

文章编号: 2095-2163(2021)08-0031-05

中图分类号: TP391.4

文献标志码: A

改进的 YOLOv4 红外图像行人检测算法

史健婷¹, 张贵强¹, 陶金², 吴林皓¹

(1 黑龙江科技大学 计算机与信息工程学院, 哈尔滨 150022; 2 黑龙江科技大学 研究生学院, 哈尔滨 150022)

摘要: 在夜视红外行人检测的定位任务中, 样本取样机器处于高点俯拍, 使得所拍摄到的行人目标体积较小。此外, 行人时刻处于活动状态, 与摄像头的距离不同, 导致检测时同类目标在图中的大小有一定的差异。基于 YOLOv4 算法, 本文提出了一种改进的 YOLOv4 红外行人检测算法, 对 YOLOv4 的网络结构进行了优化。采用形变卷积为核心组件, 构建形变特征提取模块提升对于目标特征提取的有效性; 针对形变卷积对特征提取网络模块进行优化。结果表明, 改进后的算法在整体鲁棒性、召回率、F1-Score 等评价指标方面均优于其它算法。

关键词: 红外行人检测; YOLOv4; 网络结构; 形变卷积

Pedestrian detection algorithm in infrared image based on improved YOLOv4

SHI Jianting¹, ZHANG Guiqiang¹, TAO Jin², WU Linhao¹

(1 Department of Computer and Information Engineering, Heilongjiang University of Science and Technology, Harbin 150022, China; 2 School of Graduate, Heilongjiang University of Science and Technology, Harbin 150022, China)

[Abstract] In the task of infrared pedestrian detection, the sampling machine is in the high spot top, which makes the volume of the captured pedestrian target smaller. In addition, the pedestrian is always in the active state, and the distance from the camera is different, which leads to some differences in the size of similar objects in the image. In addition, the pedestrian is always in the active state, and the distance from the camera is different, which leads to some differences in the size of similar objects in the image. Based on YOLOv4, an improved YOLOv4 infrared pedestrian detection algorithm is proposed, and the network structure is optimized. Firstly, the deformation convolution is used as the core component to construct the deformation feature extraction module to improve the effectiveness of target feature extraction; secondly, to optimize the feature extraction network module for deformation convolution. The results show that the improved algorithm is better than other algorithms in the overall robustness, recall rate, F1 score and other evaluation indicators.

[Key words] infrared pedestrian detection; YOLOv4; network structure; deformation convolution

0 引言

行人检测技术结合了数字图像处理、模式识别、计算机视觉和其他相关技术, 无论是在智慧交通, 还是在自动驾驶、人体行为分析等领域有着广阔的发展前景^[1]。行人检测技术, 是研究和判断所给的图像或在每帧视频序列中, 是否存在要检测的行人目标。近年来, 道路安全问题频繁发生, 在寻找降低交通事故发生方法的同时, 行人检测技术也受到人们的广泛关注^[2-3]。

行人检测算法可分为传统的算法和基于深度学习的算法。传统的行人检测算法有: Haar 小波特征^[4]、HOG+SVM^[5]、DPM^[6]等等。传统的行人检测主要通过人工设计方法, 对图像特征进行提取, 进行目标识别和检测, 但算法设计复杂, 权值参数难以得

到较准确的数值, 泛化能力不强。

基于深度学习的算法, 如卷积神经网络(Convolutional Neural network, CNN)^[7], 通过大量的数据, 能自动学习出目标的原始表征, 相较于手工设计的特征而言, 具有更强的判别能力和泛化能力^[8]。之后出现了一系列改进算法, 包括 Fast RCNN^[9-11]、Faster RCNN^[12]、SSD^[13]、YOLO^[14]等算法。其中, YOLOv4 网络结构简单高效, 具有易部署、运用场景广泛的特性。特别是针对红外图像下的小目标检测, 有很大的应用前景^[15]。虽然 YOLOv4 网络采用多尺度进行预测, 能够结合更好的分类器, 但是还存在识别物体的精准性差、召回率低等缺点。因此, 本文对 YOLOv4 算法进行了改进。通过与原 YOLOv4 算法进行比较, 将对红外图像行人的检测精准度(MAP)提高了 0.04%。

基金项目: 黑龙江科技大学研究生创新科研项目(YJSCX2020-212HKD)。

作者简介: 史健婷(1981-), 女, 博士, 副教授, 主要研究方向: 图像处理与模式识别、多维数据分析与可视化。

收稿日期: 2021-05-24

1 YOLOv4 算法

1.1 YOLOv4 网络结构

YOLOv4 算法以 CSPDarknet53 作为主干网络,在采用 ResNet 短跳连接的同时,增加输入输出的维度拼接,更好的实现了深浅层特征的结合,在主干网

络的顶端,加入了 SPP 模块。采用 $1 * 1, 5 * 5, 9 * 9, 13 * 13$ 最大池化的方式,进行多尺度特征融合。该模块和 PAN 结构相结合,使分辨率为 $76 * 76$ 的浅层特征向上传播,保证每个检测头都可以接受浅层特征,极大的增加了网络对小目标特征的表达力。YOLOv4 的网络结构如图 1 所示。

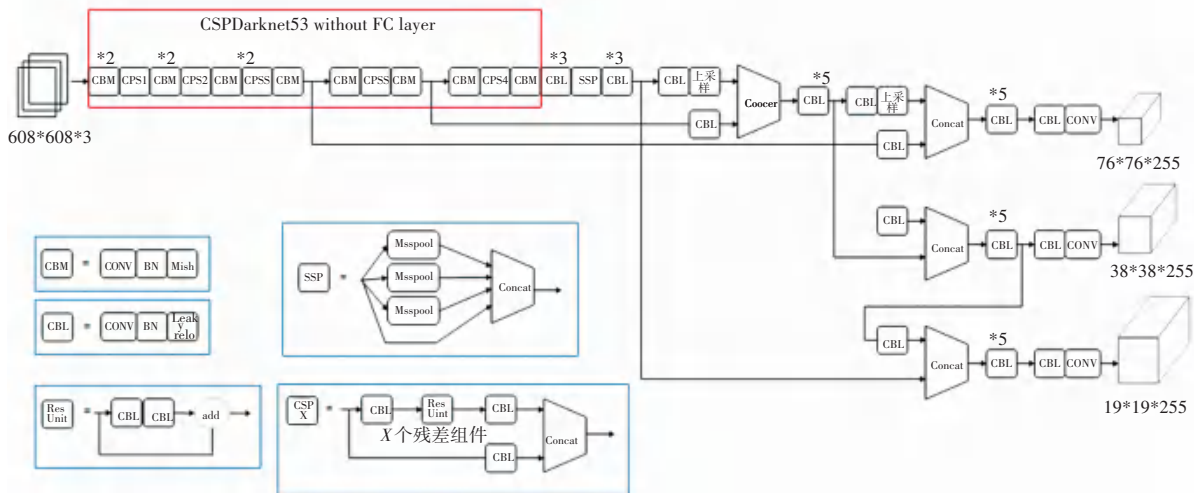


图 1 YOLOv4 网络结构图

Fig. 1 YOLOv4 network structure diagram

除了对主干网络的改进之外,YOLOv4 还引入了其它 tricks 来提升网络性能。在激活函数方面,引入 Mish 激活函数,如公式(1)所示:

$$Mish = x * \tan h(\ln(1 + e^x)), \quad (1)$$

函数图像如图 2 所示。

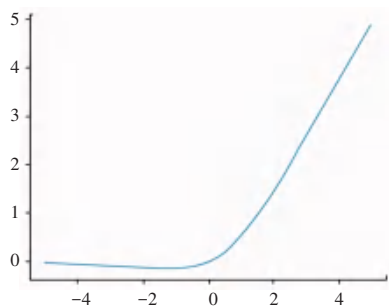


图 2 Mish 激活函数图像

Fig. 2 Mish activation function

可以看出, Mish 激活函数在 $x = 0$ 处也是光滑可导的,具有较好的泛化能力和结果的有效优化能力。在数据增强方面,采用 Mosaic 数据增强方式,对多张图片以随机缩放、随机剪裁和随机排布的方式进行拼接,大大丰富了数据集,可以让网络直接计算多张图片的数据,增加模型泛化能力。

$$CIoU = IoU - D_2^2/D_C^2 - \alpha v, \quad (2)$$

$$\alpha = v / ((1 - IoU) + v), \quad (3)$$

$$v = (4/\pi^2) (\arctan(w^g/h^g) - \arctan w/h)^2. \quad (4)$$

式中: D_2 表示预测框和目标框中心点距离, D_C 为最小外接矩形 C 的对角线距离。

1.2 注意力机制

注意力机制(Attention Mechanism)在文本分析、行人检测、外界语音处理等方面有广泛的使用。注意力机制就像人类注意力观察一样,通过相应的空间、通道等方面,从软注意力和强注意力两方面进行分析和处理。

2017 年, Jie Hu 等人通过研究,提出了一种新的框架结构——SENet (Squeeze - and - Excitation Network, 即“压缩和激励”SE 块)。SENet 通过加强所要研究的重要区域,把所要输入的图像进行卷积,然后得到 feature map 进行分析,设计出一个一维向量,作为分数来进行评价。与所要研究的图像通道一样,该向量的每个评价分数采用乘法加权方式,得到原通道的大小,这样处理提高了研究的重要区域。SE 模块的结构图如图 3 所示。

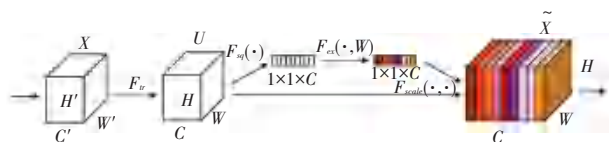


图 3 SE 模块结构图

Fig. 3 Se module structure diagram

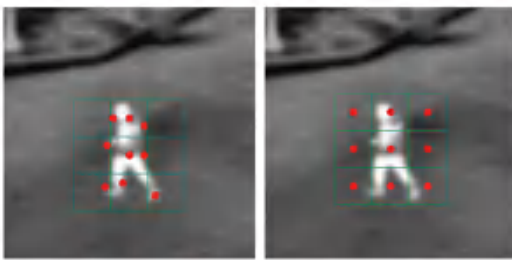
2 改进的 YOLOv4 红外行人检测算法

在 YOLOv4 中,主要采用 3x3 大小的标准卷积进行特征提取操作。标准卷积由于其感受野的形状和大小均为固定,在对小目标进行检测时,同样会对非目标区域进行特征提取操作,会导致最后卷积所提取到的特征中干扰因素较多,对检测器的预测造成较多的干扰影响。因此,基于 YOLOv4 的标准卷积思想,利用形变卷积为核心组件,构建形变特征提取模块,提升对于目标特征提取的有效性。

形变卷积与标准卷积相比,具有 3 点优势:

- (1) 感受野有效性的提升,即特征图在映射目标信息时针对性更高;
- (2) 卷积核能够适配目标位置进行采样,所提取到的特征信息与目标更匹配;
- (3) 由于形变卷积经过特征提取时,能够有效针对目标所在区域进行提取,使得特征图在网络中传递时,其稳定性(即权值参数不会突变)优于标准卷积。

形变卷积与标准卷积在进行特征提取操作时的区别如图 4 所示。



(a) 形变卷积特征提取 (b) 标准卷积特征提取
(a) Feature extraction of deformation convolution (b) Feature extraction of standard convolution

图 4 形变卷积与标准卷积特征提取对比

Fig. 4 Comparison of feature extraction between deformation convolution and standard convolution

为了增强对于目标位置信息的复用,针对 YOLOv4 的注意力机制思想,在每个尺寸的特征图,经由形变特征提取模块组后,加入 Coordinate 坐标注意力机制模块,对坐标信息进行加强。Coordinate 坐标注意力机制模块基于 SE 通道注意力机制进行优化,提取出了特征图横向与纵向的特征权值信息,再通过聚合,以达到精确的目标位置坐标显著性标记。Coordinate 坐标注意力机制模块结构如图 5 所示。

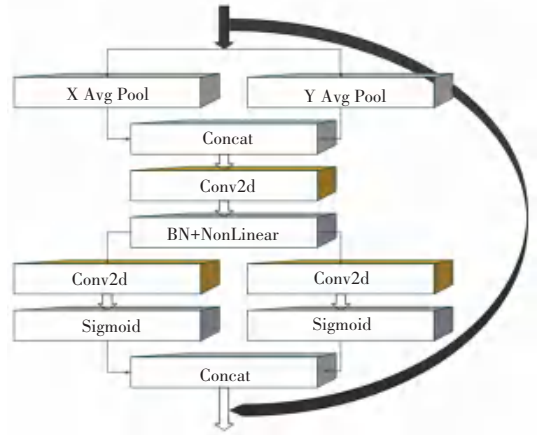


图 5 Coordinate 坐标注意力机制模块

Fig. 5 Coordinate attention mechanism module

在图 5 中,模块的工作流程主要分为两步:一是提取特征图上 X 轴与 Y 轴的特征信息;二是对提取的特征信息进行激活加权。首先,输入到模块的特征图由全局池化分解出两个方向上的一维特征,该过程基于 SE 注意力机制压缩操作进行优化。

标准全局池化计算过程为:

$$Z_c = 1/(H \times W) \sum_{i=1}^H \sum_{j=1}^W X_c(i, j), \quad (5)$$

式中: Z_c 为全局池化输出; H 为特征图的高; W 为特征图的宽; X_c 为全局池化输入。

Coordinate 坐标注意力机制则将全局池化分解为:

$$Z_c^h(w) = 1/W \sum_{0 \leq i \leq W} X_c(h, i), \quad (6)$$

$$Z_c^w(w) = 1/H \sum_{0 \leq j \leq w} X_c(j, w). \quad (7)$$

完成分解后,再对两个方向的特征图进行聚合,以此获得带有坐标信息的特征图。之后,将该特征图分别由两个二维卷积进行特征提取与激活加权计算,得到两个坐标方向的加权特征信息。将该特征信息映射到特征图中,即可反映目标在特征图中的坐标信息。

对于影响网络定位准确性的因素,最直观的表现 YOLOv4 检测器中的 anchors 组件。检测器通过 anchors 判定目标是否存在及目标的位置,即 anchors 能够对特征图的某个区域是否存在目标进行判定,同时预测目标位置。由于目标的形态大小具有不确定性,即通过手动设定的 anchors 尺寸无法准确适配目标的位置,致使在检测时存在一定的偏差。为优化 anchor 的定位准确性,在 YOLOv4 检测层中加入“Guided Anchoring”机制,通过网络自适应生成 anchors,来提高 anchors 及候选区域的质量。

不同于常规 anchors 操作,在一个坐标点上对一组 anchor 的尺寸进行预测并挑出最符合大小的一个,其值对一个 anchor 的尺寸进行预测,使得对于不规则目标的拟合性更强,召回率也更高。本文设计的网络命名为 YOLO-sd。

3 实验结果与分析

本设计实验环境配置为:软件层次上,操作系统为 Ubuntu 18.04,神经网络框架为 Darknet,CUDA 版本为 10.0,cuDNN 加速包为 7.6.4;在硬件层次上主要使用了 RTX2080ti 型号的 GPU 进行卷积计算加速。

关于红外行人检测算法评价的相关性能指标包括:交并比 IOU 、精度 ($precision$)、召回率 ($recall$) 等。

$$IOU = 2 * area(S_1 \cap S_2) / (area(S_1) + area(S_2)), \quad (8)$$

$$recall = TP / (TP + FN), \quad (9)$$

$$P = TP / (TP + FP). \quad (10)$$

式中: S_1 为红外图像预测的行人区域; S_2 为标注的行人区域; TP 为红外图像下行人区域,预测为行人正确情况; FN 为红外图像下行人区域,预测为不是行人错误情况; FP 为实际不是行人区域,但是预测此区域有行人情况。

所采用的数据集来自 OSU Thermal Pedestrian Database,通过数据清洗、预处理等操作,构成 2 100 张训练集和 500 张的测试集。将改进后的模型 YOLO-sd 与 YOLOv3、YOLOv4 以及 SSD 算法进行对比测试,测试结果见表 1。

表 1 模型检测性能对比

Tab. 1 Performance comparison of models

指标	精度 ($precision$)	召回率 ($recall$)	$F2 - 1Score$	IOU (交并比)/%	平均准确率(mAP)/%
YOLOv3	0.90	0.77	0.83	64.75	82.04
YOLO-v4	0.91	0.84	0.87	64.82	83.05
SSD	0.82	0.88	0.84	61.93	80.55
YOLO-sd(本文模型)	0.89	0.85	0.87	64.86	83.09

通过对比结果可以看出,本文提出的 YOLO-sd 算法,整体鲁棒性要优于 YOLOv3 和 YOLOv4;在召回率的对比中,YOLO-sd 优于 YOLOv3 和 YOLOv4,说明对于目标的查全率更好,且 IOU 数值也更优。YOLO-sd 与 SSD 对比,YOLO-sd 的精度、平均准确率 (map)、 $F2 - 1Score$ 要优于 SSD;其它指标,召回率和交并比略低于 SSD,综合反映了对于主干网络及检测网络部分的优化,在提升网络性能方面有巨大帮助。网络优化性能 pr 曲线如图 6 所示。

利用 YOLO-sd 的实际测试结果如图 7 所示。



图 7 测试结果

Fig. 7 Test result

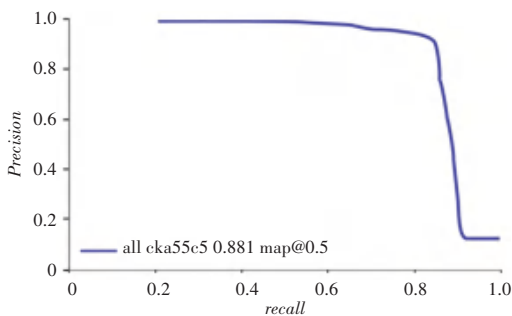


图 6 pr 曲线对比

Fig. 6 Comparison of PR curves

4 结束语

本文提出了一种基于 YOLOv4 改进的红外图像行人检测算法 YOLO-sd,优化后的 YOLO-sd 针对于灰度图及小目标的检测能力有明显提升,提高了红外检测的实用性。该算法主要应用于低像素及小目标的检测环境,主要采用形变卷积为核心组件,构

(下转第 41 页)