文章编号:1001-5078(2023)01-0057-07　　·红外技术及应用·

基于 YOLOv5 算法的红外图像行人检测研究

王晓红,陈哲奇
(上海理工大学,上海 200093)

摘　要:目标检测是自动驾驶的重要前提,是与外界信息交互的重要环节。针对夜间远处行人检测识别精度低、漏检的问题,提出一种针对检测小尺寸行人的 YOLOv5-p4 的夜间行人识别模型。首先,通过增加更小目标的检测层,引入 BiFPN 特征融合机制,防止小目标被噪声淹没,使网络模型可以更聚焦于物体的细小特征;同时使用 K-means 先验框聚类出更小目标的锚框,并且使用了多尺度的数据增强方法,增加模型的鲁棒性。使用了 MetaAcon-C 激活函数与 EIoU 回归损失函数使模型收敛效果更好,提升了算法远距离行人的检测的准确率。最后在红外行人数据集 FLIR 上验证改进后的 YOLOv5-p4 模型对于行人的检测能力,实验结果表明该方法与传统方法相比,准确率从 86.9 % 提升到 90.3 %,适合用于红外图像中的行人检测。

关键词:深度学习;目标检测;YOLOv5;特征金字塔;红外行人识别
中图分类号:TP391　　**文献标识码:**A　　**DOI:**10.3969/j.issn.1001-5078.2023.01.008

Research on pedestrian detection in infrared image based on YOLOv5 algorithm

WANG Xiao-hong, CHEN Zhe-qi
(University of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract: Target detection is an important prerequisite of automatic driving and an important link of interaction with external information. Aiming at the problem of low accuracy and missing detection of distant pedestrians at night, a nighttime pedestrian recognition model of YOLOv5-p4 for detecting small sized pedestrians is proposed in this paper. Firstly, by adding a detection layer of smaller targets and introducing a BiFPN feature fusion mechanism to prevent small targets from being drowned by noise, the network model can be more focused on the small features of the object. At the same time, K-means prior frames are used to cluster anchor frames of smaller targets, and multi-scale data enhancement method is used to increase the robustness of the model. MetaAcon-C activation function and EIoU regression loss function are used to improve the model convergence effect and improve the accuracy of long-distance pedestrian detection algorithm. Finally, the improved YOLOv5-p4 model for pedestrian detection is verified on the infrared pedestrian data set FLIR. The experimental results show that compared with the traditional methods, the accuracy of this method is improved from 86.9 % to 90.3 %, which is suitable for pedestrian detection in infrared images.

Keywords: deep learning; target detection; YOLOv5; feature pyramid; infrared pedestrian recognition

1　引　言

行人识别作为自动驾驶发展过程中的重要一环必须兼顾准确性和实时性。同时,不同时间段对应的行人检测使用的检测方法也不相同。传统

作者简介:王晓红(1971 -),女,博士,教授,彩色图像处理与机器视觉。E-mail:wang_keyan@163.com
收稿日期:2022-02-11;**修订日期:**2022-03-29

的行人检测一般是在白天通过摄像头采集图片并进行检测,但是到了夜间,由于光照因素的缺失,往往无法满足检测需求[1]。除了光照因素,远距离的行人由于本身目标较小,存在纹理细节缺失,信噪比低的情况[2],这也是当前检测阶段的重大阻碍。

红外成像设备主要是根据物体发射的红外电磁波进行成像,并且利用目标物体和背景的温度差异的原理生成图像,红外检测也可以应用于不同的场合,例如安全监测、军事制导等方面[3-4]。传统的夜间红外行人检测算法往往依赖的是人工设计的特征。具有代表性的有帧间差的方法[5]、方向梯度直方图(Histograms of Oriented Gradients,HOG)、尺度不变特征(Scale Invariant Feature Transform,SIFT)、Haar-like[6]以及局部二值模式[7](Local Binary Pattern,LBP)等。

近年来,基于深度学习的目标检测方法在飞速发展,深度学习的方法不需要太多的人工操作,同时又可以在准确度上有较大的突破,因此基于深度学习的检测方法也在慢慢取代传统方法。目前的主流深度学习检测方法主要分为两类:

(1)两阶段检测算法

在早期深度学习技术发展进程中,主要都是围绕分类问题展开研究,这是因为神经网络特有的结构将概率统计和分类问题结合,提供一种直观易行的思路。例如 Ross B. Girshick 提出 R-CNN 算法[8],其算法结构也成为后续两阶段的经典结构。在此基础上,又有 Faster R-CNN[9]、Mask R-CNN 等。

(2)单阶段检测算法

以 R-CNN 算法为代表的两阶段检测算法由于 RPN 结构的存在,虽然检测精度较高,但是其速度却遇到瓶颈,难以满足部分场景实时性的需求。因此出现一种基于回归方法的单阶段的目标检测算法。单阶段算法不需要生成候选区域,同时在保证一定准确率的前提下,具有较快的检测速度。代表算法有 YOLO[10](You Only Look Once,YOLO)算法和 SSD[11](Single Shot multibox Detector,SSD)算法。

目前行人检测需要面对的问题主要有两个,一个是由于不同环境导致特征量获取困难,另一个是由于行人远近尺度不同而出现的漏检情况[12-14]。基于行人检测要求的准确性与实时性,本文基于 YOLOv5-p4 模型来实现驾驶辅助系统中的红外行人检测[15],针对图片中存在的小目标图像,在原先的 3 个检测层的基础上上增加了一个的检测层,加深特征提取的网络深度,强化小目标的检测精度,防止出现误检、漏检的情况。同时,针对远距离行人尺度较小的特点,使用 K-means 聚类算法重新聚类出针对小目标的先验框,通过改良的特征融合机制(BiFPN[16])进行更深层次的特征融合,获取更多的特征量。考虑预测框与真实框边长的差值修改了损失函数,使用自适应激活函数使模型更易收敛,使用多尺度训练方法进行优化,最终得到 YOLOv5-p4 模型。

2 改进的 YOLOv5 网络模型

2.1 YOLOv5 特征金字塔结构改进

原版 YOLOv5 模型的特征金字塔是对输入图片进行特征提取,分别使用输入图片的 8、16、32 倍下采样的结果输入检测层进行特征检测。在注意力金字塔(Pyramid Attention Network,PAN)结构中,不仅使用了来自上一层所提取的特征,还会与特征金字塔(Feature Pyramid Networks,FPN)结构中相对应的特征层连接,从而获得更多特征。在 FPN 和 PAN 结构中,越上层的部分感受野越广,单个特征值所涵盖的像素越大,所对应的大目标的特征也越容易被学习,但是小目标也被同化到背景中甚至被当成噪声,因此该网络结构对于小目标的检测效果会较差。

在本文所使用的数据集中,由于行人在图片中的像素大小均值较小,同时也存在大量远处行人并行的情况,原版的 YOLOv5 极小目标的检测效果并不好。因此,本文在 YOLOv5 的特征提取部分多进行了一次上采样,额外生成了一个专用于极小目标检测的特征层并且将其与主干网络对应的特征层相连接,从而提高了极小目标的检出率。

随着网络的层数的增加,每次下采样,有可能会导致图像中的小目标被淹没,本文采用改进的 BiFPN 特征融合机制,具体原理如图 1 所示。在主干网络的特征提取过程中出现的不同大小的特征层会分别与 FPN 以及 PAN 结构中对应的特征层进行双向结合,最终输出的融合结果将包含高级语义以及低级语义,以达到获得更多的特

征的目的。

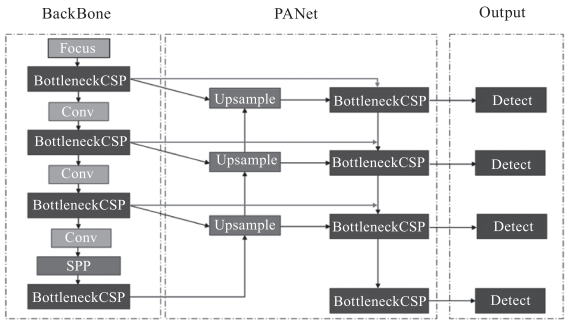


图 1　基于改进 BiFPN 改进的 YOLOv5-p4 网络结构

Fig. 1 Improved YOLOv5-p4 network structure based on improved BiFPN

原先网络用于检验的三个特征层分别是通过主干网络与 FPN、PAN 结构所获得的三个不同下采样倍率的图片，每个特征层分别对应了三个预设定的先验框，针对不同大小的物体有不同的检出效果。本文在添加了一个新的检测层和检测目标较小的基础上，使用 K-means 算法进行重新聚类，根据分类个数随机选取聚类质心点，然后计算每个点到质心点的距离，最后将点归类到距离类别质心点最近的类[17]。待每个点分类后，再重新计算每个类的质心，获得用于针对小目标检测的先验框，改进后的锚框为如表 1 所示。

表 1　K-means 聚类后锚框的宽度与高度

Tab. 1 Width and height of anchor frame after K-means clustering

检测层尺寸	(宽度,高度)	(宽度,高度)	(宽度,高度)
20 ×20	(19,65)	(29,93)	(53,174)
40 ×40	(15,24)	(15,48)	(18,38)
80 ×80	(12,27)	(12,46)	(13,35)
160 ×160	(8,19)	(9,27)	(12,20)

2.2　BottleneckCSP 部分激活函数的改进

在主干网络特征的提取过程中，BottleneckCSP 结构会将来自上一部分的内容进行三次的 Bottleneck 堆叠操作，并将其与一条残差边进行相加、标准化、激活一系列操作，最后用一个 1 ×1 的卷积核进行卷积并且输出。BottleneckCSP 的结构如图 2 所示。

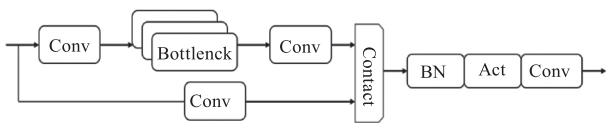


图 2　BottleneckCSP 结构

Fig. 2 BottleneckCSP structure

其中，在 BottleneckCSP 最后一个卷积前的激活函数可以提供网络的非线性建模能力。加入激活函数，深度神经网络才能具备分层的非线性映射学习能力。因此激活函数是神经网络不可或缺的一部分。原先的 YOLOv5 模型使用的是 LeakyReLu 激活函数其表达式见公式 1：

$$f(x) = \max(0.1x, x) \tag{1}$$

LeakyReLu 反应的是一种简单的线性关系，有时针对不同的输入无法做出正确的预测，因此本文使用一个自适应激活函数(Meta Activate Or Not C，MetaAcon-C[18])，该函数的通过一个 β 值来确定激活函数是线性或非线性。MetaAcon-C 的表达式见公式 2：

$$Acon-C(x) = (p_1 - p_2)x\sigma[\beta(p_1 - p_2)x] + p_2 \cdot x \tag{2}$$

其中，P_1 和 P_2 使用的是两个可学习参数，可以进行自适应的调整。对该函数求导可得：

$$\frac{d}{dx}[f_{Acon-C}(x)] = \frac{(p_1 - p_2)(1 + e^{-\beta(p_1x - p_2x)}) + \beta(p_1 - p_2)^2 e^{-\beta(p_1x - p_2x)}x}{(1 + e^{-\beta(p_1x - p_2x)})^2} + p_2 \tag{3}$$

由公式 3 可知，当 x 的值趋于正无穷或负无穷时，函数的梯度分别为 P_1 以及 P_2，这样可以有效防止神经网络出现梯度消失或者梯度爆炸的情况。

通过分别对 H，W 求平均值，然后通过两个卷积层，使得每一个通道所有像素共享一个权重，最后由 *sigmoid* 激活函数求得 β。当 β 接近 0 时，该激活函数趋近于线性，随着 β 的增大并趋于正无穷时，激活函数趋于非线性，并取值为 $\max\{x_n\}$；随着 β 的减小并趋于负无穷时，激活函数趋于非线性，并取值为 $\min\{x_n\}$。

2.3　多尺度训练

输入图片的尺寸对于网络的训练和收敛以及对于检测模型性能有相当明显的影响。多尺度训练是通过设定一个随机的缩放倍数(一般是 0.5 ~ 1.5 倍)对图片进行缩放，然后将其输入网络中进行训练。多尺度训练通过输入较大尺寸的图片可以使图片中的小特征在下采样以及卷积的过程中不容易被忽略，使网络在训练的过程中减少过拟合的情况出现，使得训练出来的模型具有较强的鲁棒性。除此

之外,使用多尺度训练后的模型在接受不同尺度的输入图片也会有检测效果。

2.4 损失函数选择

YOLOv5 原始模型使用的 IoU 损失函数为 CIoU 函数。CIoU 损失函数具有惩罚项,并且考虑到了目标与预测框之间的距离,重叠率等,使目标框的回归变得更加稳定。其中 $\rho^2(b,b^{gt})$ 表示预测框与真实框之间中心点的欧氏距离,c 表示包含预测框与真实框的对角线距离,αv 表示惩罚因子,CIoU 损失函数见公式 4:

$$L_{\mathrm{CIoU}} = 1 - IoU + \frac{\rho^2(b,b^{gt})}{c^2} + \alpha v \tag{4}$$

对于小目标的行人检测来说,一大难点是预测框是否可以在准确检测到目标的同时与目标的真实大小相一致,如果图片中存在较多小目标,有可能会出现漏检和错检的情况。CIoU 无法平衡难易样本,为了解决此问题,选择使用一个更加符合网络回归要求的损失函数 L_{EIoU},见公式 5:

$$\begin{aligned} L_{\mathrm{EIoU}} &= L_{\mathrm{IoU}} + L_{\mathrm{dis}} + L_{\mathrm{asp}} \\ &= 1 - IoU + \frac{\rho^2(b,b^{gt})}{c^2} + \frac{\rho^2(w,w^{gt})}{C_w^2} + \\ &\quad \frac{\rho^2(h,h^{gt})}{C_h^2} \end{aligned} \tag{5}$$

L_{EIoU} 主要包含三个部分:重叠损失,中心距离损失以及宽高损失。对比 CIoU,它直接将真实框与预测框的宽高进行对比,求出宽高损失,可以加速网络的收敛速度并且提高了回归精度,使得预测的结果更加准确。

3 实验与结果分析

3.1 实验数据集与预处理

本实验使用的数据来自 FLIR_ADAS_1_3 红外数据集,该数据集由红外相机拍摄,包括 10228 张图片。其中数据集中共包括行人 28151 个,汽车 46692 辆,自行车 4457 辆,以及其他的物体。本次实验使用数据集中的行人数据共 5838 张图片。在实验过程中,使用马赛克增强以及拼接的网络训练策略,丰富了检测物体的背景,同时也可以起到增加数据集的作用,使网络具有更好的泛化性。本次实验对数据集按照 8:1:1 的比例分为训练集、测试集、验证集。

3.2 实验环境

本文使用了开源的 PyTorch 深度学习框架进行训练和测试,运算平台为 64 位操作系统,16G 内存,6G 显存,NVIDIA GeForce RTX 3060 Laptop GPU 的笔记本。

3.3 评估指标

目标检测通常使用平均准确率 mAP(mean Average Precision)作为模型性能的评价指标。同时,也会参考验证集损失,防止过拟合现象出现导致模型泛化性降低。

准确率(Precision):表示模型正确检出的目标的数量与数据集中的总目标数的比值,计算见公式 6:

$$\mathrm{Precision} = \frac{TP}{TP + FP} \tag{6}$$

召回率(Recall):表示模型已检出的目标的数量与数据集中的总目标数的比值,计算见公式 7:

$$\mathrm{Recall} = \frac{TP}{TP + FN} \tag{7}$$

通过绘制训练过程的 P-R 曲线,计算 P-R 曲线与坐标轴围成的面积,该面积代表的就是这一类目标所对应的 AP 值,计算见公式 8。在本文的实验中,用于检测的目标只有一类,因此本实验的 AP 与 mAP 相同。

$$\mathrm{AP} = \int_0^1 P(R)\,dR \tag{8}$$

3.4 结果分析

3.4.1 损失函数分析

使用 tensorboard 对模型训练过程进行可视化,原版的 YOLOv5 模型和 YOLOv5-p4 模型训练损失函数结果如图 3、4 所示。

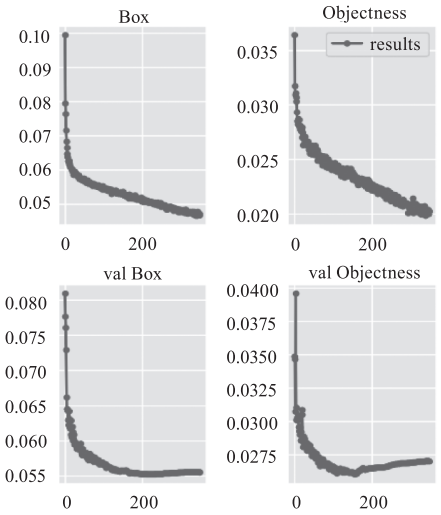


图 3 YOLOv5 的损失函数

Fig. 3 Loss function of YOLOv5

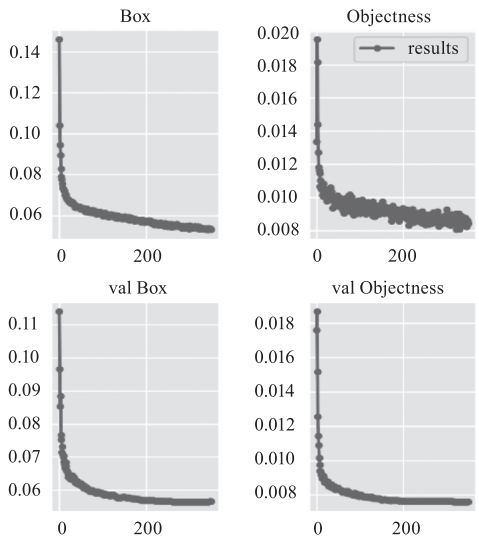


图4 YOLOv5-p4的损失函数

Fig. 4 Loss function of YOLOv5-p4

通过两个模型的损失函数进行对比和分析可以发现,两个模型在0~150 epoch训练过程中,训练损失和验证损失两者均在不断下降。原版模型的验证集的置信度损失(val Objectness)在150代之后上升,但是该模型的训练损失却仍然在下降,因此判断该模型在150代之后网络进入了过拟合的情况。模型训练进入过拟合的会导致模型的泛化性降低,因此本文在对比原模型过拟合前后的精度结果后,选择使用第160代作为模型的最终结果。对比原版的模型可以发现,YOLOv5-p4模型在150代后均未出现过拟合的情况,且置信度损失的值远低于原模型的置信度损失值。损失函数作为评价模型的优劣的标准之一,对比两个模型的损失值,可以看出优化后的模型优于原模型,特征学习的能力更强。

3.4.2 模型检测性能分析

P-R曲线中的P代表的是precision(准确率),R代表的是recall(召回率)。P-R曲线代表的是精准率与召回率的关系。由本文中准确率和召回率可以求出在0.5的交并比阈值下的平均精度mAP@0.5,在数值上等于PR曲线与坐标轴围成的面积,这是模型性能衡量的主要评价指标。P-R曲线如图5、6所示,通过该曲线可以看出原模型的mAP@0.5为86.9%,改进后的模型的mAP@0.5为90.3%。对比两个模型可以看出,YOLOv5-p4模型的平均准确率提升了3.4%,同时有更低的损失值,因此改进后的模型优于原模型。

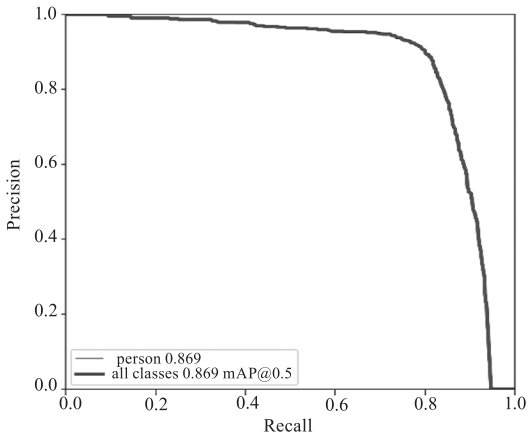


图5 YOLOv5的P-R曲线

Fig. 5 P-R curve of YOLOv5

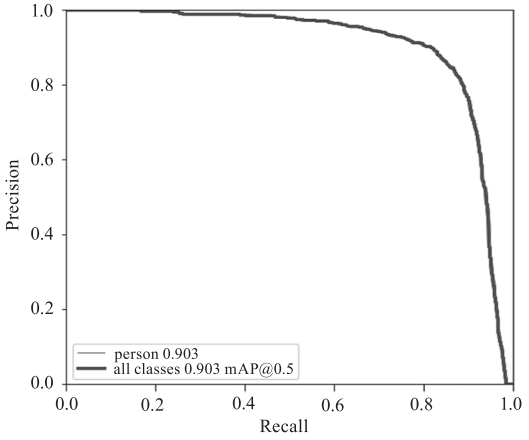


图6 YOLOv5-p4的P-R曲线

Fig. 6 P-R curve of YOLOv5-p4

3.4.3 检测效果分析

为验证模型针对不同尺度的行人的检测效果,本文在从数据集中随机抽选出200张图片,通过模型在该200张图片上的检测效果进行比较分析。将行人尺寸基于图像分辨率分为三类,分别是:小于20×20、20×20至50×50以及大于50×50。测试结果如表2、3所示。

表2 不同尺寸目标检验结果表

Tab. 2 Inspection results of different size targets

目标尺寸	算法		
	YOLOv5	YOLOv5-p4	标签标注总计
小于20×20	265	285	307
20×20~50×50	300	305	347
大于50×50	72	75	75

由表2、3可知,针对尺寸小于20×20的目标,原版YOLOv5模型的检测性能远不如YOLOv5-p4模型,YOLOv5-p4模型的检测精度相较于原

YOLOv5 模型提升了 6.5 %。针对尺寸大于 20×20 的目标,两个模型的检测性能较为接近。由此看出,YOLOv5-p4 模型在针对较小目标的检出效果远好于原模型。

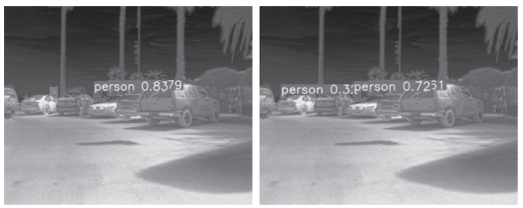
表 3 不同尺寸目标的检测精度表

Tab. 3 Table of detection accuracy of targets of different sizes

目标尺寸	算法精度	
	YOLOv5	YOLOv5-p4
小于 20×20	86.3 %	92.8 %
20×20 ~ 50×50	86.5 %	87.9 %
大于 50×50	96 %	100 %

为了更直观地看出 YOLOv5-p4 模型针对较小行人检测的有效性,本文使用同一张图片进行检测,两个模型的检验效果如图 7 所示。

(a)原图 (b)局部放大图



(c)YOLOv5检测效果图 (d) YOLOv5-p4检测效果图

图 7 检测效果图

Fig. 7 Detection effect diagram

由图 7 对于同一张图片的检测效果对比得出,原版的 YOLOv5 模型对与远处的行人存在无法检出的情况,YOLOv5-p4 模型无论是针对远处或者近处的行人均有良好的检验效果。

3.4.4 与其他算法综合对比

将本算法与 SSD 算法、YOLOv4 算法和 YOLOv5 算法在相同条件下进行对比测试,结果如表 4 所示。

由此表可以看出,相较于 SSD 算法和 YOLOv4 算法,YOLOv5-p4 模型在精度上分别提升了7.2 % 和 5.4 %,对于目标行人具有较好的检出效果。

4 结 语

本文采用改进的 YOLOv5 网络结构进行基于红外图像的行人检测。首先对 YOLOv5 网络进行改进,通过增加上采样的次数,增加一个目标检测层,提高了网络对远处较为微小的行人的识别能力;使用 BiFPN 结构,将相同尺寸的特征层相连接,提高了图片中小目标的检出率;针对小目标难以检测的情况,使用了能考虑到预测框和真实框的差距的损失函数;使用 MetaAcon-C 激活函数代替原本的激活函数,帮助模型更好的收敛;使用多尺度的模型训练方法,得到效果更好的模型。同时本文也存在一定的局限性,模型针对智能驾驶的目标检测的精度仍然未达到相当高的程度,这也是接下来的改进和研究的目标。

表 4 与其他算法精度对比表

Tab. 4 Accuracy comparison table with other algorithms

算法	Map@0.5
SSD	83.1 %
YOLOv4	84.9 %
YOLOv5	86.9 %
YOLOv5-p4	90.3 %

参考文献:

[1] Hu Junping, Sun Xi. Near infrared night pedestrian detection method based on improved Faster R-CNN[J]. Sensors and Microsystems, 2021, 40(8): 126 – 129. (in Chinese)
胡均平,孙希. 基于改进 Faster R-CNN 的近红外夜间行人检测方法[J]. 传感器与微系统,2021,40(8):126 – 129.

[2] Wang Chen, Tang Xinyi, Gao Sili. Infrared image enhancement algorithm based on human vision[J]. Laser & Infrared, 2017, 47(1): 114 – 118. (in Chinese)
王晨,汤心溢,高思莉. 基于人眼视觉的红外图像增强算法研究[J]. 激光与红外,2017,47(1):114 – 118.

[3] Gao J, Guo Y, Lin Z, et al. Infrared small target detection using multiscale gray and variance difference[C]//Chinese Conference on Pattern Recognition and Computer Vision(PRCV), Springer, Cham, 2018: 53 – 64.

[4] Zhao Ming, Zhang Haoran. An infrared target detection method based on cross domain fusion network[J]. Acta Photonica Sinica, 2021, 50(11): 331 – 341. (in Chinese)
赵明,张浩然. 一种基于跨域融合网络的红外目标检测方法[J]. 光子学报,2021,50(11):331 – 341.

[5] Su Xiaoqian, Sun Shaoyuan, Ge Man, et al. Pedestrian detection and tracking of vehicle infrared images[J]. Laser

& Infrared,2012,42(8):949 – 953. (in Chinese)
苏晓倩,孙韶媛,戈曼,等. 车载红外图像的行人检测与跟踪技术[J]. 激光与红外,2012,42(8):949 – 953.

[6] Zhang S, Bauckhage C, Cremers A B. Informed haar-like features improve pedestrian detection[C]//Computer Vision & Pattern Recognition. IEEE,2014.

[7] Watanabe T, Ito S. Two Co-occurrence histogram features using gradient orientations and local binary patterns for pedestrian detection [C]//Pattern Re-cognition. IEEE,2014.

[8] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmen-tation[J]. IEEE Computer Society,2013.

[9] Ren S, He K, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. IEEE Trans-actions on Pattern Analysis & Machine Intelligence,2017,39(6):1137 – 1149.

[10] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//Computer Vision & Pattern Recognition. IEEE,2016.

[11] Liu W. Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[J]//European Conference on Computer Vision, Springer, Cham,2016:21 – 37.

[12] Zhou Weina, Sun Lihua, Xu Zhijing. Multi-scale pedestrian real-time detection method in complex environment [J]. Journal of Electronics & Information Technology, 2021,43(7):2063 – 2070. (in Chinese)
周薇娜,孙丽华,徐志京. 复杂环境下多尺度行人实时检测方法[J]. 电子与信息学报, 2021, 43(7): 2063 – 2070.

[13] Che Kai, Xiang Zhengtao, Chen Yufeng, et al. Research on infrared image pedestrian detection based on improved Fast R-CNN [J]. Infrared Technology, 2018, 40(6): 578 – 584. (in Chinese)
车凯,向郑涛,陈宇峰,等. 基于改进 Fast R-CNN 的红外图像行人检测研究[J]. 红外技术,2018,40(6): 578 – 584.

[14] Wang Dianwei, He Yanhui, Li Daxiang, et al. Improved yolov3 infrared video image pedestrian detection algorithm [J]. Journal of Xi′an University of Posts and Telecommunications,2018,23(4):48 – 52,67. (in Chinese)
王殿伟,何衍辉,李大湘,等. 改进的 YOLOv3 红外视频图像行人检测算法[J]. 西安邮电大学学报,2018,23(4):48 – 52,67.

[15] Tan Kangxia, PING Peng, QIN Wenhu. Pedestrian detection method in infrared image based on Yolo model[J]. Laser & Infrared,2018,48(11):1436 – 1442. (in Chinese)
谭康霞,平鹏,秦文虎. 基于 YOLO 模型的红外图像行人检测方法[J]. 激光与红外, 2018, 48(11): 1436 – 1442.

[16] Tan Mingxing, Pang Ruoming, Le Quoc V. EfficientDet: scalable and efficient object detection[C]//2020 IEEE/CVF conference on Computer Vision and Poctern Recognition(CVPR). IEEE,2020.

[17] Yang Kun, Zhang Mingxin, Xian Xiaobing et al. An edge detection method based on Sobel and k-means[J]. Optical Technique,2014,40(5):394 – 398. (in Chinese)
杨昆,张明新,先晓兵,等. 一种基于 Sobel 与 K-means 的边缘检测方法[J]. 光学技术, 2014, 40(5): 394 – 398.

[18] Ma Ningning, Zhang Xiangyu, Liu Ming et al. Activate or not: learning customized activation[J]. ArXiv e-Prints, 2021,2009.04759.