



DOI: 10.12086/oe.2020.200314

复杂场景下的红外目标检测

张汝榛^{1,2,3}, 张建林^{1,2}, 祁小平^{1,2*},
左颢睿^{1,2}, 徐智勇^{1,2}

¹中国科学院光束控制重点实验室, 四川 成都 610209;

²中国科学院光电技术研究所, 四川 成都 610209;

³中国科学院大学, 北京 100049



摘要: 主流的目标检测网络在高质量 RGB 图像上的目标检测能力突出, 但应用于分辨率低的红外图像上时目标检测性能则有比较明显的下降。为了提高复杂场景下的红外目标检测识别能力, 本文采用了以下措施: 第一、借鉴领域自适应的方法, 采用合适的红外图像预处理手段, 使红外图像更接近 RGB 图像, 从而可以应用主流的目标检测网络进一步提高检测精度。第二、采用单阶段目标检测网络 YOLOv3 作为基础网络, 并用 GIOU 损失函数代替原有的 MSE 损失函数。经实验验证, 该算法在公开红外数据集 FLIR 上检测的准确率提升明显。第三、针对 FLIR 数据集存在的目标尺寸跨度大的问题, 借鉴空间金字塔思想, 加入 SPP 模块, 丰富特征图的表达能力, 扩大特征图的感受野。实验表明, 所采用的方法可以进一步提高目标检测的精度。

关键词: 红外目标检测; 深度学习; 复杂场景

中图分类号: TP391.41; TN219

文献标志码: A

引用格式: 张汝榛, 张建林, 祁小平, 等. 复杂场景下的红外目标检测[J]. 光电工程, 2020, 47(10): 200314

Infrared target detection and recognition in complex scene

Zhang Ruzhen^{1,2,3}, Zhang Jianlin^{1,2}, Qi Xiaoping^{1,2*}, Zuo Haorui^{1,2}, Xu Zhiyong^{1,2}

¹Key Laboratory of Beam Control, Chinese Academy of Sciences, Chengdu, Sichuan 610209, China;

²Institute of Optics and Electronics, Chinese Academy of Sciences, Chengdu, Sichuan 610209, China;

³University of Chinese Academy of Sciences, Beijing 100049, China

Abstract: The mainstream target detection network has outstanding target detection capability in high quality RGB images, but for infrared images with poor resolution, the target detection performance decreases significantly. In order to improve the performance of infrared target detection in complex scene, the following measures are adopted in this paper: Firstly, by referring to the field adaption and adopting the appropriate infrared image preprocessing means, the infrared image is closer to the RGB image, so that the mainstream target detection network can further improve the detection accuracy. Secondly, based on the one-stage target detection network YOLOv3, the algorithm

收稿日期: 2020-08-20; 收到修改稿日期: 2020-09-22

基金项目: 国家 863 计划资助项目(G158207)

作者简介: 张汝榛(1994-), 女, 硕士, 主要从事基于深度学习的红外目标检测的研究。E-mail: zrzheng0319@163.com

通信作者: 祁小平(1974-), 男, 副研究员, 主要从事扩展目标高精度定位技术和图像处理算法的优化及实时实现的研究。

E-mail: qixiaoping@163.com

版权所有©2020 中国科学院光电技术研究所

replaces the original MSE loss function with the GIOU loss function. It is verified by experiments that the detection accuracy on the open infrared data set the FLIR is significantly improved. Thirdly, in view of the problem of large target size span existing in FLIR dataset, the SPP module is added with reference to the idea of the spatial pyramid to enrich the expression ability of feature map, expand the receptive field of feature map, and further improve the accuracy of target detection.

Keywords: infrared target detection; deep learning; complex scenario

Citation: Zhang R Z, Zhang J L, Qi X P, *et al.* Infrared target detection and recognition in complex scene[J]. *Opto-Electronic Engineering*, 2020, 47(10): 200314

1 引言

近年来,随着计算机视觉技术的不断进步发展,计算机视觉已广泛应用到众多科研领域和民用应用。目标检测作为大量高级视觉任务的基础任务之一,在计算机视觉领域和实际应用具有重要的研究意义。目前在许多学者的共同努力下,基于深度学习的目标检测研究已经取得了显著性的进展。但是现在国内外已发表的研究成果中绝大多数应用场景是基于可见光条件,关于红外场景下的应用研究却很少。

可见光图像具有成像分辨率高、目标细节信息丰富等特点,但其相比于红外图像很容易受到光照变化的影响,这在很大程度上增加了目标识别的难度。尤其是在一些特殊天气,例如雨天、雾天、夜间和可见光光源缺少的情况下,可视距离和能见度很差,拍摄的图片根本无法正常使用,从而影响目标检测的结果。而红外成像技术具有工作距离远、抗干扰能力强、测量精度高、不受天气影响、能昼夜工作,以及穿透烟雾能力强等特点,因此红外成像技术一经提出了便得到科研领域和民用的广泛关注,市场对红外目标的检测需求也随之增加。

然而红外图像相较于其他的可见光图像也具有一些不利于目标检测的特性,例如红外图像成像模糊、分辨率差、信噪比低、对比度低,以及图像灰度分布与目标反射特征非线性关系。这些物理特性使得主流的目标检测网络在应用于红外场景时更有挑战性。并且现有的检测方法只局限于单一的红外目标检测识别,缺乏包含多类普遍目标检测的相关研究。

传统的目标检测算法最初选择不同尺度长宽比的滑动窗口产生候选区域(ROI),采用手工提取特征之后输入到分类器进行分类,主要的算法包括: Hog 特征+SVM 算法、Haar 特征+Adaboost 算法、DPM 算法。但是由于滑动窗口产生冗余窗口太多,时间复杂度高,会影响之后的特征提取及分类的速度和精度。并且传

统算法中通过手动特征提取的方法与目标特性紧密相关,而较难具有一般性,因此对于目标形态的多样性,以及背景的多样性光照变化等情况的检测效果不好。

近年来,利用卷积神经网络能够提取图像更深层、更具表示能力的语义信息,确保了目标识别时减少像素质量的影响且实现了检测结果稳定良好。2014 年 RBG 首次提出了 R-CNN 模型框架并且在目标检测上取得了巨大成功,主要的目标检测方法也随之沿着基于候选框和基于回归的目标检测算法的方向发展。在文献[1]中,作者发现将 RGB 和红外图像同时输入 Faster R-CNN 检测能够改善对行人的检测,因此设计了四种 ConvNet fusion 结构,将可见光和红外信息融合并寻找到了最优的融合方法,成功地将 Faster R-CNN 应用到 RGB-T 行人检测任务中。朱大炜等人使用 R-CNN 网络于红外飞机检测,验证了卷积神经网络进行红外目标检测的可行性,后使用速度更快的 SSD 网络并进行改进实现了对红外飞机的实时检测^[2]。侯志强等人根据 Faster R-CNN 存在目标的漏检和重复检测的问题,对 Faster R-CNN 进行了改进,提出了基于双阈值-非极大值抑制算法,提升了检测的性能^[3]。在文献[4]中,作者分析了现有的红外目标检测所遇到的困难,针对数据不足问题,分别对比了图片进行一次预处理并输入、进行多次预处理并输入以及多个预处理并行输入网络的目标检测研究,实验表明进行多次预处理并行输入的检测精度最高。SSD^[5]、DSSD^[6]、YOLOv3^[7]、YOLOv4^[8]等在多个网络层上进行检测,提高了网络在不同尺度目标的检测效果。在文献[9]中针对视频中小目标难以检测的难点改进了 YOLOv3 网络,并在 Road-garbage Dataset 数据集上成功地解决了以上问题。之后的 DetNet^[10]和 RFBNet^[11]在特征提取网络中引入空洞卷积来保证空间分辨率。实验表明,以上方法都能与主流算法媲美。

另外一个影响红外目标识别准确率的原因是缺乏高质量的数据集,针对此问题主要有两个解决方案:

1) 对数据进行增强,如翻转,裁剪等。在文献[5]中,对 VOC 数据集进行随机剪裁、翻转等预处理方法进一步提高了 SSD 对小目标的检测精度。石超等人^[12]针对红外图像的特点使用 Top-Hat 算法对红外图像进行预处理,并使用改进后的 ViBe 算法成功实现了对船舰目标检测。2) 进行 fine-tuning 网络训练,将在 RGB 图像上进行训练检测后的权重设为初始权重,将其迁移学习到红外目标的检测从而减少其对数据量的需求。在文献[13]中,作者成功应用迁移学习解决了使用少量样本进行目标跟踪的问题。

本文主要目的是实现检测精度和速度的平衡。采用 YOLOv3 作为检测的基础网络对红外目标进行检测识别,并在 YOLOv3 的基础上根据红外数据集的特点进行网络修改以提高检测的准确率。其次解决因红外数据集少所带来的训练容易过拟合的问题。本文的主要贡献有:

1) 提出一种新型的红外图像预处理方法,并在实验中验证了提出的红外图像处理方法能提高检测识别的精度。

2) 针对 YOLOv3 存在着定位不准的问题,使用 GIOU 损失函数代替原来的损失函数,用以改善该问题,实验表明目标检测准确率得以提高。

3) 借鉴空间金字塔思想方法,通过添加 SPP(spatial pyramid pooling)模块将图像的局部特征和全局特征进行融合,增大了特征图的感受视野,丰富了特征图的表达能力。

2 方法原理

2.1 红外图像预处理方法

由于主流的目标检测算法应用的场景都是基于 RGB 图像,不适于检测红外目标,因此需要将红外图像进行预处理,以使红外图像更接近 RGB 图像,使得目标检测算法精度能进一步提升。

假设有一红外图像,用 x 表示图像的像素,寻找变化函数 f 使得变化后的图像输入到目标检测算法后能提升检测精度,进行的预处理有:

1) 倒置。一般目标检测所用的 RGB 图像都是白天所摄,通常情况下背景较亮,而目标较暗。但是红外图像所成像为辐射特性,故一般背景辐射较弱而目标辐射较强。因此,采用倒置操作:

$$f: x_p = 1 - x, \quad (1)$$

其中: x 表示灰度范围为[0,1]的原图上任意一点的灰

度值。

2) 直方图均衡。与 RGB 图像的像素分布不同,红外图像的像素分布通常都是偏暗或者偏亮。为了增强红外图像的对比度,采用灰度拉伸操作在一定程度上可以均衡灰度分布:

$$f: x_p = \frac{x - x_{\min}}{x_{\max} - x_{\min}}, \quad (2)$$

其中: x_{\min} 代表像素的最小值, x_{\max} 代表像素的最大值。

3) 去噪+图片锐化。由于红外图像的信噪比较低,因此不能直接对图片进行锐化操作,需要先进行去噪提高信噪比,再进行图像锐化。先采用 5x5 大小的中值滤波器,后用高低帽操作进行锐化,其中:

高帽变换(Top-hat, $T_{\text{Top-hat}}$)定义为

$$T_{\text{Top-hat}}(f) = f - (f \circ g), \quad (3)$$

低帽变换(Bottom-hat, $T_{\text{Bottom-hat}}$)定义为

$$T_{\text{Bottom-hat}}(f) = (f \bullet g) - f, \quad (4)$$

其中: g 为结构元素, f 为原图, \circ 为图像的开操作, \bullet 为图像的闭操作。

图 1 为进行预处理后的图像样例,很容易发现倒置操作更接近 RGB 图像的灰度图。直方图均衡提高了图像的对比度,但是导致了图像过曝。去噪+图像锐化直观上看在一定程度上增加了图像的边缘,但是从这些样例中很难判断是否会使检测的效果更好。

2.2 红外目标检测网络

为了对复杂环境下的红外目标进行快速、准确地检测,需要采用多层的特征进行融合来提高网络对各种尺度目标的特征表达能力。通过对现有的主流目标检测算法的分析,结合基于回归的目标检测算法速度快且能较好保证检测精度的优势,本文采用 YOLOv3 算法来作为红外目标检测的基础网络,在其目标函数中通过加入检测目标重叠率 GIOU 进行目标位置的回归以提升系统对目标的检测定位精度,通过特征金字塔 SPP 进行多层特征的融合与多尺度目标的检测。YOLOv3 是 Redmon 等人提出的系列 YOLO 算法中的第三个版本,在 YOLOv2 的基础上借鉴了 FPN 思想,在 3 个尺度上进行预测,每个尺度对应 3 个候选框,每个候选框输出“位置偏移”、置信度以及分类结果,并且使用 Darknet-53 作为基础网络,其精度超过了大部分主流的基于候选框的算法,而且速度上有很大的优势,网络结构如图 2 所示,图 3 为 SPP 模块,图 4 为本文算法的流程图。



图 1 来自 FLIR 数据集的不同红外图片(每行)。(a) 原图; (b) 倒置; (c) 直方图均衡; (d) 去噪+图像锐化

Fig. 1 Different infrared images (per line) from the FLIR dataset.

(a) Original image; (b) Inversion; (c) Histogram equalization; (d) Denoising + image sharpening

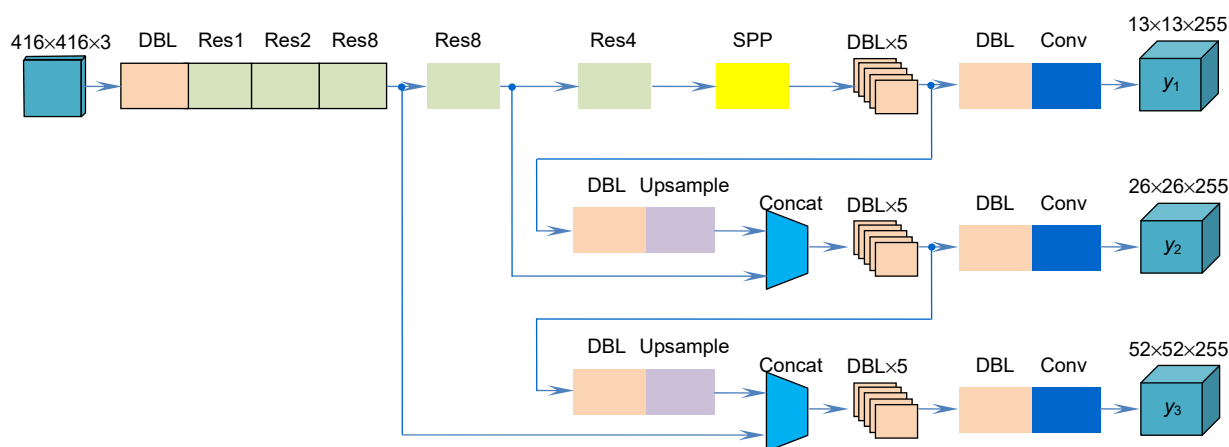


图 2 修改后的 YOLOv3 网络结构图

Fig. 2 Modified YOLOv3 network structure

在 YOLOv3 中利用均方误差(MSE)作为损失函数对 BBox 的中心点以及宽和高进行回归,但是这样做是将 BBox 的每个点的坐标值看做为自变量,没有考虑目标框的整体性,并且 $l_n - norms$ 对物体的尺度比较敏感。为了改善以上问题, Yu Jiahui 等人根据 IOU 代表真实框与 BBox 之间的交并比,提出了用 IOU 损失函数代替传统的 MSE 损失函数^[14]。之后针对 IOU 损失函数存在真实框与预测框不重合时,损失为 0,无法进行梯度回传等问题,在文献[15]中,直观地表明

了对于相同的 $l_n - norms$ 数值,检测结果的质量却不尽相同。为此,提出了 GIOU 以及 GIOU 损失函数,分别表示如下:

$$R_{GIOU} = O_{IOU} - \frac{A^c - U}{A^c}, \quad (5)$$

$$L_{GIOU} = 1 - R_{GIOU}, \quad (6)$$

其中: A^c 表示包含 BBox 与 GT 的最小框的面积, U 表示 BBox 与 GT 的总面积。GIOU 与 IOU 一样也可以看作一种距离度量,满足作为损失函数的基本要求并

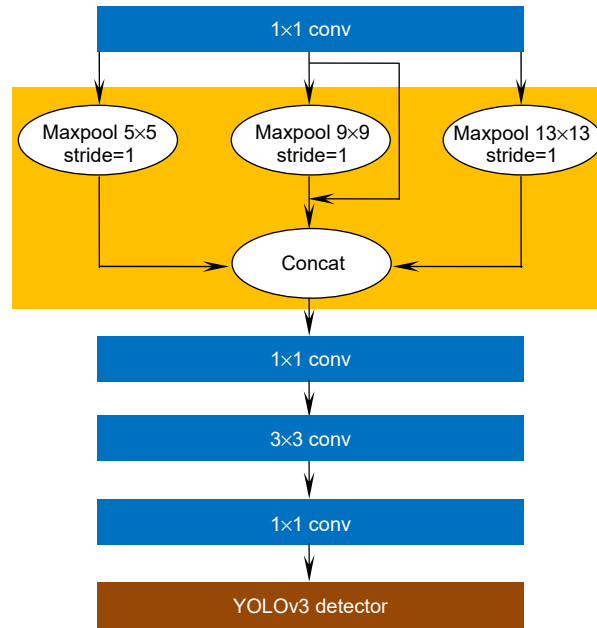


图 3 SPP 模块
Fig. 3 SPP module

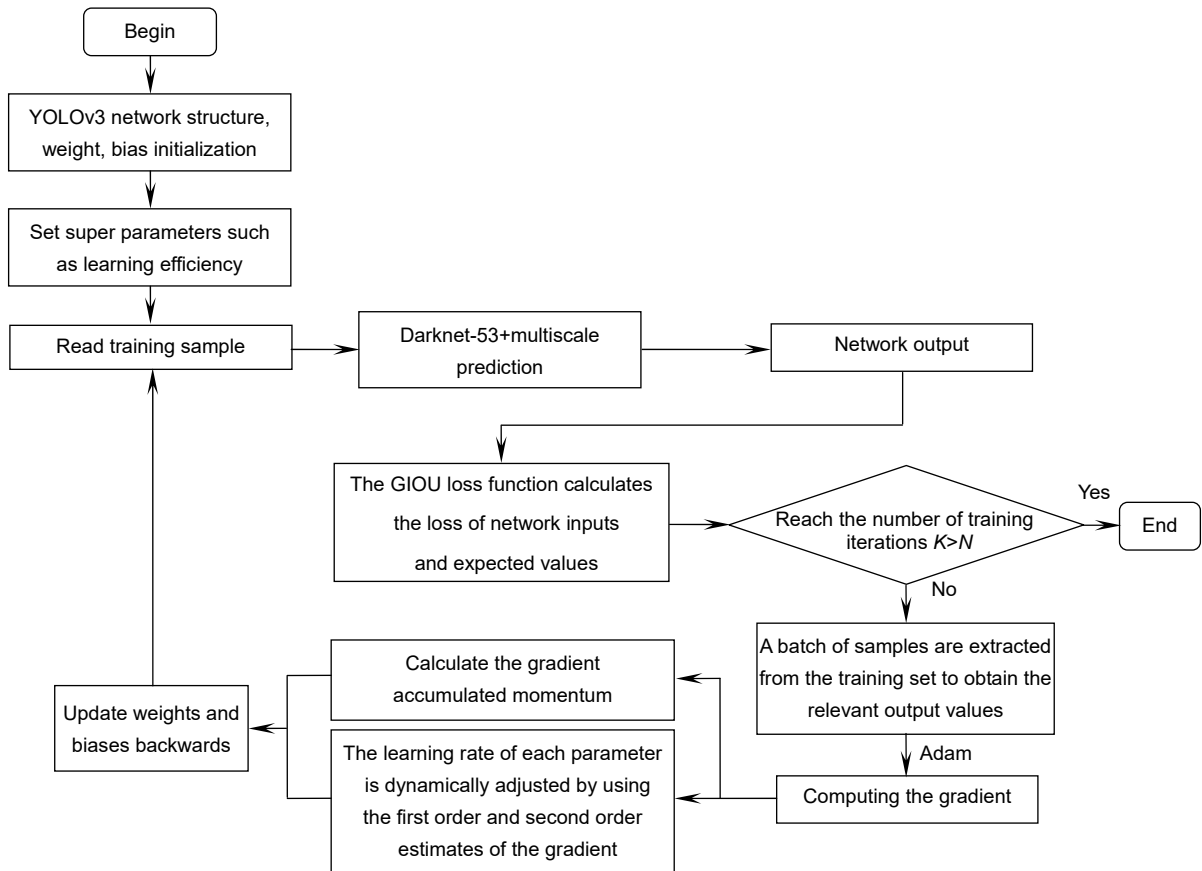


图 4 算法流程图
Fig. 4 Algorithm flow chart

且具有尺度不变性,同时改善了 IOU 作为损失函数存在的预测框和真实框不相交导致的梯度为 0 的情况为

$$\lim_{\frac{|A \cap B|}{|C|} \rightarrow 0} R_{\text{GIOU}}(A, B) = -1 \quad (7)$$

从式(7)可知,该方法能够有效改善 YOLOv3 定位不准的问题。

另外,由于红外图片包括的信息量少,为了进一步提高 YOLOv3 对特征的表达能,借鉴了空间金字塔思想,在 YOLOv3 网络中加入 SPP 模块,使原用于目标检测的特征图经过 SPP 后,将局部特征和全局特征进行融合,丰富了特征图的表达能力,扩大了特征图的感受野,有利于检测图片中目标尺寸跨度比较大的情况。由于用于目标检测的公开红外数据集比较少并且包含的数据量也很少,因此如果直接输入 YOLOv3 进行训练很容易导致过拟合。但是在通过数据集的预处理减少 RGB 与红外图像之间的差别后,使用网络对 RGB 图像进行目标检测后的权重作为检测红外目标的初始权重,可以在减少网络对数据量要求的同时还能减少网络训练的时间。工作过程为:首先,使用 RGB 图像进行训练的网络模型的权重文件作为训练红外图像的检测网络初始权重;然后用红外数据集进行训练,得到训练模型,进行保存;最后使用红外数据集中的测试集对训练好的模型进行测试,得出测试结果。

3 实验结果

3.1 实验细节

3.1.1 训练环境

算法基于 pytorch 框架进行网络搭建、训练和测试,在 Ubuntu 中用 python 进行实验。训练平台采用 Intel i7-7700K CPU, GPU 为 GTX1080Ti,内存为 12 GB。训练过程中使用 GPU 进行加速。

表 1 使用 ImageNet 和 MS COCO 数据集权重训练 FLIR 红外数据集

Table 1 Use ImageNet and MS COCO data set to train FLIR infrared data set

Dataset	mAP/%
ImageNet	49.86
MS COCO	58.02

3.1.2 数据集

本文所使用的数据集为 FLIR 公司于 2018 年 7 月发行的红外数据集,该数据集分为训练集和验证集两个部分,包含 14152 张图像,其中包含待检测目标的有 9241 张,包括的目标有人(28151 个),汽车(46692 个),自行车(4457 个),狗(240 个),由于狗的图片都包含在训练集中,测试时会导致总的检测精度降低,因此在剔除无效数据以及狗的图像后剩余 9220 张图片,并将数据集分为训练集(80%)和测试集(20%)进行检测。拍摄的内容为 11 月至次年的 5 月期间日间(60%)和夜间(40%)的加利福尼亚州圣巴巴拉市街道和公路上的车辆及行人。

3.1.3 训练参数

本文分别使用由 ImageNet 数据集和 MS COCO 数据集经过 YOLOv3 模型进行训练后得到的权重作为红外目标检测网络 YOLOv3 的初始权重进行检测,检测结果如表 1 所示,因此预训练模型的参数使用 YOLOv3 模型在 MS COCO 数据集下训练得到的参数权重作为初始化权重。优化算法使用可以加速 SGD 并且抑制振荡的 Momentum 算法,动量取值 0.9;初始学习率为 0.001; Batch size 设置为 16;周期数设置为 300。

3.2 预处理手段对比

为了严格控制实验条件,所采用的目标检测网络均为 YOLOv3,权重文件均使用 MS COCO 数据集进行预训练和初始化,实验结果如表 2 所示。

从表 2 可以看出只有倒置提高了检测准确率,原因可能是由于 FLIR 红外数据集中大部分图片都是复杂场景且目标的尺度变化比较大,因此导致常规的预处理方法可能不合适。

3.3 对比实验

为了将本文改进的网络与目前主流的目标检测算

表 2 不同的预处理方法输入到 YOLOv3 网络进行训练的检测结果

Table 2 Different pretreatment methods were input into the YOLOv3 network for training

Pretreatment methods	mAP/%
Original image	58.02
Inversion	60.71
Histogram equalization	57.26
De-noising + sharpen	57.42

法网络进行比较,采用双阶段目标检测网络中精度较高的 Faster R-CNN 来进行检测精度的比较,所有实验的网络结构在表 3 中列出。对于 Faster R-CNN 选择 256 个候选框进行训练,正负样本比例为 1:3。

从表 3 中 Faster R-CNN(VGG16)与 Faster R-CNN (Res101)的结果可以看出,拥有更深层的残差网络结

构能够提取到更多的目标信息,检测的准确率提升了 3.18%。而相对于两阶段的 Faster R-CNN 来说,YOLOv3 的检测精度可能有所下降,但测试所用时间明显缩短。用 GIOU 损失函数代替 YOLOv3 原始的损失函数,可以发现检测精度提升了 1.68%,提高了网络定位准确性。通过 Ours(YOLOv3+GIOU)与 Ours(YOLOv3+

表 3 不同框架对于 FLIR 数据集的检测结果。Faster R-CNN IOU 阈值为 0.3, YOLOv3 IOU 阈值为 0.6

Table 3 Detection results of FILR dataset by different frameworks. The Faster R-CNN IOU threshold is 0.3, and the YOLOv3 IOU threshold is 0.6

Detention method	AP/%			mAP/%	Test time/s
	car	person	bicycle		
Faster R-CNN(VGG16)	74.15	62.14	43.58	59.96	0.070
Faster R-CNN(Res101)	77.96	65.72	45.86	63.18	0.163
YOLOv3	77.69	57.47	39.74	58.02	0.026
Ours(YOLOv3+GIOU)	79.3	68.10	31.90	59.70	0.042
Ours(YOLOv3+GIOU+SPP)	81.90	72.60	49.00	66.80	0.039

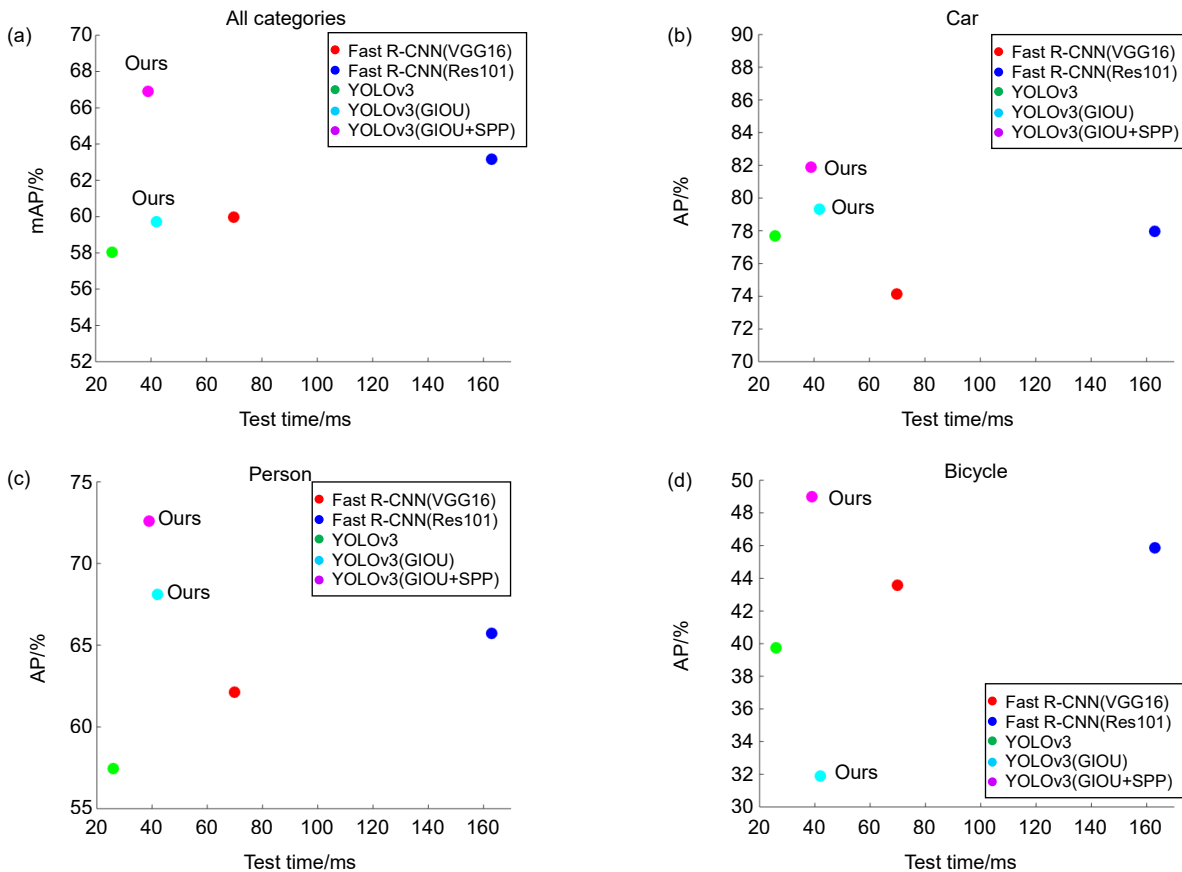


图 5 (a) 不同网络的所有类别检测速度和精度的结果; (b) 不同网络的汽车检测速度和精度的结果; (c) 不同网络的人检测速度和精度的结果; (d) 不同网络的自行车检测速度和精度的结果

Fig. 5 (a) The results of detection speed and accuracy of all categories of different networks; (b) The results of vehicle detection speed and accuracy of different networks; (c) The results of human detection speed and accuracy of different networks; (d) The results of bicycle detection speed and accuracy of different networks

GIOU+SPP)实验结果对比可知加入 SPP 模块能够提升网络的表达能力,检测精度提升了 7.1%。本文所用方法为采用 YOLOv3 框架下使用 GIOU 损失函数,并在检测器前加入 SPP 模块。经相关实验表明,在检测速度下降不是很明显的情况下,对于车、人和自行车的检测精度都有明显的提升。

图 6 为部分 FLIR 数据集检测结果图,由图 6(a)列与图 6(b)列的对比可以看出,原始的 YOLOv3 网络对红外目标进行检测时,对近距离大目标或边缘目标的检测能力较差的情况,并且在目标较小时存在着误检和漏检的问题,而 YOLOv3 的作者也指出过 YOLOv3 网络对大中型目标会存在检测精度下降的情况。通过

图 6(a)列与图 6(c)列和图 6(d)列的对比图可以看到,经过修改后的网络,对于近距离的大目标以及边缘目标的检测能力有了明显的改善,并且对于定位的准确性也有了一定的提升,网络预测的框更接近真实值。而对比图 6(c)列和图 6(d)列,能够看出在网络加了 SPP 模块后网络的误检和错检能力降低,在克服了 YOLOv3 原来的不足后,相比与只修改 GIOU 损失函数的方法能够进一步地提升目标检测精度。

4 结论

针对复杂场景下红外目标检测的成像分辨率低、对比度低的问题,提出了以下解决方法:

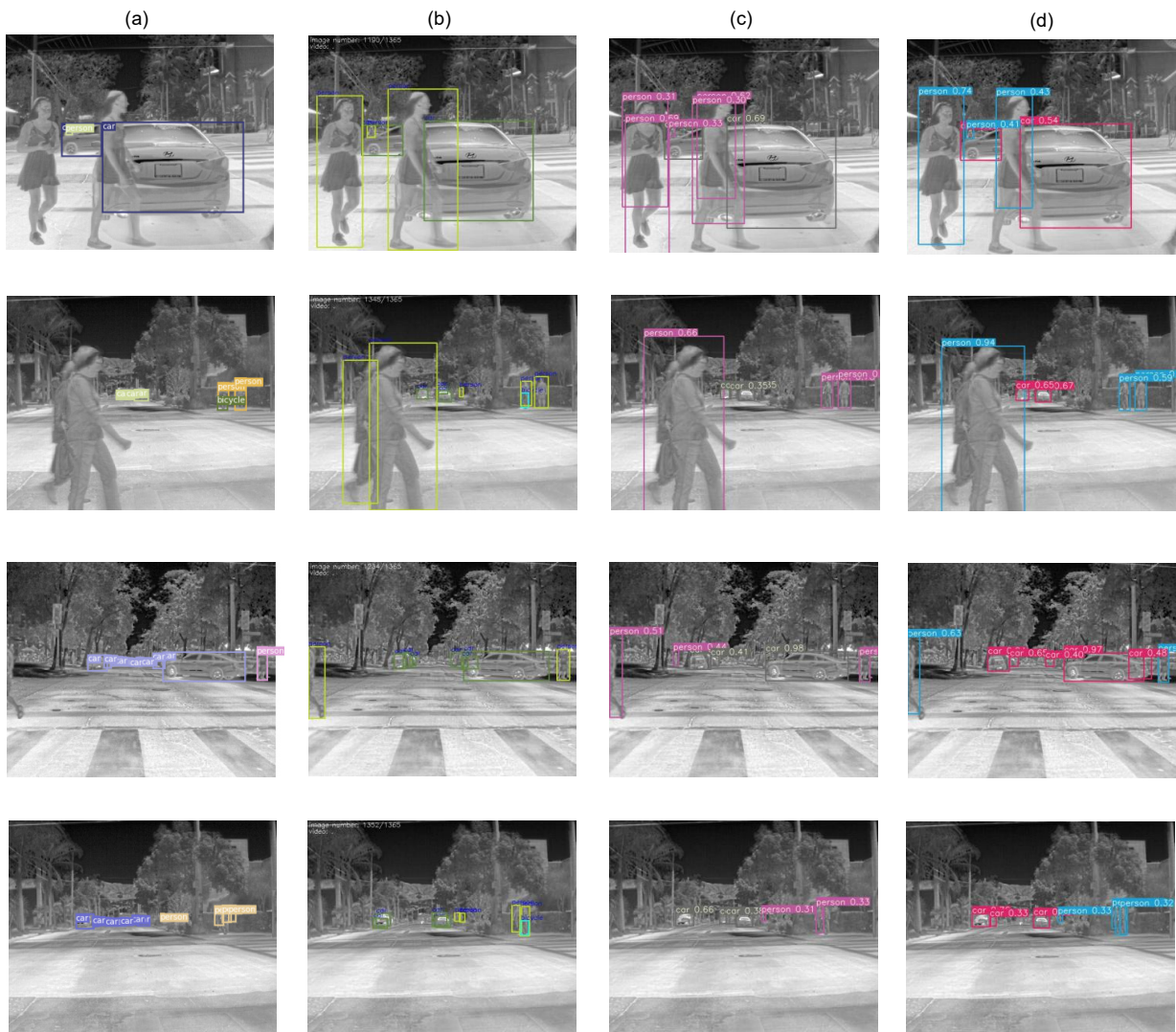


图 6 (a) YOLOv3 网络检测结果; (b) 真实值; (c) 使用 GIOU 损失函数的 YOLOv3 网络检测结果; (d) 使用 GIOU 损失函数并添加 SPP 模块的 YOLOv3 网络的检测结果

Fig. 6 (a) The network detection result of YOLOv3; (b) Ground truth; (c) The YOLOv3 network detection results using GIOU loss function; (d) Detection results for YOLOv3 network that use GIOU loss function and SPP modules

首先采用倒置的预处理手段对红外图像进行预处理,使红外图像更加接近 RGB 图像,从而能对可见光图像目标检测网络进行迁移,并采用有限的红外图像 fine-tuning 训练网络进一步提升其红外目标检测能力;考虑到 YOLOv3 网络对大中型目标的检测精度存在不足,使用 GIOU 损失函数以及 SPP 模块进一步改进了网络,提升了网络的检测能力以及定位精度,在速度损失不是很大的情况下检测精度提升了近 9 个点。当然,该网络目前仍存在部分不足,如:由于自行车类在整个数据集中存在的数量相比人和车的数量较少,数据分布不均衡的问题,因此网络对自行车类别的学习能力有限,对训练后的检测网络进行测试时存在着漏检的情况,另外对于目标重叠时检测效果也较差,接下来的工作是对这两个方面进行改进。

参考文献

- [1] Hou Y L, Song Y Y, Hao X L, et al. Multispectral pedestrian detection based on deep convolutional neural networks[C]//2017 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), 2018.
- [2] Zhu D W. Infrared image plane target detection method based on deep learning[D]. Xi'an: Xidian University, 2018.
朱大炜. 基于深度学习的红外图像飞机目标检测方法[D]. 西安: 西安电子科技大学, 2018.
- [3] Herrmann C, Ruf M, Beyerer J. CNN-based thermal infrared person detection by domain adaptation[J]. *Proceedings of SPIE*, 2018, **10643**: 1064308.
- [4] Hou Z Q, Liu X Y, Yu W S, et al. Improved algorithm of faster R-CNN based on double threshold-non-maximum suppression[J]. *Opto-Electronic Engineering*, 2019, **46**(12): 190159.
侯志强, 刘晓义, 余旺盛, 等. 基于双阈值-非极大值抑制的 Faster R-CNN 改进算法[J]. *光电工程*, 2019, **46**(12): 190159.
- [5] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[C]//*Proceedings of the 14th European Conference on Computer Vision*, 2016, **9905**: 21–37.
- [6] Fu C Y, Liu W, Ranga A, et al. DSSD: deconvolutional single shot detector[Z]. arXiv:1701.06659[cs.CV], 2017.
- [7] Redmon J, Farhadi A. YOLOv3: an incremental improvement[Z]. arXiv:1804.02767[cs.CV], 2018.
- [8] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: optimal speed and accuracy of object detection[Z]. arXiv:2004.10934[cs.CV], 2020.
- [9] Jin Y, Zhang R, Yin D. Object detection for small pixel in urban roads videos[J]. *Opto-Electronic Engineering*, 2019, **46**(9): 190053.
金瑶, 张锐, 尹东. 城市道路视频中像素小目标检测[J]. *光电工程*, 2019, **46**(9): 190053.
- [10] Li Z M, Peng C, Yu G, et al. DetNet: a backbone network for object detection[Z]. arXiv:1804.06215[cs.CV], 2018.
- [11] Liu S T, Huang D, Wang Y H. Receptive field block net for accurate and fast object detection[Z]. arXiv:1711.07767[cs.CV], 2017.
- [12] Zhao C M, Chen Z B, Zhang J L. Application of aircraft target tracking based on deep learning[J]. *Opto-Electronic Engineering*, 2019, **46**(9): 180261.
赵春梅, 陈忠碧, 张建林. 基于深度学习的飞机目标跟踪应用研究[J]. *光电工程*, 2019, **46**(9): 180261.
- [13] Shi C, Chen E Q, Qi L. Ship detection from infrared video[J]. *Opto-Electronic Engineering*, 2018, **45**(6): 170748.
石超, 陈恩庆, 齐林. 红外视频中的舰船检测[J]. *光电工程*, 2018, **45**(6): 170748.
- [14] Yu J H, Jiang Y N, Wang Z Y, et al. UnitBox: An advanced object detection network[C]//*Proceedings of the 24th ACM International Conference on Multimedia*, 2016.
- [15] Rezaatfighi H, Tsoi N, Gwak J Y, et al. Generalized intersection over union: a metric and a loss for bounding box regression[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.

Infrared target detection and recognition in complex scene

Zhang Ruzhen^{1,2,3}, Zhang Jianlin^{1,2}, Qi Xiaoping^{1,2*}, Zuo Haorui^{1,2}, Xu Zhiyong^{1,2}

¹Key Laboratory of Beam Control, Chinese Academy of Sciences, Chengdu, Sichuan 610209, China;

²Institute of Optics and Electronics, Chinese Academy of Sciences, Chengdu, Sichuan 610209, China;

³University of Chinese Academy of Sciences, Beijing 100049, China



Improved YOLOv3 network for infrared target detection

Overview: In recent years, with the continuous development of computer vision, the ability of target detection based on deep learning has been significantly improved. However, most of the images used by mainstream target detection networks are RGB images, and there are few studies on the direction of infrared target detection. Moreover, the mainstream target detection network has a prominent target detection capability in high quality RGB images, but the target detection performance in infrared images with poor resolution is significantly reduced. Compared with infrared images, visible images have higher imaging resolution and rich target detail information. However, under certain weather conditions, the visible images cannot be obtained. Infrared imaging technology has the characteristics of long range, strong anti-interference ability, high measurement accuracy, not affected by weather, able to work day and night, and strong ability to penetrate smoke. Therefore, infrared imaging technology has been widely used once it was proposed. The demand for infrared target detection is also urgent.

In order to improve the performance of infrared target detection in complex scenes, the following measures are adopted in this paper: First, referring to the field adaptive method, appropriate infrared image preprocessing means are adopted to make the infrared image closer to the RGB image, so as to further improve the detection accuracy by applying the mainstream target detection network. Secondly, mean square error (MSE), a loss function, regards the coordinate value of each point of BBox as an independent variable, which does not consider the integrity of the target frame, and l_n -norms is sensitive to the scale of the object, so the algorithm is based on the single-stage target detection network YOLOv3 and replaces the original MSE loss function with GIOU loss function. It is verified by experiments that the detection accuracy on FLIR, an open infrared data set, is significantly improved, and the problem of inaccurate location in the original network is effectively improved. Thirdly, in view of the problem of large span of target size in the FLIR data set, the SPP module is added to enrich the expression ability of feature map and expand the receptive field of feature map by referring to the idea of space pyramid. The experimental results show that the network detection error rate decreases after the addition of SPP module, and after overcoming the original deficiency of the YOLOv3, the target accuracy of detection can be further improved compared with the modification of GIOU loss function only.

Citation: Zhang R Z, Zhang J L, Qi X P, *et al.* Infrared target detection and recognition in complex scene[J]. *Opto-Electronic Engineering*, 2020, 47(10): 200314

Supported by National High Technology Research Development Program China (G158207)

* E-mail: qixiaoping@163.com