle Detection Algorithm

Li Zhihong, Xie Na, Yu Jinming, Tan Tengfei

(School of Computer and Communication, Hunan University of Engineering, Xiangtan 411104)

**Abstract:** We have designed an intelligent guide cane based on deep learning object detection algorithm to address the current issue of home travel safety for visually impaired individuals. The innovation of this guide cane lies in the improved YOLOV5 algorithm as the core, combined with an ultrasonic ranging module, to achieve accurate detection of obstacles. Then, it broadcasts the user through voice and is equipped with precise functions such as Beidou positioning, which can safely and conveniently help visually impaired people travel. So what we are designing is something more convenient for users to use, not just a cane. It can not only tell you that there is an object blocking your way, but also tell you what that object is and help you bypass it. In addition, in terms of

obstacle detection, actual traversal surveys have found that small obstacles and multi background obstacles are particularly common in places such as parks and communities. Firstly, due to the varying types and sizes of obstacles in park communities, it is easy to encounter small or medium-sized obstacles. Therefore, we added a Channel Attention Mechanism (ECANet) and combined it with an Upsampling Module (PAN) for bottom-up upsampling to capture local feature information of small obstacles, enhancing the ability to extract feature information of small obstacles. Secondly, in complex environments, it is easy to encounter scenes with the same type of obstacles and multiple backgrounds, resulting in low detection and regression accuracy and low model generalization ability. We will use an improved Decoupled Head module to perform classification and regression tasks separately, and combine it with a new SIoU bounding box loss function to improve the detection accuracy of obstacles in multiple backgrounds and the robustness of the model. The experimental results show that using the improved YOLOv5 algorithm, the obstacle detection result mAP increased by 2.2% and the Precision increased by 0.03%; Recall has increased by 0.025 and has also been significantly improved in actual scene detection performance at the hardware level.

**Keywords:** Intelligent guide cane, obstacle detection, feature information, loss function, YOLOv5

## 0 引言

随着中国社会的发展,虽然居家出行已经越来越方便了,但是就近几年来看,我国盲人或者视力障碍的人也越来越多。而近年来伴随着公园、小区等一些休闲地方障碍物的增多以及物体车辆的乱摆乱放问题,所以视障人士面对居家出行有着很大的问题,同时也增加了视障人士居家出行的安全隐患,所以解决视障人士居家出行问题显得尤为重要。

早期的导盲杖是一种帮助盲人行走时使用的长杖,没有任何电子或智能化功能。2010年,日本冈安光博开发了一种新型导盲杖,这种导盲杖装有两个超声波感应器,够感应到前方2米脸部高度处的障碍物,并通过震动手柄通知使用者。2015年, Ayat A. Nada, Mahmoud A. Fakhre 等开发了基于红外传感器的导盲杖,此类传感器可以通过发射红外光线,然后接收反射回来的红外光来检测前方2米的目标,并且此传感器精度更高,可以检测较小的障碍物。2019年, Ihab A. Satam, Mokhaled NA Al-Hamadani 等开发了一种超声波,震动电机以及蜂鸣器相结合的智能拐杖,通过超声波检测前方的障碍物,如果检测到障碍物微控制器向执行器发送信号,蜂鸣器将会发出声音警告不要靠近。2021年,由斯坦福大学智能系统实验室研发一款基于激光雷达的智能导盲杖,这款盲杖装备有一个激光雷达,一个GPS定位系统,另外还有加速计、磁力计和陀螺仪,就像一款智能汽车的简化版一样,可以实时输出用户的位置、速度、方向等信息。盲杖的顶端还安装了一个电动的、360转向的轮子,可以时刻保持与地面的接触。这个轮子通过轻轻的推动就可以实现前后左右移动,引导视力受损的用户绕过障碍物。

Ross Girshick, Jeff Donahue 等提出了 R-CNN<sup>[1]</sup> (2014) 方法,通过对原图产生感兴趣的目标候选区,将产生的候选区域 resize 到一个固定大小,再将 resize 后的图像输入到一个 CNN 网络,最后将提取到的特征输入给预先训练好的一组 SVM 分类器,识别出区域中的目标物体。Fast R-CNN<sup>[2]</sup> (2015) : Fast R-CNN 删除了 R-CNN 中的 SVM 分类器和 bbox 线性回归器,而是将两者都放在一体化网络中,通过将整张图像进行过 SS 操作后,直接输入进 CNN 层,一次性对所有的分类和回归。所以在训练速度上比 R-CNN 快了 10 倍左右。随后提出的 Faster R-CNN<sup>[3]</sup> (2015) 虽然进一步提高了 R-CNN 系列的检测精度和速度,但针对于目标障碍物检测并不能做到实时性检测。所以 Redmon 提出了单阶段目标检测算法 YOLO。YOLO 算法从 YOLOv1<sup>[4]</sup>开始就引起了广泛的关注和应用,不仅被应用在人物、车辆、建筑等场景的检测上,而且还在物体跟踪、视频监控、自动驾驶、安防等领域得到了广泛应用。同时, YOLO 算法也不断地在精度、速度、准确度等指标上得到改进,并且随

随着 YOLO 系列不断的更新, 目前已成为最热门的检测算法。YOLOv1 (2016) : YOLOv1 是最早的 YOLO 算法版本, 它通过将图像分成多个网格单元, 并在每个单元格中进行物体检测, 从而快速有效地提取图像中的物体信息。但由于 YOLOv1 仅仅在全局上单独优化了检测损失, 因此在物体小尺度物体检测、物体形状复杂等方面表现并不理想。YOLOv2<sup>[5]</sup>(2017): YOLOv2 在 YOLOv1 的基础上进行了多项改进, 包括使用 Darknet-19 替代了原来的 VGG<sup>[6]</sup> 卷积网络、引入 Anchor boxes (只在最大 mAP 时根据先验框数量进行调整)、使用 BN、卷积层修改等等。YOLOv2 可以在保持高精度的同时, 运行速度也得到了大幅提升, 为现实应用提供了更高效的检测方案。YOLOv3<sup>[7]</sup> (2018) : YOLOv3 中采取了一系列技术创新措施, 如应用多尺度预测、应用残差块、加入 FPN 技术等, 从而有效地克服了 YOLOv2 中出现的许多问题, 即在保持高速率时, 大大提高了测量准确度。另外, YOLOv3 在一定意义上还采用了 SSD<sup>[8]</sup> (2016) 算法的思路, 该网络结合了具有不同分辨率的多个特征图的预测, 以自然地处理各种尺寸的对象。但和 SSD 不同的是, YOLOv3 通过线上采样+数据融合的方法, 对多尺度数据特征实现了一定的整合。YOLOv4<sup>[9]</sup> (2020) : YOLOv4 采用更强的骨干网络, 使用 CBN<sup>[10]</sup>, SAM<sup>[11]</sup>, 近似最近邻 (PAN)<sup>[12]</sup> 等等新方法, 有效提高了检测精度。此外, YOLOv4 在训练过程中使用了各种数据增强技巧进一步提高了检测效果, 同时还提高了运行速度。伴随着深度学习算法的迅速发展, 出现了越来越多的目标检测算法。

相较于传统目标检测算法, 考虑到公园中的障碍物种类繁多, 形状大小以及障碍物背景不一的特点, 会导致目标检测算法可能会难以准确检测到它们。而过去传统的一些目标检测算法要么过于复杂运行速度慢, 或者精确度不够导致对于这种障碍物的检测不能及时反馈, 而加入改进后的 YOLOv5<sup>[13]</sup>障碍物检测算法, 拟定用更快更精确的 YOLOv5 算法, 在此基础上进行了一定的优化和改进。YOLOv5 算法虽然本身性能不错, 但由于目标障碍物种类繁多, 形状大小不一等不确定性因素, 导致 YOLOv5 算法不能精确的提取障碍物特征, 所以为了进一步提升, 我们拟在 YOLOv5 算法的基础上, 加入 ECA<sup>[14]</sup>注意力机制, 进一步提高检测精度和减少模型参数计算量。针对于公园、小区中出现多背景变化的障碍物, 但是针对于目前的目标检测算法, 难以到达对这类障碍物的特征提取。所以针对此问题, 所以我们决定将 Coupled Head<sup>[15]</sup>进行改进并结合新的 SIoU<sup>[16]</sup>边界框损失函数, 加强对多背景下目标检测效果, 增强模型的泛化能力。

# 1 智能导盲杖设计

## 1.1 功能设计

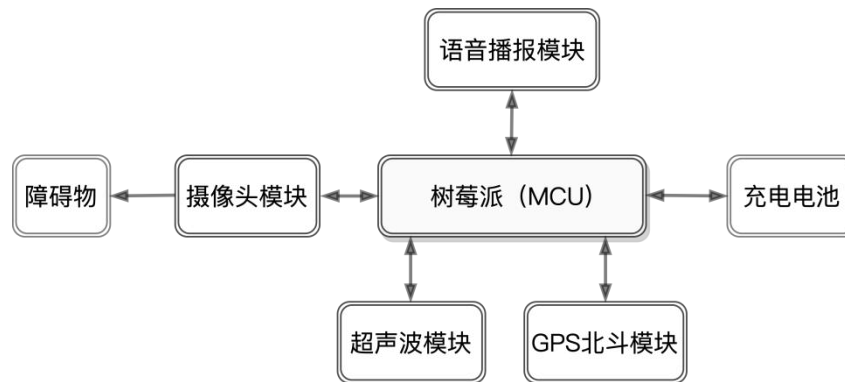


图 1：功能模块图

如上图所示，本导盲杖主要以树莓派 4B 为微控制器 (MCU)，在提供电源的情况下，我们设计的智能导盲杖主要功能是通过超声波模块结合改进后的 YOLOv5 目标检测算法实现对障碍物的精准检测。若公园、小区等场景中存在障碍物，本导盲杖先通过超声波测距，判断前方障碍物的距离，如果障碍物的距离小于设定的安全阈值距离，再调用目标检测算法对障碍物进行检测识别，若在检测的范围内同时存在多个障碍物，则蜂鸣器语音播报优先将距离最近的障碍物信息种类反馈给用户，提醒用户的注意，防止意外的发生，提高视障人士安全保障。此外，我们也在导盲杖上加入了 GPS 北斗双定位模块，当然此模块功能需要用户绑定手机才能使用，我们设计此模块主要是为了解决视障人士出行迷路失踪等问题，通过此模块可以让亲属快速精准的定位导盲杖的位置信息，便于亲属的监管，进一步保证视障人士出行的安全。

## 1.2 硬件设计

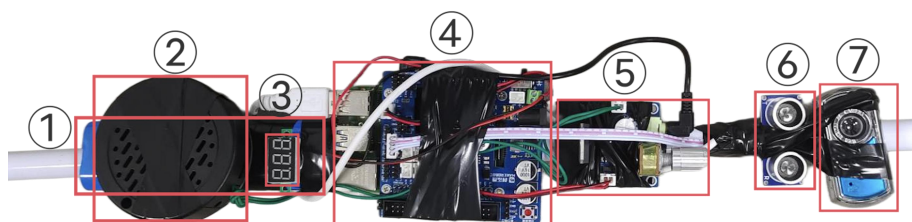


图 2：实物图

通过前述的功能模块设计，为了更好的契合智能导盲杖的功能设计，我们筛选出各个功能相应的硬件，完成了如上图 2 所示的实物制作。

首先我们通过普通的金属导盲杖为主体，初步实现了每个模块之间相应的功能。如上图所示充电电池（图标①）位于导盲杖握手端，对于语音播报功能，我们使用半径 3.5cm、高 2cm 的圆形蜂鸣器（图标②）通过连接树莓派，负责播报检测出的障碍物种类和距离。并且在电池下方，我们设计了电量显示器（图标③），方便对电量的实时监控。此外在导盲杖主体部分，我们以转接板和树莓派（图标④）为整个系统的主控中心。同时连接下方泛用性较强的 HY-SRF05 超声波传感器（图标⑥）以及 GPS 北斗定位模块（图标⑤）最后树莓派通过 USB 接口连接靠近导盲杖杖尖的摄像头（图标⑦）。初步实现了智能导盲杖的整体的硬件设计。

### 1.2.1 微处理器

为配合超声波模块实现障碍物检测的基本功能，我们将改进后的算法部署在相应的硬件开发板，以应用于实际的公园、小区等场景。我们以转接板和 Raspberry Pi（树莓派）为算法的核心部署环境，在性能上占有绝对优势，而 Raspberry Pi 是一款的微型电脑控制器其主板树莓派的尺寸只有一张信用卡大小的一个小型电脑，能够解决使用过程中图像传输及图像处理等相关问题。我们选择的型号为 Raspberry Pi 4B，树莓派 4B 的处理器有四个 ARM Cortex-A72 核心；采用 4GB 的 LPDDR4 SDRAM 作为系统内存；然后通过树莓派中的 USB 接口，完成连接树莓派摄像头，其视频输出为 2 micro HDMI ports 和 2-lane MIPI DSI display port，能够更好的部署我们改进后的模型，如下图所示。

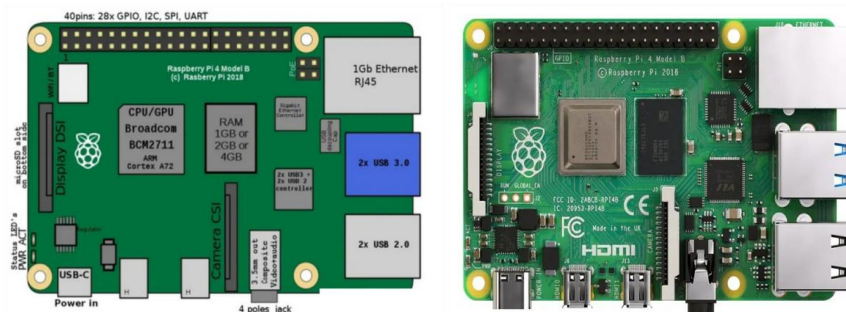


图 3：树莓派图

在算法部署阶段，为支撑算法在树莓派上的运行，我们选择 raspbian-buster-lite (4B 轻量级不带图形界面)镜像系统，并使用 minconda 软件配置 YOLOv5 模型运行的虚拟环境，然后将本地训练后的 YOLOv5 权重文件以及原模型配置到树莓派系统中，最后通过电脑 SSH 远程连接树莓派，控制模型在树莓派系统中的运行，完成改进后 YOLOv5 算法的部署，



实现障碍物检测的基本功能。

### 1.2.2 障碍物检测模块

本文设计的基于深度学习的智能导盲杖与传统的导盲杖最大的区别在于，我们在保证导盲杖基本的功能之外，加入了改进后的 YOLOv5 障碍物检测功能，通过将改进后的模型文件烧录至树莓派中，然后通过树莓派的 USB 接口连接摄像头，视频输出为 2 micro HDMI ports 和 2-lane MIPI DSI display port，实现改进后的 YOLOv5 模型在树莓派层面的部署以及视频的实时检测。当然在第二部分，我们将具体介绍 YOLOv5 如何改进。

此外在检测的同时，还需搭配超声波以及语言播报模块，在障碍物进入到超声波检测的阈值范围内，再控制树莓派打开摄像头实现障碍物的检测，最后结合蜂鸣器将障碍物的种类和距离播报出来。

### 1.2.2 超声波测距模块

超声波测距模块选择 HY-SRF05，用于测量视障人士与被检测障碍物之间的距离。模块具有 2 cm ~ 450 cm 的非接触式距离感测功能，测距精度可达高到 3 mm。如下所示。

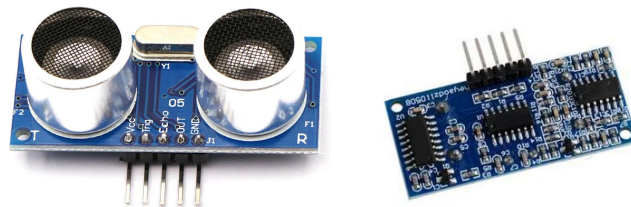


图 4：超声波实物图

模块包括超声波发射器、接收器与控制电路。若前方存在障碍物，首先采用 IO 口 Trig 触发测距，给予至少  $10\mu s$  的高电平信号；并发送 8 个 40 kHz 的方波，自动检测是否有信号返回；如果有信号返回时，通过 IO 口 Echo 输出一个高电平，高电平持续的时间即为超声波从发射到返回的时间 T，若测试距离设为 S，声速为 340 m/s，其公式为：

$$S = T * 340 / 2 \quad (1)$$

### 1.2.3 电源模块

在电池选择方面，考虑到智能导盲杖持久续航以及长期使用的情况，本智能导盲杖采

用 9 V 的可充电锂离子电池供电，相较于其他电池锂离子电池本身不仅具有较高的能量密度，能够储存更多的能量，而且能够循环充放电数百次甚至数千次，使其具有更长的使用寿命。并同时通过电压转换模块把 9 V 电压转换成各模块和传感器需要的 5.0V 和 3.3V 电压，为各个模块提供不同的电压电源。

#### 1.2.4 GPS 北斗模块

在定位方面我们可以选择 GPS 北斗双模定位模块，使用 ATGM332D，型号是 5N3X，是一款自带 GPS 和北斗定位的双模定位模块，如下图所示。ATGM332D-5N 系列模块是 12 × 16 尺寸的高性能 BDS/GNSS 全星座定位导航模块系列的总称。支持多种卫星导航系统，包括中国的 BDS（北斗卫星导航系统）、美国的 GPS 等。包含 32 个跟踪通道，可以同时接收六个卫星导航系统的 GNSS 信号，并且实现联合定位、导航与授时。相较于传统的 GPS 模块，此模块更能精准的定位导盲杖的位置信息。



图 5：实物图

#### 1.2.5 语音播报模块

在语音播报方面，由于硬件性能的局限性以及导盲杖的便捷性，我们使用半径 3.5cm、高 2cm 的圆形蜂鸣器与连接树莓派，通过使用百度 AI 开放平台提供的 API(Application Programming Interface) 应用程序编程接口，对检测出的障碍物的类别和距离，从文本文字转换成语音，最后以蜂鸣器为载体播报出来。

### 1.3 功能对比

与此同时，为了更好的体现出该智能导盲杖功能丰富与创新，在查阅目前国内智能导盲杖发展现状的相关资料后，我们会与国内流行的导盲杖作对比，如已上市的普通金属导



盲杖和流行的科峰 MTN-GC1D 型号的智能导盲杖以及未上市研究阶段的智能导盲杖作对比，得到了如下表所示结果。首先横向分析数据表，金属导盲杖的功能最为简单，用户使用完全依靠自己的判断能力；再纵向对比数据表后，可以明显发现无论已上市还是未上市的导盲杖，都没有实现障碍物的检测功能，我们将前沿性的提出将深度学习目标检测算法运用在智能导盲杖上，主要用于帮助实现对障碍物种类信息的检测。在 MCU 选择上使用性能更好的树莓派，而一般的智能导盲杖使用的是 STM32 开发板作为微控制器。然后在障碍物测距方面，相较于 HC-SR04 超声波和 HN-C90 超声波模块，我们使用的是 HY-SRF05 超声波测距模块，此模块的测量范围通常更广，精度更高，而 HCSR04 测量范围短和测量精度较低。在播报方面，我们使用蜂鸣器连接树莓派，通过使用百度 AI 开放平台提供的 API，对检测出的障碍物信息文字转换成语音，最后以蜂鸣器为载体播报出来。在定位方面，通过使用 GPS+北斗（ATGM332D）模块，具有高灵敏度、低功耗、低成本等优势，相较于传统的 GPS 模块，此模块更能精准的捕获导盲杖的位置信息。摄像头采用便捷的 USB 摄像头，而电池选择常用的 9V 可充放锂离子电子。

表 1 功能对比表

导盲杖	MCU	检测	测距	播报	定位	摄像头	电池
金属导盲杖	×	×	×	×	×	×	×
科峰 MTN-GC1D	×	×	HN-C90 超声波	×	×	×	锂电池
基于 STM32 的智能导盲 杖的设计	STM32 F103	×	HY-SRF05 超声波	XFS3031C NP-EVB 语音合成 模块	×	×	充电电 池
基于 STM32 芯片的智能 导盲杖设计	STM32 F407Z GT6	×	HC-SR04 超声波	ASR-M09 C 语音模 块	GPS (E29V-G)	×	锂离子 电池
Ours	Raspher Pi 4B	改进的 YOLOv5	HY-SRF05 超声波	有源蜂鸣 器	ATGM332D GPS+北斗	USB 摄 像头	锂离子 电池

## 2 YOLOv5 算法的结构及改进

### 2.1 使用 Mosaic 数据增强方法

采取 Mosaic<sup>[17]</sup>数据增强的方法，该数据增强是通过将含有四张目标障碍物的图片，通过裁剪、拼接的方式组成一张新的图片，如图 a 随机抽取四张数据集样图，然后将四张图片进行随机裁剪后拼接得到图 b，最后画出各自的标签框，如图 1 所示。

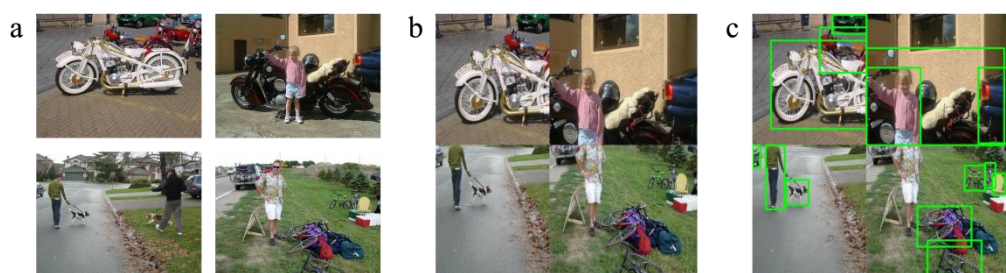


图 6: Mosaic 数据增强

当然超出分割线的部分，我们会进行边缘处理。所以该方法能够一次实现对四张不同障碍物的检测，运用该方法可以加强目标障碍物的数据集，可以让网络结构的鲁棒性更强。与此同时也可以增加处理照片的数量，减少 GPU 等硬件设施的负担。

### 2.2 加入 ECA 注意力机制

在公园小区等场景下，其障碍物大多数背景复杂，其他干扰信息较为复杂。为避免模型检测受到干扰以及出现过拟合问题，所以在 YOLOv5 模型基础上，我们引入 ECANet 注意力机制，它能够自适应地调整通道特征的权重，使得网络能够更好地关注重要障碍物的特征，抑制不重要的特征。这种机制有助于提升特征的判别能力，从而提升了网络的性能。

ECA 注意力机制的核心思想是在卷积操作中引入通道注意力机制，以捕捉不同通道之间的关系，如下图所示，通过执行卷积核大小为  $n$  的一维卷积来生成通道权重，并且每一

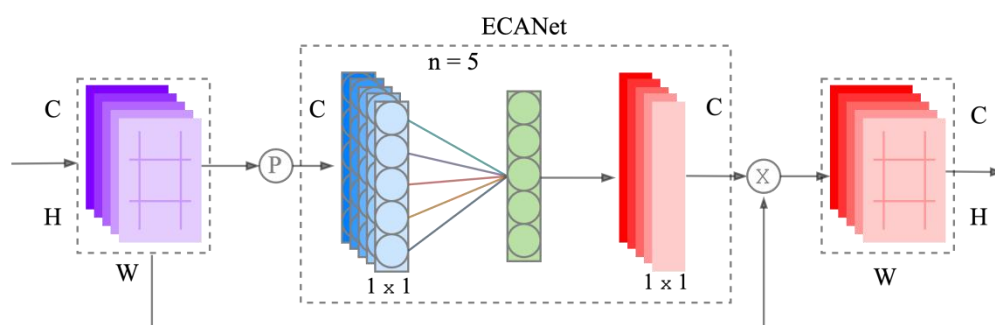


图 7: ECANet 注意力机制

次卷积过程只起部分通道的作用，即实现了适当的跨通道交互而不是像全连接层一样全通道交互。其中  $n$  值通过通道数  $C$  自适应地确定，卷积核自适应公式为：

$$n = \left\lceil \frac{\log_2^C + b}{\gamma} \right\rceil_{\text{odd}} \quad (2)$$

有效的减少了参数的计算量，其中  $C$  表示通道数， $\text{lodd}$  表示卷积核大小只能取奇数。在效果上有效的减少了 YOLOv5 模型受干扰信息的影响，其改进模型相较于原 YOLOv5 模型提升了网络模型对障碍物的检测精度，对于含有部分特征信息以及关键信息的障碍物目标，能较为准确地检测出来。

将改进后的模型经过 200 次训练迭代，其结果表明加入 ECA 注意力机制节省了整个模型计算量的时间和减少模型过拟合，加速了模型网络的收敛速度，增强了模型的训练和推理效率，所以其模型整体性能相对较出色，从而证明了本文中提出的改进后的 YOLOv5 模型性能的优越性。

## 2.3 FPN 结合 PAN 设计

FPN<sup>[19]</sup>是一个自上而下的特征金字塔，可以提高模型在多尺度检测任务上的精度，同时还可以在不影响检测速度的情况下提高检测速度。但是对于公园场景下的障碍物检测，虽然增强了障碍物的语义信息，但减少了定位信息的传递，如下图所示。

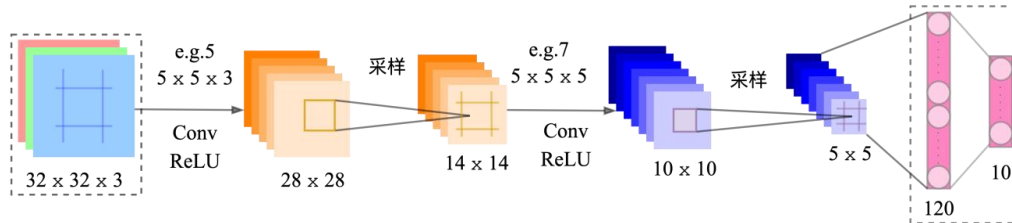


图 8: 网络层次图

随着网络深度的加深，降采样操作的增多，细节信息不断被破坏，致使小物体的检测效果逐渐变差。而大目标由于像素较多，随着层数变多，网络的感受野逐渐增大，网络对大目标的认识越来越充分，所以在网络结构加深的过程中，对于公园中小型障碍物，容易存在特征梯度消失的问题，不能实现对小型障碍物的特征信息的精确提取。这也就导致在对目标障碍物进行特征提取时，语义信息越来越多，感受野也越来越大，而局部的信息特征越来越小，这就会存在对小型障碍物检测精度偏低的问题。所以我们在 FPN 中加入 PAN 结构，下图 4 所示。

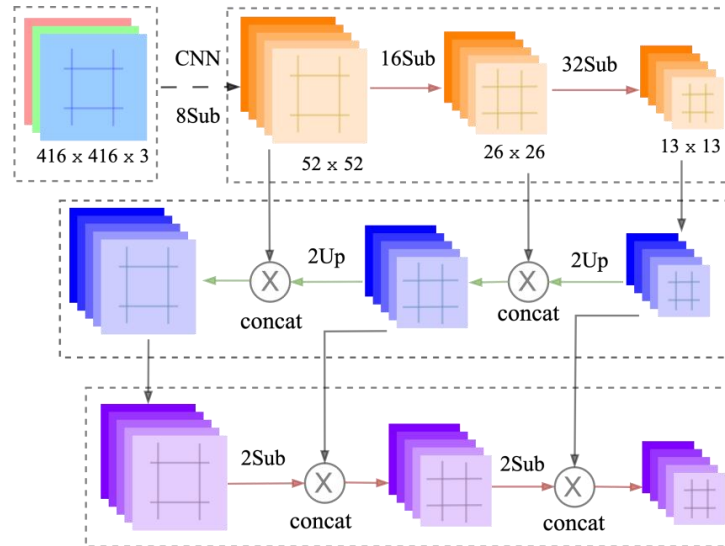


图 9: FPN+PAN 设计

PAN 网络通过自顶向下的路径聚合和自底向上的像素聚合来整合不同层级的特征，从而在保留细节信息的同时拥有更大的感受野。可以防止局部定位信息的减少并对该信息进行补充，所以从而加强网络特征融合的能力，提升特征的多样性。

深层的 feature map 携带有更强的语义特征，较弱的定位信息，而浅层的 feature map 携带有较强的位置信息，和较弱的语义特征。FPN 就是把深层的语义特征传到浅层，从而增强多个尺度上的语义表达。而 PAN 则相反把浅层的定位信息传导到深层，增强多个尺度上的定位能力。从而实现加强网络对于公园障碍物中小目标物体特征信息的提取能力，提升网络对于小型目标障碍物的检测精度。

## 2.4 改进 head 层面设计

### 2.4.1 采用改进后的 Coupled Head 检测头网络结构

在传统的 YOLOv5 的检测 head 中，通过将目标进行分类和回归分支融合共享的方式来实现障碍物的精准检测，但采用 coupled Head 将分类和回归任务放在一起，这两个任务却存在冲突性的问题。所以针对于检测任务的冲突性，我们采用改进的 Decoupled Head，如图 5 所示。

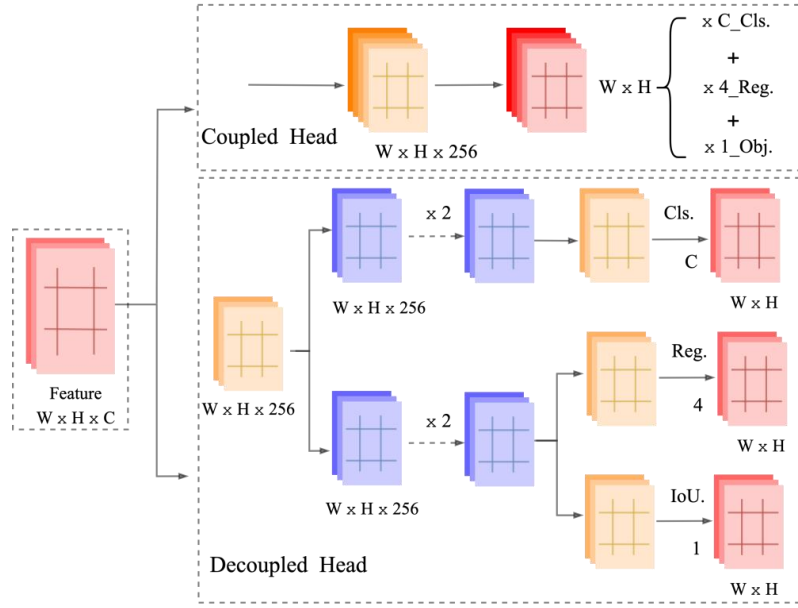


图 10: Decoupled Head 设计

通过将分类任务和回归任务通过不同的网络分支学习再融合在一起，最终得到与传统 Coupled Head 一样的 shape 网络输出。通过实验发现采用改进后 Head 层设计后，针对于多背景下的同类型障碍物检测，其模型不只是增强了模型的泛化能力，同时在精度上也得到了提高，所以使用 Decoupled Head 表达能力更好。

## 2.4.2 采用改进后的 Bounding box 边界框损失函数

在 YOLOv5 算法中采用的 Bounding box 损失函数为 CIoU<sup>[20]</sup>损失函数，它是一种用于测量预测边界框与真实边界框之间重叠程度的指标。在障碍物检测中，我们通常将 IoU<sup>[21]</sup>损失函数用来评估预测框和真实框的重叠情况，以确定预测框是否正确。其具体使用 CIoU 损失函数，公式如下。

$$L_{CIoU} = 1 - IoU + \frac{P^2(b, b^{gt})}{C^2} + \alpha v \quad (3)$$

其中 $\alpha$ 是权重参数，它的表达式为：

$$\alpha = \frac{v}{(1 - IoU) + v} \quad (4)$$

其中 CIoU 损失函数包含影响因子 $v$ ：

$$v = \frac{4}{\pi^2} \left( \arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2 \quad (5)$$

但针对于公园、小区等障碍物的场景下，因为其场景障碍物错综复杂以及障碍物自身形状不一的特点，所以算法检测模型的精准度就显格外重要。所以在传统的 YOLOv5 使用 CIoU 边界损失函数的基础上，我们使用了改进后的 YOLOv5 障碍物检测算法，采用检测精

度更高的 SIOU 边界框损失函数，此损失函数区别于传统的真实框与预测框的重叠程度、中心点距离、纵横比等因素的 IoU 计算，除外新增加了真实边界框与预测边界框之间中心角向量的计算，这样更有效降低了回归的自由度，减少了回归损失计算中的抖动误差，进一步加快网络收敛，从而提升了模型的检测精度和检测障碍物的识别速度，其 SIOU 损失函数的公式为。

$$L_{SIOU} = 1 - IoU + \frac{\alpha + \beta}{2} \quad (6)$$

其 SIOU 损失函数中的 Dis\_loss（距离损失）和 Sha\_loss（形状损失）为：

$$\alpha = \sum_{t=x,y} (1 - e^{-\nu \rho t}) \quad (7)$$

$$\beta = \sum_{t=w,h} (1 - e^{\omega t})^\theta \quad (8)$$

并且在后续模型对比实验训练中，运用改进后的 IoU 边界框损失函数，模型的类别损失和回归框损失等都得到了下降。

### 3 实验结果与分析

在实验设计阶段，为了完成本项目算法模型以及相关检测模型的训练，其实验设计采用了 anaconda3 环境与 pytorch 深度学习框架以及 Cuda11.8 版本，在硬件方面，其 CPU 大小为 15 核，GPU 为一张 NVIDIA 3090 显卡 配置 32GB 运行显存，1TB 存储硬盘。

#### 3.1 障碍物数据集选取

在数据采集方面，由于公园小区场景下的障碍物种类繁多并且形状不一，在加上适合于深度学习的障碍物数据集较多但场景较少，所以我们采取线下实体拍摄再结合 VOC<sup>[22]</sup>现有数据集的方式，构成我们项目所需的数据集。在线下采样时，通过实际采样我们发现公园小区常见的中大型障碍物主要分为：汽车、垃圾桶、摩托车、人等；此外也包含一些常见的小型障碍物：路灯、猫、狗等。针对于以上的数据情况，我们遍历了本市 10 个知名公园小区，共采集了该场景下 100 个视频段，并按照一定的帧率抽取剪切视频。同时我们也





图 11: 数据集样图

结合了部分 VOC 数据集来扩充实验数据。以达到训练模型的目的,VOC 数据集是一个广泛使用的计算机视觉数据集, 用于目标检测、分割和图像标注等任务。它包含超过 2 万张图片, 涵盖 20 个常见的对象类别, 包括人, 动物, 车辆, 家具等。每个图像都有与之关联的边界框和对象实例的分割掩码。此外,每个图像还有多个人工标注的描述, 用于图像标注任务。

针对于公园障碍物检测的场景, 我们从 VOC 数据集中提取出公园常见的七个类别, 其包括: 'person', 'chair', 'car', 'bicycle', 'motorcycle', 'cat', 'dog'等七个主要的类别, 并结合自己采集的图片数据集样本, 最终得到人、椅子、汽车、自行车、摩托车、猫、狗、消防栓、垃圾桶九个类别的障碍物, 其数量大约有 1 万张障碍物的图片。

### 3.2 模型对比和评估

我们为了验证改进后 yolov5 检测模型的真实性和准确性, 通过与传统的 YOLOv5 模型进行对照实验来实现, 其实验数据集选用同一份公园障碍物数据集, 其划分比例 train:val:test=8:1:1。

在模型训练时, 其训练次数 epochs 为 200 次、图片批次 batch 为 32 张、图片大小 size 为 640、模型权重为预训练 VOC 数据集的初始权重值, 为了实现两种模型实验结果的可视化和模型的对比性, 我们采用常见的模型训练评价指标, 其查准率 (Precision) 是指在所有被预测为正的样本中实际为正的样本的概率, 查全率 (Recall) 表示实际阳性样本中被预测为阳性样本的概率, 而 F1 参数本质是对两个率进行调和平均运算, 它被定义为查准率和查全率的调和平均数。mAP (平均精度均值) 代表每个障碍物类别下 P-R 曲线面积的平均值, 可以反应出模型在不同类别的目标障碍物检测中的准确性和鲁棒性, 其中  $AP_i$  表示第  $i$  个类别障碍物的 PR 曲线面积, 其各项指标计算公式为:

$$\text{Pre} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (9)$$

$$\text{Re} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (10)$$

$$\text{F1} = \frac{2\text{TP}}{2\text{TP} + \text{FN} + \text{FP}} \quad (11)$$

$$\text{mAP} = \frac{\sum_{i=1}^C \text{AP}_i}{C} \quad (12)$$

在检测公园障碍物的情况下，我们不只是单方面的考虑查准率和查全率，所以通过另一种以当前查全率、查准率为横轴、纵轴作图的方式就可以得到查准率和查全率曲线（P\_R 曲线），其 P\_R 图可以更加直观地反映出模型在障碍物数据集总体上的性能好坏。而对于两种检测模型，其判定指标可以通过比较模型分别对应的 P\_R 曲线下面积的大小以及每条曲线对应的平衡点（BEP）进行判断。

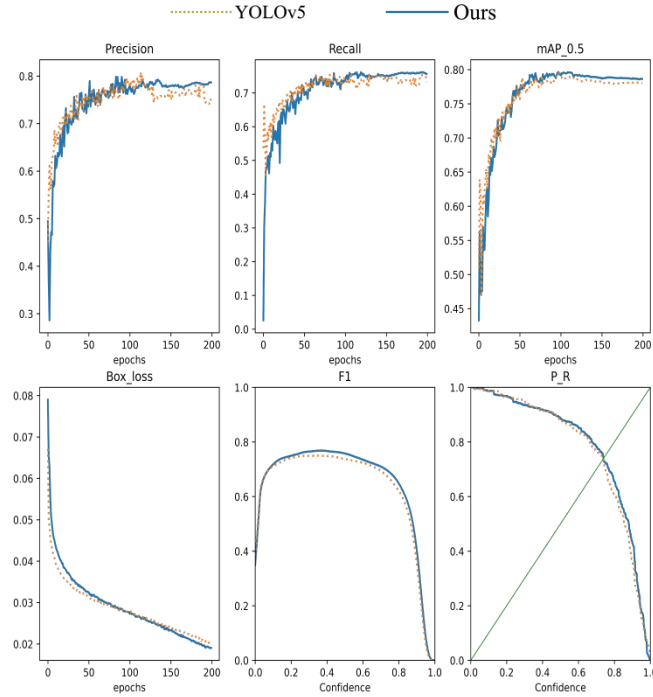


图 12: 模型指标对比图

所以对于两模型的评价，由上图 7 可以看出在各个指标上，改进后的 YOLOv5 模型相较于原 YOLOv5 模型都得到了不少的提升，在经过 200 轮模型训练后，其 Precision 为 0.788，改进前为 0.758，比改进前上升了 0.03；而 Recall 为 0.756，改进前为 0.731，比改进前的模型提升了 0.025；mAP\_0.5 为 0.793，提升了 2.2%；其 Box\_loss 为 0.019，比改进前下降了 0.002。

与此同时我们也单独对两类模型进行了数据集验证，通过调用两个模型在同一份公园障碍物数据集训练出来的模型权重，然后对同一份验证集进行模型的验证评估，便得到了

相应的评估指标 F1 曲线图、P\_R 曲线图。在图中可以看出改进后的 YOLOv5 模型在各项验证集指标上面都得到了提升，其 F1 曲线的面积相比较改进前的 YOLOv5 得到了提升；而对于 P\_R 曲线图，我们也可以发现改进后的 YOLOv5 模型在公园数据集上其曲线平衡点大于改进前的模型，可以判断改进后的模型性能更好。通过 F1 曲线图和 P\_R 曲线图的对比实验结果，可以看出在同一份障碍物数据集中与改进前的 YOLOv5 算法相比，本文中提出的改进后的 YOLOv5 算法具有较高的检测精度。

在与传统的 Faster-RCNN 和 Mask-RCNN 等双阶段目标检测模型相比，YOLOv5 模型是更加轻量级模型。为了更好地展示出我们改进后的模型性能，我们将每处的改进与改进前的 YOLOv5 模型在各个评估指标上进行对比分析, 在原模型上改用 SIoU (YOLOv5+SIoU)、改用 Decoupled\_Head (YOLOv5+Decoupled \_Head)、以及加入 ECA 注意力机制 (YOLOv5+ECA) 得到了如下的数据对比表 1。在模型检测损失指标中，定位损失 Box\_loss 用于计算预测框与真实框之间的误差值。对比 YOLOv5+SIoU 与原模型可以发现，采用改进的 SIoU 损失函数，经过 200 次训练后，Box\_loss 下降的更快，说明可快速有效地定位图像中障碍物的真

表 2 模型数据对比表

models	参数量 MB	Box_loss	Pre	F1	mAP%
YOLOv5	14.5	0.021	0.758	0.75	77.1
+SIoU	14.5	0.018	0.799	0.76	78.1
+ECA	14.5	0.020	0.758	0.75	78.8
+Decoupled_Head	29.2	0.019	0.801	0.75	77.8
Ours	29.2	0.019	0.788	0.78	79.3

值区域，从而验证了采用精确度更高的 SIoU 损失函数能够帮助模型提高查准率以及准确定位图像中的真实区域；对比 YOLOv5+ECA 模型，其 mAP 指标得到了提升，说明了加入 ECA 注意力机制后，网络模型能够增强对各类关键性障碍物的特征提取能力，更好地增强了模型的整体性能；对比 Head 层改进，我们重新构建了检测头，通过将网络中障碍物目标位置和障碍物类别信息分别提取出来，然后进行不同的分支学习再融合，实现了对障碍物种类准确分类的任务。保证 mAP 的前提下，模型的检测的查准率得到了大幅度提升。

最后对原模型，我们改进后的 YOLOv5 模型在各项性能指标上取得了大幅度的改进，

其关键的 mAP 指标对比原模型提升了 2.2%，说明改进后的 YOLOv5 在各类障碍物中都可以准确有效的提高检测精度，提高了模型检测的泛化能力。而且 Precision、F1 等指标对比分析，改进后的 YOLOv5 对于正样本的精度以及单个类别的 F1 曲线较改进前都得到了较大的提升，更好地验证模型的准确性和鲁棒性。

综合数据表格以及各类指标曲线图，都说明改进后的 YOLOv5 模型在各个指标上都得到了不错的提升，综合提升了模型对公园小型障碍物以及多背景障碍物的检测的泛化能力和精确度，从而训练出一个更加适用于检测公园障碍物的目标检测模型。

除此之外为了更加直观的展现改进后的模型检测精度，我们从公园数据集中随机抽取部分样本用于对模型的测试验证。其中 a、b、c、d 图片为改进前的测试图；A、B、C、D 图片为模型改进后的测试图。所以从对比图片中可以看出，改进后的 YOLOv5 模型具有最

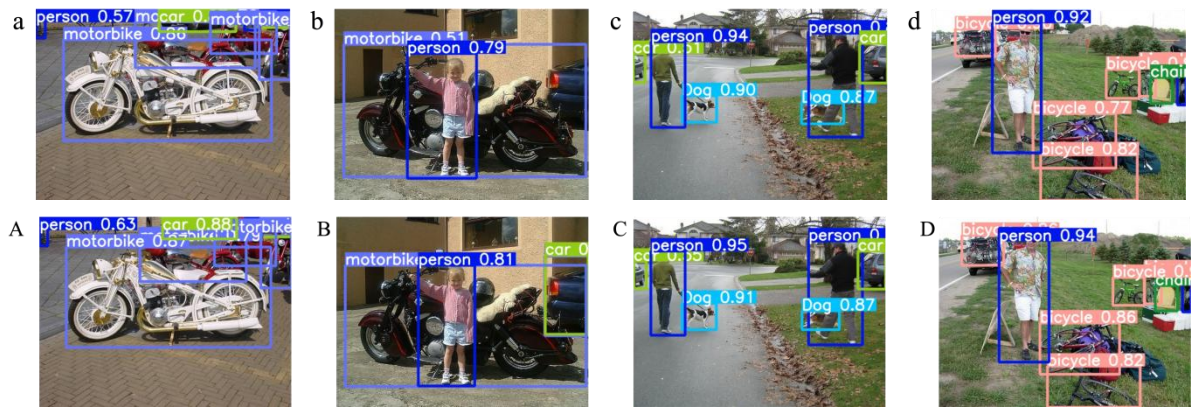


图 13：模型测试对比图





佳的检测效果，例如 b 与 B 图对比，我们发现改进前的 YOLOv5 具有缺失的检测，而改进的 YOLOv5 中加入了 ECA 注意力机制，能检测到图像边缘局部的关键障碍物。在公园中容易出现更多背景不一的同一类障碍物的情况，而采用传统的 Coupled Head 容易导致模型出现过拟合现象，所以针对于实际不同背景下的障碍物，就存在着障碍物检测精度偏低的问题。而采用改进后的 Decoupled Head 设计，通过将障碍物的位置和类别信息分别提取出来，进行不同的网络分支分别学习，可以有效的增强模型的泛化能力和鲁棒性，除此之外结合检测精度更高的 SIoU 边界框损失函数，此损失函数在模型的回归任务中，重新定义了计算损失值的任务，改进前的 YOLOv5 模型边界框之间损失较大，从而也导致了模型收敛较慢且效率较低，通过对比图可以发现，改进后的 YOLOv5 具有更加真实精准的检测，对于复杂背景下的公园障碍物有更真实的预测，更加适用于检测实际场景下的公园障碍物。






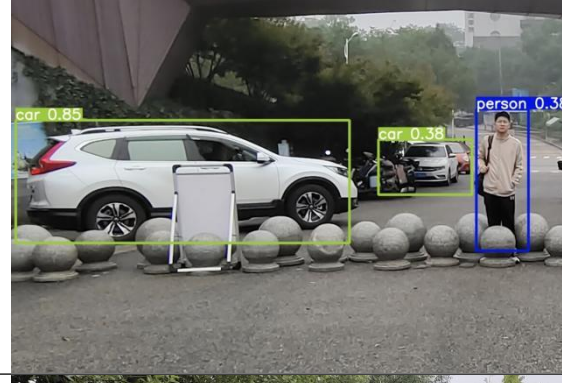
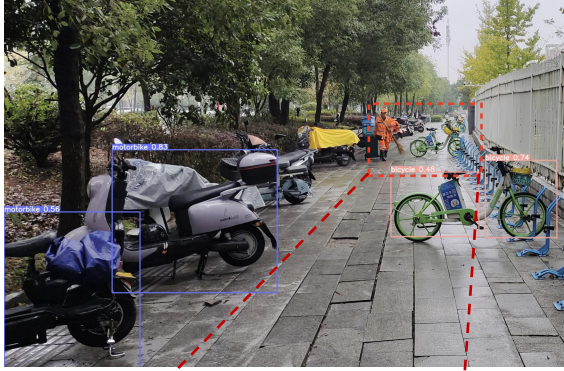

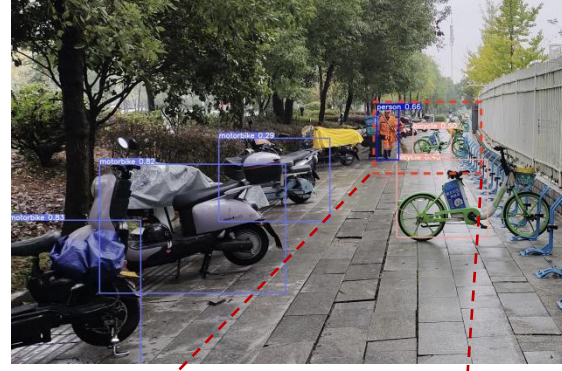

4 导盲杖场景应用

通过 SSH 远程连接将改进后的 YOLOv5 和改进前的 YOLOv5 模型部署到树莓派中，并通过 USB 摄像头将拍摄的画面实时传回本地电脑，完成智能导盲杖实际检测的准备工作。然后我们检测了公园、小区在等实际场景下，选取多种不同背景（Scene）下的障碍物，然后以超声波检测的距离为变量，其中 S\_\*代表单一背景下，最近障碍物到导盲杖距离为‘\*’；而 C\_\*代表复杂背景下，最近障碍物到导盲杖距离为‘\*’，误差为±1m。YOLOv5 代表原模型，Ours 代表改进后的模型，最终得到了下表所示结果。以距离为变量用于验证模型对于远距离小型障碍物的检测。然后通过变换目标障碍物的背景信息，验证改进后的模型对于多背景干扰信息下障碍物检测的泛化能力。

表 3 模型检测对比表

Scene	YOLOv5	Ours
S_3 (± 1m)		
S_5 (± 1m)		



C_3 (± 1m)		
C_5 (± 1m)		
--	 	 

在上表的对比实验中，首先在目标障碍物较简单情况下，图片信息较为少的情况下，我们可以发现在 3m 和 5m 处的图片，对于图片中待检测的目标障碍物而言，改进前的 YOLOv5 模型对其正确分类的障碍物检测精度较低，同时对其错误分类的障碍物检测精度较高；反观我们改进后的 YOLOv5 模型而言，基本能实现障碍物的正确分类并且其识别的精准度较高。然后在障碍物背景较为复杂情况下，我们改进后的 YOLOv5 模型能更好的提取远距离小型目标障碍物的特征信息，有效的提高了原模型对于复杂背景障碍物的检测能



力。然后对比最后一组实验图，我们可以更直观地发现对于近距离特征信息明显的障碍物，两模型检测精度几乎相似，但是对于图像边缘障碍物原模型检测精度较低。而且对于远距离的障碍物，我们改进后的模型能准确且精度较高的检测出来，所以有效解决了原模型对于远距离小目标障碍物存在漏检、误检的问题。

所以对于小型障碍物检测以及多背景复杂场景下，我们在原 YOLOv5 模型基础上，改进 Decoupled Head 检测头，同时结合 SIoU 损失函数使检测框更接近障碍物真实的边界，提高模型对复杂背景障碍物的回归分类能力，通过增加通 ECANet 注意力机制，增加小型障碍物以及边缘障碍物特征信息的提取能力。总的有效提升了模型的检测精度和泛化能力。

## 4 结束语

本文对于解决目前视障人士所存在的出行问题，设计了一款新型智能导盲杖，通过以改进后的 YOLOV5 算法为核心，在超声波测距为辅助的条件下，实现对障碍物的准确检测。最后通过语音播报的方式反馈给用户，能够安全方便的帮助视障人士出行。所以我们设计的不仅仅是一个简单的拐杖，它不但能让盲人知道前方存在障碍物，而且还能告诉你那个障碍物是什么，便于视障人士绕过它。此外为提高算法对障碍物检测的精确性，以及对于传统 YOLOv5 目标检测算法在小型等多形状以及多背景场景下障碍物存在检测精度低甚至误检的问题，我们在传统 YOLOv5 模型的基础上，加入了 ECANet 注意力机制，提高了模型对局部重要障碍物特征提取的能力，有效避免模型过拟合现象；而对于多背景下的公园障碍物，我们采取改进后的 Decoupled Head 构建多分支处理学习，并结合 SIoU 边界损失函数使其检测边框更接近真实边框，增强了模型对于复杂背景场景下检测障碍物的泛化能力。最后经过实际场景测试，改进后的 YOLOv5 模型对小障碍物以及多场景障碍物目标具有更良好的检测效果。

[1] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 580-587.

[2] Girshick R. Fast r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.

[3] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[J]. Advances in neural information processing systems, 2015, 28.

[4] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object

detection [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 779-788.

[5] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 7263-7271.

[6] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [J]. arXiv preprint arXiv:1409.1556, 2014.

[7] Redmon J, Farhadi A. Yolov3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.

[8] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C]//Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. Springer International Publishing, 2016: 21-37.

[9] Bochkovskiy A, Wang C Y, Liao H Y M. Yolov4: Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv:2004.10934, 2020.

[10] Yao Z, Cao Y, Zheng S, et al. Cross-iteration batch normalization[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 12331-12340.

[11] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.

[12] Liu S, Qi L, Qin H, et al. Path aggregation network for instance segmentation [C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 8759-8768.

[13] glenn jocher et al. yolov5. 2021. <https://github.com/ultralytics/yolov5> .

[14] Wang Q, Wu B, Zhu P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 11534-11542.

[15] Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XI 16. Springer International Publishing, 2020: 351-368.

[16] Ge Z, Liu S, Wang F, et al. YOLOX: Exceeding yolo series in 2021[J]. arXiv preprint arXiv:2107.08430, 2021.

[17] Gevorgyan Z. SIOU loss: More powerful learning for bounding box regression[J]. arXiv preprint arXiv:2205.12740, 2022.

[18] Yun S, Han D, Oh S J, et al. Cutmix: Regularization strategy to train strong classifiers with

localizable features[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 6023-6032.

[19] He K, Zhang X, Ren S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE transactions on pattern analysis and machine intelligence, 2015, 37(9): 1904-1916.

[20] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection [C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2117-2125.

[21]Zheng Z, Wang P, Liu W, et al. Distance-IoU loss: Faster and better learning for bounding box regression[C]//Proceedings of the AAAI conference on artificial intelligence. 2020, 34(07): 12993-13000.

[22] Yu J, Jiang Y, Wang Z, et al. Unitbox: An advanced object detection network [C]// Proceedings of the 24th ACM international conference on Multimedia. 2016: 516-520.

[23] Everingham M, Van Gool L, Williams C K I, et al. The pascal visual object classes (voc) challenge[J]. International journal of computer vision, 2010, 88: 303-338.