

文章编号: 2095-2163(2023)09-0189-05

中图分类号: TP391

文献标志码: A

# 基于 ECA 和 BIFPN 的低照度环境下的行人目标检测算法

相敏月, 涂振宇, 孙逸飞, 方强, 马飞

(南昌工程学院 信息工程学院, 南昌 330000)

**摘要:** 针对在低照度环境下多尺度行人目标检测准确率低的问题, 本文提出了一种基于改进 YOLOv5s 的行人目标检测模型 BE-YOLOv5s。首先, 在 YOLOv5s 的主干网络中融入 ECA 通道注意力机制, 突出目标特征同时抑制低照度环境的干扰; 其次, 引入加权双向特征金字塔 BIFPN, 增强特征融合, 提升行人检测精度; 最后, 采用可见光图像和红外图像这两组数据进行对比研究。实验结果表明, 改进后的 BE-YOLOv5s 模型在两种数据集上的平均精度均值  $mAP$  均有所提升, 同时保持了原算法的高实时性。

**关键词:** 行人检测; 注意力机制; 加权双向特征金字塔; 低照度环境; YOLOv5s

## Pedestrian target detection algorithm in low illumination environment based on ECA and BIFPN

XIANG Minyue, TU Zhenyu, SUN Yifei, FANG Qiang, MA Fei

(School of Information Engineering, Nanchang Institute of Technology, Nanchang 330000, China)

**【Abstract】** Aiming at the low accuracy of multi-scale pedestrian target detection in a low illumination environment, a pedestrian target detection model based on improved YOLOv5s is proposed. Firstly, the convolutional attention module ECA attention mechanism is integrated into YOLOv5s backbone network to highlight target characteristics and suppress interference from a low illumination environment. Then replace the original feature pyramid network, PANet, and introduce weighted bi-directional feature pyramid BIFPN to enhance feature fusion and improve pedestrian detection accuracy. Finally, the visible image and infrared image are used for comparative study. The experimental results show that the average precision mean  $mAP$  of the improved BE-YOLOv5s algorithm model on both datasets is improved, and the high real-time performance of the original algorithm is maintained.

**【Key words】** pedestrian detection; attention mechanism; weighted bidirectional feature pyramid; low illumination environment; YOLOv5s

## 0 引言

行人检测是目标检测的重要研究领域之一, 在智能交通、视频监控及无人机等方面的应用十分广泛。

传统的行人检测方法主要依赖于人工提取特征的方式, Dalal 等人<sup>[1]</sup>于 2005 年开发了一种使用方向梯度直方图 (Histogram of Oriented Gradient, HOG) 特征和支持向量机 (Support Vector Machine, SVM) 分类器的行人检测模型。该模型首先使用滑动窗口在图像中识别候选区域; 其次, 提取 HOG 特征并使用 SVM 对其进行分类; 最后, 使用极大抑制

方法将输出结果组合在一起。Felzenszwal 等人<sup>[2]</sup>在 2008 年引入了可变形零件模型 (Deformable Part Model, DPM) 用于行人检测。DPM 使用 HOG 特征并将图像分成几个部分, DPM 可以使用组件的组合来检测行人, 并能够对行人的可变形部分建模, 从而更准确地检测不同大小和形状的行人。这些模型存在明显的局限性, 手工特征提取单一, 难以适用于遮挡、姿态变化和低照度等复杂环境下的行人目标检测, 导致不同程度的漏检和误检等问题。

近年来, 深度学习在行人检测中逐渐得到广泛的应用。这种方法具有强大的表征能力, 能够解决传统方法需要人工提取特征的问题。目标检测从阶

**基金项目:** 江西省科技厅重点研发计划 (20151BBE50077); 江西省水利厅科技项目 (KT201639)。

**作者简介:** 相敏月 (1998-), 男, 硕士研究生, 主要研究方向: 机器视觉; 涂振宇 (1966-), 男, 硕士, 教授, 主要研究方向: 水信息采集与处理、信息系统开发。

**通讯作者:** 涂振宇 Email: Tzy669@126.com

**收稿日期:** 2022-10-09

段上分为两种,一阶段和二阶段。一阶段主要包括快速区域卷积神经网络(Fast Region-based Convolutional Neural Network, Fast R-CNN)和更快速的区域卷积神经网络(Faster Region-based Convolutional Neural Network, Faster R-CNN)等,这类网络预先回归一次目标候选框,再利用网络对候选框进行分类和回归,虽然精度较高但检测时间过长。二阶段主要包括单激发多框探测器(Single Shot MultiBoxDetector, SSD)和YOLO(You Only Look Once)等为代表,只进行一次分类和定位,大大提高了检测速度,但同时也导致了精度较差。何自芬<sup>[3]</sup>等针对辅助驾驶中夜间小目标红外行人检测精度低的问题,提出在网络中添加空间金字塔池化模块与更小的感受野的检测层,来增强网络输出特征图的表征能力;郝帅<sup>[4]</sup>等通过构建分层注意力映射模块来增强行人特征表达能力;李传东<sup>[5]</sup>以轻量级LFFD(Light and Fast Face Detector)网络为基础,由两级改进网络组合,提高了检测精度。但是在低照度环境下,这些研究依旧存在不同程度的漏检问题。

深度学习的行人检测方法大多应用于可见光下的场景,针对低照度等复杂环境下的检测,往往效果

较差。可见光图像的优势在于依据物体的反射率的不同进行成像,光谱信息较多,分辨率较高,图像背景比较丰富,但易受到外界环境因素的影响,在低照度等复杂环境下不能正常工作。而在红外图像中,受光照条件的影响较少,更容易识别出行人位置。

本文改进YOLOv5s的主干网络,加入通道注意力机制ECA(Efficient Channel Attention),加强网络对行人特征的初步提取;在颈部网络中引入加权双向特征金字塔(Bidirectional Feature Pyramid Network, BIFPN),通过残差连接增强特征的融合能力;最后,采用公开的韩国科学技术院KAIST多光谱行人检测数据集作为实验数据,进行模型性能测试,并与YOLOv5其他模型进行对比。

## 1 YOLOv5s 模型原理

YOLOv5通过调整两个参数,即网络深度和特征图宽度划分出多个模型,其中YOLOv5s深度最小,特征图的宽度最小,是当前一种实时性和准确性俱佳的行人检测模型,并且在多尺度目标检测中具有良好的效果。所以本文采用YOLOv5s模型,模型结构如图1所示。

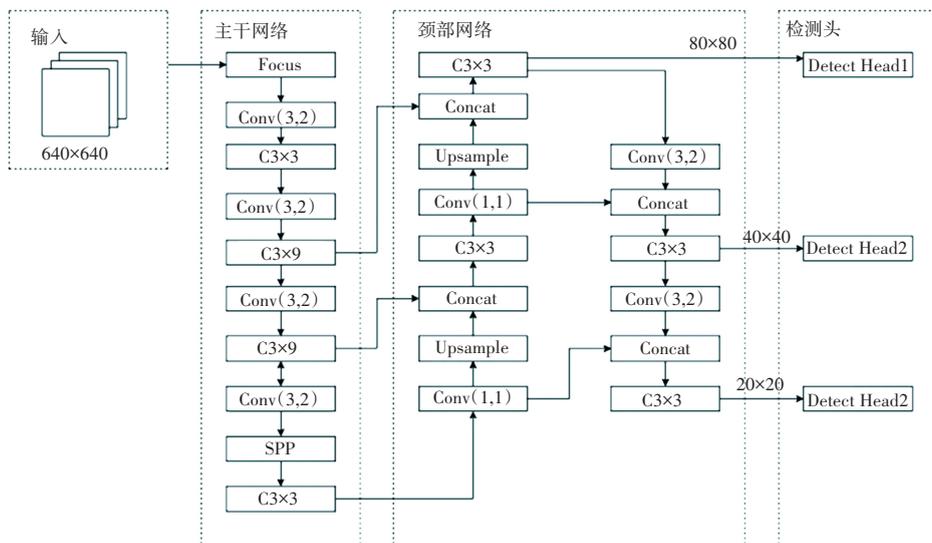


图1 YOLOv5s 模型结构

Fig. 1 YOLOv5s model structure

整个模型结构主要包含4个部位,分别为输入端(Input)、主干网络(Backbone)、颈部网络(Neck)和头部(Head)检测模块。检测模块相对于YOLOv3和YOLOv4没有变化;主干网络(Backbone)主要是用于提取输入图像的特征,Foucs模块对图片进行切片操作,使网络提取到更加充分的特征信息;颈部网络主要用于生成特征金字塔,增强网络模型对不

同尺度物体的检测能力,实现对同一物体不同尺寸和尺度的识别。YOLOv5s在特征金字塔网络(Feature Pyramid Networks, FPN)结构的基础上参考路径聚合网络(Path Aggregation Network, PANet),实现了多尺度特征融合,增强了特征的表达能力。

## 2 改进 YOLOv5s 模型

在低照度环境下,行人检测或多或少会存在漏检和误检的问题,本文在 YOLOv5s 的基础上,在主干网络中插入 ECA 通道注意力机制,提升模型对低照度环境下行人细节的提取能力;在颈部网络中,用 BIFPN 网络来代替 PANet 网络,使得模型可以更精确的识别行人目标,加快特征融合。改进后的 BE-YOLOv5s 模型结构如图 2 所示。

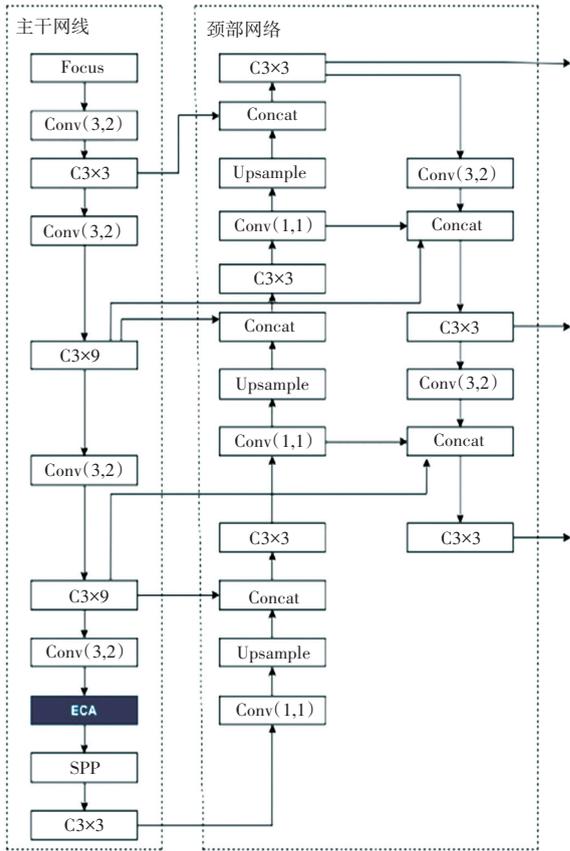


图 2 改进后的 BE-YOLOv5s 模型结构

Fig. 2 Improved BE-YOLOv5s module structure

### 2.1 主干网络改进

在神经网络中加入不同的通道注意力机制,可以提升模型的检测精度,更准确的识别和定位在低照度环境下的行人目标。注意力机制的原理是根据权重系数,重新加权求和。注意力机制的本质在于对不同的任务可以根据输入进行特征匹配,ECA 通道注意力机制有效的减少了参数计算量,提升了检测速度。

ECA 通道注意力机制的工作原理如图 3 所示。首先,剔除原来的压缩和激励 (Squeeze - and Excitation, SE) 模块中的全连接层,将输入特征图进行全局平均化池操作;其次,进行卷积核大小为  $k$  的

一维卷积操作,使用 Sigmoid 激活函数生成通道权重;最后,将特征图与通道权重相乘,得到输出特征图。同时 ECA 通道注意力机制将原来 SE 模块中的多层感知机模块转变为一维卷积形式,降低了参数计算量,实现了跨通道交互,用更少的计算成本提高检测网络的性能。

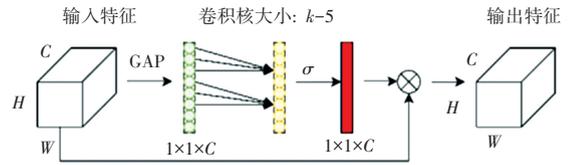


图 3 ECA 通道注意力机制

Fig. 3 ECA Channel attention mechanism

在 YOLOv5 提取行人的初始特征过程中,由于受到低照度环境的影响,特征显示不足,本文在主干网络的最后一个 CSP (Cross Stage Partial) 模块后加入 ECA 通道注意力机制,控制了参数量且增强了对行人特征的提取能力。

### 2.2 特征金字塔改进

引入 BIFPN 加权双向特征金字塔,该结构多次使用特征网络层,进行加权特征融合。对于不同分辨率特征的融合,BIFPN 为每个输入添加额外的权重,并让网络区分不同特征的重要程度,结构设计如图 4 所示。

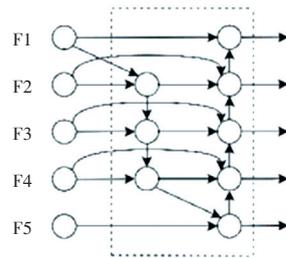


图 4 加权双向特征金字塔

Fig. 4 Bidirectional Feature Pyramid

本文在 YOLOv5s 的颈部采用 BIFPN 网络,快速进行多尺度特征融合,提升检测效果。

## 3 实验与结果分析

### 3.1 实验环境

实验采用 Pytorch 深度学习框架进行网络模型部署,整体基于 Windows10 操作系统,CPU 为 AMD Ryzen 5 3600X 处理器,显卡为 NVIDIA GeForce RTX 2070S(8 G)。

### 3.2 实验设计

本文网络模型训练所用实验数据来源于韩国科学技术院公开的 KAIST 数据集,抽取 2 000 张可见

光图像,以及与之对应的 2 000 张红外图像作为数据集,按照 8 : 1 : 1 的比例,划分训练集、验证集和测试集。

### 3.3 实验评价指标

本文主要采用准确率 ( $P$ , Precision)、召回率 ( $R$ , Recall)、平均精度均值 ( $mAP$ , mean Average Precision) 以及推理时间作为模型评价指标。 $P$  和  $R$  的计算公式如式(1)和式(2):

$$P = \frac{TP}{TP + FP} \quad (1)$$

$$R = \frac{TP}{TP + FN} \quad (2)$$

其中,  $TP$  表示正例被正确预测;  $FP$  表示负例被错误预测为正例;  $FN$  表示正例被错误预测。

$mAP$  是对  $P$  和  $R$  的一种综合处理指标,表示  $PR$  曲线下的面积。推理时间代表检测每个图像需要消耗的时间。

### 3.4 实验数据分析

与 YOLOv5n、YOLOv5l、YOLOv5s 模型进行实验对比,在可见光数据集和红外数据集上的实验结果见表 1、表 2。

表 1 可见光数据集性能指标对比

Tab. 1 Comparison of performance indicators for visible light datasets

Model	$P$	$R$	$mAP_{0.5}$	$mAP_{0.5:0.95}$	时间/s
YOLOv5n	0.679	0.856	0.702	0.285	0.011
YOLOv5l	0.849	0.837	0.738	0.341	0.021
YOLOv5s	0.826	0.838	0.742	0.370	0.011
BE-YOLOv5s	0.924	0.816	0.780	0.377	0.013

表 2 红外数据集性能指标对比

Tab. 2 Comparison of performance indicators for infrared datasets

Model	$P$	$R$	$mAP_{0.5}$	$mAP_{0.5:0.95}$	时间/s
YOLOv5n	0.753	0.876	0.729	0.312	0.011
YOLOv5l	0.869	0.873	0.757	0.359	0.021
YOLOv5s	0.857	0.880	0.796	0.388	0.011
BE-YOLOv5s	0.968	0.858	0.823	0.412	0.013

由表 1 和表 2 可见,在两种不同的数据集上,改进后的 BE-YOLOv5s 模型相比于改进前,均大幅提升了检测准确率  $P$ ,由于  $P$  和  $R$  之间存在一定的相关性,所以难以避免地会使检测召回率  $R$  稍有降低,改进后的模型在两种数据集上均显著提升了  $mAP$ 。推理时间方面,改进后模型的检测时间相比于改进前虽有所提高,但仍满足实时性要求。YOLOv5l 和 YOLOv5n 是通过调整 YOLOv5 不同的网络深度和宽度这两个参数得到的模型,YOLOv5n

的两个参数小于 YOLOv5s,其检测速度更快,但精度更差。YOLOv5l 的两个参数均大于 YOLOv5s,其检测速度更慢,但精度更高。由此可见,BE-YOLOv5s 模型在提升行人检测准确性的同时,保持了原模型的检测速度。在两种数据集上训练时,4 种模型在验证集上的  $mAP_{0.5:0.95}$  对比如图 5 所示。

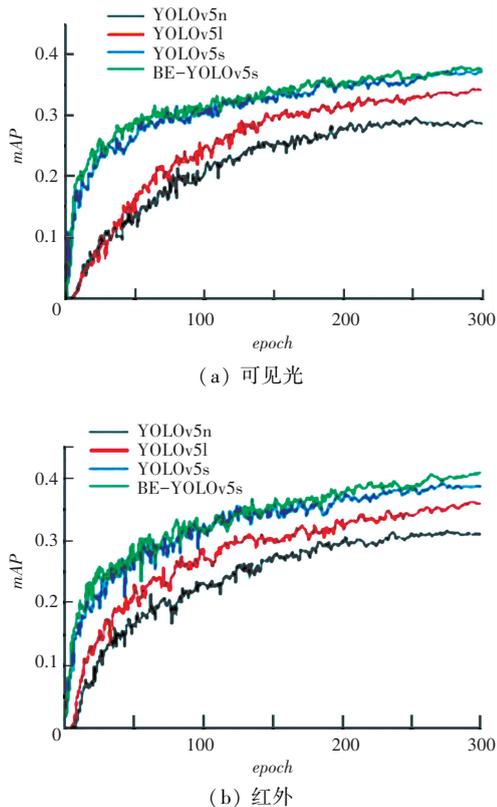


图 5 4 个模型  $mAP_{0.5:0.95}$  对比

Fig. 5 Comparison of four models  $mAP_{0.5:0.95}$

另一方面,从表 2 可见,红外数据集上的各项精度指标均高于可见光数据集,推理时间满足实时性要求。由此可见,红外图像增强了行人目标与背景信息之间的特征差异,提升行人检测的准确性。在 YOLOv5s 模型的部分测试集的可见光图像与红外图像的行人检测结果如图 6 所示,目标框上的数字表示置信度。

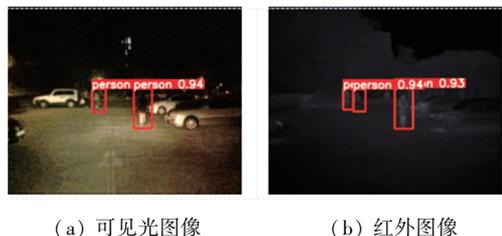


图 6 YOLOv5s 模型的检测结果

Fig. 6 YOLOv5s model detection performance

由图6可见,由于夜晚光照条件不足,YOLOv5s模型在可见光图像上检测效果较差,可见光图像中的3个位于光线较暗处的行人均未检测出,而在红外图像中,依旧有两位行人未检测出。改进后的BE-YOLOv5s模型下的行人检测结果如图7所示,可见两种图像中均可准确检测出所有行人。综合分析,BE-YOLOv5s模型在低照度的环境下检测效果较好。



(a) 可见光图像 (b) 红外图像

图7 BE-YOLOv5s模型的检测结果

Fig. 7 BE-YOLOv5s model detection results

## 4 结束语

本文针对YOLOv5s模型在低照度环境下对多尺度行人检测准确率低的问题,引入ECA通道注意力机制,提高模型对行人特征的提取,将原PANet

网络替换为BIFPN网络,加强了不同尺度的特征融合,得到了准确性和实时性俱佳的BE-YOLOv5s目标检测模型。在可见光数据集和红外数据集上分别进行测试,并与YOLOv5s、YOLOv5l、YOLOv5n模型进行对比,实验结果表明:改进后的BE-YOLOv5s模型在两种数据集上的*mAP*值均高于原模型,并且保持了原模型高实时性,有效提升了行人检测的精度。未来将融合可见光图像和红外图像各自的优势,不断提升检测精度。

## 参考文献

- [1] DALAL N, TRIGGS B. Histograms of oriented gradient for human detection[C]//IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, USA, 2005:20-25.
- [2] FELZENSZWALB P, MCALLESTER D, RAMANAN D. A discriminatively Trained, Multi-scale, Deformable part model [C]//IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, Alaska, USA, 2008:23-28.
- [3] 何自芬,陈光晨,陈俊松,等. 多尺度特征融合轻量化夜间红外行人实时检测[J]. 中国激光,2022,49(17):130-139.
- [4] 郝帅,高山,马旭,等. 基于跨尺度特征聚合与分层注意力映射的红外行人检测[J]. 光子学报,2022,51(6):419-435.
- [5] 李传东,徐望明,伍世虔. 基于弱显著图的实时热红外图像行人检测[J]. 红外技术,2021,43(7):658-664.
- [6] tracking [C]//Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2017: 42-49.
- [7] GATYS L A, ECKER A S, BETHGE M. Image style transfer using convolutional neural networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 2414-2423.
- [8] PARK D Y, LEE K H. Arbitrary style transfer with style-attentive networks [C]//proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 5880-5888.
- [9] LI Y, FANG C, YANG J, et al. Diversified texture synthesis with feed-forward networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 3920-3928.
- [10] GATYS L A, ECKER A S, BETHGE M, et al. Controlling perceptual factors in neural style transfer[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 3985-3993.
- [11] 余元超,雷刚,陈小旋,等. 基于EnsNet与MCGAN级联处理的字符样本扩充方法[J]. 微电子学与计算机,2022,39(6):69-78.
- [12] 刘雪峰,刘佳明,付民. 生成对抗网络扩充样本用于高光图像分类[J]. 电子测量技术,2022,45(3):146-152.
- [13] DAVIS J W, KECK M. A two-stage template approach to person detection in thermal imagery[C]//2005 Seventh IEEE Workshops on Applications of Computer Vision (WACV/MOTION'05) - Volume 1. IEEE, 2005:364-369.
- [1] 刘箴,吴馨远,陈士超,等. 典型红外精确制导武器发展分析[J]. 火箭与制导学报,2022,42(1):19-27.
- [2] 范晋祥. 精确制导对红外成像探测的新需求和新挑战[C]//2018年光学技术与应用研讨会暨交叉学科论坛论文集,2018:62-68.
- [3] 彭明松. 地面复杂环境与典型目标光学特性的数字仿真[D]. 北京:北京理工大学,2016.
- [4] KAI Z, JUN M, SILIANG S. Dynamic infrared target simulator driving and controlling system design [J]. Laser and Infrared, 2011, 41(1): 59-61.
- [5] 钱育龙,王治乐,张成标. 红外双波段目标模拟器的国内外发展现状[J]. 航空兵器,2014,283(5):23-27,36.
- [6] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [J]. arXiv preprint arXiv:1409.1556, 2014.
- [7] FAN H, LING H. Sanet: Structure-aware network for visual

(上接第188页)

容结构间实现了迁移,从而实现了大范围地面场景红外图像的快速生成。实验结果表明,所生成图像评价指标方面接近于实测红外图像,具有较高的置信度,在现有的仿真实验中具有一定实际意义。但如何对输入图像进行更有效的预处理及对生成图像进行更为准确的二次标校,仍需做为进一步研究的方向。

## 参考文献