

文章编号: 2095-2163(2022)06-0033-06

中图分类号: TP391.4

文献标志码: A

基于改进型 YoloV5s 的热红外道路车辆及行人检测方法

李 阳¹, 赵 娟², 严运兵¹

(1 武汉科技大学 汽车与交通工程学院, 武汉 430065; 2 中国科学院 深圳先进技术研究院集成所, 广东 深圳 518055)

摘要: 在基于机器视觉的 ADAS(Advanced Driver Assistance System)系统中,基于传统 RGB 可见光成像的障碍物检测方法极易受到异常天气、光学折射等原因影响。相比之下,被动式热红外成像技术有着不受天气干扰、不受光线干扰等优势,对未来的 ADAS 系统是一种潜在的解决方案。本文基于热红外图像的特性,对 YoloV5s 检测网络进行针对性优化,提出基于改进型 YoloV5s 的热红外道路车辆及行人检测方法。该方法针对热红外图像的特性以及实际部署中嵌入式平台的算力问题,使用卷积下采样层替换 YoloV5s 中低效的 Focus Layer。针对实际部署环境中遇到的遮挡、重叠等易出现误检、漏检的情况,以 DIoU_NMS 替换加权 NMS 作为候选框非极大值抑制方法。使用 FLIR 热红外交通检测数据集进行了算法验证,该方法参数量为 7.4 M,计算量为 17.5 GFLOPs,检测速度达到 213 FPS,同时 mAP 达到 93.4%。该方法在保证检测精度的同时能满足 ADAS 系统对实时性的要求。

关键词: 目标检测; 行人检测; 卷积神经网络; YoloV5s

Thermal infrared road vehicle and pedestrian detection method based on improved YoloV5s

LI Yang¹, ZHAO Juan², YAN Yunbing¹

(1 College of Automobile and Traffic Engineering, Wuhan University of Science and Technology, Wuhan 430065, China;

2 Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen Guangdong 518055, China)

[Abstract] In an ADAS system based on machine vision, obstacle detection methods based on traditional RGB visible light imaging are extremely susceptible to abnormal weather, optical refraction and other reasons. In contrast, passive thermal infrared imaging technology has the advantages of being immune to weather interference and light interference, and is a potential solution for future ADAS systems. Based on the characteristics of the thermal infrared image, this paper optimizes the YoloV5s detection network and proposes a thermal infrared road vehicle and pedestrian detection method based on the improved YoloV5s. Faced with the characteristics of thermal infrared images and the computing power of embedded platforms in actual deployment, the method uses a convolutional downsampling layer to replace the inefficient Focus Layer in YoloV5s. Aiming at the occlusion, overlap and other situations that are prone to false detection and missed detection in the actual deployment environment, the weighted NMS is replaced by DIoU_NMS as the candidate frame non-maximum suppression method. The algorithm is verified using FLIR thermal infrared traffic detection data set. The method parameter is 7.4 M, the calculation is 17.5 GFLOPs, the detection speed reaches 213 FPS, and the mAP reaches 93.4%. This method can meet the real-time performance required by the ADAS system while ensuring the detection accuracy.

[Key words] object detection; pedestrian detection; Convolutional Neural Network; YoloV5s

0 引言

近十年来,汽车夜视辅助驾驶系统的研究取得了快速发展,相关研究表明车载红外夜视仪在减少夜间行车事故中有很好的效果^[1-3]。

与传统的激光雷达成像以及可见光成像等探测技术相比,红外探测技术具有其独特优势。对此可分述如下:

(1) 被动式红外热成像传感器接收前方热辐射信号并输出为图像帧,能够捕捉更清晰的目标轮廓

信息,输出图像梯度更大,有利于特征提取。

(2) 红外热成像技术能够实现全天候工作。可见光探测器多数只能够在白天或光源充足的情况下进行探测工作,而红外热成像探测器能够不受白天、夜晚的限制,实现昼夜工作。

(3) 热红外成像探测技术抗干扰能力强。主动雷达成像探测的效率容易受到目标表面材料吸波性能的影响;可见光成像探测容易受到雨天、雾霾、风沙等对自然光有强干扰的天气的影响。由于红外热成像探测属于被动探测,不需要向外界发送信号,故

基金项目: 国家自然科学基金(51975428)。

作者简介: 李 阳(1996-),男,硕士研究生,主要研究方向:计算机视觉与智能驾驶。

收稿日期: 2022-01-08

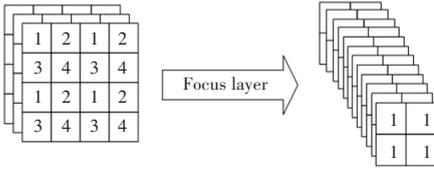


图 3 Focus 层原理示意图

Fig. 3 Schematic diagram of the Focus layer

相比于传统的 Conv 下采样, Focus 层的优势在于不丢失信息, 而是将原有的空间信息转化为深度信息。反面观之, 相比于传统的 Conv 下采样, Focus 层的计算量大幅度增加, 得到的效果相比于传统的 Conv 下采样提升并不大, 又由于热红外图像本身代表的是热量梯度, 输出与单通道灰度图类似, 在三通道输出时, 每个通道的图像矩阵元素相等, 其空间信息在三通道上表现并不明显, 通过 Focus 层将三通道图像的特征转化为 12 通道图像特征时, 对特征信息的提取效果不佳, 反而占用了计算资源, 减慢了推理速度, 见表 1。

表 1 Focus 层对算法的影响

Tab. 1 The impact of the Focus layer on the algorithm

算法	mAP/ %	Recall/ %	FPS
YoloV5s	87.7	75.8	222
YoloV5s-Conv	88.6	75.9	227

1.2 上采样层优化

在原版 YoloV5 的 neck 结构中采用的上采样方式为最近邻插值上采样法, 如图 4 所示。

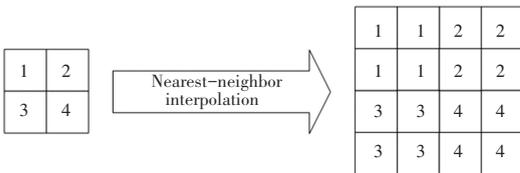


图 4 最近邻插值算法

Fig. 4 Nearest neighbor interpolation algorithm

算法原理为计算新特征图的坐标点像素值对应原图坐标点的像素值, 其对应关系如式 (1):

$$\begin{aligned} srcX &= dstX \times (srcWidth/dstWidth) \\ srcY &= dstY \times (srcHeight/dstHeight) \end{aligned} \quad (1)$$

其中, 原特征图坐标定义为 $(srcX, srcY)$, 新特

征图坐标定义为 $(dstX, dstY)$, 长宽同理。

当带入新特征图像素点坐标 $(dstX, dstY)$, 求出原特征图像素点坐标 $(srcX, srcY)$ 不为整数时, 最近邻插值法采用的方法为四舍五入选取最接近的整数。

最近邻插值上采样法的优点是不需要计算, 只需要寻找, 所以速度最快, 但缺点是生成的新特征图破坏了原特征图像素点之间的渐变关系。

针对此问题, 基于热红外图像的特性, 本文提出转置卷积上采样方法替代传统最近邻插值算法。

对于正向卷积下采样操作, 假设图像矩阵 $input$ 为:

$$input = \begin{Bmatrix} x_{0,0} & x_{0,1} & x_{0,2} & x_{0,3} \\ x_{1,0} & x_{1,1} & x_{1,2} & x_{1,3} \\ x_{2,0} & x_{2,1} & x_{2,2} & x_{2,3} \\ x_{3,0} & x_{3,1} & x_{3,2} & x_{3,3} \end{Bmatrix} \quad (2)$$

假设卷积核 $kernel$ 为:

$$kernel = \begin{Bmatrix} w_{0,0} & w_{0,1} & w_{0,2} \\ w_{1,0} & w_{1,1} & w_{1,2} \\ w_{2,0} & w_{2,1} & w_{2,2} \end{Bmatrix} \quad (3)$$

当步长 $stride = 1$, 填充 $padding = 0$ 时, 根据卷积运算边长计算公式:

$$output = \lfloor \frac{input + 2 \times padding - kernel}{strides} \rfloor + 1 \quad (4)$$

可得输出矩阵 $output$ 大小为 2×2 。假设输出矩阵为:

$$output = \begin{Bmatrix} y_{0,0} & y_{0,1} \\ y_{1,0} & y_{1,1} \end{Bmatrix} \quad (5)$$

将 $input$ 矩阵与 $output$ 矩阵按行展开为列向量可得:

$$input = \{x_{0,0} \ x_{0,1} \ \dots \ x_{3,2} \ x_{3,3}\}^T \quad (6)$$

$$output = \{y_{0,0} \ y_{0,1} \ y_{1,0} \ y_{1,1}\}^T$$

故卷积下采样运算可表示为:

$$output = Cinput \quad (7)$$

推导可得稀疏矩阵 C 为:

$$C = \begin{Bmatrix} w_{0,0} & w_{0,1} & w_{0,2} & 0 & w_{1,0} & w_{1,1} & w_{1,2} & 0 & w_{2,0} & w_{2,1} & w_{2,2} & 0 & 0 & 0 & 0 & 0 \\ 0 & w_{0,0} & w_{0,1} & w_{0,2} & 0 & w_{1,0} & w_{1,1} & w_{1,2} & 0 & w_{2,0} & w_{2,1} & w_{2,2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & w_{0,0} & w_{0,1} & w_{0,2} & 0 & w_{1,0} & w_{1,1} & w_{1,2} & 0 & w_{2,0} & w_{2,1} & w_{2,2} & 0 \\ 0 & 0 & 0 & 0 & 0 & w_{0,0} & w_{0,1} & w_{0,2} & 0 & w_{1,0} & w_{1,1} & w_{1,2} & 0 & w_{2,0} & w_{2,1} & w_{2,2} \end{Bmatrix} \quad (8)$$

故转置卷积上采样操作可近似表示为:

$$C^T \text{output} = \text{input} \quad (9)$$

由上式可得,转置卷积操作可将 *output* 矩阵特征的尺寸信息恢复到下采样前矩阵 *input* 的尺寸信息,但无法恢复 *input* 矩阵的每个元素值。基于 YoloV5s 的 Neck 结构中 PANet 的结构特性,低维度特征上采样后与高维度特征结合,且在前向推理的过程中,转置卷积上采样卷积核参数为可训练参数,相比于传统最近邻插值上采样,转置卷积上采样在一定程度上增强了网络推理能力。

最近邻插值上采样层与转置卷积层的对比见表 2。由表 2 分析可知,相对于最近邻插值上采样方法,转置卷积上采样能够有效提高网络识别准确率,消除移去 Focus 层带来的影响。

表 2 最近邻插值上采样层与转置卷积层的对比

Tab. 2 Comparison of nearest neighbor interpolation upsampling layer and transposed convolutional layer

算法	mAP/%	Recall/%	FPS
YoloV5s-Upsample	88.6	75.9	227
YoloV5s-ConvTranspose	90.7	77.7	223

1.3 NMS 非极大值抑制方法优化

在 YoloV5s 的后处理中,使用加权 NMS^[7] 算法来进行检测框的非极大值抑制。YoloV5 加权 NMS 原理如下式:

$$\text{box} = \frac{\sum_{i=1}^n w_i \times B_i}{\sum_{i=1}^n w_i} \quad (10)$$

$$w_i = C_i \times \text{IoU}(B_i, B_{\arg\max C_i})$$

其中, *box* 表示每个候选框经过加权平均后的候选框; w_i 表示权重; C_i 表示每个候选框的置信度; $\text{IoU}(\cdot)$ 表示第 i 个候选框和最大置信度候选框进行 IoU 计算。

加权 NMS 的算法在通常情况下,相对于传统 NMS 算法通常能够获得更高的 *Precision* 值以及 *Recall* 值,但其缺点也很明显,就是加权因子取决于候选框的 IoU 与得分。在 YoloV5s 中,候选框的 IoU 只考虑 2 个框的重叠面积,对于候选框的重叠关系表示并不全面,而每个候选框的得分又会有定位与得分不一致的问题,进而导致在使用加权 NMS 算法作为后处理方法时出现对遮挡、重叠目标识别精度不佳的情况。

为此,提出以 DIOU_NMS ^[8] 方法来优化候选框后处理方法。相比于传统 IoU 方法以及加入权重项

的 Weighted_IoU 方法, DIOU_NMS 方法加入了以候选框与最大置信度候选框的中心距为惩罚项的惩罚机制,原理如图 5 所示。

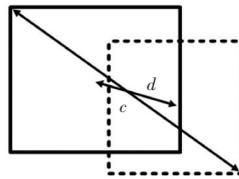


图 5 DIOU_NMS 原理

Fig. 5 Principle of DIOU_NMS method

研究推得的数学表达式为:

$$\text{DIOU} = \text{IoU}(B_i, B_{\arg\max}) \left(\frac{d^2}{c^2} \right)^\beta \quad (11)$$

其中, d 为 $B_i, B_{\arg\max}$ 中心点的欧式距离; c 为 $B_i, B_{\arg\max}$ 的最小外接矩形的对角线长度; β 用于控制惩罚幅度。

在实际应用时,使用 DIOU_NMS 算法表示为:

$$s_i = \begin{cases} s_i & \text{IoU}(B_i, B_{\arg\max}) - \left(\frac{d^2}{c^2} \right)^\beta < \varepsilon \\ 0 & \text{IoU}(B_i, B_{\arg\max}) - \left(\frac{d^2}{c^2} \right)^\beta \geq \varepsilon \end{cases} \quad (12)$$

其中, s_i 为 B_i 的分类置信度。

由公式可得出,当 $\beta \rightarrow \infty$ 时, DIOU 退化为 IoU , 此时 DIOU_NMS 与传统 NMS 效果相当。当 $\beta \rightarrow 0$ 时,几乎所有中心点不与 $B_{\arg\max}$ 重合的框都被保留。

经试验得到,在不增加计算成本的情况下,相对于传统 IoU_NMS 与加权 IoU_NMS , DIOU_NMS 方法可以使得相关性能得到提升,其实验结果如图 6 所示。修改 NMS 方法对算法的影响见表 3。

表 3 修改 NMS 方法对算法的影响

Tab. 3 The influence of different NMS methods on the algorithm

算法	mAP/%	Recall/%	FPS
YoloV5s	87.7	75.8	222
YoloV5s-DIOU_NMS	90.3	74.9	217

同时,由图 6 可知,对于遮挡、重叠目标的检测效果, DIOU_NMS 方法要优于传统 IoU_NMS 与加权 IoU_NMS 。



(a) 加权 NMS

(b) DIOU_NMS

图 6 加权 NMS 与 DIOU_NMS

Fig. 6 Weighted NMS and DIOU_NMS

2 实验及结果分析

2.1 实验数据集

实验数据集为 FLIR Thermal Dataset,数据集采集参数见表4。

表4 FLIR 热红外道路信息数据集参数

Tab. 4 FLIR thermal infrared road dataset information

参数名	参数解释
图像	大于 14 K 总图像,其中大于 10 K 来自短视频片段和随机图像样本,以及来自 140 s 视频的大于 4 K 额外图像
帧采集频率	以 30 Hz 录制。以 2 帧/s 或 1 帧/s 的速度采样的数据集序列。视频注释以 30 帧/s 的速度记录
Label 标签总数	<p>总共 10 228 帧; 9 214 帧为带边界框标注的正样本帧。</p> <ol style="list-style-type: none"> 1. Person (28, 151) 2. Car (46, 692) 3. Bicycle (4, 457) 4. Dog (240) 5. Other Vehicle (2, 228)
驾驶环境	<p>时间:当年 11 月至次年 5 月,昼间 (60%) 和夜间 (40%)。</p> <p>地点:加利福尼亚州圣巴巴拉地区的街道和高速公路上行驶。</p> <p>天气:晴朗至多云</p>
相机参数	<p>相机型号:IR Tau2 13 mm f/1.0</p> <p>分辨率:640×512</p> <p>水平视场角 $HFOV$: 45°</p> <p>垂直视场角 $VFOV$: 37°</p>
数据集格式	<p>图像格式:8-bit JPEG</p> <p>标注格式:JSON (MSCOCO format)</p>

2.2 训练系统及训练参数

训练平台参数见表5。

表5 训练平台参数

Tab. 5 Parameters of training platform

Parts	Model
CPU	E5 2620 v4
RAM	128 G (32 G * 4)
GPU	1080Ti * 4
VRAM	44 G (11 G * 4)
Deep learning framework	PyTorch 1.8

模型训练参数见表6。

表6 模型训练参数

Tab. 6 Training parameters of the model

参数名	参数值
Batch size	128
Epochs	300
Input shape	640 * 512
Training set	7 532
Validation set	1 330
Test set	1 366
Initial learning rate	0.01
Momentum	0.937
Optimizer weight decay	5e-4
IoU training threshold	0.2
Class loss gain	0.5

2.3 测试平台

Hi3519AV100 开发平台是一颗面向监控 IP 摄像机、运动相机、全景相机、ADAS 系统、航拍等多个领域推出的高性能、低功耗的 SoC 开发平台。该芯片平台集成了神经网络硬件加速引擎 (Neural Network Inference Engine, NNIE), 是海思媒体 SoC 中专门针对神经网络、特别是深度学习卷积神经网络进行加速处理的硬件单元,并在计算能力达到 1.7 TFLOPS 的同时,支持现有大部分的公开网络结构,如 AlexNet、VGG16、GoogleNet、ResNet18、ResNet50 等分类网络, Faster R-CNN、YOLO、SSD、RFCN 等检测网络以及 SegNet、FCN 等场景分割网络。该芯片集成了海思第四代 ISP, 支持 WDR、多级降噪、六轴防抖及多种图像增强和矫正算法。

综上所述,该研究基于 12 nm 制程的高性能、低功耗架构设计,相比于同类型产品大大降低了散热需求,有利于应用部署在车载系统常见的狭小、密闭环境中。

因此,本文以 Hi3519AV100 为基础搭建实验平台,进行对照组实验以及多模型对照实验。

2.4 实验结果分析

表7为不同检测算法在 FLIR 数据集中进行对照实验后得到的检测精度结果。

从表7可以看出,本文的优化方法针对热红外图像数据集,相比于原版 YoloV5 在推理帧数相接近的情况下,运行后在 mAP 上得到了 5.7% 的提高,召回率 $Recall$ 上得到了 4.1% 的提高。相对于常应用在嵌入式开发平台上的 Tiny-YoloV3、Tiny-YoloV4 算法而言,本文算法在 mAP 上获得了压倒性的优

势,大幅提高了识别精度。相对于工业领域得到广泛应用的 YoloV3、YoloV4 算法,在 mAP 指标接近的情况下,大幅降低了所需计算量,提高了帧数的处理能力。

表7 主流检测算法对照实验

Tab. 7 Comparison experiments of mainstream detection algorithms

算法	GfLOPs	mAP/ %	Recall/ %	FPS/ 帧
Tiny-YoloV3	5.6	58.8	47.3	208
Tiny-YoloV4	6.9	65.1	55.9	178
YoloV3	156.4	95.0	81.1	64
YoloV4	128.5	96.3	83.4	62
YoloV5s	16.3	87.7	75.8	222
YoloV5s6	16.7	88.4	75.7	208
本文算法	17.5	93.4	79.9	213

由上述实验结果可以看出,本文方法针对热红外条件下的图像进行优化,相对于传统的算法,性能指标得到较大提高。

3 结束语

本文基于 YoloV5s,探讨了传统基于可见光成像的图像算法在热红外图像上的检测性能不佳的问题,并结合热红外图像的特性,进行针对性优化,在 FLIR 数据集上进行了实验验证,实验后 mAP 达到 93.4%、 $Recall$ 达到 79.9% 的同时, FPS 达到 213 帧,不仅提高了检测精度,而且也保证了 ADAS 系统要求的实时性。

虽然本方法达到基本预期,但本文方法仍存在以下几点问题:

(1) 本文检测方法的 head 部分仍是基于传统的 Anchor Based 方法,在优化潜力与前向推理效率方面与新兴的 Anchor Free 模型相比存在一定差距。

(2) 本文方法所针对的任务较为单一,且在不同的检测任务,如全景分割、实例分割、目标跟踪等任务中的表现欠佳。

后续工作将围绕如下 2 方面进行改进:在网络结构方面,参考 Anchor Based 思想对网络结构进行后续改进。在任务方面,针对热红外图像的实例分割、全景分割等视觉任务进行更深入的优化。

参考文献

- [1] 孙文财,孙德雨,李梦茹,等. 一种基于红外热成像的夜间行车安全辅助系统:中国, CN212422974U [P]. 2021-01-29.
- [2] 王国华,刘琼,庄家俊. 基于局部特征的车载红外行人检测方法研究[J]. 电子学报, 2015, 43(07): 1444-1448.
- [3] 王军. 车载红外行人检测系统设计与实现[D]. 西安:西安电子科技大学, 2017.
- [4] LIU Shu, QI Lu, QIN Haifang, et al. Path aggregation network for instance segmentation [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018: 8759-8768.
- [5] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA: IEEE, 2017: 936-944.
- [6] WANG C Y, LIAO H Y M, WU Y H, et al. CSPNet: A new backbone that can enhance learning capability of CNN [C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Seattle, WA, USA: IEEE, 2020: 1571-1580.
- [7] NING Chengcheng, ZHOU Huajun, SONG Yan, et al. Inception single shot multibox detector for object detection [C]// 2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW). Hong Kong, China, 2017: 549-554.
- [8] ZHENG Zhaohui, WANG Ping, LIU Wei, et al. Distance-IoU loss: Faster and better learning for bounding box regression [C]// Proceedings of the AAAI Conference on Artificial Intelligence. AAAI, 2020, 34(07): 12993-13000.
- [3] GUO Hao, SUN Siwei, SHI Danping, et al. Differential attacks on CRAFT exploiting the involutory S-boxes and Tweak additions [J]. IACR Transactions on Symmetric Cryptology, 2020, 2020(3): 119-151.
- [4] HADIPOUR H, SADEGHI S, NIKNAM M M, et al. Comprehensive security analysis of CRAFT [J]. IACR Transactions on Symmetric Cryptology, 2020, 2019(4): 290-317.
- [5] HADIPOUR H, BAGHERI N, SONG Ling. Improved rectangle attacks on SKINNY and CRAFT [J]. IACR Transactions on Symmetric Cryptology, 2021(2): 140-198.
- [6] LI Wei, LIAO Linfeng, GU Dawu, et al. Ciphertext-only fault analysis on the LED lightweight cryptosystem in the Internet of Things [J]. Science China Information Sciences, 2020, 63: 139112.
- [7] De SANTIS F, GUILLEN O M, SAKIC E, et al. Ciphertext-only fault attacks on PRESENT [C]// EISEENBARTH T, ÖZTÜRK E. Lightweight cryptography for security and privacy. LightSec 2014. Lecture Notes in Computer Science. Cham: Springer, 2015, 8898: 85-108.
- [8] FUHR T, JAULMES E, LOMME V, et al. Fault attacks on AES with faulty ciphertexts only [C]//Workshop on Fault Diagnosis and Tolerance in Cryptography. Los Alamitos, CA, USA: IEEE, 2013: 108-118.
- [9] TRICHINA E, KORKIKYAN R. Multi fault laser attacks on protected CRT-RSA [C]//Workshop on Fault Diagnosis and Tolerance in Cryptography. Santa Barbara, CA, USA: IEEE, 2010: 75-86.
- [10] REED I. A class of multiple-error-correcting codes and the decoding scheme [J]. Transactions of the IRE Professional Group on Information Theory, 1954, 4(4): 38-49.

(上接第 32 页)