

融合图像显著性特征的轻量级目标检测算法*

马雷, 杨顺清, 王欢欢, 翟家琛, 徐健傲

(燕山大学车辆与能源学院, 秦皇岛 066004)

[摘要] 针对智能车辆在实际交通环境中面临的目标密集、边缘严重遮挡和前景背景模糊的问题, 本文提出了一种融合图像显著性特征的轻量级目标检测算法。首先基于灰度图像提取出显著性特征图, 和彩色图像分别输入卷积神经网络。其次采用轻量化模块(ghost model)搭建轻量级融合网络, 并使用EIoU优化模型的边框定位损失。在网络后端将非极大值抑制算法进行改进, 以此提高网络对同类别遮挡目标的检测准确率。最后在KITTI数据集上进行训练和测试。实验表明, 改进后的网络 mAP 达到92.7%, 相比原始网络YOLOv5平均精度提高3.8%, 精确率和召回率分别提高3%和6.2%。

关键词: 目标检测; 多特征融合; 轻量级网络; YOLOv5

Lightweight Object Detection Algorithm Based on Image Saliency Feature Fusion

Ma Lei, Yang Shunqing, Wang Huanhuan, Zhai Jiachen & Xu Jianao

College of Vehicle and Energy, Yanshan University, Qinhuangdao 066004

[Abstract] For the problems of dense targets, severe edge occlusion, and blurred foreground and background that intelligent vehicles face in actual traffic environments, a lightweight object detection algorithm based on image saliency feature fusion is proposed in this paper. Firstly, salient feature maps are extracted based on gray-scale images, and input into convolutional neural networks with color images. Secondly, a lightweight fusion network is constructed using the Ghost Model, and the EIoU is used to optimize the model's border localization loss. In order to enhance the detection accuracy of similar occluded targets, non-maximum suppression algorithm is improved on the backend of the network. Finally, the KITTI dataset is used for training and testing. The experiment shows that the improved detection mAP value of the network reaches 92.7%, with an average accuracy improvement of 3.8% compared to the original network YOLOv5. The accuracy and recall rates are increased by 3% and 6.2%.

Keywords: target detection; multi-feature fusion; lightweight network; YOLOv5

前言

目标检测是智能车辆研究热点之一, 国内外学者进行了较为深入的研究。目前基于多模态和显著性多特征的融合算法逐渐成为目标检测领域主流方法^[1]。红外图像的显著性特征和可见光图像融合, 既保留红外图像不受光照影响的特点, 又保留可见光图像的细节信息。因此文献[2]中首次提出红外

图像和可见光图像在深层特征进行融合, 提高了对行人的检测精度。文献[3]中进一步研究红外图像和可见光图像在特征级层面的融合, 提出性能更好的中间层特征融合算法。文献[4]和文献[5]中基于Faster R-CNN算法设计多模态深度卷积神经网络, 对比研究融合彩色图像、偏振图像和红外图像, 得出同时融合多种模态网络检测精度优于任意单一模态。文献[6]中研究在Faster R-CNN检测算法的不同阶段分别融合可见光和红外图像特征, 提出一种

* 国家重点研发计划项目(2021YFB3202204)资助。

原稿收到日期为2023年05月31日, 修改稿收到日期为2023年07月08日。

通信作者: 马雷, 教授, E-mail: malei97yan@163.com。

感知光照导向的融合算法,完成对目标的精确检测。文献[7]中为提升暗光场景下的检测能力,将红外图像显著性特征与可见光图像动态权值融合,并改进网络结构,增强网络对小目标以及密集目标的检测能力。文献[8]中提出一种自适应空间像素特征融合算法,设计出自适应特征提取、融合模块,主动完成可见光图像和红外图像的多特征融合。

将图像深度信息和特定卷积网络获得的显著性特征分别与彩色图像融合也是多特征融合研究的重点。文献[9]中基于颜色和深度显著图提出了融合显著性深度特征的RGB-D图像显著目标检测。文献[10]中基于卷积神经网络生成粗糙显著图并使用残差学习细化显著性特征,使网络保证了足够的检测精度。文献[11]中采用金字塔注意力模块对目标位置特征进行下采样,强化了显著性特征,但对噪声更敏感。文献[12]中设计了编码器-解码器结构,融合注意力机制算法,捕捉图像中不同尺度的显著性目标,提高了复杂背景下预测显著性目标边缘精度。

融合后的网络通常是多个主干模型,导致整个检测模型参数量和计算量庞大,运行速度缓慢,不利于应用到车载芯片等微处理器。文献[13]中提出一种小型化的改进YOLOv3实时车辆检测及跟踪方法。采用较少的重复残差结构单元构建深度更小的基础卷积网络,实现目标的实时检测及跟踪。文献[14]中通过深度可分离卷积替换检测模型SSD的骨干结构,搭建轻量级的Mobile Net-SSD检测模型,使模型更适合车内驾驶员的特定目标识别。

以上融合算法结构复杂,且对外部传感器有较高需求,至少需要两种或以上传感器,并在数据处理和融合上消耗大量时间,导致整个检测系统响应数据输入缓慢,空余大量计算资源。为此,本文提出一种融合图像显著性特征的轻量级目标检测算法,仅针对一种视觉传感器采集的图像数据研究融合算法。首先在数据输入端,基于原始灰度图像提取出边缘特征图、熵特征图和二值特征图,组合成多维度、多模态的特征矩阵。随后和彩色图像分别输入检测网络,在空间上进行多特征深层次融合。在模型结构端,为精简融合网络体积,提高推理速度,将融合网络的卷积模块进行轻量化处理,采用Ghost模块进行算法重构,并改进网络的边框定位损失函数,加快收敛。最后在模型处理后端,采用改进的非极大值抑制算法,提高网络对同类密集遮挡目标的检测性能。

1 显著性特征提取

在客观世界中,人眼主要依靠目标的外部边缘区分物体,再根据其内在纹理细节,做到准确识别。对于计算机视觉而言,目标的边缘信息也作为一种重要的潜在特征。边缘即寻找图像中不同区域突变的像素点,通过计算图像相邻像素的1阶或2阶导数完成,图像 f 在 (x, y) 坐标处的导数 ∇f 可表示为

$$\nabla f \equiv \text{grad}(f) \equiv \begin{bmatrix} g_x \\ g_y \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} \quad (1)$$

边缘特征的获取主要通过特定算子实现,如Sobel算子^[15]、Canny算子^[16]等。本文在数据预处理时,首先将彩色图像转换为32位浮点数灰度图像,再调用Opencv提供的Sobel算子接口,完成边缘特征图的提取,如图1(b)所示。

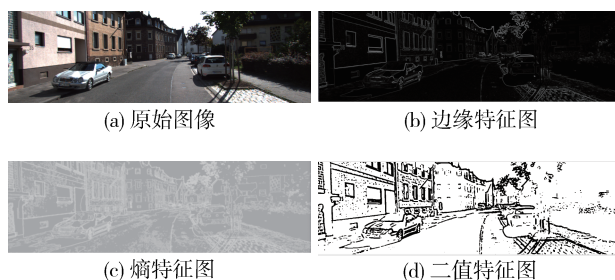


图1 显著性特征图对比

除目标的边缘特征外,目标内在纹理特征也是一类重要的图像特征,反映邻域像素的同质度信息。图像邻域纹理计算方法主要有统计方法、结构方法和频谱方法,考虑到计算资源消耗,故采用基于统计方法的灰度共生矩阵^[17](GLMC)表征图像纹理。首先,计算图像灰度共生矩阵:

$$C(i, j) = \text{Num}(F(x, y) = i \cap F(x + \Delta m, y + \Delta n) = j) \quad (2)$$

式中: C 为灰度共生矩阵; i 和 j 为不相等的灰度级; (x, y) 为某一个像素点的坐标; Δm 为两个像素点的横向距离; Δn 为两个像素点的纵向距离。

其次提取灰度共生矩阵的特定描述子对纹理特征进行量化,通常采用能量、对比度、相关度和熵等描述子表征图像的纹理特征。各个描述子具有不同的特性,本文检测任务所需图像纹理较丰富,故提取

信息量最多的熵特征,通常图像越复杂,熵值越大,信息越多,熵特征公式为

$$\sigma_{ent} = -\sum_r \sum_c C(r, c) \log C(r, c) \quad (3)$$

式中: r 和 c 分别为灰度共生矩阵的行和列。将所得熵特征图可视化如图1(c)所示,充分保留了原始图像信息。

对于道路目标检测任务,主要面向地面目标识别,而从摄像头获取的图像中包含许多区域,如天空、建筑等。故本文采用图像分割技术将目标和背景进行区分,利用图像二值化方法将图像中的目标进行分割。首先使用高斯函数加权方法提取出窗口大小为 15×15 的阈值,将高于阈值的像素设置为

255,再使用中值滤波算法去除二值图像中的部分噪点,得到二值特征图,如图1(d)所示。

将获取的3个特征图使用8位整型保存,在通道上进行拼接,组合成三通道的矩阵,再转化为RGB格式图像,可视化如图2融合网络输入所示。

2 网络构建

基于一阶段YOLOv5算法构建显著性特征融合网络。通过重构特征提取网络、损失函数优化和模型轻量化,设计出本文的轻量级显著性特征融合网络,最终的网络设计方案如图2所示。

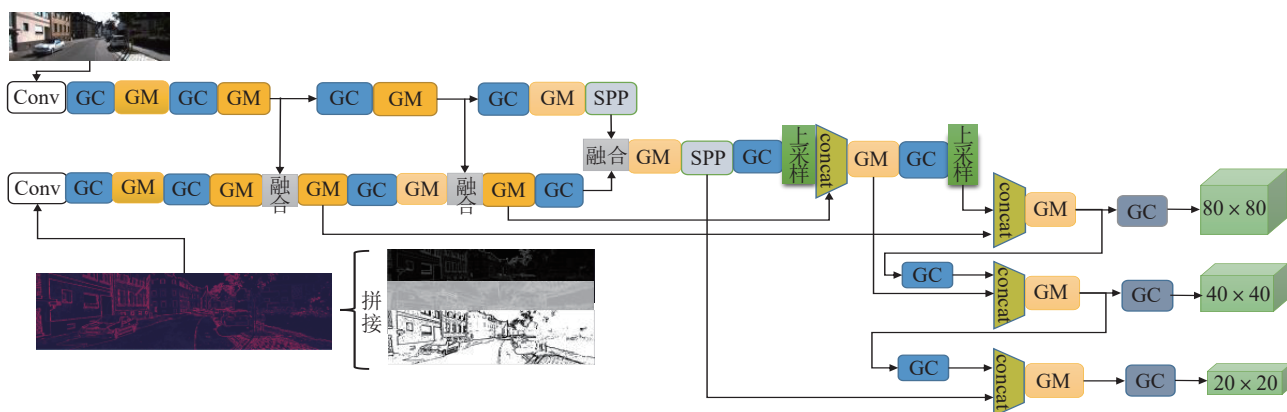


图2 融合网络结构图

设计两路特征提取网络,其中一个分支输入RGB彩色图像,另一个分支输入显著性特征图像。在输入图像分辨率为 $N \times N$ 时,两个分支先后在下采样8倍、16倍和32倍的特征图上进行融合,充分保证网络兼顾彩色图像特征和显著性特征。融合的特征图经过卷积网络调整通道数,进行深层特征的提取。在网络颈部保留原始网络的构建方式,即将浅层次定位信息和深层次语义信息相联系,如将下采样32倍特征图采用最近邻插值算法上采样,和16倍下采样特征图通道拼接。最后在检测头部得到3个大小分别为 $\frac{N}{8} \times \frac{N}{8}$ 、 $\frac{N}{16} \times \frac{N}{16}$ 、 $\frac{N}{32} \times \frac{N}{32}$ 的特征图,分别映射到原始图像从低级到高级的不同层次特征信息。

在本文特征融合模块,针对相同尺度的显著性特征和彩色图像特征,采用基于特征层的融合网络,即将相应的特征图在通道上进行拼接,再利用后续卷积调整融合通道,融合操作如图3所示。

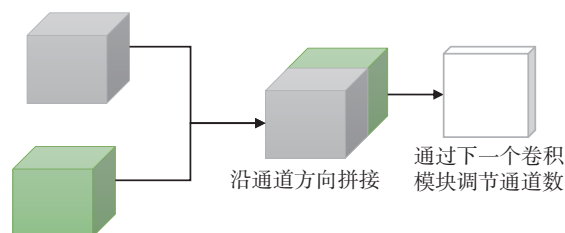


图3 多特征融合模块

2.1 轻量级Ghost模块

文献[18]中提出,冗余特征在卷积网络学习特征时也非常关键,因此基于对冗余特征的高效应用提出Ghost轻量级卷积模块。标准卷积和Ghost轻量级卷积示意图如图4所示。

标准卷积生成许多近似特征,占用大量计算资源,若采用更廉价的线性变换完成冗余特征的生成,可极大地降低卷积运算成本。Ghost模块根据该思想完成相似特征的计算。首先,采用标准卷积生成

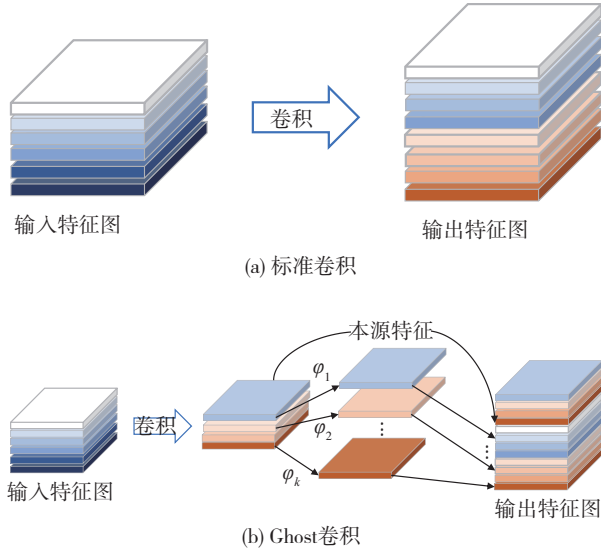


图4 标准卷积与Ghost卷积

m 层本源特征,计算表达式为

$$Y' = X * f + b \quad (4)$$

式中: $Y' \in \mathbb{R}^{h' \times w' \times m}$ 为输出特征图; b 表示偏置; $*$ 表示卷积运算。随后对 Y' 进行廉价映射:

$$y_{ij} = \varphi_{i,j}(y'_i), \forall i = 1, \dots, m, j = 1, \dots, s \quad (5)$$

式中: $y'_i \in Y'$; $\varphi_{i,j}$ 表示本源特征 i 的第 j 次线性变换。可见, $\varphi_{i,j}$ 生成对应的多个Ghost特征 y_{ij} 。

标准卷积浮点运算为 $n \times h' \times w' \times c \times k \times k$,其中, c 为输入通道数,而Ghost卷积在标准卷积的基础上结合 $m(s-1) = n/s(s-1)$ 次线性计算,线性变换卷积核尺寸为 $d \times d$,所有二者计算比率为

$$R = \frac{nh'w'cck}{(n/s)h'w'cck + (s-1)(n/s)h'w'dd} = \frac{cck}{s^{-1}cck + ((s-1)/s)dd} = \frac{sc}{s+c-1} \approx s \quad (6)$$

其中 $d \times d = k \times k$, $s \ll c$,所以Ghost卷积相比标准卷积理论运算次数提升了 s 倍。本文将基于轻量级Ghost模块重构整个融合网络,旨在降低融合网络的参数量和减少运算成本。

2.2 损失函数优化

YOLOv5中损失函数分为置信度损失函数、分类损失函数和边框定位损失函数。其使用 $GIoU$ ^[19]表征定位损失,解决了普通IoU无法优化二者重叠的问题,计算表达式为

$$GIoU = IoU - \frac{A_c - C}{A_c} \quad (7)$$

式中: A_c 为预测框与真实框最小外接矩形的面积; C

为预测框与真实框的相交面积。

研究者发现在实际使用中仍存在不足,在模型训练过程中会优先放大预测框的面积而不使预测框无限逼近真实框。为解决这一问题,学者们提出了 $DIoU$ ^[20](distance IoU),公式为

$$DIoU = IoU - \frac{\sigma^2(b, b^{gt})}{c^2} \quad (8)$$

式中: $\sigma^2(b, b^{gt})$ 为预测框与真实框中心点的欧式距离; c^2 为两框最小外接矩形对角线长度。

在IoU系列的研究中,没有限制预测框的长宽比,因此又有学者提出了 $CIoU$ (complete IoU),其直接限制了最小外接矩形的面积,使网络训练更加偏向于移动预测框,加快模型收敛,计算公式为

$$CIoU = IoU - \frac{\sigma^2(b, b^{gt})}{c^2} - \alpha v \quad (9)$$

式中: α 为权重系数; v 为两框长宽比的差异。

$CIoU$ 计算了预测框和真实框的相交面积、中心点尺度和纵横比,由于纵横比是一个模糊值,不能准确反映实际大小。因此本文将宽高比做拆分,采用 $EIoU$ (efficient IoU)作为回归损失,其公式为

$$l_{EIoU} = l_{IoU} + l_{dis} + l_{asp} = 1 - IoU + \frac{\sigma^2(b, b^{gt})}{c^2} + \frac{\sigma^2(w, w^{gt})}{c_w^2} + \frac{\sigma^2(h, h^{gt})}{c_h^2} \quad (10)$$

式中: c_w 为预测框与真实框最小外接矩形宽度; c_h 为预测框与真实框最小外接矩形高度。 $EIoU$ 损失包括3个部分,分别是预测框与真实框的交集面积、中心点距离和宽高欧氏距离。本文从多方面考虑边框回归,旨在加快模型收敛过程。

3 非极大值抑制算法改进

基于锚框的目标检测,通常依据一些先验知识对目标大小进行假设,再在图像上将假设的锚框构建出来,进行回归任务。YOLOv5模型在特征提取后对每个特征图构建3个不同尺度的锚框,故在模型输出上回归出大量重叠的检测框。

非极大值抑制算法应用在模型后端,即剔除某一目标置信度低的边框,保留目标边框的局部最优值,如图5所示。首先对目标的所有边框按照置信度排序,取出置信度最大的边框设为局部最优 M ,随后计算其余边框与 M 的交并比(IoU),若交并比大于给定阈值,表明二者重合度大,为同一目标,则剔除该框。重复上述步骤,直至得到唯一检测框。

标准的非极大值抑制算法在同种目标严重遮挡

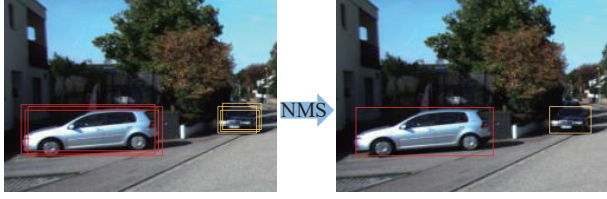


图5 非极大值抑制示意图

时,容易将遮挡目标直接删除,导致检测查全率偏低,如图6所示。



图6 严重遮挡示意图

标准NMS抑制函数为

$$s_i = \begin{cases} s_i, & iou(M, b_i) < N_i \\ 0, & iou(M, b_i) \geq N_i \end{cases} \quad (11)$$

式中: b_i 为预测边框; M 为置信度最大边框; N_i 为给定IoU阈值; s_i 为局部最优边框置信度得分。

本文将采用软化的非极大值抑制(Soft NMS)算法改善模型性能。将交并比大于阈值的预测框不直接剔除,而是进行惩罚衰减,使交并比大的具有较高惩罚,小的具有较低惩罚,中间呈过渡状态。因此Soft NMS抑制函数为

$$s_i = \begin{cases} s_i, & iou(M, b_i) < N_i \\ s_i e^{-\frac{iou(M, b_i)^2}{\sigma}}, & iou(M, b_i) \geq N_i \end{cases} \quad (12)$$

使用高斯惩罚函数对标准抑制函数改进,使交并比大于阈值的检测框得分衰减,保留检测框。

4 实验设计与验证

4.1 实验设置

将轻量级特征融合网络在KITTI数据集上进行训练和测试。实验基于ubuntu20.04系统搭建,配置Python=3.8,Pytorch=1.8。处理器Intel i5-12400H,显卡NVIDIA GTX3050,内存4GB。

训练中使用迁移学习方法对网络进行训练,首先设置加载的预训练权重路径,再设置部分超参数,其中训练批次设置为9,训练轮数设置为100,采用

随机梯度SGD优化算法,预热训练设置初始学习率,使用余弦退火算法逐步更新学习率。

4.2 评价指标

为验证融合网络的轻量化效果,使用参数量(Parameters)、计算量(FLOPs)和每秒检测帧数(FPS)对算法进行评价。同时,为验证融合网络的检测性能,使用精确率 P (precision)、召回率 R (recall)、 AP (average precision)、 mAP (mean average precision)对算法进行定量分析。指标计算公式为

$$P = \frac{TP}{TP + FP} \quad (13)$$

$$R = \frac{TP}{TP + FN} \quad (14)$$

$$AP = \int_0^1 P(r)dr \quad (15)$$

$$mAP = \frac{\sum_{i=1}^k AP_i}{k} \quad (16)$$

式中: TP 为检测为正样本的数量; FN 为检测为负样本的正样本数量; FP 为检测为正样本的负样本数量; P 为查准率; R 为查全率; k 为检测类别数。通常模型性能须综合分析多个指标。

4.3 定量分析

训练过程中的损失曲线、平均精度、精确率、召回率如图7所示。

由图7可见,本文轻量级融合网络的精确率 P 、召回率 R 、 $mAP@0.5$ 相比原始模型均有明显提升。由图7(a)损失曲线可见,本文算法优化边框损失后,明显提高网络收敛速度。再采用改进的非极大值抑制算法,网络的召回率也显著提高。本文融合网络的参数指标YOLOv5对比如表1所示。

由表1可见,本文融合算法在两支特征提取网络下,参数量和计算量略高于YOLOv5,每秒检测帧率降低2.19,但仍保持高帧率,以上表明本文算法轻量化效果显著。性能详细对比如表2所示。

由表2可见,本文算法在融合图像显著性特征后检测平均精度 $mAP@0.5$ 、 $mAP@0.5:0.95$ 分别高于YOLOv5算法3.8%、5.2%。同时,使用优化的边框损失函数和改进的非极大值抑制算法将网络的检测精确率和召回率分别提高3%、6.2%。

4.4 定性分析

将密集行人、车辆严重遮挡和目标模糊等复杂环境,分别使用YOLOv5算法和本文算法检测,可视化分析如图8所示。

由图8(a)可见,出现大量小目标的行人,在图中

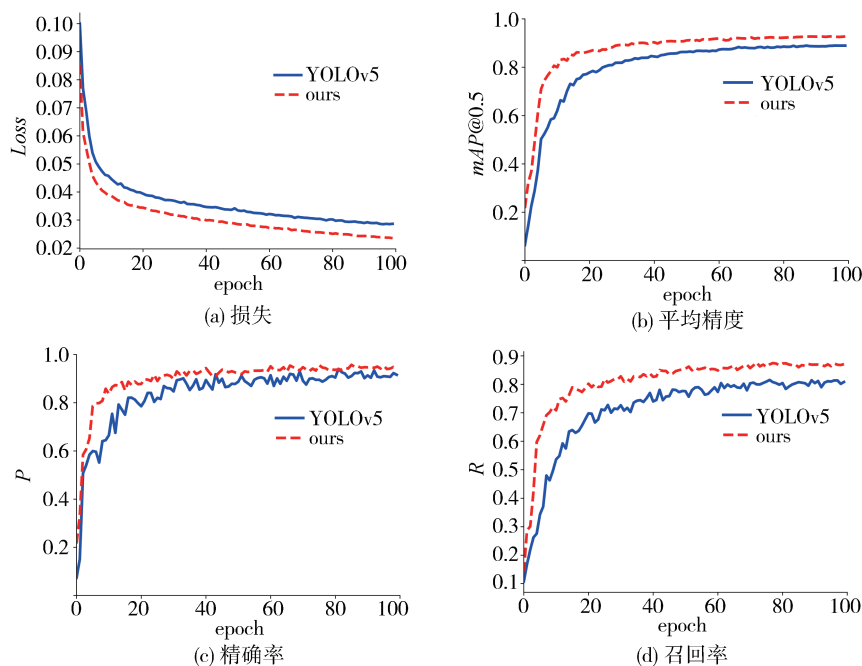


图7 Loss、R、P、mAP曲线

表1 模型参数对比

模型	参数量/M	计算量/M	FPS/(帧·s ⁻¹)
YOLOv5	7.03	16.0	62.51
本文算法	7.40	17.0	60.32

表2 模型性能对比

模型	P/%	R/%	mAP@0.5/%	mAP@0.5:0.95/%
YOLOv5	91.5	80.7	88.9	61.4
本文算法	94.5	86.9	92.7	66.6

蓝色箭头处存在部分遮挡, YOLOv5算法未能检测, 而图8(b)在本文算法融合图像显著性特征下, 将两个遮挡的行人正确检测。由图8(c)可见, 在蓝色箭头处, 车辆遮挡90%以上, 仅露出小部分车头, YOLOv5算法不能将该车辆检测, 而图8(d)中在本文改进的非极大值抑制算法下, 将该车辆正确识别, 大大提高了多个相同类别相互遮挡的检测性能。且对于图8(c)中绿色箭头处目标, 本文算法的边框回归也优于YOLOv5算法。由图8(e)可见, 图中蓝色箭头处, 出现两辆车, 颜色相同且部分区域重叠, 导致第2辆车与第1辆边界模糊, 与图像背景极为相似。使用YOLOv5算法不能将二者准确识别。而本文算法在图像显著性特征加持下, 结合边框优化算法, 将目标模糊的车辆准确识别。

这3张图像中一共有3处细节, 检测难度高于普通场景, 导致通用目标检测算法YOLOv5均未能准

确识别, 与其对比的本文算法针对特定检测任务优化算法, 改进效果明显。

4.5 消融实验

为深入探究本文改进算法对YOLOv5网络的提升效果, 表3逐点对比了各算法的消融实验分析。A表示使用融合显著性特征算法, B表示使用Ghost模块轻量化重构模型, C表示使用EIoU优化边框定位损失函数, D表示网络后端使用Soft-NMS算法。

由表3可见, 使用本文改进算法对YOLOv5检测性能均有小幅提高。其中, 算法1和算法2呈明显互补特性, 算法1使用显著性特征融合算法将网络精度提升至93.0%, 效果最为显著, 但是牺牲了检测速度。算法2使用轻量级模块重构网络, 检测速度达到最佳, 检测精度略显不足。算法3使用优化的边框损失函数, 在训练中网络收敛更快。算法4采用优化的非极大值抑制算法, 检测精度略微提升。由算法3和算法4检测效果可见, YOLOv5网络在现存卷积计算下, 模型拟合效果已最优。因此, 本文算法将算法1和算法2相结合, 同时加入算法3和算法4, 构建出融合图像显著性特征的轻量级目标检测算法, 在检测精度和检测速度达到最佳。

4.6 与其他目标检测算法性能对比

为进一步验证本文算法的轻量化效果和检测性能, 将本文算法和其他算法横向对比。如表4所示, 分别与Faster R-CNN、SSD、YOLOv3、YOLOv4等经



图8 YOLOv5和本文算法可视化对比

表3 在KITTI数据集上的消融实验

算法	改进算法				mAP@0.5/%	FPS/(帧·s ⁻¹)
	A	B	C	D		
YOLOv5					88.9	62.51
算法1	√				93.0	45.70
算法2		√			86.7	90.10
算法3			√		90.1	63.50
算法4				√	90.8	60.60
本文算法	√	√	√	√	92.7	60.32

表4 经典目标检测算法测试结果对比

模型	mAP@0.5/%	FPS/(帧·s ⁻¹)
Faster R-CNN	78.7	12.49
SSD	67.3	37.02
YOLOv3	75.6	56.78
YOLOv4	83.6	60.29
YOLOX	85.3	61.50
TPH-YOLOv5	89.6	58.30
本文算法	92.7	60.32

典目标检测框架对比。使用 $mAP@0.5$ 评估检测性能,使用 FPS 评估检测速度以及模型的复杂程度。由表4可见,在YOLOv4之前,双阶段检测算法Faster R-CNN检测精度最好,但是检测速度远远落后于单阶段算法。而对于新兴网络YOLOX和TPH-YOLOv5检测精度及检测速度显著提高。与之相比

本文算法在检测速度精度取得更好的平衡。可见本文检测网络在模型轻量化的同时,提高了检测精度,满足车辆检测需求。

5 结论

以经典目标检测算法YOLOv5作为研究的基本框架,针对实际交通环境中出现的密集小目标、严重遮挡和前景背景模糊的问题,构建了融合图像显著性特征的轻量级网络,研究结论如下。

(1)数据输入,在保留彩色图像基础上加入灰度特征,如边缘特征图、熵特征图和二值特征图,提高了网络对前景背景模糊目标的检测能力。

(2)模型结构,为使网络充分学习彩色图像和显著性的特征,重构模型骨干网络,搭建两个分支提取不同语义特征,并将其融合。考虑到融合网络结构庞大,使用Ghost模块将网络轻量化改进,并优化边框回归损失函数,加快了模型收敛。

(3)网络后端,针对标准非极大值抑制算法不能良好区分同类目标严重重叠的问题,采用添加惩罚的方式,改进了重叠最优目标框的选取。

实验表明,本文算法在与YOLOv5保持较少数量和计算量的条件下,将网络精度提高3.8%,将精确率和召回率分别提高3%和6.2%,且对于密集遮挡和前景背景模糊目标的检测效果改善显著。

参考文献

- [1] BISWAS S K, MILANFAR P. Linear support tensor machine with LSK channels: pedestrian detection in thermal infrared images [J]. *IEEE Transactions on Image Processing*, 2017, 26 (9): 4229-4242.
- [2] LIU J, ZHANG S, WANG S, et al. Multispectral deep neural networks for pedestrian detection [J]. *arXiv preprint arXiv: 1611.02644*, 2016.
- [3] 李晓艳, 符惠桐, 牛文涛, 等. 基于深度学习的多模态行人检测算法[J]. *西安交通大学学报*, 2022, 56(10): 61-70.
LI X Y, FU H T, NIU W T, et al. Multi-modal pedestrian detection algorithm based on deep learning [J]. *Journal of Xi'an Jiaotong University*, 2022, 56(10): 61-70.
- [4] KONIG D, ADAM M, JARVERS C, et al. Fully convolutional region proposal networks for multispectral person detection [C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, USA: IEEE*, 2017: 49-56.
- [5] 邹伟, 殷国栋, 刘昊吉, 等. 基于多模态特征融合的自动驾驶车辆低辨识度目标检测方法[J]. *中国机械工程*, 2021, 32(9): 1114-1125.
ZOU W, YIN G, LIU H J, et al. Low-observable target detection method for autonomous vehicles based on multi-modal feature fusion [J]. *China Mechanical Engineering*, 2021, 32(9): 1114-1125.
- [6] ZHANG L, LIU Z, ZHANG S, et al. Cross-modality interactive attention network for multispectral pedestrian detection [J]. *Information Fusion*, 2019, 50: 20-29.
- [7] 刘子龙, 沈祥飞. 融合Lite-HRNet的Yolo v5 双模态自动驾驶小目标检测方法[J]. *汽车工程*, 2022, 44(10): 1511-1520.
LIU Z L, SHEN X F. Small target detection method for dual-modal autonomous driving with Yolo v5 and Lite-HRNet fusion [J]. *Automotive Engineering*, 2022, 44(10): 1511-1520.
- [8] FU L, GU W, AI Y, et al. Adaptive spatial pixel-level feature fusion network for multispectral pedestrian detection [J]. *Infrared Physics & Technology*, 2021, 116: 103770.
- [9] 吴建国, 邵婷, 刘政怡. 融合显著深度特征的RGB-D图像显著目标检测[J]. *电子与信息学报*, 2017, 39(9): 2148-2154.
WU J G, SHAO T, LIU Z Y. RGB-D saliency detection based on integration feature of color and depth saliency map [J]. *Journal of Electronics & Information Technology*, 2017, 39(9): 2148-2154.
- [10] CHEN S, TAN X, WANG B, et al. Reverse attention for salient object detection [C]. *Proceedings of the European Conference on Computer Vision. Heidelberg: Springer*, 2018: 234-250.
- [11] WANG W, ZHAO S, SHEN J, et al. Salient object detection with pyramid attention and salient edges [C]. *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE*, 2019: 1448-1457.
- [12] 刘迪, 郭继昌, 汪显东, 等. 融合注意力机制的多尺度显著性目标检测网络[J]. *西安电子科技大学学报*, 2022, 49(4): 118-126.
LIU D, GUO J C, WANG Y D, et al. Multi-scale salient object detection network combining an attention mechanism [J]. *Journal of Xidian University*, 2022, 49(4): 118-126.
- [13] 许小伟, 陈乾坤, 钱枫, 等. 基于小型化YOLOv3的实时车辆检测及跟踪算法[J]. *公路交通科技*, 2020, 37(8): 149-158.
XU X W, CHEN Q K, QIAN F, et al. REAL-time vehicle detection and tracking based on miniaturized YOLOv3 [J]. *Journal of Highway and Transportation Research and Development*, 2020, 37(8): 149-158.
- [14] 杜虢龙, 余华平. 基于改进Mobile Net-SSD网络的驾驶员分心行为检测[J]. *公路交通科技*, 2022, 39(3): 160-166.
DU X L, YU H P. Detecting driver's distracted behavior based on improved mobile Net-SSD network [J]. *Journal of Highway and Transportation Research and Development*, 2022, 39(3): 160-166.
- [15] SOBEL I, FELDMAN G. A 3x3 isotropic gradient operator for image processing [J]. *A Talk at the Stanford Artificial Project in*, 1968: 271-272.
- [16] CANNY J. A computational approach to edge detection [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1986 (6): 679-698.
- [17] WANG J S, REN X D. GLCM based extraction of flame image texture features and KPCA-GLVQ recognition method for rotary kiln combustion working conditions [J]. *International Journal of Automation and Computing*, 2014, 11(1): 72-77.
- [18] HAN K, WANG Y, TIAN Q, et al. Ghostnet: more features from cheap operations [C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual: IEEE*, 2020: 1580-1589.
- [19] REZATOFIGHI H, TSOI N, GWAK J Y, et al. Generalized intersection over union: a metric and a loss for bounding box regression [C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach: IEEE*, 2019: 658-666.
- [20] ZHENG Z, WANG P, LIU W, et al. Distance-IoU loss: faster and better learning for bounding box regression [C]. *Proceedings of the AAAI Conference on Artificial Intelligence, New York: AAAI*, 2020, 34(7): 12993-13000.