



计算机工程与应用
Computer Engineering and Applications
ISSN 1002-8331, CN 11-2127/TP

《计算机工程与应用》网络首发论文

题目： 基于改进 YOLOv4-tiny 轻量化校内行人目标检测算法
作者： 孙好，董兴法，王军，陈致远
网络首发日期： 2022-11-03
引用格式： 孙好，董兴法，王军，陈致远. 基于改进 YOLOv4-tiny 轻量化校内行人目标检测算法[J/OL]. 计算机工程与应用.
<https://kns.cnki.net/kcms/detail/11.2127.TP.20221102.1646.009.html>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

基于改进 YOLOv4-tiny 轻量化校内行人目标检测算法

孙好¹, 董兴法^{1*}, 王军^{1,2,3}, 陈致远¹

1. 苏州科技大学电子与信息工程学院, 江苏 苏州 215009

2. 中国科学院长春光学精密机械与物理研究所, 长春 130033

3. 中国白城兵器试验中心, 吉林 白城 137001

摘要: 深度学习常用于行人检测, 为了在嵌入式设备上应用复杂的传统卷积神经网络, 网络的轻量化是必然趋势, 但难以兼顾速度和精度。为解决这个问题, 该文设计了一种基于改进 YOLOv4-tiny 的轻量化校内行人目标检测算法。首先, 提出了一种多尺度空洞卷积模块的改进 Ghost 卷积特征提取模块, 同时普通卷积用深度可分离卷积代替, 降低了模型复杂度, 增加特征提取的多样性; 其次, 构建了一种空洞深度可分离卷积的改进空间金字塔池化结构, 增强上下文特征的融合, 提高检测精度, 减少网络参数; 最后, 再引入 Soft-NMS 取代传统非极大值抑制, 降低漏检率。实验表明, 该算法在多个数据集和硬件平台上, 其具有精度高、速度快、模型参数少和体积小等特点, 可以应用于嵌入式设备。

关键词: 校内行人; 深度学习; YOLOv4-tiny; Ghost 卷积; 非极大值抑制

文献标志码: A 中图分类号: TP389.1 doi: 10.3778/j.issn.1002-8331.2207-0313

Based on the improved YOLOv4-tiny lightweight pedestrian in school target detection algorithm

SUN Hao¹, DONG Xingfa^{1*}, WANG Jun^{1,2,3}, CHEN Zhiyuan¹

1. School of Electronic and Information Engineering, Suzhou University of Science and Technology, Suzhou, Jiangsu 215009, China

2. Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Science, Changchun 130033, China

3. Center of Arms Experiment of Baicheng, Baicheng, Jiangsu 137001, China

Abstract: Deep learning is often used for pedestrian detection. In order to apply complex traditional convolutional neural networks on embedded devices, the lightweight of the network is an inevitable trend, but it is difficult to balance speed and accuracy. To solve this problem, this paper designs a lightweight intra-school pedestrian target detection algorithm based on improved YOLOv4-tiny. First, an improved Ghost convolution feature extraction module is proposed, which is a multi-scale hole convolution module. At the same time, ordinary convolution is replaced by depthwise separable convolution, which reduces the complexity of the model and increases the diversity of feature extraction. Second, An improved spatial pyramid pooling structure with depthwise separable convolution of holes enhances the fusion of contextual features, improves detection accuracy, and reduces network parameters. Finally, Soft-NMS is introduced to replace traditional non-maximum suppression to reduce the missed detection rate. Experiments show that the algorithm has the characteristics of high accuracy, fast speed, few model parameters and small size on multiple data sets and hardware platforms, and can be applied to embedded devices.

Key words: pedestrians on campus; deep learning; YOLOv4-tiny; ghost convolution; non-maximum suppression

目前机器视觉目标检测是热门的研究方向之一, 从传统的机器学习到人工神经网络, 再到深度恶习,

基金项目: 国家自然科学基金资助项目(11804244); 江苏省“十三五”重点学科项目(20168765)。

作者简介: 孙好(1994—),男,研究生,图像处理与光电信息检测,E-mail: 1557403968@qq.com;董兴法(1963—),男,博士,教授,光电信息处理和光电测控技术。

技术在不停的进步。现在深度学习被应用在各行各业,越来越成熟。目标检测的深度学习主要分为两大类:以 R-CNN^[1]为代表的两阶段 (two-stage) 算法与以 YOLO^[2] (You Only Look Once)、SSD^[3] (Single Shot multibox Detetor) 和 EfficientDet^[4] 为代表的单阶段算法 (one-stage)。两阶段算法检测中,先提取候选框,然后对候选框进行回归和分类,特点是精度高,速度慢。单阶段是直接将整张图片作为输入,通过回归获得目标分类和位置,特点是速度快,精度没有两阶段高。

在行人目标检测领域,为了提高人员检测的准确率。Liu 等人^[5]提出渐近定位拟合模块,该模块堆叠了一系列的预测因子,得到了具有 SSD 的速度和 Faster R-CNN 的精度网络。2020 年张明伟等人^[6]通过密集连接卷积网络改进 YOLOv3,实现了模型的压缩与网络参数和特征的复用。2021 年董小伟等人^[7]提出了用加权的方式将不同尺度的特征层融合,在客流量大量增加的地铁站实现了精准实时检测。2021 年苏杨等人^[8]提出了以 MobileNet 改进 SSD 网络,用轻量化的网络实现了机房内人员的实时检测。2021 年 Wang 等人^[9]在提出了可以运行在低端 GPU 设备的 YOLOv4-tiny 网络,缩放了 YOLOv4^[10] 的网络结构,优势检测速度更快,但精度下降了。2022 年方梦瑞等人^[11]提出了高性能轻量化的茶叶嫩芽检测模型 YOLOv4-tiny-Tea,引入了双向特征金字塔网络、卷积块注意力机制等方法,保持轻量化计算成本的基础上,提高了检测精度。2022 年郭奕裕等人^[12]提出了一种轻量化的 DT-YOLO 模型,增加了空间金字塔池化模块,使用通道剪枝压缩模型大小和计算量,提高了推理速度和减小了模型体积。2022 年王斌等人^[13]提出了一种基于 YOLOv4-tiny 的轻量化抗干扰火焰检测深度网络,引入 Mosaic 数据集增强、深度可分离卷积和引出多检测头等方法,提高了准确率和缩短了执行时间。2022 年郭克友等人^[14]提出了利用 MobileNet v3 作为主干网络的轻量化 YOLOv4 检测模型,结合深度可分离卷积、标签平滑和退火余弦算法等改进该模型,降低了权重文件大小和提高了检测速率。2021 年张明路等人^[15]针对电子元器件检测设计了一种具有线性瓶颈的逆残差结构 YOLOv4 模型,相比于原算法提高了精度和速度。2022 年程浩然等人^[16]提出了以两个 Resblock-D 模块代替 CSP 模块,同时引入 SPP 和 CA 注意力机制模块的模型,有效提高了检测速度。2022 年王伟彤等人^[17]设计了一种基于 YOLOv4-tiny 的 CenterNet 网络算法,极大的提升了推

理速度。

在实际生活中,由于经典目标检测算法运行的设备算力加速有限,难以达到预期的精度和速度,导致无法满足工作中需求。为了在有限性能的设备上运行检测算法,获得较高的精度和速度且能应用于嵌入式设备。因此,为了轻量化目标检测网络以及较高的准确率,选择 YOLOv4-tiny 网络并对其进行改进,改进主要分为 3 部分:(1)以多尺度空洞卷积模块为基础改进的轻量化 Ghost 卷积^[18]特征提取网络作为网络的主干,用于精简网络的结构;(2)加入用空洞深度可分离卷积改进的空间金字塔池化模块 (Spatial Pyramid Pooling, SPP)^[19],加强特征的融合能力,提高检测精度,使用空洞^[20]深度可分离卷积对其进行优化,减少网络参数;(3)引入 Soft-NMS 取代原非极大值抑制,降低漏检率。为了测试该算法的性能,分别在多个数据集和测试平台测试,从精度、速度、模型参数和体积等方面的表现,可以看出速度和精度达到最优。

1 相关网络和模块

1.1 Yolov4-tiny 算法

YOLOv4-tiny 网络主要由三部分组成:主干 (CSPDarknet53-tiny)、颈部 (FPNet)、检测头 (YOLO Head)。多个 3×3 普通卷积、3 个跨阶段连接 (Cross Stage Partial, CSP) 和 3 个最大池化 (MaxPooling) 对图像进行降采样,构成了主干特征提取网络。颈部特征金字塔 (Feature Pyramid Networks, FPN) 的作用是融合两个不同尺度的特征。检测头中的预测框数量主要由类别 N 决定,类别越多,预测框数量越多,输出的大小分别是 $13 \times 13 \times n \times (5+N)$ 和 $26 \times 26 \times n \times (5+N)$, n 为先验框的个数,一般设为 3, 5 包含了目标位置坐标信息 (x, y, w, h) 和置信度 (confidence)。通过预先设定的置信度大小和非极大值抑制算法 (Non Maximum Suppression, NMS) 过滤掉置信度低的预测框,获得最终预测框。

1.2 Ghost 卷积和 Inception 模块

2020 年 Han 等人^[18]发现冗余的特征图在神经网络架构中是非常重要的特征,提出 Ghost 卷积来替代普通卷积生成特征图。如图 1 所示, Ghost 卷积主要

由两个步骤组成,第1步正常使用普通卷积提取本源特征图,第2步在本源特征图的基础上使用简单的线性操作提取更多的幻影特征。最后将本源特征图与提取的幻影特征结合得到最终要输出的特征图。2014年Szegedy等人^[21]提出增加了网络的宽度和深度的

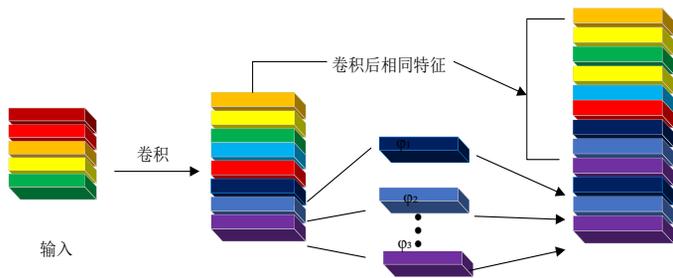


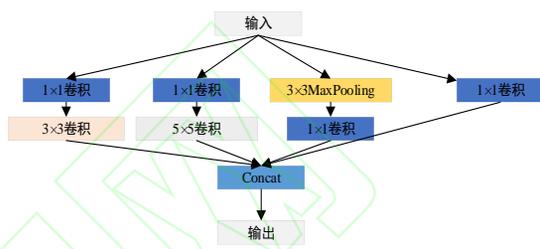
图1 Ghost卷积和Inception结构

Fig.1 Ghost convolution and Inception structure

2 改进 Yolov4-tiny 的轻量化检测网络

为了实现网络的轻量化,以及提高准确率的需求下,在YOLOv4-tiny网络的主干、颈部和检测头三部分进行改进。假定输入一张大小为 $416 \times 416 \times 3$ 的图像,经过两次深度可分离卷积后,进入基于Ghost多尺度空洞卷积模块(Ghost Multi-scale Dilated Convolution Module, Ghost-MDCM)的主干网络。Ghost-MDCM模块有四个分支,将特征图 $104 \times 104 \times 64$ 输入扩张率分别为2,3,4的空洞卷积替代的普通卷积分支中,防止因通道数降低特征提取能力,在每个分支引入Ghost模块,使用 1×1 卷积调整通道数,共需通过三个Ghost-MDCM模块。然后,在主干与颈部之间添加了空洞深度可分离卷积空间金字塔池化模块(Dilated Depthwise Sparable Convolution Spatial Pyramid Pooling, DDSCSPP)。特征图 $13 \times 13 \times 512$ 通过扩张率分别为3,6,9的空洞卷积DDSCSPP模块,,增强了上下特征之间的融合,使用 1×1 卷积调整通道数输出特征图 $13 \times 13 \times 512$,通过特征融合颈部网络,下采样系数设为2,得到大小为 $13 \times 13 \times 18$ 、 $26 \times 26 \times 18$ 的特征图。最后检测头采用Soft-NMS算法进行

Inception结构。如图2所示,主要分为2步骤,第1步在卷积核前面增加 1×1 卷积,减少参数的数量,也可以增加特征中的非线性组合,第二步就是采用不同的卷积,如 3×3 、 5×5 、 7×7 等。最后将不同尺度的特征进行拼接得到输出的特征图。



预测,网络结构如图2所示。首先,为了减少参数数量和计算量,将普通卷积 3×3 换成了深度可分离卷积,特征提取的主干则是以轻量化的Ghost卷积模块(Ghost Block)为基础,以少量的卷积操作结合线性变换的方式生成特征图,同时借鉴了Inception的思想,由单一分支机构改成多分支结构,使用分解卷积取代标准卷积,还融入了空洞卷积,不增加参数数量的条件下,增大感受野,获取多尺度信息,融合更大范围的上下文特征信息,形成了Ghost-MDCM的CSP结构特征提取网络,其次,特征融合网络由传统的空间金字塔池化模块改为DDSCSPP模块,提取多尺度特征后融合特征的同时,减少了计算量和参数量;最后,传统的非极大值抑制算法需要求出同类置信度最大的预测框和同类中剩余预测框的交并比(Intersection over Union, IoU),交并比的值超过一定数值后删除该预测框,在这不断重复的过程中,会删除大量置信度低、重叠度高的预测框,尤其是校内行人行走过程中,成群结队的情况居多,造成漏检、误检的情况,为此本算法采用Soft-NMS进行改进,提高检测准确率。

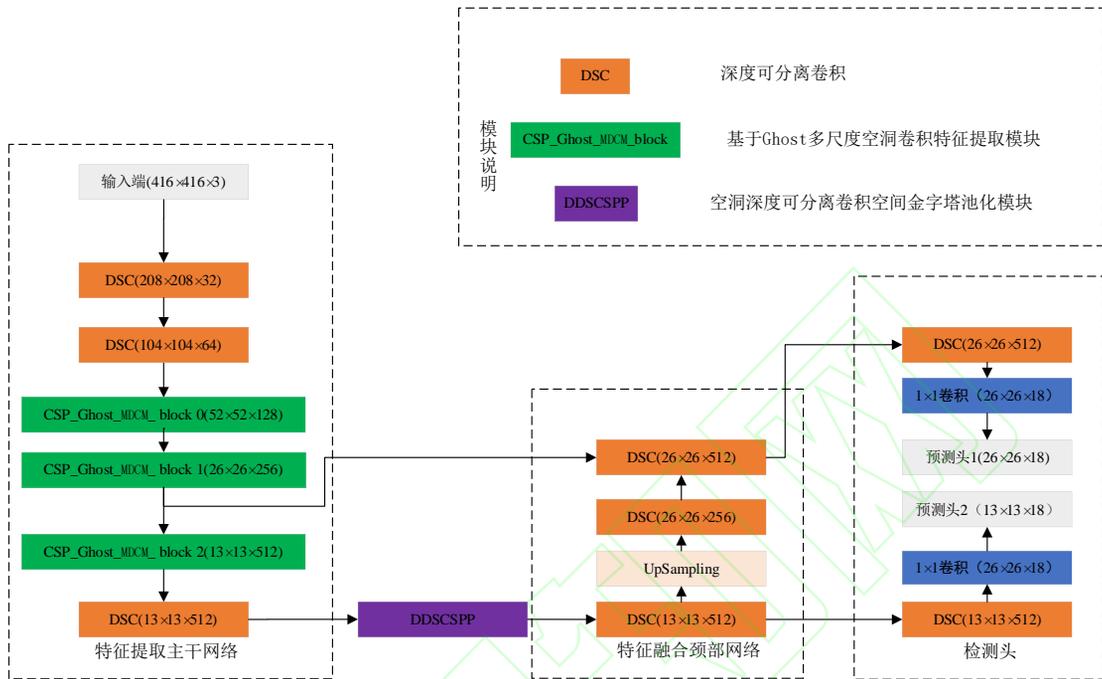


图 2 改进 YOLOv4-tiny 的轻量化校内行人检测网络结构图

Fig.2 Improved YOLOv4-tiny's lightweight intra-school pedestrian detection network structure

2.1 基于 Ghost-MDCM 的特征提取主干网络设计

YOLOv4-tiny 的主干特征提取网络是以 CSP 结构为基础的, CSP 结构将原输入的特征图分为两个分支, 使得通道数减半, 一支直接与输出相连接, 另一支进行多个重复的普通卷积操作。如图 3 所示, 本文保留了 CSP 结构, 但考虑到参数量集中于此, 对重复进行普通卷积操作的分支进行了优化。本文借鉴了 Inception 的思想增加分支数, 但是根据倒置残差结构^[15]可以发现, 通道数的下降会导致网络提取特征的能力变弱。为了解决这个问题, 引入了 Ghost 卷积, Ghost

卷积的特点就是消耗少的计算资源获取更多的通道数量, 所以将 Ghost 卷积 1 扩展通道数量, Ghost 卷积 2 压缩通道数量, 最后使用 1×1 卷积调整输出的通道数。但是 Inception 模块中如果卷积核的大小相同, 则 Inception 会损失提取多个尺度融合和减少冗余特征的优势。为了不增加参数量, 引入了扩张率不同的空洞卷积, 获取不同尺度的特征。当扩张率分别 $D=2$ 或 3 时, 对应的感受野是 5 或 7。相比于普通卷积 3×3 , 该结构解决了特征提取过程中“大而密”的问题, 加快了训练的速度, 降低了参数量, 提高了特征表达能力。

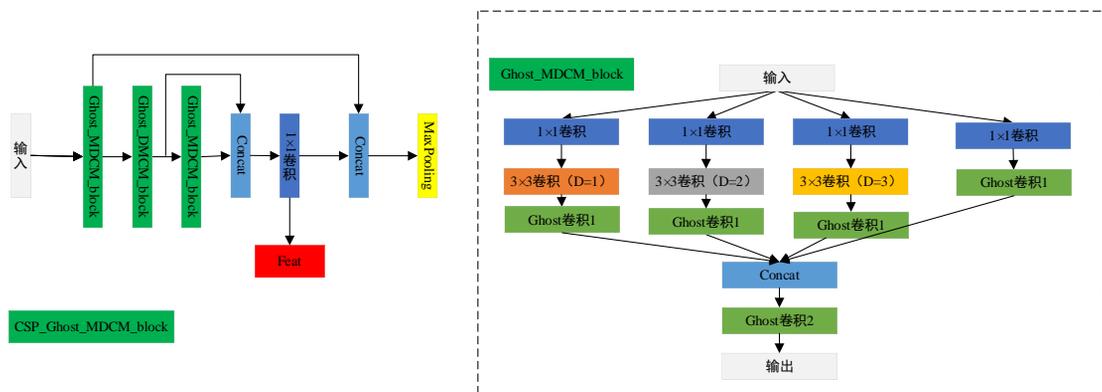


图 3 CSP_Ghost 多尺度空洞卷积特征提取网络模块

Fig.3 CSP_Ghost multi-scale atrous convolution feature extraction network module

2.2 改进的空间金字塔池化特征融合模块

YOLOv4 中的空间金字塔池化结构的作用就是增强上下文特征融合能力,但是在 Yolov4-tiny 中却未使用该结构,使得精度下降了一些。受到不同扩张率的空洞卷积启发,利用它来改进空间金字塔池化(Spatial Pyramid Pooling, SSP)^[22],目的是增大感受野和融合不同感受野的特征。如图 4 所示,为了避免增加网络

的复杂度,采用 1×1 的卷积降低通道数,以及在不同扩张率的空洞卷积后面增加深度可分离卷积中的 1×1 点卷积,同时降低参数量,避免产生特征的冗余,最后将各个分支的特征进行合并,使用 1×1 卷积进行特征融合,增强信息之间的交流,降低通道数至指定的大小。该结构放置的位置同 YOLOv4 一样,位于主干特征提取网络的后面。

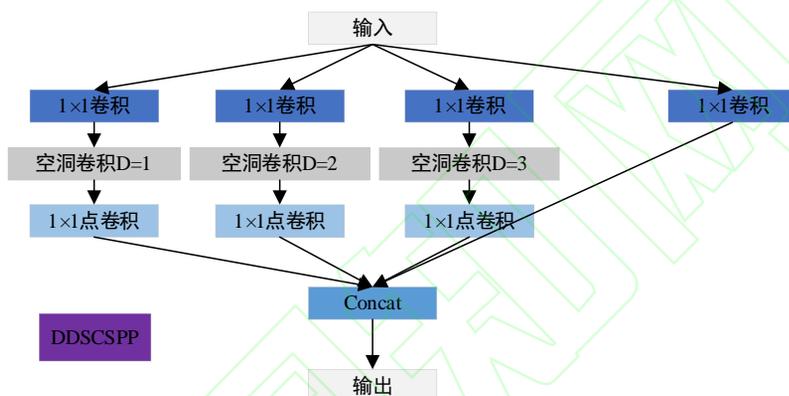


图 4 改进的 DDSSPP 结构

Fig.4 Improved DDSSPP structure

2.3 Soft-NMS

在传统 NMS 算法执行过程中,由于其机制的原因,会强行删除一些置信度低且重叠度高的预测框,容易造成漏检或者误检的情况。传统 NMS 算法主要是根据预先设定的 IoU 的大小决定是否删除预测框,其计算公式如 (1) 所示。

$$S_i = \begin{cases} T_i, iou(A, B_i) < N_i \\ 0, iou(A, B_i) \geq N_i \end{cases}, \quad (1)$$

相比于传统 NMS 直接将置信度值设为 0 的做法,Soft-NMS 改进的地方就是逐步降低置信度值,其公式如 2 所示。

$$S_i = \begin{cases} T_i, iou(A, B_i) < N_i \\ T_i(1 - iou(A, B_i)), iou(A, B_i) \geq N_i \end{cases}, \quad (2)$$

考虑这个分段函数并不连续,所以采用高斯重置函数对其进行改进,解决因跳变影响检测效果的问题,其公式如 (3) 所示。

$$S_i = T_i \exp\left[-\frac{iou(A, B_i)^2}{\sigma}\right]; \forall B_i \notin D, \quad (3)$$

因此,本文针对传统 NMS 存在的问题,采用高

斯重置函数优化的 Soft-NMS 取代传统 NMS 算法,调整了预测框置信度值得改变方式,减少因强行将置信度值设为 0 而删除该预测框造成的目标漏检、误检问题,从而提高检测的精确度。

3 实验结果与分析

3.1 数据集采用与评价指标

为了检测所提算法的性能,分别选取了四个数据集:标准化的 PASCAL VOC 数据集^[23]、行人检测 INRIA 数据集^[24]、行人检测 CUHK Occlusion 数据集^[25]和自建数据集。PASCAL VOC (The Pascal Visual Object Classes) 为图像检测提供了一个非常标准化的数据集,通常自建数据集会按照 PASCAL VOC 数据集的格式。它的样本类别中数量比较多就是 Person,将 PASCAL VOC 2007 和 2012 中的 Person 图像提取,构建 Person 数据集用于检测,该数据集用于训练的共 9583 张,用于测试的 5126 张。自建数据集的图片来自真实场景下校园实验室,分别在不同的时间段和天气条件下提取,共计 15230 张图片,利用 LabelImg 软件进行对图片中的人标注,用于训练的共 12640 张,用于测试的共 9620 张,用于真实环境下的检测评估。

通常使用精确率、召回率、平均精度均值 3 个指标评估所提算法网络的精度。评估网络模型轻量化指标有多个; 参数的数量 (Params)、每帧图片检测消耗的时间、每秒检测图片数量 (Frames Per Sec-

ond,FPS)、浮点型计算量 (Floating Point Operations,FLOPs) 和模型的体积。

3.2 实验设置

实验的具体资源配置如表 1 所示。

表 1 各设备信息

Table 1 Information about each device

设备 1	设备 2	设备 3	设备 4
I5 10400F CPU	I5-8250U	NVIDIA JETSON XAVIETNX	NVIDIAJETSON NANO
NVIDIA RTX2070SUPER GPU	CPU	嵌入式设备	嵌入式设备
内存 16GB	内存 8GB	ARM V8.2 架构 CPU	Cortex-A57 CPU
Windows10	平板电脑	NVIDIA Volta 架构 GPU	内存 4GB
Python 3.6		内存 8GB	
Pytorch 1.2			
CUDA10.0			

网络使用 Adam(Adaptive momentum estimation) 优化器梯度下降法训练, 学习率预设置为 10-4 并用余弦退火衰减(Cosine Annealing LR)对学习率进行调整。以 PASCAL VOC Person 数据集为基准训练, Epoch 设置为 2000。为了尽快训练速度, 在网络结构保持不变的情况下, 引入迁移学习, 预载入 PASCAL VOC Person 训练得到的权重, Epoch 设置为 500, 其他设置不发生变化。

3.3 实验结果分析

3.3.1 模块参数实验

改进的 Ghost 多尺度空洞卷积特征提取模块

(Ghost-MDCM) 中含有 3 个不同扩张率的空洞卷积, 空洞深度可分离卷积空间金字塔池化模块(DDSCSPP) 也是。为了确定最佳的扩张率数值, 将 3 个扩张率组成一组, 选取了多个取值不同的扩张率组进行试验, 使用的数据集是 PASCAL VOC Person, 具体结果如表 2 所示。当 3 个扩张率相同时, 实验的结果都不是最好的, 因为单一扩张率的空洞卷积, 像素之间过于稀疏, 会产生网格效应^[26], 同时也会让基于 Inception 思想的结构失去它的优势。当 3 个扩张率不同时, Ghost 多尺度空洞卷积特征提取模块和空洞深度可分离卷积空间金字塔模块的最佳取值分别为[2,3,4]、[3,6,9]。

表 2 不同扩张率下实验结果

Table 2 Results at different expansion rates

Ghost-多尺度空洞卷积特征提取模块				空洞深度可分离空间金字塔池化模块			
扩张率	精确率 (%)	召回率 (%)	mAP (%)	扩张率	精确率 (%)	召回率 (%)	mAP (%)
[2,2,2]	76.53	68.18	81.64	[3,3,3]	79.33	78.99	82.15
[4,4,4]	79.41	67.90	81.06	[9,9,9]	78.84	79.35	82.24
[2,3,4]	76.25	80.39	82.89	[2,4,6]	80.21	78.56	82.43
[3,2,4]	79.17	80.76	82.45	[3,6,9]	80.19	80.38	83.97
[4,5,6]	79.36	78.42	82.15	[12,14,18]	79.36	77.47	82.35

3.3.2 模块验证实验

为了验证提出的 Ghost 多尺度空洞卷积 (Ghost-MDCM)、空洞深度可分离卷积金字塔池化 (DDSCSPP) 和 Soft-NMS 这 3 个改进点对网络的影响, 使用的数据集为 PASCAL VOC Person, 具体结果如表 3 所示。可以发现, 网络依次添加 3 个模块后, 参数量、FLOPs 和模型体积上涨一点外, 精确率由

76.63% 上升到 80.26%, 召回率由 74.98% 上升至 80.38%, mAP 由 81.88% 上升至 83.97%。同时, 为了验证每个模块的有效性。当只加入 Ghost-MDCM 模块时, mAP 由 74.73% 上升至 81.88%; 当只加入 DDSCSPP 模块时, mAP 由 74.73% 上升至 77.27%; 当只加入 Soft-NMS 是, mAP 由 74.73% 上升至 75.28%。由此证明了每个改进模块的可行性。

表 3 模块验证结果

Table 3 Module Verification Results

名称	Ghost-MDCM	DDSCSPP	Soft-NMS	参数量 (M)	FLOPs (G)	模型体积 (MB)	精确率 (%)	召回率 (%)	mAP (%)
模型 A				5.91	3.43	22.5	76.54	62.86	74.73
模型 B	√			1.25	1.24	5.5	76.63	74.98	81.88
模型 C		√		5.72	3.28	21.6	76.59	65.52	77.27
模型 D			√	5.71	3.35	22.4	76.52	63.90	75.28
模型 E	√	√		1.44	1.39	6.4	79.68	79.74	83.42
模型 F	√	√	√	1.64	1.48	6.5	80.19	80.38	83.97

3.3.3 多环境算法精度对比实验

为了验证所提算法在不同环境中的性能,选取了四个不同环境的数据集进行实验,具体结果如表 4 所示。对比所提算法和 YOLOv4-tiny 在四个数据集上的实验结果,可以发现的是所提算法大部分的精度值高于 YOLOv4-tiny 的精度值,说明所提算法具有泛化性能,可以应对不同的环境下的检测。如图 5 所示,分别在四个数据集中选取的图片作为实验效果的对比。

比较图 a(1)、a(2)和 b(1)、b(2)可以发现,在人数较少

的情况下,所提算法没有漏检、误检的情况;比较图 a(3)、a(4)、a(10)和 b(3)、b(4)、b(10)可以发现,在人数较多且有遮挡、重叠的情况下,所提算法检测出行人效果更好;比较图 a(5)、a(9)和 b(5)、b(9)可以发现,当人的尺度大小差距较大的情况,所提算法能检测出小目标;比较图 a(6)、a(7)和 b(6)、b(7)可以发现,当背景颜色与人的衣服颜色相近或姿势不同的时候,所提算法检测目标的效果较好。

表 4 多个数据集下检测效果对比

Table 4 Comparison of detection effects under multiple datasets

数据集名称	评价指标	Yolov4-tiny	本文算法
PASCAL VOC Person 数据集	精确率	76.75	80.19
	召回率	62.81	80.38
	mAP	74.67	83.97
INRIA 数据集	精确率	90.83	98.34
	召回率	75.12	79.45
	mAP	88.79	91.89
CUHK Occlusion 数据集	精确率	90.85	89.95
	召回率	73.91	72.60
	mAP	82.50	86.75
自建数据集	精确率	74.61	95.85
	召回率	96.12	90.29
	mAP	95.73	94.98



图5 本文算法与 YOLOv4-tiny 检测效果对比图

Fig.5 Comparison of the detection effect between this algorithm and YOLOv4-tiny

3.3.4 不同网络模型对比分析

为了能够展现所提算法的整体性能，在数据集 PASCAL VOC Person 上分别测出各个网络的参数数量、计算量、模型体积和模型精度数据，具体结果如表 5 所示。选取了一些常用目标检测网络，如 YOLOv4、SSD、EfficientDet、YOLOv4-tiny。在精度损失较小的情况下，本文算法相比于其他通用性模型，参数量、计算量、模型体积大幅度降低，例如与精度最高的 YOLOv4 相比，参数数量下降了 97.4%，所提算法拥有参数量小、计算量低、模型体积小的优势，参数量下降了 97.4%，计算量下降了 95.0%，模型体积 97.6%，精度只少了 2.76%。其次，选取了一些轻量化网络，如 Mobilenet-SSDv2、YOLOv4-MobileNet V1、

YOLOv4-MobileNet V2、YOLOv4-MobileNet V3、YOLOv4-GhostNet。只比通用目标检测网络中 YOLOv4 的检测精度低，说明该算法具有良好的整体性能，与轻量化网络 YOLOv4-tiny 相比，参数数量下降了 72.2%，计算量下降了 56.8%，模型体积下降了 71.1%，精度却提升了 9.24%。所提算法在轻量化网络中整体性能都比较突出，特别是基于 YOLOv4 改进的轻量化网络，产生的预测框多于本文算法，但提升效果并不明显。除了比较通用算法外，还选取了有关行人检测的网络模型 Faster R-CNN、Cascade R-CNN 和 RetinaNet 与本文算法进行比较。同样环境下，三者中只有 Cascade R-CNN 的 mAP 要比本文算法高出 0.27%，但在参数量、FLOPs 和模型体积上，本文算法有明显优势。

表 5 不同网络模型结果对比

Table 5 Comparison of the results of different network models

模型类别	模型名称	参数量 (M)	FLOPs (G)	模型体积 (MB)	精确率 (%)	召回率 (%)	mAP (%)
通用目标 检测网络	YOLOv4 ^[10]	64.36	30.16	277.7	76.32	84.53	86.73
	SSD ^[3]	26.15	59.52	90.7	69.07	71.20	72.21
轻量化网 络	EfficientDet ^[4]	3.87	2.55	14.9	79.94	70.85	82.30
	YOLOv4-tiny ^[9]	5.91	3.43	22.5	76.54	62.86	74.73
	Mobilenet-SSDv2 ^[27]	6.07	1.55	14.5	76.31	64.55	75.91
	YOLOv4-MobileNet V1 ^[28]	12.26	4.98	51.4	75.32	80.06	82.01
	YOLOv4-MobileNet V2 ^[29]	10.37	3.78	46.8	75.89	80.00	83.04
	YOLOv4-MobileNet V3 ^[30]	11.30	3.51	54.1	70.80	73.57	82.45
行人检测 网络	YOLOv4-GhostNet ^[31]	11.00	3.25	42.7	77.54	78.10	83.11
	本文算法	1.64	1.48	6.5	75.91	80.28	83.97
	Faster R-CNN ^[32]	278	60.51	204	70.26	76.45	74.13
	Cascade R-CNN ^[33]	704	69.2	216	76.87	82.33	84.24
	RetinaNet ^[34]	64.4	30.2	244	75.64	75.62	81.54

3.3.5 不同网络模型推理速度对比分析

因为所提算法不仅需要良好的整体性能,还需要在能够运行在嵌入式的设备中,这就要求对网络模型的推理速度进行评测。为此分别准备四个性能各不相同的设备进行实验,具体图表 6 所示。在设备相同的情况下,所提算法 FPS 和图片推理耗时都是最突出的,相比 YOLOv4, FPS 提升了 70%~90%左右,推理耗时提升了 10%~30%左右。与轻量化 YOLOv4-tiny 相比,

FPS 和推理耗时略有提升。此外,同样选取了与行人检测相关 Faster R-CNN、Cascade R-CNN 和 RetinaNet 网络模型相比较。比较每秒检测图像的帧数数量和每一帧图片推理消耗的时间,本文算法领先三者较大。从总体上看该算法在四个不同硬件平台的性能拥有一定的优势。

表 6 不同性能设备推理速度对比

Table 6 Comparison of inference speed of different performance devices

模型 类型	模型名称	FPS (帧/s)				帧图片推理耗时 (ms)			
		GPU 环境		CPU 环境		GPU 环境		CPU 环境	
		RTX2070	I5-8200U	Jetson Nx	Jetson Nano	RTX2070	I5-8200U	Jetson Nx	Jetson Nano
通用 目标 检测 网络	Yolov4 ^[10]	25	0.01	5.27	1.36	37	49760	197	682
	SSD ^[3]	65	0.35	12.80	2.86	14	2854	927	344
轻量化网 络	EfficientDet ^[4]	16	0.24	5.80	3.46	54	7032	217	288
	Yolov4-tiny ^[9]	103	4.41	27.00	12.48	8.90	259	45	78
	Mobilenet-SSDv2 ^[27]	74	2.53	20.00	14.47	11	425	514	60
	Yolov4-MobileNet V1 ^[28]	51	1.30	16.30	6.03	17	827	66	191
	Yolov4-MobileNet V2 ^[29]	43	1.07	14.2	6.35	20	850	78	201
	Yolov4-MobileNet V3 ^[30]	38	1.76	10.90	6.71	25	795	89	191
行人 检测 网络	Yolov4-GhostNet ^[31]	32	1.47	9.90	5.22	31	790	112	234
	本文算法	104	8.32	28.00	15.01	10.32	113	38	64
	Faster R-CNN ^[32]	8.6	0.003	1.81	0.46	107.3	144304	571	1978
	Cascade R-CNN ^[33]	5.6	0.0019	1.18	0.30	162	217899	862	2987
	RetinaNet ^[34]	19	0.34	6.98	8.75	7.86	65593	260	299

3.3.6 轻量化网络模型在不同环境综合分析

为了能够更加全面评估本文算法在不同环境下

的整体性能,以参数量、计算量、模型体积和 FPS 这 4 精度指标为基准,对 4 个数据集 PASCAL VOC Person、

INRIA、CUHK Occlusion 以及自建数据集分别测量 mAP 值,具体图 6 所示对比图 1 表示 4 个数据集分别测得的数据,对比图 2 表示 4 个数据集测得的数据取平均。为了方便观察,图中以红色五角星表示本文算

法。从图中可以发现,本文算法在各个指标上表现优异,4 个数据集上测得 mAP 值同样领先。本文算法在 4 个数据集上的 INRIA 和 CUHK Occlusion 上测得的效果最佳。

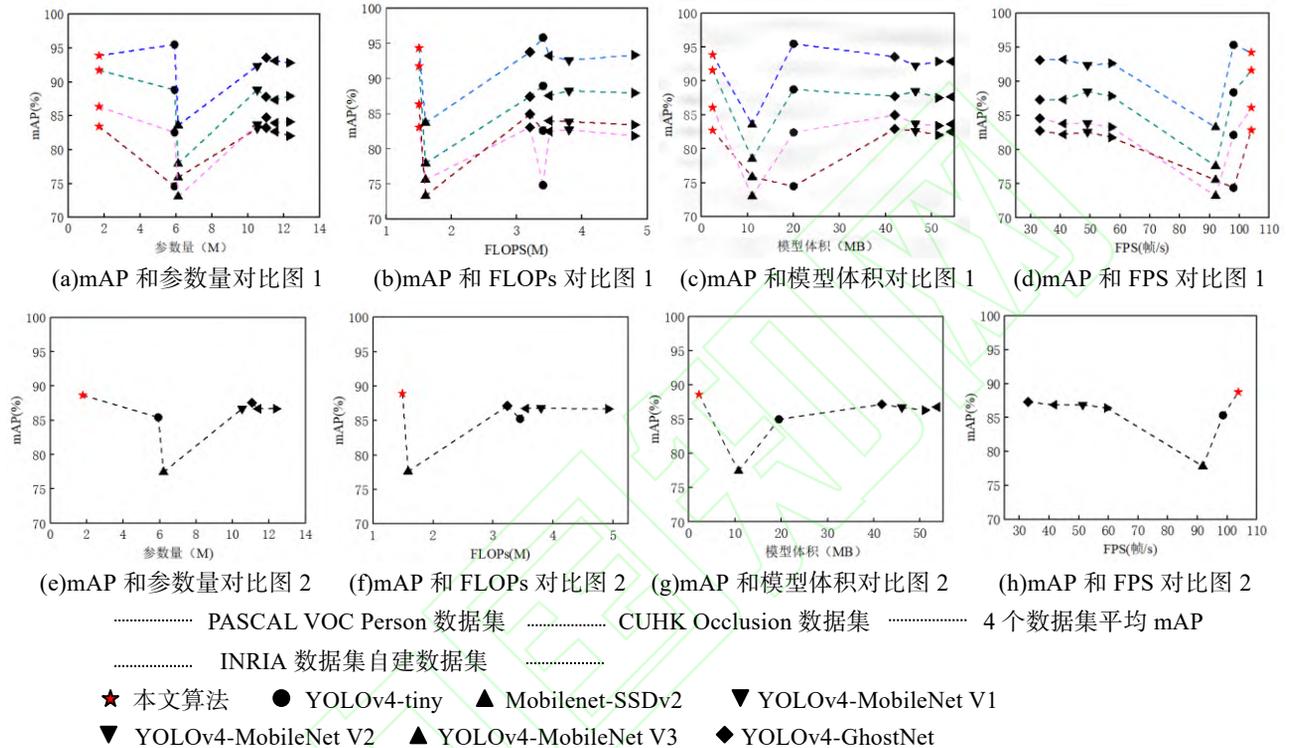


图 6 不同环境下多指标综合对比图

Fig.6 Comprehensive comparison chart of multiple indicators in different environments

4 结束语

校园实验室环境下的人员检测具有实际运用的价值和研究前景。本文算法以 YOLOv4-tiny 为基础网络进行改进,设计了一种轻量化的检测网络。一是基于 Ghost 卷积的多尺度空洞卷积模块来构建 CSP 结构作为网络特征提取的主干,既丰富了特征的提取,又实现了网络的轻量化;二是在原有网络的基础上添加了用空洞深度可分离卷积改进的空间金字塔池化模块,增强了特征之间的融合;三是用 Soft-NMS 替代了传统的 NMS 算法,降低漏检、误检率。通过在多个不同的硬件平台和数据集测得的数据分析表明,在不同环境下的整体性能相比于其他网络更加突出。因此,该算法可以运用于嵌入式设备具有突出的优势。

参考文献:

- [1] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, USA, 2014: 580-587.
- [2] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 779-788.
- [3] LIU Wei, ANGELOV D, ERHAN D, et al. SSD: Single shot MultiBox detector[C]. Proceedings of the 14th European Conference on Computer Vision, Amsterdam, The Netherlands, 2016: 21-37.
- [4] Tan Mingxing, Pang Ruoming, Le Q V. Efficientdet: Scalable and efficient object detection[C]. Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, USA, 2020: 10778-10787.
- [5] LIU W, LIAO S C, HU W D, et al. Learning efficient single-stage pedestrian detectors by asymptotic localization fitting[C]. Proceedings of the 15th European Conference on Computer Vision, Munich, Germany, 2018: 643-659.
- [6] 张明伟, 蔡坚勇, 李科, 等. 基于DE-YOLO的室内人

- 员检测方法[J]. 计算机系统应用, 2020, 29(1): 203–208.
- ZHANG M W, CAI J Y, LI K, et al. Indoor person detection method based on DE-YOLO [J]. Computer System Application, 2020, 29(1): 203–208.
- [7] 董小伟, 韩悦, 张正, 等. 基于多尺度加权特征融合网络的地铁行人目标检测算法[J]. 电子与信息学报, 2021, 43(7): 2113–2120. DONG X W, HAN Y, ZHANG Z, et al. Subway pedestrian object detection algorithm based on multi-scale weighted feature fusion network[J]. Journal of Electronics and Information, 2021, 43(7): 2113–2120.
- [8] 苏杨, 卢翔, 李琨, 等. 基于轻量深度学习网络的机房人物检测研究[J]. 工业仪表与自动化装置, 2021(1): 100–103. SU Y, LU X, LI K, et al. Research on Person Detection in Computer Room Based on Lightweight Deep Learning Network [J]. Industrial Instrumentation and Automation Device, 2021(1): 100–103.
- [9] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. Scaled-YOLOv4: Scaling cross stage partial network[C]. Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, USA, 2021: 13024–13033.
- [10] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection[EB/OL]. <https://arxiv.org/abs/2004.10934v1>, 2020.
- [11] 方梦瑞, 吕军, 阮建云, 等. 基于改进YOLOv4-tiny的茶叶嫩芽检测模型[J]. 茶叶科学, 2022, 42(04): 549–560. FANG M R, LYU J, RUAN J Y, et al. A tea sprout detection model based on improved YOLOv4-tiny[J]. Tea Science, 2022, 42(04): 549–560.
- [12] 郭奕裕, 周笋鱼. 安全帽佩戴检测网络模型的轻量化设计[J/OL]. 计算机工程: 1-12[2022-09-24]. 郭奕裕, 周笋鱼. 安全帽佩戴检测网络模型的轻量化设计[J/OL]. 计算机工程: 1-12[2022-09-24].
GUO Y Y, ZHOU L Y. Lightweight design of helmet wearing detection network model[J/OL]. Computer Engineering: 1-12[2022-09-24].
- [13] 王斌, 李靖, 赵康, 等. 面向火焰快速检测的轻量化深度网络研究 [J]. 计算机工程与应用, 2022, 58(17): 256–262. WANG B, LI J, ZHAO K, et al. Research on lightweight deep network for rapid flame detection[J]. Computer Engineering and Applications, 2022, 58(17): 256–262.
- [14] 郭克友, 李雪, 杨民. 基于轻量化YOLOv4的交通信息实时检测方法 [J/OL]. 计算机应用 : 1-9[2022-09-24]. <http://kns.cnki.net/kcms/detail/51.1307.TP.20220418.1024.003.html>
GUO K Y, LI X, YANG M. Real-time detection method of traffic information based on lightweight YOLOv4 [J/OL]. Computer Applications: 1-9[2022-09-24]. <http://kns.cnki.net/kcms/detail/51.1307.TP.20220418.1024.003.html>
- [15] 张明路, 郭策, 吕晓玲, 等. 改进的轻量化YOLOv4用于电子元器件检测 [J]. 电子测量与仪器学报, 2021, 35(10): 17–23.
ZHANG M L, GUO C, LYU X L, et al. Improved lightweight YOLOv4 for electronic component testing[J]. Journal of Electronic Measurement and Instrumentation, 2021, 35(10): 17–23.
- [16] 程浩然, 王薪陶, 李俊燃, 等. 改进YOLOv4-tiny的疫情协同口罩佩戴检测方法 [J/OL]. 计算机工程与应用 : 1-16[2022-09-24]. <http://kns.cnki.net/kcms/detail/11.2127.TP.20220812.1448.010.html>
CHENG H R, WANG X T, LI J R, et al. Improving YOLOv4-tiny's epidemic collaborative mask wearing detection method[J/OL]. Computer Engineering and Applications: 1-16[2022-09-24]. <http://kns.cnki.net/kcms/detail/11.2127.TP.20220812.1448.010.html>
- [17] 万伟彤, 李长峰, 朱华波, 等. 轻量化CenterNet网络的二维条码定位算法 [J]. 电子测量与仪器学报, 2022, 36(05): 128–135. WAN W T, LI C F, ZHU H B, et al. Two-dimensional barcode positioning algorithm for lightweight CenterNet network[J]. Journal of Electronic Measurement and Instrumentation, 2022, 36(05): 128–135.
- [18] HAN K, WANG Y H, TIAN Q, et al. GhostNet: more features from cheap operations[C]. Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, USA, 2020: 1577–1586.
- [19] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(4): 834–848.
- [20] YU F, KOLTUN V, FUNKHOUSER T. Dilated residual networks[C]. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 636–644.
- [21] Szegedy C, Liu W, Jia Y, et al. Going Deeper with Convolutions[J]. IEEE Computer Society, 2014.
- [22] SANDLER M, HOWARD A, ZHU Menglong, et al. MobileNetV2: Inverted residuals and linear bottlenecks[C]. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 4510–4520.
- [23] HE K M, ZHANG X Y, REN S Q, 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.
- [24] 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(2): 303–338.
- [25] DALAL N and TRIGGS B. Histograms of oriented gradients for human detection[C]. Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, USA, 2005: 886–893.

- [26] OUYANG W L, WANG X G. A discriminative deep model for pedestrian detection with occlusion handling[C]. Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, USA, 2012: 3258–3265.
- [27] CHIU Y C, TSAI C Y, RUAN M D, et al. MobilenetSSDv2: An improved object detection model for embedded systems[C]. Proceedings of 2020 International Conference on System Science and Engineering, Kagawa, Japan, 2020: 1–5.
- [28] LIU J, LIU L Z. Helmet wearing detection based on YOLOv4-MT[J]. Proceedings of the 2021 4th International Conference on Robotics, Control and Automation Engineering, Wuhan, China, 2021: 1–5.
- [29] FANG L F, WU Y Q, LI Y H, et al. Ginger seeding detection and shoot orientation discrimination using an improved YOLOv4-LITE network[J]. Agronomy, 2021, 11(11): 2328.
- [30] WANG S Y, CHEN T, LV X Y, et al. Forest fire detection based on lightweight Yolo[C]. Proceedings of the 2021 33rd Chinese Control and Decision Conference, Kunming, China, 2021: 1560–1565.
- [31] WANG H X, GE H Y, LI M X. PFGYOLO: A safety helmet detection based on YOLOv4[C]. Proceedings of the 2021 IEEE 5th Information Technology, Networking, Electronic and Automation Control Conference, Xi'an, China, 2021: 1242–1246.
- [32] 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.
- [33] Cai Z, Vasconcelos N. Cascade R-CNN: Delving into high quality object detection[J]. Computer Vision and Pattern Recognition, 2017(6): 798–802.
- [34] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, PP(99): 2999–3007.