

文章编号: 2095-2163(2023)11-0220-07

中图分类号: TP394.1

文献标志码: A

基于 YOLOv5s 的导盲系统障碍物检测算法

刘昕斐, 张荣芬, 刘宇红, 刘源, 程娜娜, 杨双

(贵州大学 大数据与信息工程学院, 贵阳 550025)

摘要: 由于盲人缺乏视觉感知能力,因此在户外独立出行时具有较大的风险。为了增强盲人户外场景下的环境感知能力,本文针对导盲系统的实际应用,提出一种基于 YOLOv5s 改进的导盲系统障碍物检测算法。首先,为了降低整体模型的计算量,使用 MobileNetV3 代替原网络的主干特征提取网络;然后,引入 CA 注意力机制使模型更好地关注训练过程中的有效特征;最后,采用 $EIoU$ 边界框损失函数替换原模型的 $CIoU$,优化了预测框的回归速度与精度。在服务器上进行模型验证,实验结果表明本文所提算法相较原模型计算量降低了 59%,参数量降低了 49.3%,同时 mAP 提高了 2.3%,具有一定的实用价值。

关键词: YOLOv5s; 轻量化; 注意力机制; 障碍物检测

An obstacle detection algorithm for guide system based on YOLOv5s

LIU Xinfei, ZHANG Rongfen, LIU Yuhong, LIU Yuan, CHENG Nana, YANG Shuang

(College of Big Data and Information Engineering, Guizhou University, Guiyang 550025, China)

Abstract: Because blind people lack visual perception, they are at greater risk when traveling outdoors independently. In order to enhance the environment perception ability of blind people in outdoor scenes, this paper proposes an obstacle detection algorithm for guide system based on YOLOv5s for the practical application of guide system. Firstly, in order to reduce the calculation amount of the overall model, MobileNetV3 is used instead of the backbone feature extraction network of the original network. Then, the CA attention mechanism is introduced to make the model pay better attention to the effective features in the training process. Finally, the $EIoU$ bounding box loss function is used to replace the $CIoU$ of the original model, and the regression speed and accuracy of the prediction box are optimized. Compared with the original model, the experimental results show that compared with the original model, the algorithm proposed in this paper reduces the calculation amount by 59%, the number of parameters is reduced by 49.3%, and the mAP is increased by 2.3%, which has certain practical value.

Key words: YOLOv5s; lightweight; attention mechanism; obstacle detection

0 引言

根据世界卫生组织 (WHO) 的调查,全世界约有 2.85 亿人患有视力疾病^[1]。目前,国内约有 500 万盲人,且盲人数量正在以每年约 40 万的速度增加^[2]。视力障碍人群对日常生活辅助服务的需求不断加强。

视力障碍人士在独自出行的过程中需要足够外部环境信息提示以避免发生碰撞,这些信息包括路面凸起(如石头)、隔离桩、随意停放的自行车和摩托车等障碍物,以及斑马线、盲道、路面坑洼等路面情况,由于很多城市在建设过程中没有充分考虑盲人的出行需求,因此当前存在盲道设置不科学与盲

道维护不及时的问题,这就在一定程度上限制了盲人在室外独自出行的活动。而当前导盲辅助服务、如专人陪同或导盲犬对使用者的经济水平有较高要求。传统的盲人出行辅助器材大多基于超声波、红外传感器,很难满足当前的盲人出行需求,随着深度学习技术的迅速发展,基于计算机视觉领域的目标检测研究为导盲算法提供了新的发展方向。

2012 年, Krizhevsky 等学者^[3]在 ImageNet 图像分类竞赛中使用了深度卷积神经网络 (CNN) 模型,大幅度超越了传统的机器学习算法,这也成功标志着深度学习在计算机视觉领域的应用开始进入快速发展的阶段。2015 年, Girshick 等学者^[4]提出了

基金项目: 贵州省基础研究(自然科学)项目黔科合基础-ZK[2021]重点 001 资助。

作者简介: 刘昕斐(1996-),男,硕士研究生,主要研究方向:目标检测;刘宇红(1963-),男,教授,主要研究方向:计算机视觉、智能图像处理。

通讯作者: 张荣芬(1977-),女,博士,教授,主要研究方向:机器视觉、智能算法、智能硬件。Email: rfzhang@gzu.edu.cn

收稿日期: 2023-01-07

RCNN 模型,该模型使用基于候选的方法,显著提高了目标检测的准确率。随后,各种优秀的算法如 Fast RCNN^[5]、Faster RCNN^[6]、Mask RCNN^[7] 等模型算法相继被提出。在 2016 年,Redmon 等学者^[8]首次提出了 YOLO(You Only Look Once)检测模型,使得目标检测算法的精度进一步上升,且模型运行的计算量需求有了大幅度下降。

2018 年,李林等学者^[9]使用 MobileNet 网络基于迁移学习方法对盲道障碍物图片进行分类。2022 年白俊卿等学者^[10]使用 ECA 注意力结合 YOLOv4 进行无人机障碍物检测。刘力等学者^[11]使用 YOLOv4 模型对铁路上的入侵障碍物进行检测,取得了良好的效果。

本文针对视力障碍人群在出行时可能碰到的各种情况,提出了一种改进型 YOLOv5s 障碍物检测算法,来解决导盲系统使用过程中的障碍物感知问题。

1 YOLOv5s 目标检测模型及改进

1.1 YOLOv5 网络介绍

YOLOv5 是目前流行的目标检测算法之一,

YOLOv5 的网络结构如图 1 所示。YOLOv5 针对不同的部署环境提出了 4 种模型结构,分别是 YOLOv5s、YOLOv5m、YOLOv5l 和 YOLOv5x,其中 YOLOv5s 网络参数量最少,另外 3 种网络以此为基础进行不同程度的加深加宽,精度相应地有一定的提升,但是对计算资源的需求也逐渐提高。

YOLOv5s 网络结构主要分为 3 个部分:主干网络(Backbone)、颈部(Neck)和检测头(Head)。其中,Backbone 主要负责提取特征,由 CBS、CSP1 和 SPPF 三部分组成。研究可知,CBS 是由卷积(Conv)、批量归一化(Batch Normalization, BN)和 SiLU 激活函数构成;CSP1 是一种残差结构^[12],可以使计算过程中的参数量变小,速度更快,并且通过残差模块可以控制模型的深度,CSP1_X,CSP2_X 的 X 表示该模块使用的串接次数、即深度;SPPF 的作用是对特征图进行多次池化,对高层特征提取并融合,比 SPP-Net 拥有更快的推理速度。Neck 采用 PANet^[13]结构,主要作用是进行特征融合,PANet 由 CBS、上采样(Upsample)、CSP2 组成。

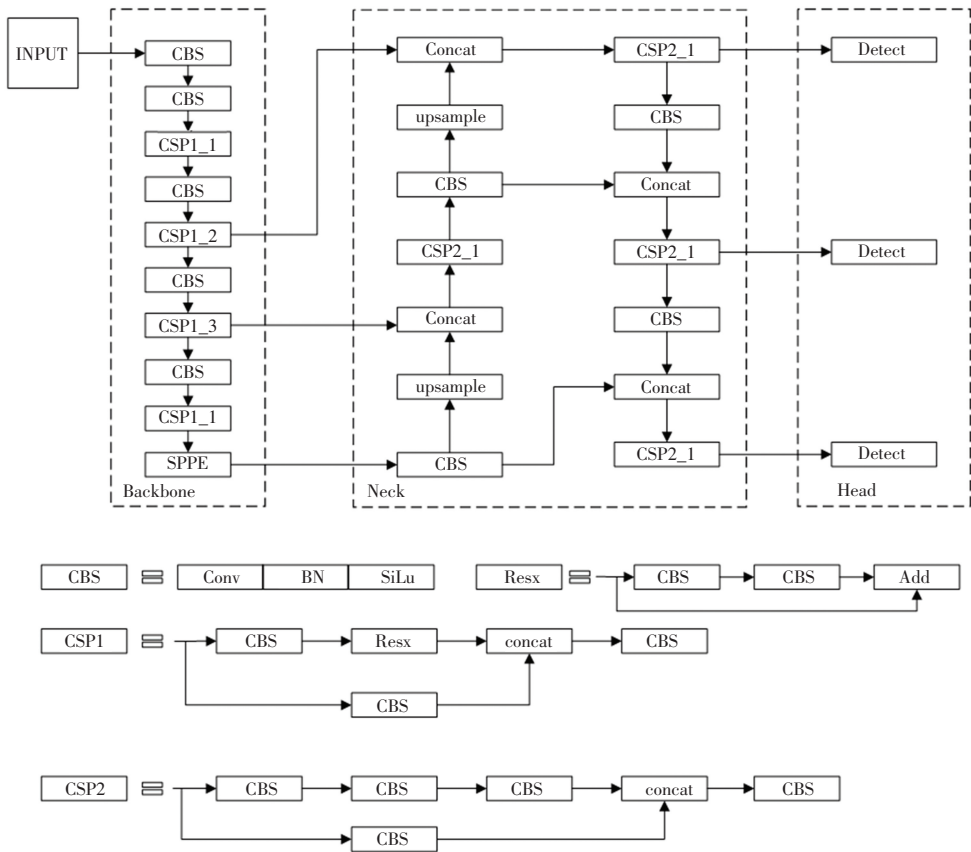


图 1 YOLOv5 网络结构

Fig. 1 YOLOv5 network architecture

1.2 改进后的模型整体网络

本文使用 MobileNetV3^[14] 网络替换 YOLOv5s 的主干特征提取网络,以减少参数量,降低计算量,提高运算速度。在主干特征提取网络和特征融合网络

中插入 CA 注意力,使模型更好地聚焦于有效特征。使用 EIoU 边界框损失函数替换原网络的 CIoU 边界框损失函数,提高了模型的回归精度。改进后的模型整体网络结构如图 2 所示。

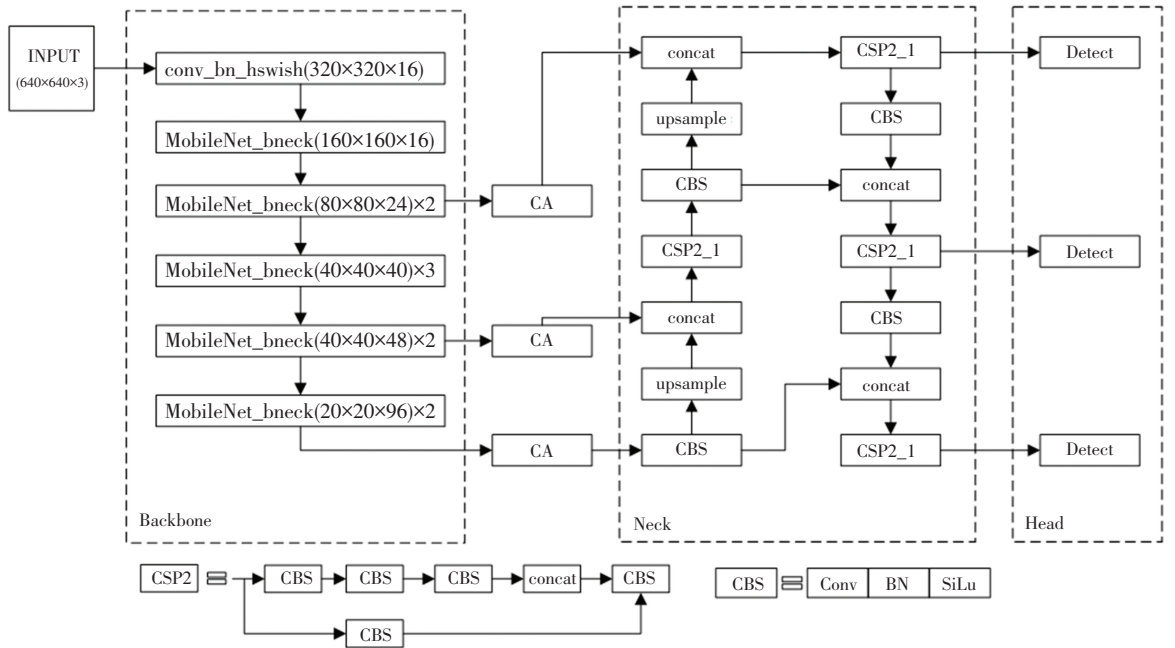


图 2 改进后的模型整体网络结构

Fig. 2 The overall network structure of the improved model

1.3 MobileNetV3 轻量化计算网络

MobileNetV3 是由 Google 团队在 2019 年提出的一种轻量级卷积神经网络,被广泛应用于移动设备等计算资源有限的场景中。相比于以前的版本,MobileNetV3 在速度和精度上都有着显著提升。

因此在逐通道卷积的过程中无法提高通道数,可以使用逐点卷积对逐通道卷积后的信息进行整合。

MobileNetV3 的设计思路主要有 3 个:减少计算量和内存占用、优化神经网络架构、增加非线性变换。MobileNetV3 的在具体实现上表现在 3 个方面。

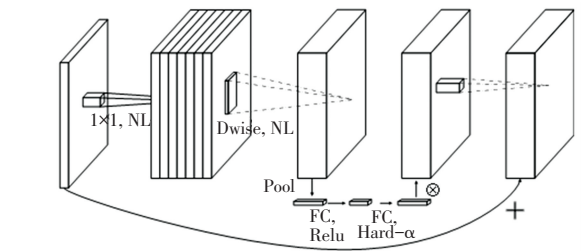


图 3 MobileNetV3 block

Fig. 3 MobileNetV3 block

(1) MobileNetV3 引入了“深度可分离卷积”(Depthwise Separable Convolution)来代替标准的卷积操作,减少了网络的计算量。深度可分离卷积将标准卷积分解为逐通道和逐点卷积两层,前者用于在通道维度上处理输入特征图,后者用于在空间维度上处理特征图。通过使用深度可分离卷积,MobileNetV3 可以显著减少参数量和计算量,并提高网络的运行速度。MobileNetV3 的 block 组成如图 3 所示。

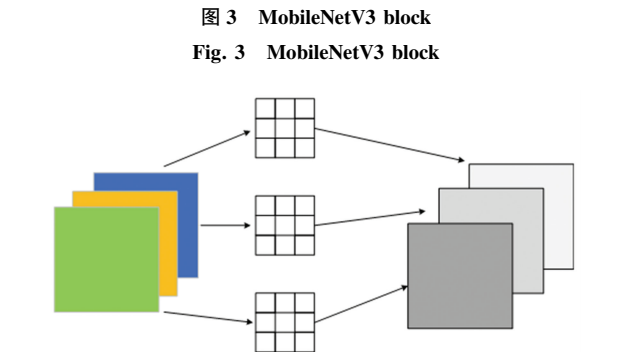


图 4 逐通道卷积

Fig. 4 Channel by channel convolution

在深度可分离卷积中逐通道卷积是通过一个一维的卷积核对一个通道进行卷积操作后再对卷积后的结果进行汇总,如图 4 所示。一张三通道的彩色图片通过逐通道卷积运算后可以得到 3 张特征图,

逐点卷积的卷积核大小为 $1 \times 1 \times M$,其中 M 为输入数据的维度,逐点卷积可以通过加权组合的方

式对逐点卷积形成的特征图进行信息提取并生成新的特征图,如图 5 所示。一张 3 通道的彩色图片用 4 个 $1 \times 1 \times 3$ 的逐点卷积进行计算后可以形成 4 个新的特征图。

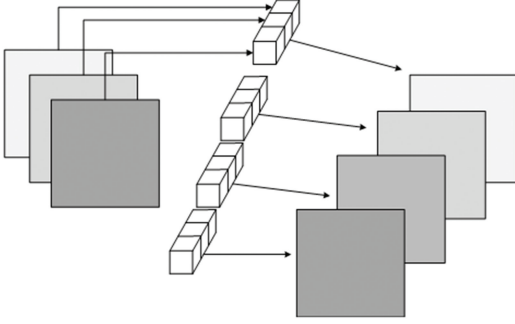


图 5 逐点卷积

Fig. 5 Point by point convolution

在使用标准卷积计算的情况下,设输入数据为 $D_F \times D_F \times M$, 卷积核为 $D_K \times D_K \times N$, 卷积步长为 1 时,标准卷积的参数量为:

$$Params = D_K \times D_K \times M \times N \quad (1)$$

对应的计算量为:

$$Flops = D_F \times D_F \times D_K \times D_K \times M \times N \quad (2)$$

在同样的情况下使用深度可分离卷积时对应的参数量为:

$$Params = D_K \times D_K \times M + M \times N \quad (3)$$

此时的计算量为:

$$Flops = D_F \times D_F \times D_K \times D_K \times M + D_F \times D_F \times M \times N \quad (4)$$

因此,深度可分离卷积与标准卷积的参数量与计算量之比均为:

$$F = \frac{1}{N} + \frac{1}{D_K} \quad (5)$$

因此可知,使用深度可分离卷积可以极大地降低参数量与计算量。

(2) MobileNetV3 使用了非线性激活函数 h -swish 和 h -sigmoid,能够显著减少模型的计算量,同时保持较高的准确率。 h -swish 是一个可微的连续函数,其形式类似于 ReLU 激活函数,但是比 ReLU 要更加平滑,从而可以减少梯度爆炸和梯度消失的问题。而 h -sigmoid 则是对 sigmoid 函数的一种变体,可以减少运算量并提高网络的计算效率。

(3) MobileNetV3 使用了 Squeeze and Excitation (SE) 注意力模块,可以自适应地对不同的通道、进行加权,加强重要的通道而减弱不重要的通道。通过使用 SE 模块,MobileNetV3 可以更加有效地利用有限的计算资源,提高网络的精度和效率。

1.4 CA 注意力机制

Hou 等学者^[15]在 2021 年提出了 CA (Coordinate Attention) 注意力机制。CA 注意力模块结构如图 6 所示。CA 注意力机制可以在基本没有增大计算开销的情况下插入神经网络中,提高网络检测性能。CA 注意力机制相比当前流行的 SE^[16]、CBAM^[17] 有显著的优点,既关注了通道维度又关注了空间维度,且解决了长距离依赖问题。

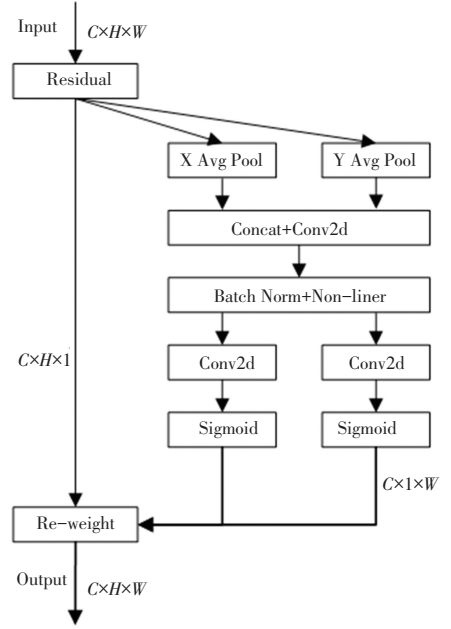


图 6 CA 注意力模块结构

Fig. 6 CA attention module structure

CA 注意力机制可以对网络中的任意中间特征张量:

$$X = [x_1, x_2, \dots, x_c] \in \mathbb{R}^{H \times W \times C} \quad (6)$$

进行转化后输出同样尺寸的张量;

$$Y = [y_1, y_2, \dots, y_c] \in \mathbb{R}^{H \times W \times C} \quad (7)$$

CA 注意力机制对通道关系和空间关系进行编码的过程可以分为坐标信息嵌入和注意力生成两个阶段。在进行坐标信息嵌入时,对输入的特征图在 X 和 Y 两个方向进行池化操作,用以保留特征图的空间结构信息。

因此高度为 h 的第 c 个通道可以表示为:

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq i < W} x_c(h, i) \quad (8)$$

同样,宽度为 w 的第 c 通道输出可以写成

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq j < H} x_c(j, w) \quad (9)$$

接着, X 和 Y 方向的特征图进行拼接,再对其进行卷积操作,使其维度降低为原来的 $\frac{C}{r}$, 然后将经

过批量归一化处理后的特征图 F_1 送入 *Sigmoid* 激活函数得到形如 $1 \times (W + H) \times \frac{C}{r}$ 的特征图 f , 计算公式如下:

$$f = \delta(F_1([z^h, z^w])) \quad (10)$$

在此基础上, 将特征图 f 按照输入数据的高度和宽度进行的卷积, 分别得到通道数与原来一样的特征图 F_h 和 F_w , 经过 σ 激活函数后分别得到特征图在高度和宽度上的注意力权重 g^h 和在宽度方向的注意力权重 g^w 。其数学公式可写为:

$$g^h = \sigma(F_h(f^h)) \quad (11)$$

$$g^w = \sigma(F_w(f^w)) \quad (12)$$

经过上述计算后将得到输入特征图在高度方向的注意力权重和在宽度方向的注意力权重。最后, 在原始特征图上通过乘法加权计算, 得到最终在宽度和高度方向上带有注意力权重的特征图, 如式 (13) 所示:

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j) \quad (13)$$

1.5 边界框损失函数改进

在 YOLOv5s 网络中, 边界框回归损失函数使用的是 *CIoU* 损失函数, *CIoU Loss* 虽然考虑了边界框回归的重叠面积、中心点距离、纵横比, 但是通过在计算过程中只考虑了纵横比的差异, 而忽略了宽高分别与其置信度的真实差异。针对这一问题, 本文使用 *EIoU*^[18] 边界框损失函数替代原模型使用的 *CIoU* 边界框损失函数, 用来加快模型的收敛速度, 提高模型的精度。*EIoU* 损失函数的公式为:

$$Loss_{EIoU} = 1 - IoU + \frac{\rho^2(b, b_{gt})}{c^2} + \frac{\rho^2(w, w_{gt})}{c_w^2} + \frac{\rho^2(h, h_{gt})}{c_h^2} \quad (14)$$

其中, $Loss_{EIoU}$ 为 *EIoU* 损失函数的值; b, b_{gt} 为预测框和真实框的中心点; ρ 为计算 2 个中心点之间的欧氏距离; w 为框的宽度; h 为框的高度; c 为能够同时包含预测框和真实框的最小外接矩形的对角线距离; c_h, c_w 为以 2 个中心点构成的矩形的高和宽。

2 实验与结果分析

2.1 实验数据集

本文针对盲人出行时常见的障碍物数据集进行收集, 具体包括 3 种类型的障碍物, 分别是: 路面情况, 如因年久失修造成的路面坑洼; 人为设置的路面的障碍, 如隔离桩、三角锥和石墩等; 以及路面上出现的随机障碍, 如随意停放的自行车、路上的行人或街道上常见的猫、狗等。本文的数据集从互联网、实

地拍摄人行道的障碍物以及 VOC 等公共数据集上进行收集, 并对收集的图片采用 *labelimg* 图像注释工具进行数据标注。实验数据集将检测障碍物分为 20 类, 共计 26 872 张图片。各类别具体数量见表 1。训练集和验证集按 9 : 1 的比例随机进行划分。

表 1 数据集的种类与数量

Tab. 1 Types and quantity of datasets

种类	数量	种类	数量
自行车	2 016	隔离柱	1 657
石球	820	杆	1 540
长椅	1 903	垃圾桶	1 205
台阶	1 498	楼梯	920
轿车	1 680	路锥	1 120
摩托车	1 209	坑	1 203
禁止横栏	1 130	人	2 230
消防栓	1 060	狗	1 180
公交车	1 098	猫	1 180
斑马线	1 559	红绿灯	923

2.2 实验细节

本文实验均在服务器 Ubuntu 20.04 操作系统下运行, 计算机处理器型号为 AMD 3900X, 显卡型号为 NVIDIA GTX 3090, 内存为 24 G。采用 Pytorch 1.7.1 框架, 所使用的编程语言 Python 3.7。模型训练时使用 *sgd* 优化器, 设定 *batch size* 为 32, 初始学习率为 0.01, 最小学习率为 0.000 1, 动量因子为 0.937。设置训练轮数为 300。

2.3 实验结果分析

2.3.1 实验评价指标

本文采用准确率 P (Precision) 和召回率 R (Recall) 计算出所有检测类别的平均精度 (mAP) 来对模型的检测效果进行评估, 使用计算量 ($Flops$) 和参数量 ($Params$) 两个指标来整体评估模型对计算资源的占用程度。

其中, AP 与 mAP 的计算公式为:

$$P = \frac{TP}{TP + FP} \quad (15)$$

$$R = \frac{TP}{TP + FN} \quad (16)$$

$$AP = \int_0^1 P(R) dR \quad (17)$$

$$mAP = \frac{\sum_{i=1}^n AP_i}{n} \quad (18)$$

其中, TP 、 FP 、 FN 分别表示正确检测的数量、错误检测的数量、没有检测出的数量。

2.3.2 不同模型实验数据对比

为了验证本文算法的检测性能,将 Faster RCNN^[19]、Centernet^[20]、YOLOv3^[21]、YOLOv4^[22]、YOLOv5s 和本文算法在同一数据集下进行对比实验,结果见表 2。

表 2 不同算法对比

Tab. 2 Comparison of different algorithms

模型	$mAP@0.5 / \%$	Params / M	Flops / G	FPS
Faster RCNN	44.6	28.5	470.1	25
Centernet	43.3	32.7	22.1	67
YOLOv4	74.1	64.1	29.9	90
YOLOv5s	74.8	7.1	16.1	96
本文算法	77.1	3.6	6.6	116

2.3.3 消融实验

为验证本文改进算法的有效性,对本文算法的

表 3 不同模块对模型整体影响

Tab. 3 The overall impact of different modules on the model

Model	MobileNetV3	CA	EIoU	$mAP@0.5 / \%$	Params / M	Flops / G	FPS
1				74.8	7.1	16.1	96
2	✓			73.9	3.6	6.5	125
3		✓		75.6	7.3	16.7	89
4			✓	75.1	7.1	16.1	95
5	✓	✓		76.9	3.6	6.6	118
6	✓		✓	74.7	3.6	6.5	123
7	✓	✓	✓	77.1	3.6	6.6	116

2.3.4 模型运行效果图

改进前与改进后的模型检测效果如图 7 所示。改进后的模型检测精度有所上升,在被检测物体之间存在遮挡情况下,因为改进后模型的特征提取能力较强,可以检测到原模型的部分漏检情况。



(a) YOLOv5s



(b) 本文算法

图 7 YOLOv5s 模型与改进后的模型检测效果对比

Fig. 7 Comparison of detection performance of YOLOv5s model and improved performance model

改进部分,分别进行消融实验得到表 3。由表 3 的实验结果可知,使用 MobileNetV3 轻量化主干特征提取网络之后计算量和参数量明显下降,计算精度轻微下降,表明采用 MobileNetV3 轻量化主干网络可以有效实现网络的轻量化,在添加 CA 注意力与改进边界框损失函数后,计算精度有所上升,表明 CA 注意力机制可以使模型有效地聚焦于被检测目标的有效特征,与此同时计算量和参数量轻微上升。

由对比实验与消融实验可知,在盲人出行道路障碍物检测精度上可以达到 77.1%,本文模型精度上优于 Faster RCNN、Centernet、YOLOv4、YOLOv5s 等模型,参数量和计算量明显小于对比算法,计算速度有所提升。

3 结束语

为了解决导盲系统的实际需求,本文提出了一种基于 YOLOv5s 的改进模型。通过将主干特征提取网络替换为 MobileNetV3,显著降低了网络的计算量和参数量,在网络中融入 CA 注意力机制,有效地提升了检测模型的精度;采用 EIoU 边界框损失函数,使得对目标的定位更加精准。

实验结果表明,本文算法在速度上满足了实时检测的需求,检测目标的准确率也优于现有的 YOLOv4、YOLOv5s 等算法, mAP 达到了 77.1%,单张检测速度达到了 116 FPS。

由于条件有限,本文的研究还有部分不足之处。一,数据集多为光照条件良好时拍摄的照片,因此模型在夜晚的识别能力有所下降;二,模型算法仍需要 6.6 GFlops 的计算量,对部分算力不高的边缘计算设备仍存在一定的压力。后续将对数据集扩充部分夜间拍摄的图片,以及采用模型剪枝、知识蒸馏等措施对模型进行进一步压缩,实现算法在边缘计算设备上的流畅运行。

参考文献

- [1] ACKLAND P, RESNIKOFF S, BOURNE R. World blindness and isual impairment; despite many successes, the problem is rowing [J]. Community Eye Health, 2017, 30: 71-73.
- [2] 武翌晗, 荣学文, 范永. 导盲机器人研究现状综述[J]. 计算机工程与应用, 2020, 56(14): 1-13.
- [3] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [J]. Advances in Neural Information Processing Systems, 2012, 25(2): 1097-1105.
- [4] GIRSHICK R, DONAHUE J, DARRELL T, et al. Region-based convolutional networks for accurate object detection and segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 38(1): 142-158.
- [5] GIRSHICK R. Fast R-CNN [C]//Proceedings of the IEEE International Conference on Computer Vision (ICCV). Santiago: IEEE, 2015.
- [6] REN Shaoqing, HE Kaiming, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region Proposal networks [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6): 1137-1149.
- [7] HE Kaiming, GEORGIA G, PIOTR D, et al. Mask R-CNN [J]. arXiv preprint arXiv:1703.06870, 2017.
- [8] 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. Las Vegas: IEEE, 2016: 779-788.
- [9] 李林, 李小舜, 吴少智. 基于迁移学习和显著性检测的盲道识别 [J]. 计算机工程与应用, 2018, 54(11): 8-14, 29.
- [10] 白俊卿, 张文静. 基于 YOLOv4 优化的轻量级无人机障碍物检测方法 [J]. 电子测量技术, 2022, 45(22): 87-91.
- [11] 刘力, 苟军年. 基于 YOLO v4 的铁道侵限障碍物检测方法研究 [J]. 铁道科学与工程学报, 2022, 19(2): 528-536.
- [12] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2017: 936-944.
- [13] LIU Shu, QI Lu, QIN Haifang, et al. Path aggregation network for instance segmentation [C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2018: 8759-8768.
- [14] HOWARD A, SANDLER M, CHEN B, et al. Searching for MobileNetV3 [C]//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul: IEEE, 2019: 1314-1324.
- [15] HOU Qibin, ZHOU Daquan, FENG Jiashi. Coordinate attention forefficient mobile network design [C] // Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, TN, USA. New York: IEEE Press, 2021: 13708-13717.
- [16] HU Jie, SHEN Li, ALBANIE S, et al. Squeeze-and-excitation networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 42(8): 2011-2023.
- [17] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module [C] //Proceedings of the European Conference on Computer Vision (ECCV). Munich, Germany: Springer, 2018: 3-19.
- [18] ZHANG Yifan, REN Weiqiang, ZHANG Zhang, et al. Focal and efficient IOU loss for accurate bounding box regression [J]. arXiv preprint arXiv:2101.08158v1, 2021.
- [19] REN Shaoqing, HE Kaiming, GIRSHICK R, et al. Faster-RCNN: Towards real-time object detection with region proposal networks [C]//Proceedings of International Conference on Neural Information Processing Systems. MIT Press, 2015: 91-99.
- [20] ZHOU X Y, WANG D Q, KRAHENBUHL P. Objects as points [C]//Proceedings of IEEE Conference on Computer Vision and Patern Recognition. Long Beach: IEEE, 2019: 7263-7271.
- [21] REDMON J, FARHADI A. YOLOv3: An incremental improvement [J]. arXiv preprint arXiv:1804.02767, 2018.
- [22] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection [J]. arXiv preprint arXiv:2004.10934, 2020.