



中国科技期刊卓越行动计划项目入选期刊

控制与决策

CONTROL AND DECISION



基于改进YOLOv4的轻量化路侧视角多目标检测算法

金立生, 张舜然, 郭柏苍, 王欢欢, 韩卓桐, 刘星辰

引用本文:

金立生,张舜然,郭柏苍,王欢欢,韩卓桐,刘星辰. 基于改进YOLOv4的轻量化路侧视角多目标检测算法[J]. 控制与决策, 2024, 39(9): 2885–2893.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2023.0545>

您可能感兴趣的其他文章

Articles you may be interested in

多目标小尺度车辆目标检测方法

Multi-target and small-scale vehicle target detection method

控制与决策. 2021, 36(11): 2707–2712 <https://doi.org/10.13195/j.kzyjc.2020.0635>

基于MobileNet的多目标跟踪深度学习算法

Deep learning algorithm based on MobileNet for multi-target tracking

控制与决策. 2021, 36(8): 1991–1996 <https://doi.org/10.13195/j.kzyjc.2019.1424>

改进YOLOv2的端到端自然场景中文字符检测

End-to-end Chinese character detection in natural scene based on improved YOLOv2

控制与决策. 2021, 36(10): 2483–2489 <https://doi.org/10.13195/j.kzyjc.2020.0270>

基于双分支特征融合的场景文本检测方法

A scene text detection based on dual-path feature fusion

控制与决策. 2021, 36(9): 2179–2186 <https://doi.org/10.13195/j.kzyjc.2020.0002>

复杂背景下全景视频运动小目标检测算法

Panoramic video motion small target detection algorithm in complex background

控制与决策. 2021, 36(1): 249–256 <https://doi.org/10.13195/j.kzyjc.2019.0686>

基于改进YOLOv4的轻量化路侧视角多目标检测算法

金立生^{1,2}, 张舜然², 郭柏苍^{1,2†}, 王欢欢¹, 韩卓桐¹, 刘星辰¹

(1. 燕山大学 车辆与能源学院, 河北 秦皇岛 066000; 2. 吉林大学 交通学院, 长春 130022)

摘要: 面向道路交通场景中多类别、可变规模的目标车辆检测需求, 如何有效地以低算力构建结构化数据, 实现超视距感知并解决单车视距限制, 是自动驾驶汽车环境感知技术领域亟待解决的重要问题之一. 为此, 提出兼顾精度和实时性的轻量化路侧视角多目标检测算法. 首先, 以嵌入通道域注意力机制的倒残差网络结构代替单阶段检测算法特征提取网络作为网络骨干部分, 利用深度可分离卷积降低特征提取网络参数量; 其次, 采用空间金字塔池化 (spatial pyramid pooling, SPP) 处理深层网络输出特征图, 选取轻量化后骨干网络不同深度特征图的输出, 利用路径聚合网络 (path aggregation network, PANet) 融合深层语义信息与浅层表面信息构成检测模型颈部; 最后, 在检测模型头部设置3种不同特征图大小的网络输出, 使同一图像信息不同尺寸目标在适宜网络深度进行目标信息回归, 提出轻量化路侧视角多目标检测算法M3-YOLOv4. 实验结果表明, M3-YOLOv4在数据集RS-UA表现的mAP值为0.906, 相较YOLOv4的mAP值下降1.1%, M3-YOLOv4模型参数量缩减为YOLOv4的10%, 同平台下模型前向推理速度也具备明显优势.

关键词: 自动驾驶; 环境感知; 路侧视角; 多目标检测; YOLOv4; 深度学习

中图分类号: U463.6

文献标志码: A

DOI: 10.13195/j.kzyjc.2023.0545

引用格式: 金立生, 张舜然, 郭柏苍, 等. 基于改进YOLOv4的轻量化路侧视角多目标检测算法[J]. 控制与决策, 2024, 39(9): 2885-2893.

A lightweight multiple object detection algorithm for roadside perspective based on improved YOLOv4

JIN Li-sheng^{1,2}, ZHANG Shun-ran², GUO Bai-cang^{1,2†}, WANG Huan-huan¹, HAN Zhuo-tong¹, LIU Xing-chen¹

(1. College of Vehicles and Energy, Yanshan University, Qinhuangdao 066000, China; 2. College of Transportation, Jilin University, Changchun 130022, China)

Abstract: Facing the detection requirements of multi category and variable scale vehicles in the road traffic scene, how to effectively construct structured data with low computational power to achieve beyond sight distance perception, and overcome the limitation of single vehicle sight distance is one of the important problems to be solved in the field of autonomous vehicle environment perception technology. In this paper, we propose a lightweight roadside perspective based multi object detection algorithm that balances accuracy and real-time performance. First, a reverse residual network structure embedded in the channel domain attention mechanism is used as the backbone of the network, replacing the single stage detection algorithm feature extraction network with a deep separable convolution to reduce the number of feature extraction network parameters. Second, spatial pyramid pooling (SPP) is used to process the output feature map of deep networks, then we select maps of different depth feature in the lightweight backbone network to output, and use the path aggregation network (PANet) to fuse deep semantic information and shallow superficial information to form the neck of the detection model. Finally, at appropriate network depth, three different network outputs of feature map sizes are set at the head of the detection model to regress the target information of different sizes of targets in the same image. A lightweight detection model M3-YOLOv4 is established. The experimental results show that the mAP of M3-YOLOv4 on RS-UA dataset is 0.906, which performs 1.1% decrease compared to the YOLOv4. The parameter quantity of the M3-YOLOv4 model is reduced to 10% of the YOLOv4, and the forward inference speed of the model on the same platform also shows significant advantages.

Keywords: autonomous vehicle; environmental perception; roadside perspective; multiple object detection; YOLOv4; deep learning

收稿日期: 2023-04-24; 录用日期: 2023-09-07.

基金项目: 国家自然科学基金项目 (52072333); 河北省省级科技计划项目 (21340801D); 河北省高等学校科学技术研究项目 (BJK2023026).

†通讯作者. E-mail: guobaicang@ysu.edu.cn.

0 引言

车路协同自动驾驶的关键技术在朝着以各类传感器为基础的多模态感知技术、基于云平台技术的智能网联交通分布式云平台、融合网联化智能技术的自动驾驶技术以及车路一体化的智能交通系统优化技术等方向融合发展^[1-2]。环境感知是车路协同与自动驾驶相关领域的关键,准确检测复杂交通场景中的交通参与者,既是实现安全驾驶的基础,也是汽车产业智能化发展所需要攻克的关键问题。

在感知系统构建方面,相较于车载视角信息采集方案存在的感知视距有限、传感器组成方案成本高昂、系统内各感知模块间的信息冗余等问题,路侧感知系统(roadside sensing system, RSS)^[3]作为智能交通系统中的关键组成部分,已成为车路协同融合感知的重要解决方案。路侧感知系统通过将多源异构传感器安装于路侧高处,能够获得大范围的感知距离以减少感知视野盲区,通过提高道路利用率以发挥车路协同的优势。路侧感知系统拓宽了自动驾驶车辆的感知范围,能够为网联场景内的交通参与者与交通指挥中心提供全域覆盖、全天候感知、实时有效的路况信息。

目标检测任务是自动驾驶环境感知技术领域的底层任务,该任务需要找出感知信息中感兴趣的待检测目标,定位并识别目标属性信息^[4]。然而,由于路侧设备和边缘计算平台视角固定的问题^[5],基于路侧视角进行物体检测一直是一项具有挑战性的任务,主要挑战如下:

1) 图像空间位置信息的长距离依赖性。同类车辆目标在路侧视角采集的图像信息中,由于空间位置不同导致成像后呈现不同尺度的问题。

2) 轻量化与准确性。目标检测任务需要在路边的边缘计算平台上完成,这对轻量级、实时、高精度的检测模型提出了很高的要求。

现阶段,目标检测主要分为基于传统方法的车辆目标检测和基于深度学习的车辆目标检测^[6]。传统的目标检测技术通过车辆的外形特征对目标物进行识别和检测,并以此将前景与背景进行划分,在外形识别的检测器中,检测器对车辆位置进行假定并后续通过验证该假定是否正确来实现对车辆目标物的检测^[7-8]。文献[9]通过HOG特征与支持向量机(support vector machines, SVM)算法结合的方式实现了对行人的检测;文献[10]同样利用了HOG特征与SVM结合的方式实现了对自行车的跟踪与检测;文献[7]提取了分层梯度方向直方图(pyramid histogram

of oriented gradients, PHOG)特征并利用主成分分析减小特征的维数,并结合SVM以及遗传算法最终进行了验证;文献[11]使用real Adaboost分类器将Haar特征与边缘梯度灰度图特征相结合,实现了对图像中目标物候选区域的分类。上述几种传统方法的实现过程较为复杂且计算量较大,其检测过程分为两步,当第1阶段中出现一定的提取特征错误时会导致第2阶段误差更为明显。

基于深度学习的目标检测方法在一定程度上解决了传统方法计算过程低效和复杂等缺陷^[12]。Redmon等^[13]提出了单阶段检测算法YOLO(you only look once),将图片作为深度学习的神经网络输入,输出为检测结果。在此基础上,Liu等^[14]根据YOLO中的小目标问题做出补充和改进,提出了SSD(single shot multibox detector)目标检测算法,进一步融合了SSD和FPN(feature pyramid networks)的优点,提出了DSSD(feature fusion single shot multibox detector)目标检测算法,在检测小目标物体时提升了检测的精度和准确率。文献[15]采用轻量级骨干网络MobileNetV2替换YOLOv3的原始骨干网络Darknet-53进行特征提取,提出了适用于嵌入式系统的改进YOLOv3目标检测算法。在算力受限的车载计算平台上,金立生等^[16-17]针对原始YOLOv3在实时检测方面的挑战,使用轻量级骨干网络MobileNetV2替代Darknet-53进行特征提取,提高了算法的性能和效率。YOLOv4修改了主干网络CSPDarknet53,并使用空间金字塔池化(spatial pyramid pooling, SPP)提高了目标检测速度^[18]。YOLOv5^[19]与前4版的YOLO算法不同之处在于,YOLOv5的backbone采用了Focus结构和CSP结构^[20]。该检测算法在分类时扩大了正样本的锚框,这种改进手段对模型训练时的收敛速度有所提升,但其本身算法结构已与两阶段更为近似。综上,现有文献多以算法耗时和检测精度为优化目标,但考虑算法耗时和检测精度互相耦合,使得算法实时性和准确性难以兼顾。

针对前述问题,考虑到算法的泛化性和路侧视角的数据特点,本文提出基于YOLOv4的轻量化单阶段多目标检测算法。依据模型检测任务通过深度可分离卷积降低特征提取网络参数量,提出基于MobileNetV3的骨干网络结构;采用空间金字塔池化处理深层网络输出特征图,使网络对各类别多尺寸目标特征表达各具鲁棒性;合理选取轻量化后骨干网络不同深度特征图输出,通过路径聚合方法实现深层

语义信息与浅层目标表现信息融合,并在网络检测头部分设置3种不同特征图大小的网络输出,使同一图像信息不同尺寸目标在适宜网络深度进行目标信息回归,并最终解码为可视化表达形式。

1 路侧视角下多目标检测数据集建立

1.1 路侧视角下的公开数据集UA-DETRAC

UA-DETRAC数据集是用以对检测和多目标跟踪算法效果进行评估的视频数据集,其中含有超过140 000帧的国内真实交通场景,约1 210 000个可用于不同类别车辆目标检测研究的目标包围框。该数

据集的全部视频数据均由安置于道路两旁与道路上方的视觉摄像机采集,均为路侧视角下的图像数据。

UA-DETRAC数据集包含训练集视频流共83 791帧,标注共577 899处包围框;测试集视频流共56 340帧,标注共632 270处包围框。训练集注释了8 250辆车,训练集5 936辆车,即轿车5 177辆,公交车106辆,厢式货车610辆,其他车辆43辆;测试集共2 314辆车,即轿车1 961辆,公交车199辆,厢式货车123辆,其他车辆31辆。数据集的部分注释图像可视化信息如图1所示^[21]。

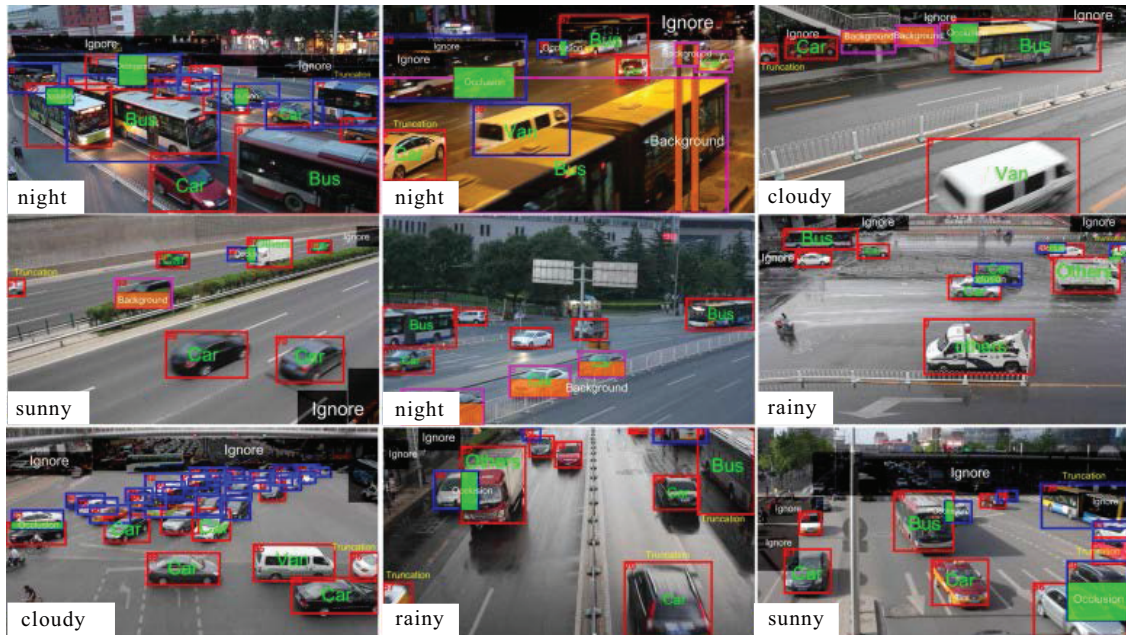


图1 UA-DETRAC数据集部分注释图像可视化信息

图1中包围框的颜色表示遮挡级别,完全可见为红色包围框,部分被其他车辆遮挡为蓝色包围框,部分被背景遮挡为粉色包围框。黑色不透明区域是数据集标注中未采用的背景区域,绿色不透明区域是被其他车辆遮挡的区域,橙色不透明区域是被背景杂波所遮挡的区域。照明条件见每帧图左下角的文字。

UA-DETRAC最初被设计用以多目标跟踪深度学习模型训练与评价任务,数据信息通过数据流保存,数据集形式为每秒25帧的图像序列,连续序列间存在大量相似的图像特征。为此,需要对原始数据采用随机采样方法降低数据分布相似度,防止模型训练过程中出现对特定场景过拟合,避免出现对差异性较大的道路场景泛化能力下降问题。此外,UA-DETRAC对光照条件进行分类,分别为sunny、night、rainy和cloudy,而在实际场景中,由光源直射导致成像过曝,由雪天气候导致道路结构化特点改变,或积雪覆盖路面反射光线影响成像质量等问题未被该数

据集覆盖。

针对上述问题,本文结合场景需求,对路侧视角下交通场景进行实地图像信息采集,完成图像预处理并进行手工标注,扩充抽样后样本量减少的原有公开数据集UA-DETRAC,使数据集样本更加符合本文场景要求。

1.2 自主采集数据集的建立

选取人行天桥作为图像采集位置,其具有较好的视野高度,符合路侧视角图像采集特点,同时可保证实验人员操作安全性。图2(a)采集地点选取为城市内快速路人行天桥处,该地点道路为主干道,交通流量大,道路两侧设有公交车站,存在大型车辆(如公交车)频繁通过,交通参与者种类更为复杂,符合本文对多类车辆检测需求。图2(b)采集路段为双向六车道城市环线道路人行天桥处,采集时气候条件为雨天场景,具有特殊性,可一定程度扩充具有雨天气候特点的数据样本数量。



(a) 东北地区城市快速路



(b) 华北地区城市路段

图2 采集地点卫星图

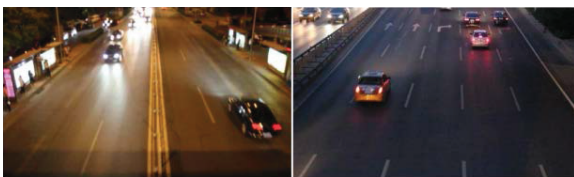
在东北和华北地区分别采集 15 min 视频流数据, 包括不同时间点的光照条件(照度正常、逆光过曝、照度不足等)、不同拍摄角度的交通场景图像信息, 图像分辨率为 1920×1080 px, 对焦模式采用 AF-C 连续自动对焦模式, 每秒传输帧率 30, 采集视频流数据以 *.jpg 格式进行解析与保存, 编写脚本进行样本抽样, 降低连续帧中相同或相似交通场景出现频次. 东北地区选取 642 张图像样本, 华北地区选取 360 张图像样本, 部分展示如图 3 所示.



(a) 照度正常



(b) 逆光过曝



(c) 夜间照度不足

图3 自采集样本示意图

对自主采集的图像序列进行预处理与标注工作, 共标注图像 1 500 张, 目标 13 951 个, 完成可训练的自主采集数据集制作.

1.3 路侧视角下多目标检测数据集 RS-UA

对路侧视角下公开数据集 UA-DETRAC 进行数据解析与标签提取, 同时进行实地数据采集并进行前期预处理与手工标注. 由于已经对两者数据集标注方式与标注存储格式进行了统一, 可对两类数据集进行融合, 得到本文最终使用的路侧视角下多目标检测数据集. 因为本数据集基于公开数据集 UA-DETRAC 建立, 补充样本具备路侧感知特点, 故命名为 RS-UA 数据集. 相较于原公开路侧数据集 UA-DETRAC, 本文采用的 RS-UA 数据集训练样本量由原有公开数据集的 10 820 帧(完成采样处理后)与自主采集选取图像样本中 1 000 帧图像构成, 测试样本量由原有公开数据集中抽取的 6 760 帧与自主采集数据集中 500 帧图像构成.

2 基于 YOLOv4 的轻量化多目标检测模型构建

2.1 轻量化单阶段多目标检测模型网络结构设计

YOLOv4 的骨干网络采用 CSPDarkNet53^[22], 由于本文针对路侧边缘计算平台的模型运行场景进行模型设计, 需对模型提出降低模型计算量、减少训练参数量并能对多尺度目标与小尺度目标实现实时检测功能的要求. 因此, 本文结合 MobileNet 系列网络轻量化的特点, 采用 MobileNetV3 的特征提取网络替换 CSPDarkNet53 作为本文模型的骨干网络, 完成对图像进行卷积并提取车辆目标特征信息的任务, 极大地减少了该检测模型的计算量与参数量. 为保证对不同深度输出的特征层进行融合, 实现对路侧视角下多尺度目标与小尺度目标特征信息的正确表达, 在特征图输出部分采用原网络中空间金字塔池化 (spatial pyramid pooling, SPP) 与路径聚合网络 (path aggregation network, PANet) 对网络不同深度特征信息进行处理^[23-24].

SPP 在原骨干网络 CSPDarkNet53 的最后一个特征层进行 3 次由卷积、批标准化和采用 Leaky ReLU 函数激活构成的运算单元后, 利用 4 个不同尺度的最大池化进行处理, 由于其本身的结构特点, 使得无论输入图像尺寸多大, 都可输出固定大小的特征图^[25], 提升了不同高宽比图像的尺度不变性. 同时, 由于训练过程中多尺度训练样本相较于单一尺寸样本更易使网络收敛, 不同尺度池化能够极大增加感受野, 能更好地分离出最显著的上下文特征.

路径聚合网络的一个重要特点是特征的反复提取. PANet的网络结构如图4所示, (a)处为传统的特征金字塔结构,在完成特征金字塔从下到上的特征提取后,还需要实现(b)中从上到下的特征提取. 该网络结构旨在沟通不同深度特征层信息. 由于本文应用场景为路侧视角下的车辆目标检测,远处目标成像尺寸小,因卷积网络特点,在深层特征图上尺寸可能小于单独像素点所代表的感受野,无法得到较好的表达,而同类目标因其不同空间位置成像尺寸大小各异,需在不同网络深度进行输出,因此路径聚合网络

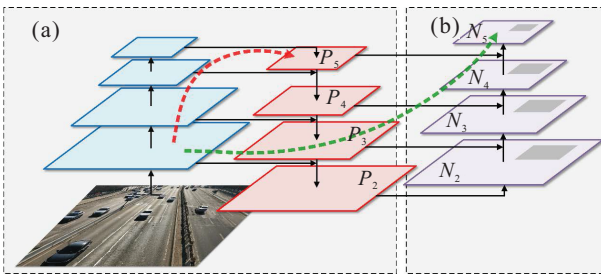


图4 路径聚合网络前端结构示意图

通过沟通网络深浅层信息,使提取变尺寸目标与小尺寸目标特征的表现和语义信息表达更具鲁棒性.

根据对网络应用需求与算法层面可行性分析,本文设计的轻量化单阶段多目标检测模型 MobileNetV3-YOLOv4 (M3-YOLOv4) 结构示意图如图5所示. 以输入图像尺寸 416×516 为例,对本文设计的检测网络进行描述. 图像先经过由二维卷积运算、批标准化与非线性函数激活组合成的 Conv2d 模块,后通过 MobileNetV3 骨干网络所应用的倒残差模块 bneck 进行特征提取与图像尺寸的降采样. 进而,采用 YOLOv4 的空间金字塔池化层使深层特征层获得更广阔的感受野,将骨干网络第6、第12及经过 SPP 结构处理的特征层输入 PANet 网络结构,实现低层图像表现特征与高层特征层语义信息的充分融合,增强网络对多尺度目标与小尺寸目标的表达能力. 最终,通过 YOLOv4 所使用的检测头进行目标位置与类别信息的回归处理,实现对路侧视角下车辆目标的检测任务.

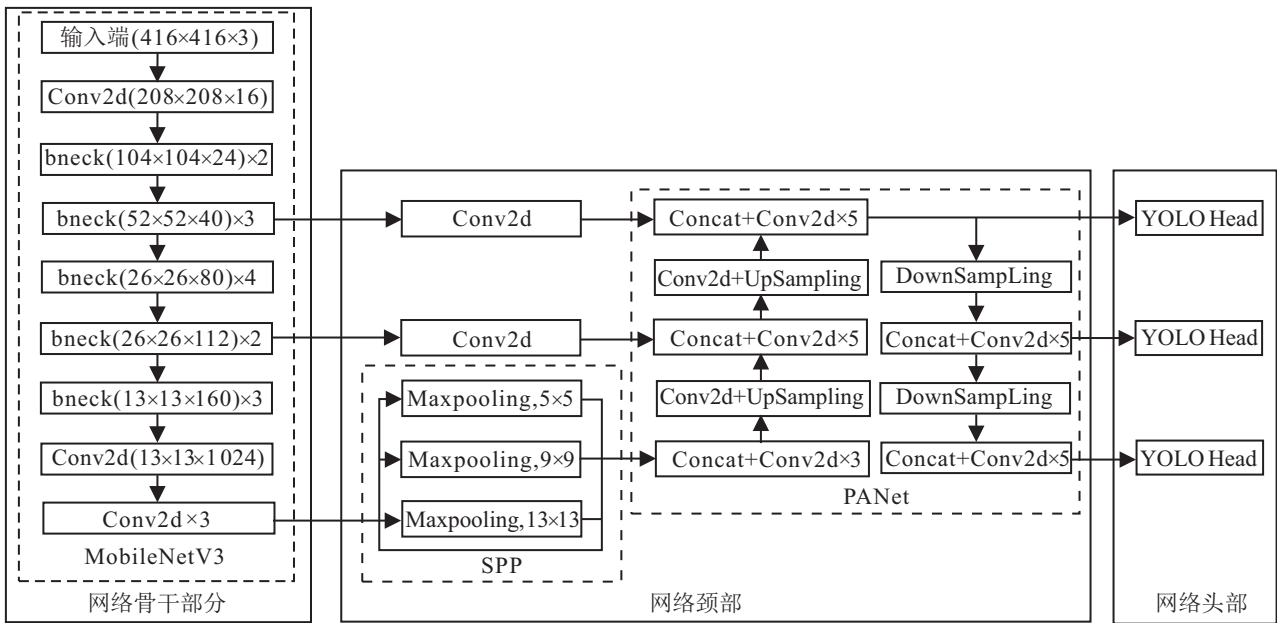


图5 M3-YOLOv4网络结构

2.2 M3-YOLOv4模型损失函数设计

M3-YOLOv4 模型总损失由置信度损失、类别损失与定位损失构成. 置信度损失采用二值交叉熵损失进行定义,其原理在于该函数可衡量对比两者间分布差异,从而使得训练模型所学到的分布趋于训练数据分布,该损失函数如下:

$$L_{\text{conf}}(o, c) = - \frac{\sum (o_i \ln(\hat{c}_i) + (1 - o_i) \ln(1 - \hat{c}_i))}{N} \quad (1)$$

其中: $o_i \in [0, 1]$ 为预测目标包围框与真值包围框的交并比, c_i 为预测值, \hat{c}_i 为依据预测值 c_i 计算得出的预测置信度, N 为正负样本个数.

类别预测部分同样使用二值交叉熵损失构建损失函数,如下所示:

$$L_{\text{cls}}(O, C) = \frac{\sum_{i \in \text{pos}} \sum_{j \in \text{cls}} (O_{ij} \ln(\hat{C}_{ij}) + (1 - O_{ij}) \ln(1 - \hat{C}_{ij}))}{N_{\text{pos}}} \quad (2)$$

其中: $O_{ij} \in \{0, 1\}$ 表示预测包围框 i (此处为正样本)

中是否存在 j 类目标, 是则取 1, 否则取 0; C_{ij} 为网络预测值, 通过 sigmoid 函数激活得到 \hat{C}_{ij} , 表示为 j 类概率; N_{pos} 为正样本个数.

定位损失部分, 本文网络包围框可训练参数如图 6 所示^[26]. 图中为某一预测输出特征层, 虚线包围框为网络对数据集自适应预设先验锚框, 在本文训练小节会对该部分进行计算与设置, 其中心坐标即 (c_x, c_y) , 在该层输出层中高为 p_h 、宽为 p_w , 网络训练参数分别为 t_x, t_y, t_h, t_w . 其中: $b_x = \sigma(t_x) + c_x$ 和 $b_y = \sigma(t_y) + c_y$ 分别表示预测包围框中心点的纵横坐标位置, $b_w = p_w e^{t_w}$ 和 $b_h = p_h e^{t_h}$ 分别表示该包围框的宽与高. 显然, 训练参数 t_x, t_y 指导中心点坐标位置的生成, t_h, t_w 指导包围框高与宽的生成, 分别进行以 e 为底的指数运算, 与预设锚框高宽相乘进行调整, 得到预测包围框.

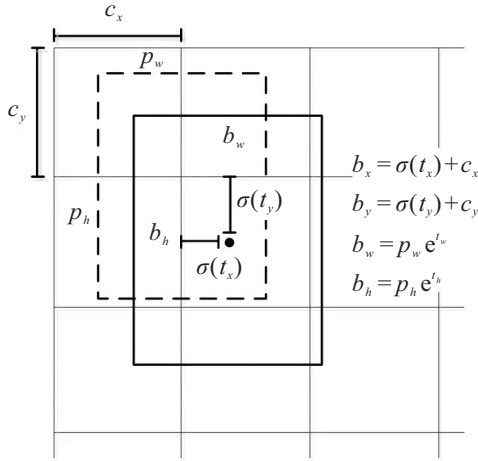


图 6 网络包围框可训练参数

在设置定位损失时, 需将真值包围框信息进行转换, 从而生成损失

$$L_{\text{loc}}(l, g) = \frac{\sum_{i \in \text{pos}} \sum_{m \in \{x, y, w, h\}} (\hat{l}_i^m - \hat{g}_i^m)^2}{N_{\text{pos}}}; \quad (3)$$

$$\hat{l}_i^x = \sigma(t_x), \hat{l}_i^y = \sigma(t_y); \quad (4)$$

$$\hat{l}_i^w = t_w, \hat{l}_i^h = t_h; \quad (5)$$

$$\hat{g}_i^x = g_i^x - c_i^x, \hat{g}_i^y = g_i^y - c_i^y; \quad (6)$$

$$\hat{g}_i^w = \ln(g_i^w / p_i^w), \hat{g}_i^h = \ln(g_i^h / p_i^h). \quad (7)$$

其中: 真值包围框信息 $g_i^x, g_i^y, g_i^w, g_i^h$ 通过式(6)、(7)转换为 \hat{g}_i^m , 与训练参数 \hat{l}_i^m 统一形式, 通过欧氏距离公式完成损失生成.

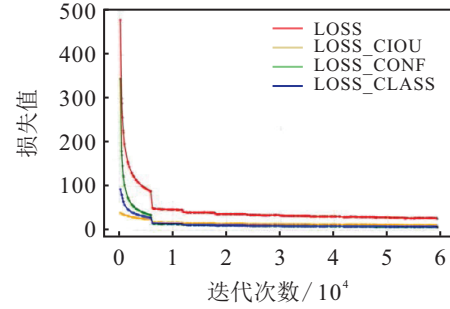
综上, 本文模型训练总损失函数如下:

$$L(o, c, O, C, l, g) = \lambda_1 L_{\text{conf}}(o, c) + \lambda_2 L_{\text{cls}}(O, C) + \lambda_3 L_{\text{loc}}(l, g), \quad (8)$$

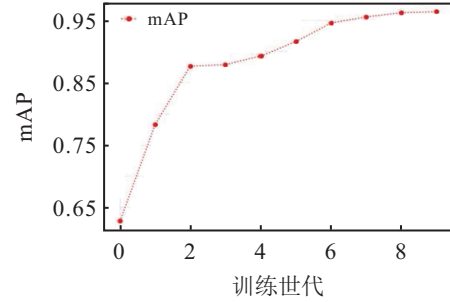
其中 $\lambda_1, \lambda_2, \lambda_3$ 为平衡系数.

3 M3-YOLOv4模型的训练与测试

对训练过程产生的交并比损失、置信度损失、类别损失与训练总损失 4 项损失的下陷情况进行记录, 同时, 每完成一次所有数据的训练, 同步对当前模型权重在验证集上进行验证, 计算得 mAP 值并绘制曲线图. YOLOv4 与本文设计的 M3-YOLOv4 网络训练过程可视化如图 7 和图 8 所示. 损失图中红色折线为训练总损失, 黄色折线为交并比损失, 绿色折线为置信度损失, 蓝色折线为类别损失.

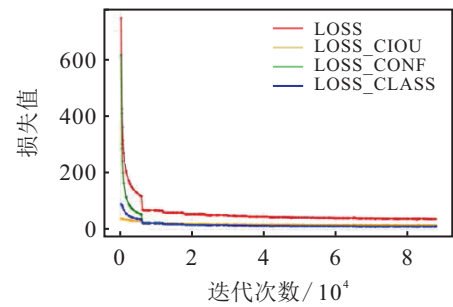


(a) YOLOv4 网络训练损失

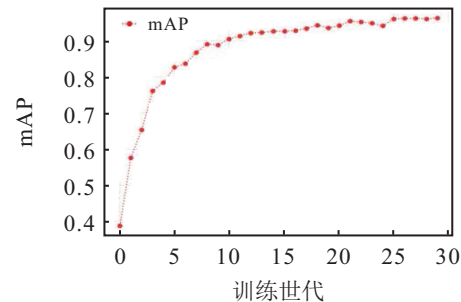


(b) YOLOv4 网络训练验证集 mAP 变化

图 7 YOLOv4 训练过程损失图与验证集 mAP 曲线



(a) M3-YOLOv4 网络训练损失



(b) M3-YOLOv4 网络训练验证集 mAP 变化

图 8 M3-YOLOv4 训练过程损失图与验证集 mAP 曲线

YOLOv4 训练开始时, 模型训练总损失值为 480.30, 经过约 60 000 次迭代, 最终损失稳定在 (26, 27) 区间内, 第 59 540 次迭代所得损失值为 26.74, 为本次训练总损失的最小值. mAP 曲线图则较为直观地反映出深度学习网络对数据集的拟合效果, 可以发现, mAP 值最终稳定在 96%, 在第 10 epoch 处取得 mAP 最大值为 96.36%.

M3-YOLOv4 训练开始时, M3-YOLOv4 未使用预训练权重, 训练总损失值较大为 746.70, 经过约 180 000 次迭代, 最终训练损失稳定在 32 附近, 第 177 300 次迭代所得损失值为 31.92, 为训练损失的最小值. mAP 值稳定在 96% 附近, 说明 M3-YOLOv4 与 YOLOv4 网络对本文训练数据集拟合性能均较好, 能够在对模型进行轻量化处理后保证对数据集的学习能力, 其在 28 epoch 处取得 mAP 最大值 96.32%. YOLOv4 与 M3-YOLOv4 的 AP 值和 mAP 值对比如表 1 所示.

表 1 YOLOv4 与 M3-YOLOv4 的 AP 值和 mAP 值对比

模型	AP				mAP
	轿车	公交车	厢式货车	其他	
YOLOv4	0.959	0.970	0.844	0.896	0.917
M3-YOLOv4	0.947	0.975	0.813	0.890	0.906

经过对模型特征提取网络轻量化处理, 轿车、厢式货车与其他车辆类别的检测 AP 值均有所下降, 但下降幅度较小, M3-YOLOv4 相较 YOLOv4, AP 值平均下降幅度约为后者 AP 值的 1.9%, 而公交车检测 AP 值上升 0.5%, M3-YOLOv4 检测结果的 mAP 仅下降 1.1%, 基本维持了原有网络 YOLOv4 对多目标较好的检测效果.

在检测速度与模型体积方面, 如图 9 所示, M3-YOLOv4 模型在本文实验验证硬件平台 CPU 下的 inference time 为 88.26 ms, 在 GPU 下则能达到 21.23 ms, 模型参数量大小为 31.13 M. 作为对照, 基础模型 YOLOv4 在 CPU 下的 inference time 为 286.00 ms, GPU 下该指标为 33.52 ms, 模型参数量大小为 264.22 M. 轻量化改进网络 M3-YOLOv4 在损失较小程度各类别检测精度的前提下, 模型参数量大约缩减为原有基础网络参数量大小的 10%, 网络前向推理速度相较 YOLOv4 网络也具有较为明显的优势. 本模型针对路侧视角多目标检测所提出的实时性、轻量化要求达到较好的效果.

检测性能需要从精度和实时性等方面对模型进行评估, 与其他 SOTA 模型对比分析结果如表 2 所示, 其中加粗字体表示该指标达到最优. 从表 2 可以看

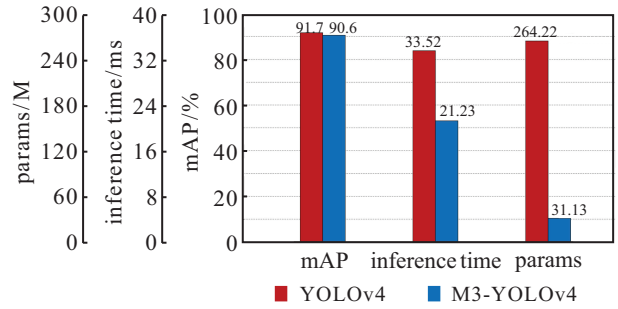


图 9 YOLOv4 与 M3-YOLOv4 的 3 项指标对比

出, 本文提出的方法在检测精度上虽然降低了 1.1%, 但是在模型的参数量和实时性有较大提升, 相比之下均达到了当前最优. 综合来看, 本文提出的模型在不过度损失精度的前提下, 显著地提升了算法实时性, 实现了模型的轻量化改进. 选取一组图像样本输入检测模型进行车辆目标检测, 实际检测效果可视化如图 10 所示.

表 2 不同模型的性能对比

模型	mAP↑	Params/M↓	FPS/Hz↑
RCNN ^[27]	0.593	—	0.1
Faster R-CNN ^[28]	0.827	79.32	17
SSD ^[28]	0.684	68.53	34
YOLOv3 ^[28]	0.747	65.47	21
YOLOv4-Tiny ^[28]	0.795	34.94	28
YOLO9000 ^[26]	0.833	—	—
EB ^[27]	0.897	—	9
MVD ^[29]	0.889	—	59
YOLOv4	0.917	264.22	30
M3-YOLOv4	0.906	31.13	47



(a) 原图 (b) 检测可视化示意图

图 10 M3-YOLOv4 检测结果可视化

4 结论

本文针对交通道路场景中的多类别、多尺度车辆目标,提出了兼顾准确性和实时性的轻量化多目标检测算法,其良好的轻量化性能有助于在嵌入式设备、边缘计算设备等部署应用,对提高环境感知系统综合性能有较好的支撑. 研究内容包括以下部分:

1) 使用 MobileNetV3 替代 CSPDarkNet53 作为检测模型的主干网络,通过深度可分离卷积降低特征提取网络参数.

2) 使用 SPP 处理深层网络的输出特征图,选择轻量级骨干网络不同深度特征图的输出,通过 PANet 融合深层语义信息和浅层表现信息,形成检测模型颈部.

3) 在检测模型头设置 3 个不同特征地图大小的网络输出,构建了轻量级 M3-YOLOv4 检测模型. 通过模型训练和测试,验证了模型的轻量化和实时性效果,也论证了该模型在同一平台下的前向推理时间具有明显的优势.

参考文献(References)

- [1] 李克强, 戴一凡, 李升波, 等. 智能网联汽车(ICV)技术的发展现状及趋势[J]. 汽车安全与节能学报, 2017, 8(1): 1-14.
(Li K Q, Dai Y F, Li S B, et al. State-of-the-art and technical trends of intelligent and connected vehicles[J]. Journal of Automotive Safety and Energy, 2017, 8(1): 1-14.)
- [2] 金沙沙, 龙伟, 胡灵犀, 等. 多目标检测与跟踪算法在智能交通监控系统中的研究进展[J]. 控制与决策, 2023, 38(4): 890-901.
(Jin S S, Long W, Hu L X, et al. Research progress of detection and multi-object tracking algorithm in intelligent traffic monitoring system[J]. Control and Decision, 2023, 38(4): 890-901.)
- [3] Taghvaeeyan S, Rajamani R. Portable roadside sensors for vehicle counting, classification, and speed measurement[J]. IEEE Transactions on Intelligent Transportation Systems, 2014, 15(1): 73-83.
- [4] 杜豫川, 师钰鹏, 都州扬, 等. 智能网联环境下路侧感知单元数据质量在线监测框架[J]. 中国公路学报, 2022, 25(3): 273-285.
(Du Y C, Shi Y P, Du Z Y, et al. An online monitoring framework for data quality of roadside perception units in intelligent and connected environment[J]. China Journal of Highway and Transport, 2022, 25(3): 273-285.)
- [5] 李克强, 常雪阳, 李家文, 等. 智能网联汽车云控系统及其实现[J]. 汽车工程, 2020, 42(12): 1595-1605.
(Li K Q, Chang X Y, Li J W, et al. Cloud control system for intelligent and connected vehicles and its application[J]. Automotive Engineering, 2020, 42(12): 1595-1605.)
- [6] 尹宏鹏, 陈波, 柴毅, 等. 基于视觉的目标检测与跟踪综述[J]. 自动化学报, 2016, 42(10): 1466-1489.
(Yin H P, Chen B, Chai Y, et al. Vision-based object detection and tracking: A review[J]. Acta Automatica Sinica, 2016, 42(10): 1466-1489.)
- [7] Sun Z H, Bebis G, Miller R. Monocular precrash vehicle detection: Features and classifiers[J]. IEEE Transactions on Image Processing, 2006, 15(7): 2019-2034.
- [8] Sun Z H, Bebis G, Miller R. On-road vehicle detection: A review[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(5): 694-711.
- [9] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego, 2005: 886-893.
- [10] Platt J C. A fast algorithm for training support vector machines[J]. Journal of Information Technology, 1998, 2(5): 1-28.
- [11] Cho H, Rybski P E, Zhang W D. Vision-based bicycle detection and tracking using a deformable part model and an EKF algorithm[C]. 13th International IEEE Conference on Intelligent Transportation Systems. Funchal, 2010: 1875-1880.
- [12] Khairdoost N, Amirhassan Monadjemi S, Jamshidi K. Front and rear vehicle detection using hypothesis generation and verification[J]. Signal & Image Processing: An International Journal, 2013, 4(4): 31-50.
- [13] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, 2016: 779-788.
- [14] Liu W, Anguelov D, Erhan D, et al. SSD: Single shot MultiBox detector[C]. Computer Vision—ECCV 2016. Cham: Springer International Publishing, 2016: 21-37.
- [15] Gerónimo D, Sappa A D, López A, et al. Adaptive image sampling and windows classification for on-board pedestrian detection[C]. Proceedings of the 5th International Conference on Computer Vision Systems. Bielefeld, 2007.
- [16] 金立生, 华强, 郭柏苍, 等. 基于优化 DeepSort 的前方车辆多目标跟踪[J]. 浙江大学学报: 工学版, 2021, 55(6): 1056-1064.
(Jin L S, Hua Q, Guo B C, et al. Multi-target tracking of vehicles based on optimized DeepSort[J]. Journal of Zhejiang University: Engineering Science, 2021, 55(6): 1056-1064.)

- [17] 金立生, 郭柏苍, 王芳荣, 等. 基于改进YOLOv3的车辆前方动态多目标检测算法[J]. 吉林大学学报: 工学版, 2021, 51(4): 1427-1436.
(Jin L S, Guo B C, Wang F R, et al. Dynamic multi-target detection algorithm in front of vehicles based on improved YOLOv3[J]. Journal of Jilin University: Engineering and Technology Edition, 2021, 51(4): 1427-1436.)
- [18] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: Optimal speed and accuracy of object detection[J/OL]. 2020, arXiv: 2004.10934.
- [19] Zhu X K, Lyu S C, Wang X, et al. TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios[C]. 2021 IEEE/CVF International Conference on Computer Vision Workshops. Montreal, 2021: 2778-2788.
- [20] 刘建骐, 闫河, 王潇棠, 等. 改进金字塔和跳跃连接的YOLOv5目标检测网络[J]. 控制与决策, 2023, 38(6): 1730-1736.
(Liu J Q, Yan H, Wang X T, et al. YOLOv5 object detection network with improved pyramid and skip connection[J]. Control and Decision, 2023, 38(6): 1730-1736.)
- [21] Wen L Y, Du D W, Cai Z W, et al. UA-DETRAC: A new benchmark and protocol for multi-object detection and tracking[J]. Computer Vision and Image Understanding, 2020, 193: 102907.
- [22] Wu D H, Lv S C, Jiang M, et al. Using channel pruning-based YOLOv4 deep learning algorithm for the real-time and accurate detection of apple flowers in natural environments[J]. Computers and Electronics in Agriculture, 2020, 178: 105742.
- [23] He K M, Zhang X Y, Ren S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.
- [24] Yue J, Mao S J, Li M. A deep learning framework for hyperspectral image classification using spatial pyramid pooling[J]. Remote Sensing Letters, 2016, 7(9): 875-884.
- [25] He K M, Zhang X Y, Ren S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.
- [26] Redmon J, Farhadi A. YOLO9000: Better, faster, stronger[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, 2017: 6517-6525.
- [27] Li Wang, Yao Lu, Hong Wang, et al. Evolving boxes for fast vehicle detection[J/OL]. 2017, arXiv: 1702.00254v3.
- [28] 廖航, 霍爱清, 王泽文. 轻量化YOLOv4的车辆目标检测研究[J]. 工业控制计算机, 2023, 36(3): 98-100.
(Liao H, Huo A Q, Wang Z W. Research on vehicle target detection of lightweight YOLOv4[J]. Industrial Control Computer, 2023, 36(3): 98-100.)
- [29] Fu H Y, Guan J, Jing F, et al. A real-time multi-vehicle tracking framework in intelligent vehicular networks[J]. China Communications, 2021, 18(6): 89-99.

作者简介

金立生(1975—), 男, 教授, 博士生导师, 从事智能车辆感知决策与控制、人车路协同与车联网技术等研究, E-mail: jinls@ysu.edu.cn;

张舜然(1995—), 男, 硕士, 从事智能车辆环境感知等研究, E-mail: zhang.shunran@connect.umac.mo;

郭柏苍(1991—), 男, 副教授, 博士, 从事自动驾驶人机协作测试与评价等研究, E-mail: guobaicang@ysu.edu.cn;

王欢欢(1995—), 女, 博士生, 从事智能车辆环境感知技术等研究, E-mail: 853200505@qq.com;

韩卓桐(1999—), 男, 硕士生, 从事智能网联汽车测试与评价等研究, E-mail: hzhuotong@163.com;

刘星辰(1999—), 男, 硕士生, 从事智能车辆环境感知、驾驶行为分析等研究, E-mail: liuxc999@163.com.