

DOI:10.19651/j.cnki.emt.2210286

基于YOLOv5的轻量化端到端手机检测方法^{*}

刘星^{1,2} 蔡乐才² 陈波杰¹ 成奎² 高祥¹ 段少松¹

(1.四川轻化工大学自动化与信息工程学院 宜宾 644002; 2.宜宾学院三江人工智能与机器人研究院 宜宾 644000)

摘要: 针对监控图像中手机尺寸较小、分辨率低且特征不明显等问题,给检测算法研究带来了困难。提出了一种改进的YOLOv5网络模型方法用来识别手机的使用。改进的检测算法引入轻量级网络GhostNet作为主干提取网络,将GhostConv模块、C3Ghost模块分别代替主干网络中的Conv基本卷积模块和C3模块,减小网络参数和复杂度;同时,将注意力机制CBAM引入到主干网络中,减少融合后冗余特征的影响,提取到目标区域中更加关键的特征信息;使用四尺度特征检测,在原算法基础上对应的增加检测层,用以提高更小目标的检测精度。实验结果表明,改进后的YOLOv5算法准确率达到95.7%,平均精度达到97.1%,比改进前训练的准确率和平均精度分别提升了2.5%和1.8%,运算量和参数量较改进前分别减少了14.3%和24.5%。改进的YOLOv5算法不仅具有轻量化优势,同时保证了准确率和平均精度。该方法为智能监测技术行业违规使用手机提供了理论依据和技术参考。

关键词: 手机检测;轻量化;YOLOv5;GhostNet

中图分类号: TP391 **文献标识码:** A **国家标准学科分类代码:** 520.6040

Lightweight end to end mobile phone detection method based on YOLOv5

Liu Xing^{1,2} Cai Lecai² Chen Bojie¹ Cheng Kui² Gao Xiang¹ Duan Shaosong¹

(1. School of Automation and Information Engineering, Sichuan University of Science & Engineering, Yibin 644002, China;

2. Sanjiang Artificial Intelligence and Robot Research Institute, Yibin University, Yibin 644000, China)

Abstract: For problems such as small size, low resolution and not obvious features of mobile phones in the monitoring images, it has brought difficulties to the study of the detection algorithm. This article proposes an improved YOLOv5 network model method to identify the use of mobile phones. The improved detection algorithm introduces the lightweight network GhostNet as the main extraction network, and the GhostConv module, the C3Ghost module instead of the Conv basic convolution module and the C3 module in the main network to reduce the network parameters and complexity; at the same time, the attention mechanism CBAM is introduced into the main network, reducing the effects of redundant characteristics after fusion, and extraction of more critical feature information in the target area; using four scale feature detection, the corresponding increase of the detection layer on the basis of the algorithm, to improve the detection accuracy of smaller targets. The experimental results show that, the accuracy of the improved YOLOv5 algorithm is 95.7%, and the mAP is 97.1%, the accuracy and mAP of the training increased by 2.5% and 1.8%, the calculation and parameters were reduced by 14.3% and 24.5%. The improved YOLOv5 algorithm not only has the advantages of lightweight, but also ensures the mAP and accuracy. This method provides theoretical basis and technical reference for the use of mobile phones in the intelligent monitoring technology industry.

Keywords: mobile phone detection; lightweight; YOLOv5; GhostNet

0 引言

手机作为21世纪人类最伟大的发明,改变了人们的生产、生活和交往方式等各个方面^[1]。让人们的沟通变得轻松方便的同时,手机也逐渐给我们的生活增加了许多负面

影响。例如,驾驶员在驾驶过程中分心玩手机,导致交通事故^[2];行人玩手机误入车行道被车撞^[3]。加油站、机场等特殊场合玩手机是被严令禁止的,还有一些办公场景中,工作人员因为在办公时间里长时间玩手机而造成不良后果。为提高工作人员的工作效率,节约人力监督成本;同时,将玩

收稿日期:2022-06-11

^{*} 基金项目:四川省科技厅重点研发项目(2019YFN0104)、四川轻化工大学研究生创新基金(y2021076)、宜宾学院校级培育项目(2019PY39)资助

手机识别算法应用到驾驶员或课堂纪律上,也能有效避免玩手机造成的交通事故,以及学生玩手机导致的成绩下降等后果。因此,准确、高效地识别手机违规使用问题十分必要。

传统目标检测主要是通过手工设计的方法来获取物体特征,但存在较多缺陷。2012 年之后,随着计算机运算能力的不断提升,目标检测逐渐从基于传统手工特征的检测算法,迅速发展为基于深度学习的检测算法^[4]。基于深度学习的目标检测在步骤处理阶段可分为单阶段(one-stage)方法和两阶段(two-stage)两种框架。两阶段方法有 RCNN 系列^[5-8];单阶段方法有 SSD 系列^[9]和 YOLO 系列^[10-13]等。相较于两阶段,单阶段方法缺少了候选区域的提取过程,但检测速度大大提升。熊群芳等^[14]利用校准网络算法和 YOLOv3 进行驾驶员打电话行为监测;张宏群等^[15]利用 YOLOv5 优化损失函数和目标框方式,提高了遮挡中的小目标检测效果;王泽杰等^[16]利用一种人体姿态估计和融合 YOLOv3 目标检测玩手机行为,该方法计算复杂

且资源消耗大,不利于部署。

目前,大部分研究只对打电话行为进行识别,而且大多数都还是针对驾驶员。鉴于目前现有检测“玩手机”的相关研究稀少,且难以简单迁移使用,适应场景有限难以满足多样的现实需求。

综上,本文提出了一种基于 YOLOv5 网络模型,选择 YOLOv5 的最小化模型 YOLOv5n。因其简洁、轻量、高效等优点和具有易移植性,可装配至嵌入式设备平台进行目标检测的任务,成为本文的首选。

1 YOLO 系列网络

YOLOv5 目标检测算法是 Ultralytics 公司与 2020 年发布的^[17]。整体结构如图 1 所示。最初的 YOLOv5 有 4 个版本,分别为 YOLOv5s、YOLOv5m、YOLOv5l、YOLOv5x,这 4 个版本,根据模型的宽度和深度依次增加。随着 YOLOv5 发布的第 6 个版本,推出支持移动端的最小模型 YOLOv5n。

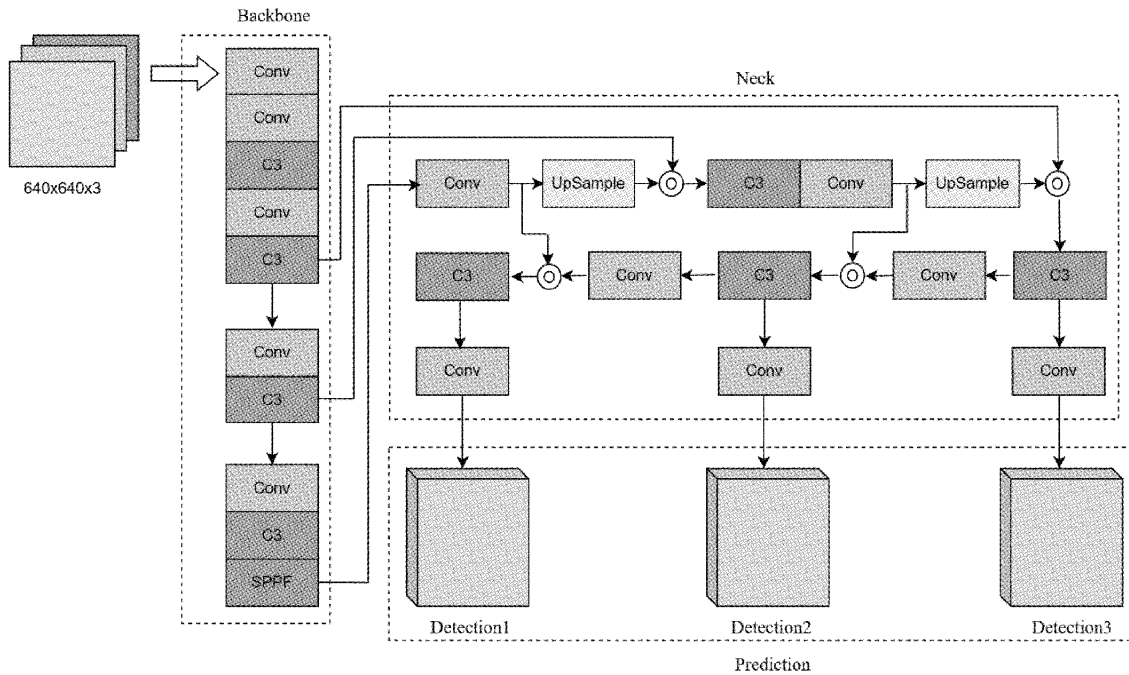


图 1 YOLOv5 整体结构

它们均由输入端、Backbone、Neck、Prediction 四部分构成。输入端引入自适应图片缩放机制,减少了信息冗余;backbone 中把原来的 Focus 结构换成了一个 6×6 大小的卷积层,其中的 SPP^[18]结构换成了 SPPF 结构,SPPF 的计算速度比原有的 SPP 更高。其中的激活函数使用了 SiLU。Neck 中网络使用 FPN+PAN^[19]的结构,结合下采样和上采样,生成特征金字塔,加强网络特征融合的能力;Prediction 端采用了与 YOLOv3 一样的结构 1×1 卷积结构,并有 3 组输出。采用 GIOU_loss 作为边界框的损失函数,引入了预测框和真实框的最小外接矩阵,增加了相交

尺度的衡量。

由图 1 可以看出,YOLOv5 的网络结构是由 Conv 模块、Bottleneck 模块、C3 模块和 SPPF 模块五大模块组成。

2 改进的 YOLOv5 模型

小目标是指目标尺寸的长宽是原图像尺寸的 10% 以下,或者尺寸小于 32×32 像素的目标,具有低分辨率的特点^[20]。手机的属性特点是小、包含视觉信息少且特征表达较弱。本文在 YOLOv5 模型的基础上进行改进,由于 YOLOv5 网络的骨干网络中结构复杂,冗余较大,因此本

文将其替换成结构更简单、运算复杂度更低的 GhostNet 网络,在保证全面特征信息的同时减少了计算量;增加多尺度检测,提高小目标的识别精度,单一的添加检测层易

出现目标漏检等情况,因此同时引入注意力机制来进一步提高网络的性能。改进后的网络结构如图 2 所示。创新点主要包括以下 3 个方面:

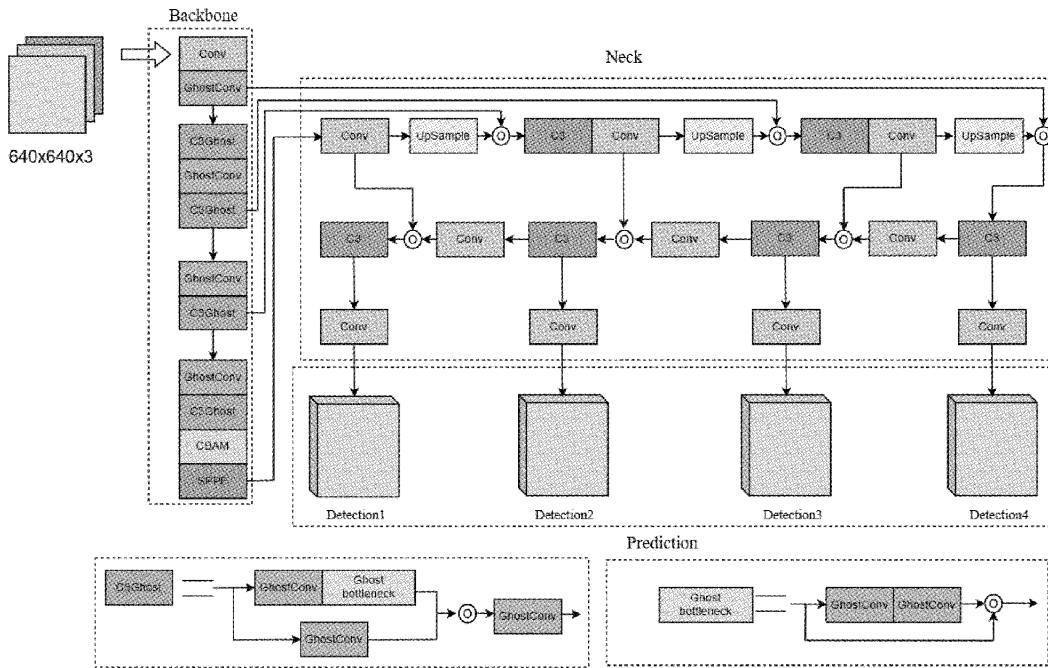


图 2 改进后的网络结构

2.1 使用 GhostNet 作为主干提取网络

为了便于智能算法在低算力设备上的部署,卷积神经网络的轻量性和低延性性格外重要。华为诺亚方舟实验室发布的新型 GhostNet 网络更加关注特征之间的冗余性和相关性,在同等参数量的情况下,精度远高于之前的轻量级网络。

Han 等^[21]提出了 Ghost 基本单元,用于减少特征提取

的计算量,构建出的 GhostNet 网络。GhostNet 网络主要由 Ghost 模块组成的 Ghost Bottleneck 模块搭建而成的神经网络模型,其思想是将传统的卷积分两步进行操作:第一步使用较少的卷积核进行普通卷积操作生成一部分特征图,然后对第一步生成的特征图使用简单的计算获得另一部分特征图,最后将两组特征图拼接在一起,从而增加通道和扩充特征。普通卷积和 Ghost 卷积过程如图 3 所示。

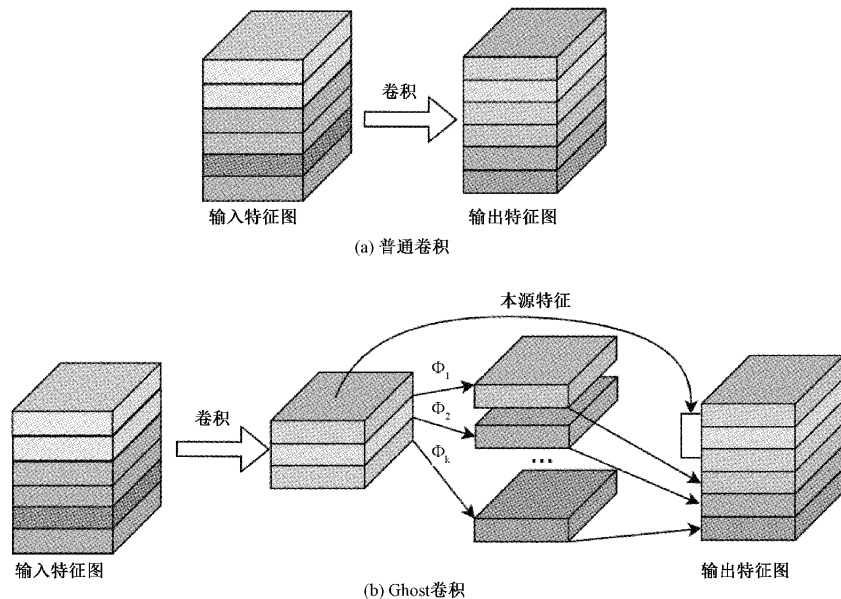


图 3 普通卷积和 Ghost 卷积

对于任意卷积运算都表示如式(1):

$$Y = X * f + b \quad (1)$$

其中,“*”代表卷积运算符, b 为偏差项, $Y \in h' \times w' \times n$, $X \in c \times h \times w$, $f \in c \times k \times k \times n$, c 为卷积通道数, h 表示特征图高度, w 表示输出数据特征图宽度, n 表示特征图个数, $k \times k$ 表示卷积核的尺寸大小,通常卷积层 n 和 c 的数量较多,由 FLOPs 的计算公式可知: $c \times k \times k \times n \times h' \times w'$,网络的计算参数量较大。

相较于普通卷积运算,在 Ghost 模块中而是通过少数原始特征图,然后再通过廉价的线性变换生成一些幻象层。生成的固有特征图公式为式(2):

$$Y' = X * f' \quad (2)$$

其中, $Y' \in h' \times w' \times m$,表示的是具有 m 个通道的固有特征图, $f' \in c \times k \times k \times m$,表示的是具有 m 个通道尺寸大小为 $k \times k$ 的卷积核,由于 $m < n$,将偏置项简化。由已得到的 m 维,通过进行一系列的线性变换生成 s 维来得到 n 维特征图。其公式为式(3):

$$y_{ij} = \Phi_{i,j}(y'_j), \forall i = 1, 2, 3 \dots m, j = 1, 2, 3 \dots s \quad (3)$$

式中: y'_i 为固有特征图中的第 i 个特征图; Φ 表示的是线性转换操作, $\Phi_{i,j}$ 表示的是第 j 个线性转换操作,用于生成第 j 个幻影特征图 y_{ij} ,经过公式的变换即可得到 $n = m \times s$ 个特征图 $Y = [y_{11}, y_{12}, \dots, y_{ms}]$ 。使用 Ghost 模块获得理论加速比计算式(4)为:

$$r_s = \frac{n \cdot h' \cdot w' \cdot c \cdot k \cdot k}{\frac{n}{s} \cdot h' \cdot w' \cdot c \cdot k \cdot k + (s-1) \cdot \frac{n}{s} \cdot h' \cdot w' \cdot d \cdot d} = \frac{c \cdot k \cdot k}{\frac{1}{s} \cdot c \cdot k \cdot k + \frac{s-1}{s} \cdot d \cdot d} \approx \frac{s \cdot c}{s+c-1} \approx s \quad (4)$$

其中, $d \times d$ 与 $k \times k$ 的大小接近, $s \ll c$,其参数压缩比式(5)为:

$$r_c = \frac{n \cdot c \cdot k \cdot k}{\frac{n}{s} \cdot c \cdot k \cdot k + (s-1) \cdot \frac{n}{s} \cdot d \cdot d} \approx \frac{s \cdot c}{s+c-1} \approx s \quad (5)$$

作者利用 Ghost 模块参数量和计算量少的优点,以 Ghost 为基础,搭建与残差网络相似的结构—GhostNet。如图 4 所示,其中每一个 Ghost 模块后面紧接着批量归一

化层和 ReLU 激活函数。在 stride=2 结构中,在其中间加了一个深度卷积(depthwise convolution, DWConv),该结构会进行两倍的降采样,通过此操作将输入的特征矩阵降为原来的 1/2,同时跳跃连接上的通路也需要下采样层,这样降低了参数规模,进一步提高了网络的效率。

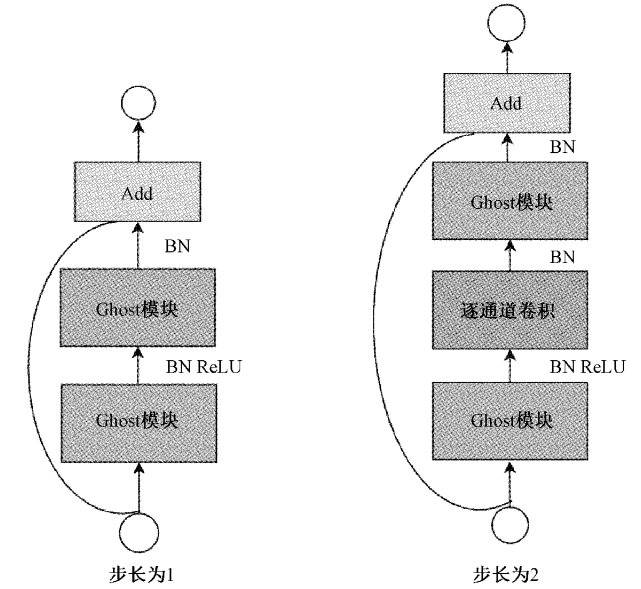


图 4 不同步长的 Ghost bottleneck 结构

本文将输入尺寸为 640×640 的图像在使用 Ghost 系列模块替换中,主干网络中将 GhostNet 分别替换第 2 层、第 4 层、第 6 层和第 8 层进行提取,同时使用轻量 Ghost 卷积代替一般卷积。

2.2 引入注意力机制

原有 YOLOv5 网络结构中并未添加注意力模块,导致网络不能很好的将注意力集中在重要特征上,本研究在 YOLOv5 的主干网络中加入视觉注意力机制,从众多冗余信息中选择出关键信息。

CBAM(convolutional block attention module)注意力机制由 Woo 等^[22]提出,它是一种结合空间和通道的注意力机制模块,空间模块赋予的是生物特征和序列特征对于同一维度的不同权重,而通道注意力对卷积之后的特征图通道赋予不同权重。一个完整的 CBAM 模块结构如图 5 所示,其具有结构简单、易于部署等优点,本文将将其嵌入到 YOLOv5 模型中。

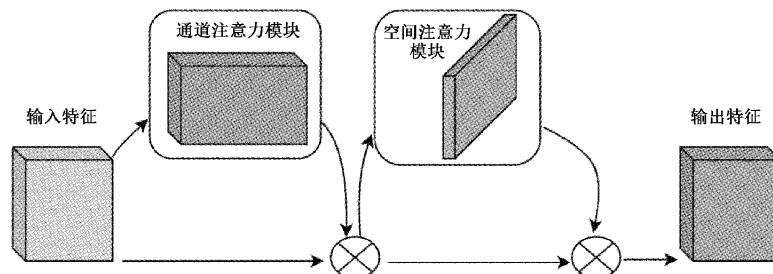


图 5 CBAM 注意力机制

该模块将第一层 CNN 处理特征图 $F \in R^{C \times H \times W}$ 作为输入,通过通道注意力机制 $M_c(F)$ 和空间注意力机制 $M_s(F)$ 来赋予特征图不同权重。计算过程概括为式(6):

$$\begin{cases} F' = M_c(F) \times F \\ F'' = M_s(F') \times F' \end{cases} \quad (6)$$

为了有效提取目标的一些轮廓信息特征,引入通道关注权重 M_c 。其计算过程为式(7):

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) = \sigma(W_1(W_0(F_{avg}^c)) + W_1(W_0(F_{max}^c))) \quad (7)$$

其中, F 代表输入特征; MLP 代表多层感知机模型; σ 表示激活函数; F_{avg}^c 和 F_{max}^c 均共享 MLP 的权重 W_1 和 W_0 。

同样的为了精准定位到检测目标的位置,利用空间注意力模块是对通道进行压缩,得到空间关注特征图。计算过程为式(8):

$$M_s(F) = \sigma(f^{7 \times 7}([AvgPool(F); MaxPool(F)])) \quad (8)$$

$f^{7 \times 7}$ 表示滤波器大小为 7×7 的卷积运算。 σ 为 sigmoid 激活函数。

2.3 多尺度检测的改进

在原始的 YOLOv5 网络结构中,使用 3 种不同的尺寸特征图来检测不同大小的目标。分别利用 8 倍下采样、16 倍下采样、32 倍下采样的特征图来检测目标,得到 20×20 、 40×40 和 80×80 三种不同尺寸大小的特征图,实现了大、中、小目标的检测和位置预测。它在目标较大的 coco 数据集^[23]上检测效果较好,但对于图像大小小于 8×8 的目标,检测将会变得困难^[24],易导致错检率与漏检率提高。因此,我们在原有三层的检测层的基础上,增加一个 4 倍下采样的过程,到特征融合网络得到新尺寸 160×160 小目标检测层,改进多尺度检测结构从而提升手机检测的检测率。改进后的尺度结构如图 6 所示。

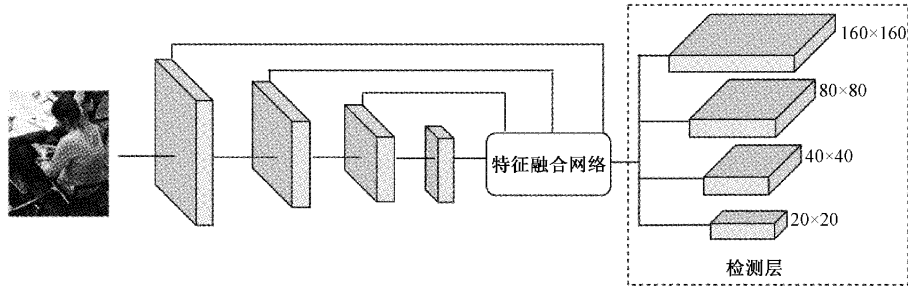


图 6 多尺度改进后的特征提取网络

YOLOv5 使用自适应先验框算法,利用 K-Means 聚类算法和遗传学习算法对数据集所有目标边界框进行聚类和学习^[23],从而避免人工设计先验框尺寸带来的误差,保留了最佳召回率的结果。本文数据集通过遗传学习和 K-Means 聚类后的小目标 Anchor 分配如表 1 所示。

表 1 Anchor box 分配表

特征图尺寸	20×20	40×40	80×80	160×160
感受野大小	大	较大	中	小
	(146,178)	(125,60)	(96,54)	(77,62)
先验框尺寸	(148,89)	(125,115)	(111,82)	(84,104)
	(169,65)	(114,41)	(113,46)	(90,40)

3 实验环境与数据集

3.1 实验环境

本实验研究的平台软件为 Ubuntu18.04,CPU 采用的是 Intel Xeon E-2276M,GPU 采用的是英伟达的 Quadro RTX5000,内存为 32 G,训练的所用的深度学习框架为 Pytorch,CUDA 和 cuDNN 版本分别为 11.1 和 8.0.5。

3.2 数据集

本文建立了一个包含 8 201 张图像以及详细标注的手

机检测数据集,用于手机检测任务的模型训练以及测试,一部分用 Python 脚本从网络中下载,另一部分从视频帧获取,该数据集包含了多种真实的社会背景。采用开源标注工具 Labelimg 进行标注,标注完成后会产生一个 xml 文件,图 7 为标注和数据制作过程。其中,将 80% 的数据划分为训练集,将 10% 的数据作为验证集,剩下的 10% 作为测试集。



图 7 数据集标注制作

4 实验评估与分析

4.1 评估参数

为了准确评估模型测试效果,本文选取准确率 P(Precision)、

召回率 R(Recall)和平均精度 mAP(mean average precision)。准确率 P 和召回率 R 的计算过程为式(9)、(10):

$$P = \frac{TP}{TP + FP} \times 100\% \quad (9)$$

$$R = \frac{TP}{TP + FN} \times 100\% \quad (10)$$

其中, TP 为将正样本预测为正的数量; FN 为将正样本预测为负的数量, FP 为将负样本预测为正的, TN 为将负样本预测为负的数量。

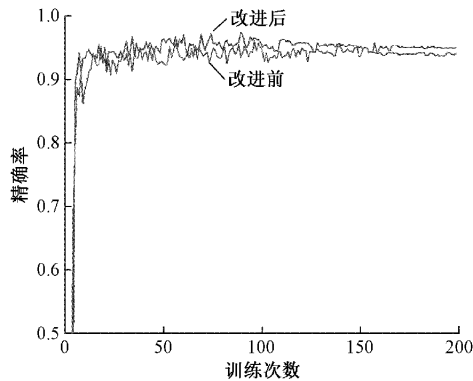
为了衡量模型对不同类别的目标的检测整体效果,使用评价指标 mAP,其表示地是所有目标类别的平均精确率的均值,可以衡量整个模型的性能。其计算过程为式(11):

$$mAP = \frac{\sum_{j=1}^c AP_j}{C} \quad (11)$$

其中, C 为验证集中目标类别的数量, AP_j 表示第 j 类的 AP 值。

4.2 实验改进分析

为了确保实验的可靠性,在所有训练过程中,选取相同的配置参数环境。采用 SGD 优化器,设置训练批次(batch size)为 32,学习率(Learning_rate)初始值设置为 0.01,使用余弦退火衰减学习法对网络学习率进行调整,



学习率梯度动量(Momentum)设置为 0.937,权重衰减系数(Weight_decay)设置为 0.0005,统一选取 640×640 大小的图像作为输入,模型迭代(Epochs)为 200 次。经过训练之后,模型网络的复杂度可以用参数量(Parameters)和浮点运算数 GFLOPs(floating point operations)来衡量,参数量和浮点运算数值越小,其模型复杂度越低。改进前后模型参数对比如表 2 所示。

表 2 改进前后参数对比

模型	浮点运算数	参数量	层数
改进前	4.2	1 765 270	270
改进后	3.6	1 333 174	450

由上表可知,改进前的层数相较于改进后的层数,增加了 40%;但模型的运算量和总参数量较改进后的分别减少了 14.3% 和 24.5%,大大降低了计算量和网络参数量。层数增加的同时实现了对网络复杂度的降低。

同样得到训练数据集对比结果如图 8 所示,根据训练过程中保存的日志文件绘出准确率和召回率曲线图。左边为准确率对比图,右边则为召回率对比图,横坐标为 Epoch 训练次数,纵坐标分别为 Precision 和 Recall,均使用相同的参数环境设置。其中绿色曲线为改进算法后的曲线,红色曲线为改进算法前的曲线。

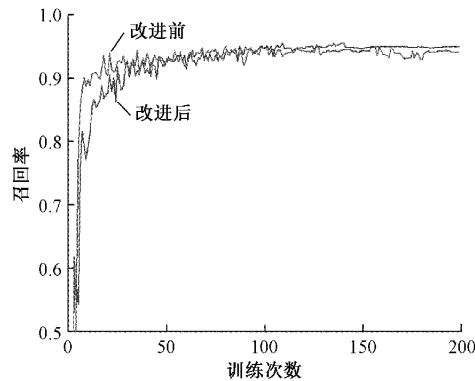


图 8 训练结果对比

由图 8 可知,模型训练在 0~50 epochs 时,两者都趋于增长趋势,改进后的准确率略高于改进前,而召回率相反;在 50~150 epochs 时,两者逐渐趋于平稳,而改进后的准确率和召回率均高于改进前且更加稳定;最后在 150~200 epochs 时,两者均保持相对稳定状态。改进前的算法准确率稳定在 0.93 附近,而改进后的精度稳定在 0.95 附近;改进前召回率稳定在 0.94 附近,而改进后的稳定在 0.95 附近。改进后的模型比改进前获得了更高更稳定的性能。

4.3 对比实验

为了增强和验证实验的对比效果,除了与原 YOLOv5 进行对比的同时,还将改进后的方法分别与现有的 YOLOv3、YOLOv3-tiny、YOLOv3-spp、YOLOv4-tiny 和 YOLOv5s 模型

进行性能指标比较。使用相同的数据集,分别训练 200 epochs。将上述算法的检测平均精确度、Recall、检测速率等参数进行对比统计。实验对比结果如表 3 所示。

由实验表格结果可知,本文改进的算法准确率和平均精度得到了提高,在模型参数方面较 YOLOv3 和 YOLOv4 系列具有更佳的轻量化优势,但相比于原算法的检测速度,略微减缓了,在比原算法每秒少传输 3.2 帧。结合模型的准确率和参数大小综合考虑,本文改进的算法更适合对手机小目标的检测任务。

4.4 消融实验

消融对比实验是为了验证各个改进模块的优化作用,使用相同的超参数以及训练参数环境,在输入尺寸大小统

表 3 实验对比结果

算法	精确率/%	模型/MB	召回率/%	帧率/(frame·s ⁻¹)	平均精度/%	浮点运算数
YOLOv3	89.5	123.4	90.3	39.4	81.1	156.3
YOLOv3-tiny	88.3	17.4	91.2	55.2	80.8	13.2
YOLOv3-spp	90.9	125.5	92.3	38.6	86.6	157.1
YOLOv4-tiny	86.8	23.6	94.7	66.4	89.3	16.2
YOLOv5s	90.8	14.3	85.3	98.3	91.4	16.5
YOLOv5	93.2	3.8	94.4	126.6	95.3	4.2
本文算法	95.7	3.2	95.2	123.4	97.1	3.6

一为 640×640 的条件下,本文在各个改进点上对 YOLOv5 进行了消融实验,实验结果如表 4 所示。改进 1

表示引入了 CBAM 注意力机制模块;改进 2 表示引入了轻量卷积模块;改进 3 表示引入多尺度检测。

表 4 消融实验对比

模型	改进 1	改进 2	改进 3	浮点运算数	模型/MB	平均精度/%	精确率/%	召回率/%	参数量
YOLOv5	×	×	×	4.2	3.8	95.3	93.2	94.4	1 765 270
模型 1	√	×	×	15.9	14.4	96.0	95.5	92.2	7 055 192
模型 2	×	√	×	2.9	2.9	96.8	94.1	94.7	1 288 258
模型 3	×	×	√	4.9	4.1	96.3	93.0	93.4	1 801 896
本文算法	√	√	√	3.6	3.2	97.1	95.7	95.2	1 333 174

由表 4 可知,引入改进 1 方法上看,可以提升平均精度和准确率,分别相应的提高了 0.7 个百分点和 2.3 个百分点,但是模型大小、参数量和运算量大大地增加了,没有达到轻量化效果。引入改进 2 方法,模型大小减小了 0.9 MB,参数减少了 27%,表明引入轻量化卷积神经网络降低了一定的参数。引入改进 3 的方法,平均精度提高了 1.0 个百分点,在引入改进 1 的方法基础上提升了 0.3 个百分点,但是略微降低了准确率。综合来看,通过引入三类改进点后,虽然运算量没有改进 2 方法的低,但是其它指标都有较大的提升,相较于原算法,改进后的算法在小目标检测表现出了更优越的性能。

4.5 测试结果分析

为了更直观的视野感受改进算法与原算法的检测区别,将检测脚本中的 line_thickness 统一设置为 1,此值越小检测框越细。下面随机选取 10% 中测试集的图片进行检测结果分析,分别选取了复杂背景场景下、弱

光线场景下和密集场景下的检测进行对比,结果如图 9 所示,左侧为改进前检测效果,右侧为改进后的效果。在图 9(a)中,可以看到存在一些干扰项目标,其中图片中一共存在 4 个手机目标,原算法检测到 3 个,本文改进的算法检测到 4 个,且都没有错检。在图 9(b)中,对比弱光线场景下的检测,可以看到目标不太明显,其中原算法远处的小目标没有检测出来,相反本文改进的算法准确地检测出小目标。在图 9(c)中,照片清晰度差,且存在多个小目标遮挡情况,原算法虽然全部准确检测到多个正确的小目标,但是识别结果的置信度却没有改进后算法高,而改进的算法多检测出一个,但其识别置信度较高。综合以上场景的结果来看,本文改进的算法在不同场景下有很好的鲁棒性,在性能方面较原算法表现出了更好的优越性以及更高的置信度,而原算法在一些场景下存在小目标错检、漏检的现象,不能够很好的满足实际检测需求。



(a) 复杂背景场景下的检测



图 9 检测效果对比图

5 结 论

针对一些场合对违规使用手机检测的需求,以及现有小目标手机检测的空缺,本文所提出了基于最小化模型的 YOLOv5 版本检测算法。通过在 YOLOv5 的基础上融合了三种改进,引入注意力机制;引入轻量级卷积神经网络;采用多检测层检测,增加 4 倍上采样结构层。对 YOLO 系列算法进行对比实验、与改进前算法对比,实验结果表明:改进后的平均精度比改进前提高了 1.8%,召回率和准确率分别提高了 0.8%和 2.5%,模型的计算量和总参数量较改进后的分别减少了 14.3%和 24.5%。能够在保证召回率的基础上,同时实现准确率和平均精度的最佳平衡,减少了计算量,降低了计算复杂度。能够满足对资源不足、计算能力低的嵌入式部署成本。

参考文献

- [1] 刘志远,张有光. 世界尽在掌中:智能手机改变生活[J]. 科技导报,2016,34(9):27-31.
- [2] 杨奕枫. 基于轻量级卷积神经网络的驾驶行为识别[D]. 重庆:重庆邮电大学,2021.
- [3] 刘军,范长军,瞿崇晓. 基于人体姿态估计的手机使用状态监控[J]. 计算机系统应用,2021,30(3):164-170.
- [4] 李智伟,杨亚莉,钟卫军,等. 基于改进的 SSD 模型手机违规使用目标检测[J]. 电子测量与仪器学报,2021,35(1):120-127.
- [5] 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, 2013:580-587.
- [6] GIRSHICK R. Fast R-CNN [C]. Proceedings of the IEEE International Conference on Computer Vision, 2015:1440-1448.
- [7] REN S, HE K, 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.
- [8] HE K, GKIOXARI G, DOLLAR P, et al. Mask R-CNN [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2020, 42(2):386-397.
- [9] LIU W, ANGUOLOV D, ERHAN D, et al. SSD: Single shot multibox detector [C]. European Conference on Computer Vision, Springer, Cham, 2016:21-37.
- [10] 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.
- [11] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017:6517-6525.
- [12] REDMON J, FARHADI A. YOLOv3: An incremental improvement [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018:89-95.
- [13] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection [J]. ArXiv Preprint, 2020, ArXiv:2004.10934.

- [14] 熊群芳,林军,岳伟,等.基于深度学习的驾驶员打电话行为检测方法[J].控制与信息技术,2019,(6):53-56.
- [15] 张宏群,班勇苗,郭玲玲,等.基于YOLOv5的遥感图像舰船的检测方法[J].电子测量技术,2021,44(8):87-92.
- [16] 王泽杰,沈超敏,赵春,等.融合人体姿态估计和目标检测的学生课堂行为识别[J].华东师范大学学报(自然科学版),2022,(2):55-66.
- [17] 王玲敏,段军,辛立伟.引入注意力机制的YOLOv5安全帽佩戴检测方法[J].计算机工程与应用,2022,58(9):303-312.
- [18] HE K M, ZHANG X Y. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence,2014,37(9):1904-1916.
- [19] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation [C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR),2018:8759-8768.
- [20] 戴腾,张珂,尹东.一种用于驾驶场景下手机检测的端到端的神经网络[J].光电工程,2021,48(4):71-80.
- [21] HAN K, WANG Y, TIAN Q, et al. GhostNet: More features from cheap operations[C]. IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2020:1577-1586.
- [22] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module [J]. Springer, Cham,2018:3-19.
- [23] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: common objects in context [C]. 13th European Conference on Computer Vision, Cham: Springer, 2014:740-755.
- [24] 杨其晟,李文宽,杨晓峰,等.改进YOLOv5的苹果花生长状态检测方法[J].计算机工程与应用,2022,58(4):237-246.
- [25] 蒋寰游.基于YOLOV5的无人机航拍输电线路绝缘子多缺陷检测[D].广州:广东工业大学,2021.

作者简介

刘星,硕士研究生,主要研究方向为计算机视觉与嵌入式技术。

蔡乐才,教授,主要研究方向为智能检测与控制、智能信息处理等。

陈波杰,硕士研究生,主要研究方向为大数据分析等等。

成奎,讲师,硕士研究生,主要研究方向为智能控制、图像处理等。

高祥,讲师,硕士研究生,主要研究方向为物联网技术等。

段少松,硕士研究生,主要研究方向为数据挖掘。

E-mail:1610830030@qq.com