

基于改进 YOLOv5 的车辆目标检测研究

章程军^{1,2}, 胡晓兵^{1,2}, 牛洪超^{1,2}

(1. 四川大学机械工程学院, 成都 610065; 2. 四川大学宜宾园区, 宜宾 644000)

摘要: 针对现有目标检测算法在自动驾驶等领域的车辆目标检测中存在检测精度不高, 实时性和鲁棒性较差等问题, 本文提出一种基于 YOLOv5 的车辆目标检测方法. 本文在 YOLOv5s 网络模型框架下, 引入一次性聚合(OSA)模块优化主干网络结构, 提升网络特征提取能力; 并采用非局部注意力机制进行特征增强; 同时利用加权非极大值抑制方法实现检测框筛选. 实验结果表明, 在自制车辆检测数据集上, 改进网络模型与原 YOLOv5s 模型相比, 平均准确率均值(mAP)提升 3%, 不同目标类检测的平均准确率(AP)均得到提升, 且检测速度满足实时性要求, 对于密集车辆和不同光照条件下均能较好实现车辆目标检测.

关键词: 车辆检测; 多层特征融合; 特征增强; 非极大值抑制

中图分类号: TP39 **文献标识码:** A **DOI:** 10.19907/j.0490-6756.2022.053001

Vehicle object detection based on improved YOLOv5 method

ZHANG Cheng-Jun^{1,2}, HU Xiao-Bing^{1,2}, NIU Hong-Chao^{1,2}

(1. School of Mechanical Engineering, Sichuan University, Chengdu 610065, China;

2. Yibin R&D Park of Sichuan University, Yibin 644000, China)

Abstract: In view of the problems of low detection accuracy, poor real-time and robustness of existing target detection algorithms in vehicle target detection in autonomous driving fields, a vehicle target detection method based on YOLOv5 is proposed. With the framework of YOLOv5s network model, a one-shot aggregation (OSA) module is introduced to optimize the backbone network structure and improve the network feature extraction capability. Non-local attention mechanism is used for feature enhancement. At the same time, the weighted non-maximum suppression method is used to filter the detection frame. The experimental results show that compared with the original YOLOv5s model, the mAP of the improved network model is improved by 3%, and the AP of different target detection classes is improved, and the detection speed meets the real-time requirements. For dense vehicles and under different illumination conditions, vehicle target detection can be better achieved.

Keywords: Vehicle detection; Multi-layer feature fusion; Feature enhancement; Non-maximum suppression

1 引言

随着人工智能技术的快速发展, 智能交通系统以及自动驾驶等领域受到研究者的广泛关注. 车辆

检测作为其前端数据采集的重要模块, 能够快速准确地完成对车辆的检测可以为整体系统提供有效的数据保障. 所以, 设计强实时性与鲁棒性的车辆目标检测算法有着重要意义.

收稿日期: 2021-10-14

基金项目: 川大-宜宾校市合作项目(2020CDYB-3)

作者简介: 章程军(1998-), 男, 安徽池州人, 硕士研究生, 研究方向为机器视觉, 图像处理等. E-mail: 2481379172@qq.com

通讯作者: 胡晓兵. E-mail: scuhxb@163.com

在车辆检测的实际应用工程中,环境复杂多样及检测目标尺寸种类不一等难题一直困扰着研究人员.传统方式的目标检测算法主要包括图像预处理、区域选择、候选区域特征提取和分类等 4 个环节^[1],虽然算法检测准确度较高,但其泛化能力较差,特征往往需要人为进行设计,研究人员对应用领域的了解以及经验直接影响到算法的优劣,无法满足车辆检测领域复杂环境的要求.

近年来,基于深度学习的车辆检测算法逐渐受到研究者的青睐,相比于传统方式的目标检测,有较强的鲁棒性以及泛化能力.其能够自主完成对待检目标的特征识别,不需要人为进行特征设计,受人为因素影响较小.目前主流算法分为以 YOLO (You Only Look Once)、SSD(Single Shot Multi-Box Detector)系列为代表的单阶段检测算法和以 R-CNN(Region-CNN)系列为代表的二阶段检测算法两类,二者在精确度以及检测速度方面各有优劣.二阶段的目标检测算法是基于候选区域思想,首先利用滤波窗口在图像上滑动提取可能含有目标的感兴趣区域,减少无关区域对目标检测的影响;然后利用卷积神经网络对提取区域的目标进行分类和定位. Tang 等^[2]以 Faster-RCNN 网络作为基础网络,采用超区域提议网络(Hyper Region Proposal Network, HRPN)结合层次特征图提取类车目标,增强网络对小尺寸车辆目标的检测能力. Luo 等^[3]将神经网络架构搜索(Neural Architecture Search, NAS)和特征增强与 Faster-RCNN 网络相结合,实现了多尺度车辆目标的检测.基于候选区域的二阶段目标检测算法虽能取得较高的检测精度,但实时性较差,不能满足自动驾驶领域车辆检测速度要求.单阶段的目标检测算法基于回归思想,不再需要提取候选区域,将待检图片直接输入网络,在输出层得到待检图片中目标边界框和类别信息,此类算法虽检测精度比基于候选区域思想算法有所降低,但是检测速度基本满足实时要求. Sri 等^[4]以 YOLOv3-Tiny 为基础,通过融合空间金字塔池化、特征拼接,提高网络的学习能力,使车辆检测速度与精度都得到提升. 马永杰等^[1]将 YOLOv3 模型和 Deep-SORT 算法相结合,兼顾了对不同尺寸目标的检测,同时保证了网络的检测速度满足实时性要求. 宋世奇等^[5]基于 SSD 网络,通过在其特征提取模块融合具有抗混叠效应的滤波层,提高了网络检测精度和小尺寸目标的检测效果.但 SSD 算法的许多参数需要人为设置,调试过

程非常依赖研究人员经验,图像分辨率对网络输出结果影响较大,实际应用存在一定限制.

综合检测速度,精确度以及算法使用难易程度等方面考虑,选择整体检测性能较好的 YOLO 系列算法,开展其在车辆检测领域的研究工作.目前 YOLO 系列算法已经发展到 YOLOv5 版本,其性能相比于初代已经得到很大提升,因此本文以 YOLOv5 算法为基础模型,对其进行相关改进:(1)主干网络(Backbone):将 CSP 模块与 VOV-Net^[6]中 OSA 模块进行融合提出 CSP-OSA 模块,替代原网络中 BottlenCSP1 结构,提高网络特征提取能力,减少计算成本;(2)颈部(Neck)网络:引入注意力机制 NonLocal^[7]进行特征增强,增强对小目标检测能力;(3)后处理:使用加权 NMS^[8]替代原有算法,增强网络对遮挡目标的检测.本文利用自制数据集进行实验验证,检测改进算法的性能.

2 YOLOv5 网络结构

YOLO^[9]系列算法将输入图片划分为 $S \times S$ 个网格,如果待检目标中心落在某个网格内,则这个网格负责预测该目标,每个网格预测 B 个边界框和一个类别信息,边界框包含目标位置和置信度信息.基于此检测方式,YOLOv5 网络对每个网格预测 3 个边界框(Bounding Box),每个边界框包含坐标信息 (x, y, w, h) 、1 个置信度和 C 个条件类别概率信息.目前 YOLOv5 包含 YOLOv5s、YOLOv5m、YOLOv5l 和 YOLOv5x 等 4 种结构基本相同的模型,在网络宽度和深度上依次增加.

YOLOv5 网络结构由输入端(Input)、主干网络(Backbone)、颈部网络(Neck)和输出端(Head)等 4 部分组成,具体网络结构图如图 1 所示.

输入端对输入图像进行预处理,对图片进行自适应缩放统一标准尺寸,增加了自适应锚架计算功能.在每次训练过程中,算法自适应地计算不同训练集中最优锚框的值.该方法使用了 Mosaic^[10]数据增强方法,在训练过程中,随机选取 4 张图片进行旋转缩放等操作,进行随机剪裁后拼接形成新的训练数据.这种方法丰富了数据集,增强了网络鲁棒性,在小目标检测方面表现更好,同时计算 4 张图片数据,提高了模型训练速度和检测能力. Mosaic 数据增强效果如图 2 所示.

主干网络(Backbone)由 Focus、CBL、CSP 和 SPP 等 4 种模块组成,用以图像中目标特征提取. CSP 模块借鉴跨阶段局部网络(Cross Stage Par-

tial Network, CSPNet^[11]), 从网络结构设计角度解决推理过程中网络优化中的梯度信息重复问题, 增强了网络的学习能力以及运行成本. 增加 Focus 结构实现切片操作, 在图片进入主干网络前进行 2 倍下采样获得 4 张特征图, 图片的 W 、 H 信息集中到空间通道, 该过程并不会造成图片特征信息丢

失, 可以使网络提取到更加充分的特征信息. YOLOv5s 中将原始的 $608 \times 608 \times 3$ 图像输入到 Focus 结构中, 通过切片操作形成 $304 \times 304 \times 12$ 的特征图, 然后对 32 个卷积核进行卷积操作, 最终构造出 $320 \times 320 \times 32$ 的特征图. Focus 模块切片操作示意图如图 3 所示.

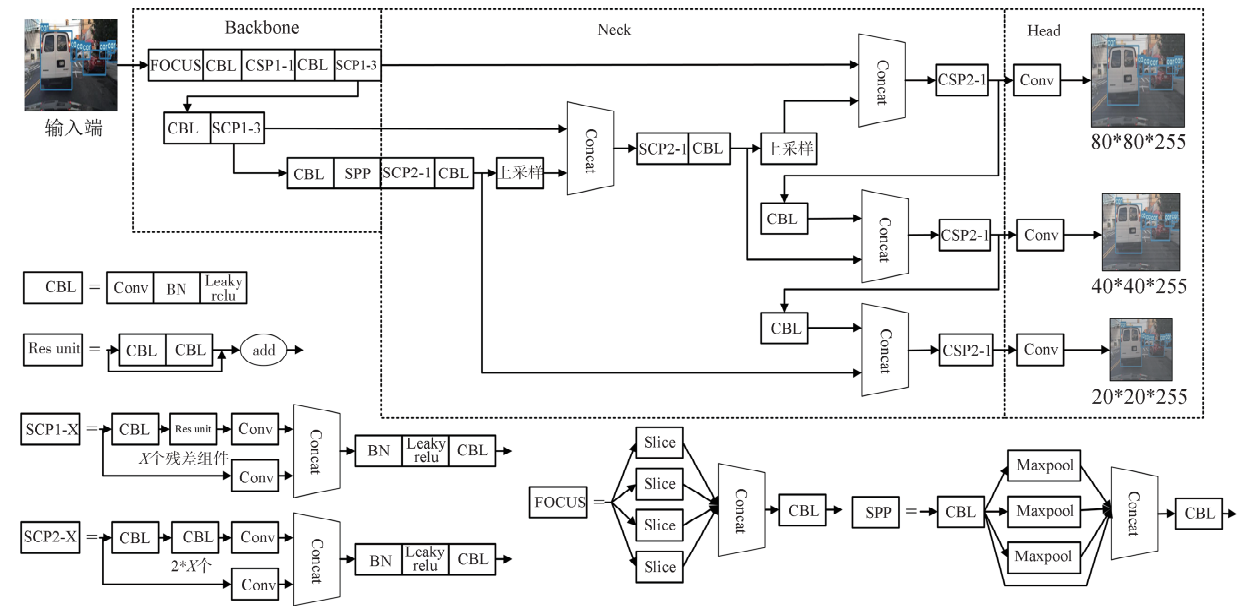


图 1 YOLOv5 网络结构
Fig.1 YOLOv5 model structure

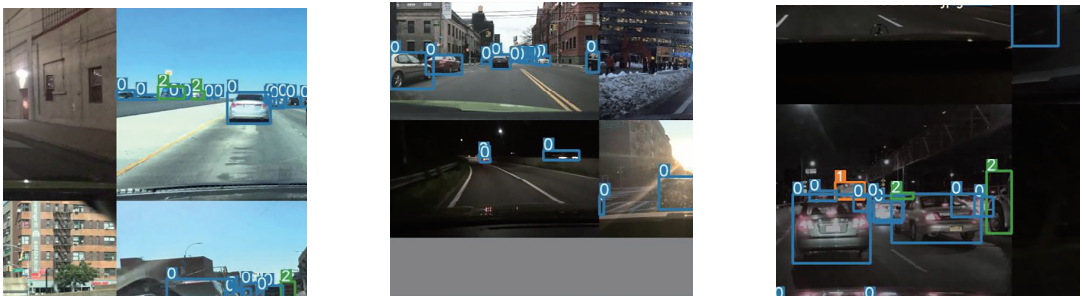


图 2 部分 Mosaic 数据增强效果
Fig.2 Part of the mosaic enhanced

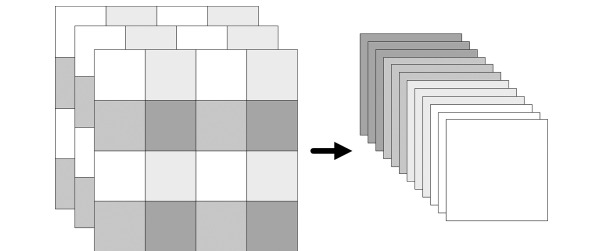


图 3 Focus 切片操作示意图
Fig.3 Schematic diagram of slicing operation of focus module

颈部(Neck)网络主要用于生成特征金字塔, 增强网络模型对不同尺度物体的检测能力, 实现对

同一物体不同尺寸和尺度的识别. 其采用 FPN-PAN 网络结构, FPN^[12]通过上采样将深层语义信息与底层目标位置信息进行融合, 增强网络对特征学习能力, 在此基础上融合 PAN^[13]结构, 自底向上传达强定位特征, 全面提升网络对语义信息和定位信息的学习性能. FPN-PAN 结构图如图 4 所示.

输出层(Head)3 个检测头分别对原图像进行了 8 倍、16 倍和 32 倍下采样, 生成 3 个不同大小的特征向量, 用以预测图像特征、生成边界框和预测图像中目标类别及置信度.

YOLOv5 在目标检测领域, 整体检测性能较

强,但由于其并不是专门针对车辆目标的检测,仍然存在检测效果不佳的情况,特别是对于车辆目标密集,环境光线不足的情况下. 为保证在自动驾驶等领域能够实时精确完成对车辆目标检测,同时满足网络轻量化问题,选用 YOLOv5s 网络,对其进行改进,提升网络对车辆目标检测精度.

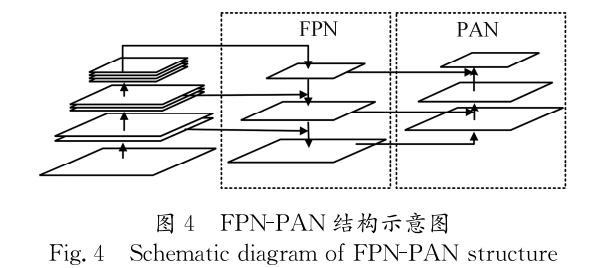


图 4 FPN-PAN 结构示意图
Fig. 4 Schematic diagram of FPN-PAN structure

3 基于 YOLOv5s 车辆检测算法改进

3.1 主干网络(Backbone)的改进

基于深度学习的目标检测模型需将卷积神经网络(Convolutional Neural Network, CNN)作为主干网络提取目标特征,本文借鉴 VOVNet 网络结构,设计 CSP-OSA 模块替代原网络 CSP1 结构.

VOVNet 网络模型是对 DenseNet^[14] 网络模型一种改进,相比于目前在目标检测领域常用的主干网络 ResNet^[15],DenseNet 网络特征提取能力更强,具有更多样化的特征表达能力,其网络参数以及计算量(FLOPs)也更少. 主要原因是 ResNet 与 DenseNet 网络模型聚合特征的方式不相同. ResNet 网络采用求和方式从浅层聚合特征,随着特征图逐渐累加,早期特征图包含信息会被稀释,使得 ResNet 网络模型特征提取能力相对较弱. 不同于此,DenseNet 网络通过密集连接(concat)进行特征聚合. 这种方式可以保留和积累不同层感受野特征,深层特征与浅层特征进行特征融合时不会稀释浅层特征信息,信息会持续存在,使得每一层都具有不同层感受野特征,可以在不同尺度上捕获目标特征信息,更适用于目标检测中特征提取. 但由于采用密集连接方式会使后面每层都会聚合前面层特征,造成特征冗余,导致内存访问成本与能量消耗升高,网络预测速度变慢,效率降低. 针对以上问题,VOVNet 网络在保留 DenseNet 网络特征提取能力的同时,能有效地解决内存访问成本和 GPU 运行效率问题. 在 DenseNet 基础上,VOVNet 提出一次性聚合(One-Shot Aggregation, OSA)模块,删除密集连接方式,将浅层特征直接聚合到最

后一层,这种方式使 VOVNet 网络性能优于 DenseNet 和 ResNet 网络模型,对于小目标检测性能也有了显著提高. OSA 模块结构图如图 5 所示.

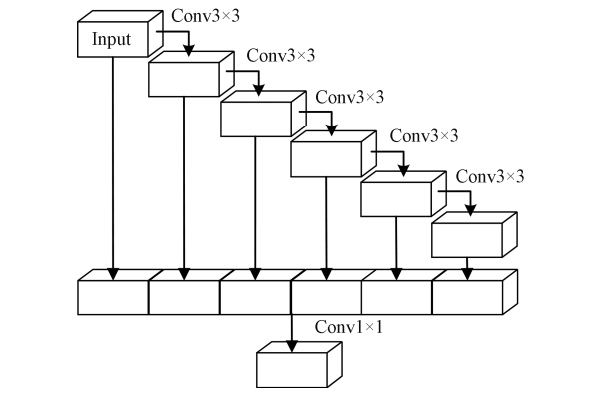


图 5 OSA 模块结构
Fig. 5 OSA module structure

本文设计的 CSP-OSA 模块在 OSA 模块基础上添加 CSPNet 结构,将基础层特征图分为两部分,一部分通过卷积直接连接到阶段末尾;另一部分通过 OSA 模块. 本文将二者进行特征融合,能够实现更加丰富的梯度组合,提高网络特征提取能力. CSP-OSA 模块结构图如图 6 所示.

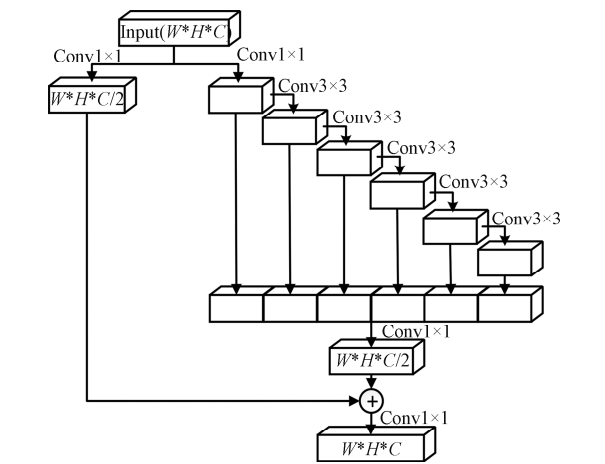


图 6 CSP-OSA 模块结构
Fig. 6 CSP-OSA module structure

3.2 颈部(Neck)网络的改进

颈部(Neck)网络连接主干网络和输出层,将主干网络提取的特征通过特征增强模块进行特征增强处理后输出到输出层,提高网络检测精度. 目前在深度学习领域使用较广泛的特征增强模块为 FPN(特征金字塔网络)模块,可以灵活运用在目标检测和实例分割等任务中. 在 FPN 机构提出之前,绝大多数目标检测算法都是采用自底向上的前向过程,只对顶层特征做预测,可以利用高层特征丰

富的语义信息,但对底层特征目标位置信息利用较少,使得检测目标定位不够准确. FPN 模块在此基础上添加自顶向下结构,二者之间采用横向连接,将深层语义信息传送到浅层中,并且对每个融合后的特征层进行单独预测,这种结构实现多尺度目标检测,有效提高网络对小目标的检测能力. 在 YOLOv3 中已经使用 FPN 结构, YOLOv5 中对 FPN 结构进行改进,在 FPN 结构基础上融入 PANet 网络结构,添加了自底向上的特征金字塔结构,将浅层丰富位置信息传递到网络深层,充分使高层语义信息与浅层位置信息融合,提高了网络整体检测性能. 但由于该方法侧重于将相邻层间的语义和位置信息进行融合加强特征,对于非相邻层之间没有进行信息交流,在一定程度上限制了网络对小目标的检测能力. 针对上述情况,本文提出一种可以实现多特征层间信息融合的特征增强模型. 其结构如图 7 所示.

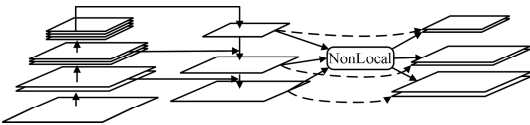


图 7 改进 FPN-PAN 结构
Fig. 7 Improved FPN-PAN model structure

改进 FPN-PAN 结构不再使用 PANet 模型中自底向上结构,其将 FPN 中所有特征层所包含的深层语义信息和浅层位置信息进行聚合,通过注意力机制进行特征增强. 本文采用注意力机制相比于原非局部注意力机制(NonLocal),在保证整体性能未减少的情况下进行了简化,去除一条分支,其结构如图 8 所示.

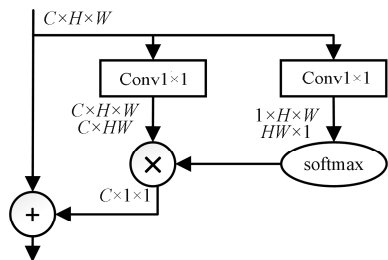


图 8 简化的非局部注意力机制
Fig. 8 Simplified non-local attention mechanism

通过简化后注意力机制进行特征增强,可以直接捕获长距离依赖关系,特征信息不再只关注局部位置关系. 如图 7 中虚线所示,再将经过注意力机制增强的特征与 FPN 中每层特征分别进行信息融合得到新的特征,以此来保证非相邻特征层之间的

信息交流. 这种方式也在一定程度上缓解了原网络结构中相邻层间进行特征融合时存在的语义间隙问题,对于网络小目标的检测能力有所增强,使网络特征更加丰富,便于后面输出层的预测.

3.3 后处理算法的改进

非极大值抑制(Non-Maximum Suppression, NMS)算法是目标检测后处理过程,目的是为了对目标框进行筛选,消除多余的检测框,找到最佳的物体检测的位置. 通过将所有检测框得分进行排序选择得分最高检测框,计算其余检测框与当前得分最高检测框交并比(Intersection Over Union, IOU)来对重叠的预测候选框进行筛选,当 IOU 值高于阈值时,一组检测框被标识为同一对象. 其公式可以表示为

box = B_{\arg \max C_i} \tag{1}

其中, B 代表检测框; C 代表第 i 个检测框的置信度; box 表示具有最高置信度的检测框,作为输出结果. 但被抑制的非最大值检测结果可能包含目标特征最大值,直接忽略非最大值检测框并不合理,会造成目标定位不准确的问题. 对于自动驾驶领域中车辆检测,目标车辆存在密集和相互遮挡等复杂情况,定位不准确的检测框会严重影响到目标车辆与自身车辆的相对位置关系的判断. 针对以上问题,本文采用加权 NMS 对检测框进行筛选,使得输出边界框定位更加准确,从而提升算法整体性能. 加权 NMS 算法公式如下.

$$box = \frac{\sum_i w_i B_i}{\sum_i w_i}, B_i \in \{B \mid IOU(M, B) \geq$$

$$thresh\} \cup \{box\}, w_i = C_i \times IOU(box, B_i)$$
(2)

其中, n 为检测框数量; $IOU(M, B)$ 为第 i 个检测框与置信度最高检测框的 IOU 值; w 是每个检测框权重. 若假设所有预测的检测框都来自同一个目标,通过考虑所有的预测结果来充分利用对象信息,并不是简单选择置信度最高的检测框作为输出结果. 加权 NMS 算法伪代码如下.

输入 $B = \{(B_n, C_n)\}_{n=1 \text{ to } N}$, 其中 s_n 是 B_n 得分;
 $D = \emptyset, E = \emptyset$.

Step1: B 中选择最大得分框 M

Step2: for B_i in B

if $IOU(M, B_i) \geq NMS_threshold$

将 B_i 及其得分添加进 C 中, 为 B 去掉 B_i

及其得分.

```
end if
end for
Step3: for  $C_j$  in  $C$ 
    按式(2)计算新检测框  $d$  添加  $D$  中
end for
    去掉  $C$  中所有检测框及其得分
Step4: 重复 Step1~3,直至  $B=\varnothing$ 
Step5: 输出  $D$ 
```

先对回归产生的 N 个检测框根据置信度分值大小进行排序,选择置信度分值最大的检测框,计算其与所有检测框的 IOU 值;将 IOU 值大于阈值的检测框按照式(2)进行加权处理重新计算得到新的检测框 d 并剔除大于阈值的检测框;剩下检测框重复操作直到处理完全部检测框;计算得到的新检测框集合 D 即为检测结果.加权 NMS 并没有将大于阈值的检测框直接剔除,而是根据网络置信度进行加权处理,获得定位更加准确新的检测框,在一定程度上提高网络检测的精确度.

4 实验过程与分析

4.1 实验数据集

本文使用的数据集是在伯克利大学发布的 BDD100K^[16] 自动驾驶数据集基础上进行挑选修改的自制数据集,包含 7 万帧图片以及所对应的标注信息,数据信息丰富多样,包含不同地理,环境,天气等因素,数据集中包含 car, bus 和 truck 等 3 种车辆目标类别,将数据集按 7 : 2 : 1 比例分配为训练集,测试集和验证集.

为了使模型具有更高检测精度以及提升模型泛化能力和鲁棒性.采用 Z-Score 标准化方法对数据进行归一化处理,使图像数据符合标准正态分布,减少因数据差异太大对模型训练时收敛速度和检测精度影响.使用 K-means 聚类方法更新网络 anchors,避免因为数据集改变造成原先 anchors 与当前数据集 anchors 不匹配情况.并通过调整图像数据饱和度、曝光量以及色调来增加训练样本.

4.2 实验环境与评价指标

本实验的硬件配置为:CPU: Core(TM) 1.80 GHz; GPU: NVIDIA GeForce RTX 2080Ti; 内存:11 G. 软件环境为: Linux 操作系统、python 3.7、Torch 1.6.0. 为提升网络收敛速度,在网络模

型训练过程使用随机梯度下降法(Stochastic Gradient Descent,SGD)进行学习和更新网络参数.设置网络超参数具体如表 1 所示.

表 1 网络训练超参数
Tab.1 Training Hyper parameter

训练参数名称	参数值
初始学习率	0.01
学习率下降参数	0.0001
权重衰减系数	0.0005
动量(momentum)	0.937
批处理大小(batch)	8
训练批次(epoch)	200
图像输入尺寸	640×640

本文实验结果从召回率(recall)、查准率(precision)、平均准确率均值(mAP)、各个类别的平均准确率(AP)以及推理时间来对改进模型的性能进行评估.召回率与查准率的计算公式如下.

$$\text{recall} = \frac{TP}{TP + FN}$$
 (3)

$$\text{precision} = \frac{TP}{TP + FP}$$
 (4)

其中,TP(True Positives)为正样本数据被正确识别出的目标数量,即被正确识别的车辆目标;FN(False Negatives)为正样本数据没有被正确识别的目标数量,即漏检的车辆目标;FP(False Positives)为将非正样本数据识别为正样本目标数量,即误检的车辆目标.

平均准确率 AP(Average Precision)与平均准确率均值 mAP(mean Average Precision)可以用来衡量网络模型的优劣,其需要根据模型的召回率(recall)和查准率(precision)计算得出,计算公式如下.

$$AP = \int_0^1 P(R) dR$$
 (5)

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i$$
 (6)

4.3 各模块改进对比分析

为了验证改进模块对网络整体性能的影响,本文采用递增方式逐步修改每个改进模块,观察其是否能够提升网络性能.实验结果如表 2 所示. I-YOLOv5s 只对网络主干进行改进,将改进的 CSP-OSA 模块替代原网络中的 CSP 模块,网络特征提取能力更强,相比于 YOLOv5s 网络 mAP 值提升

2.46%; II-YOLOv5s 对主干网络和颈部网络都进行改进, 在颈部网络引入非局部注意力机制进行特征增强, 相比于 I-YOLOv5s 网络 mAP 值提升 0.23%; III-YOLOv5s 在 II-YOLOv5s 网络基础上使用加权非极大值抑制的后处理方式, 相比于 II-YOLOv5s 网络 mAP 值提升 0.31%.

表 2 改进模块对网络性能影响
Tab. 2 Influence of improved modules on network performance

网络	mAP, 5	mAP, 5 : . 95	推理时间/s
YOLOv5s	0.6701	0.4534	0.008
I-YOLOv5s	0.6947	0.477	0.010
II-YOLOv5s	0.697	0.4795	0.011
III-YOLOv5s	0.7001	0.4806	0.011

从表 2 数据可以看出, 随着对网络每个模块改进, mAP. 5 和 mAP. 5 : . 95 值均逐步提升, 主干网络修改对网络性能提升最大, 并且随着每个模块改进, 推理时间并没有显著增加.

4.4 改进前后 YOLOv5 模型的对比分析

为了了解改进网络 III-YOLOv5s 性能与原网络差异, 对两模型训练结果进行对比, 如表 3 和表 4 所示.

表 3 改进前后网络性能对比
Tab. 3 Comparison of two algorithms

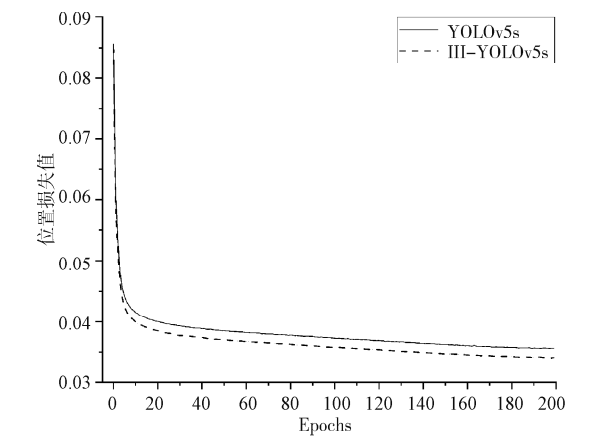
网络模型	precision	recall	mAP	推理时间/s
YOLOv5s	0.4314	0.7277	0.6701	0.008
III-YOLOv5s	0.452	0.7537	0.7001	0.011

表 4 目标的平均准确率(AP)对比
Tab. 4 Comparison of AP of targets

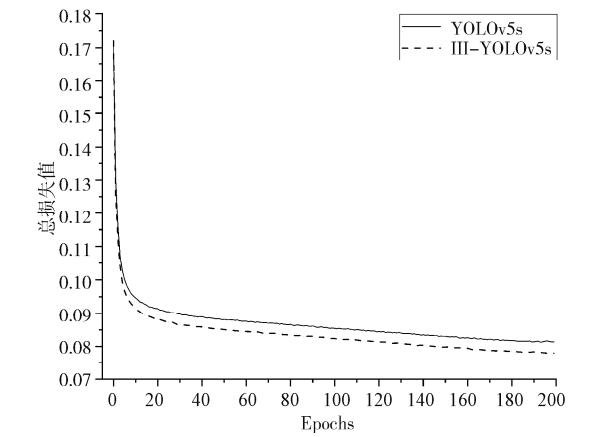
目标类别	YOLOv5s	III-YOLOv5s
car	0.769	0.781
bus	0.614	0.648
truck	0.627	0.655

从表 3 可以看出, 改进网络模型 III-YOLOv5s 平均准确率均值 (mAP) 达到 70.01%, 相比于 YOLOv5s 网络 mAP 值提升 3%; 在查准率和召回率分别提高 2.06% 和 2.6%; 推理时间为 0.011 s, 相比于原网络 0.008 s 虽然有所升高, 但仍满足实时性要求. 表 4 为选择的三个目标类 (car、bus、truck) 平均准确率对比表, 改进网络模型 III-YOLOv5s 在三个目标类上检测的 AP 值均有所提

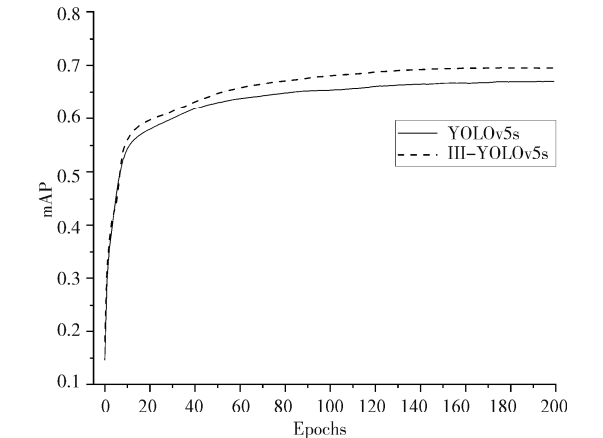
升, 其中在 bus 目标类中提升幅度最高, 达到 3.4%.



(a) 位置损失值对比



(b) 总损失值对比



(c) mAP 值对比

图 9 YOLOv5s 与 III-YOLOv5 损失值和 mAP 对比
Fig. 9 Comparison of two algorithms loss values andmAP

图 9 为改进网络与 YOLOv5s 网络损失值和平均准确率均值 (mAP) 对比曲线. 从图 9a 和 9b 中损失值对比曲线可以看出, 随着训练批次增加, 总损失值和位置损失不断下降, 在迭代 50 步左右后

趋于稳定,改进模型损失值整体低于 YOLOv5s 网络损失值。从图 9c 中 mAP 值对比曲线可以看出,在网络趋于稳定后,改进模型的 mAP 值优于 YOLOv5s 网络。

图 10 为改进网络 III-YOLOv5s 检测效果图,

从图 10 可以看出,对于不同车辆密度,改进网络 III-YOLOv5s 在不同光照强度下均可对车辆目标进行较好的识别。面对车辆目标尺寸不同,光照不同的复杂环境条件下,改进网络有较好的鲁棒性,能够成功完成目标车辆检测,检测能力较强。

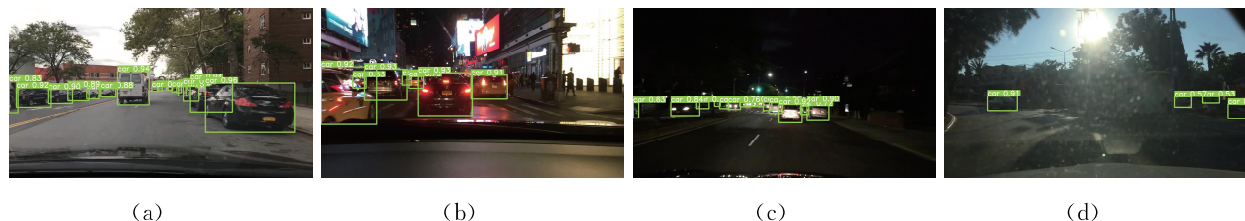


图 10 III-YOLOv5s 实际检测效果
Fig. 10 Actual detection effect of III-YOLOV5S

5 结 论

本文在网络 YOLOv5s 的基础上改进使其更适合自动驾驶等领域中车辆目标检测。改进网络 III-YOLOv5s 在主干网络中使用 CSP-OSA 结构,提高网络的特征提取能力;使用非局部注意力机制进行特征增强,丰富网络提取特征;并通过加权非极大值抑制来获得定位更加准确检测框。在自动驾驶数据集 BDD100K 上实验表明,改进网络平均准确率均值相比 YOLOv5s 有所提高,总损失和位置损失均有一定降低,并且网络检测速度仍满足实时性要求;对于不同密度和光照情况,III-YOLOv5s 有较强鲁棒性,能够完成对目标车辆检测。相比于原网络 YOLOv5s,改进网络 III-YOLOv5s 整体检测性能得到提升。

参考文献:

- [1] 马永杰, 马芸婷, 程时升, 等. 基于改进 YOLOv3 模型与 Deep-SORT 算法的道路车辆检测方法[J]. 交通运输工程学报, 2021, 21: 222.
- [2] Tang T, Zhou S, Deng Z, *et al.* Vehicle detection in aerial images based on region convolutional neural networks and hard negative example mining [J]. Sensors, 2017, 17: 336.
- [3] Luo J Q, Fang H S, Shao F M. Multi-scale traffic vehicle detection based on faster R-CNN with NAS optimization and feature enrichment [J]. Defence Technol, 2021, 17: 1542.
- [4] Sri J S, Esther R P. Little YOLO-SPP: a delicate real-time vehicle detection algorithm [J]. Optik, 2020, 225: 165818.
- [5] 宋世奇, 李旭, 祝雪芬, 等. 基于改进 ssd 的航拍城

- 市道路车辆检测方法[J]. 传感器与微系统, 2021, 40: 114.
- [6] Lee Y, Hwang J W, Lee S, *et al.* An energy and GPU-computation efficient backbone network for real-time object detection [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. [S. l.]: IEEE, 2019.
- [7] Wang X L, Ross G, Abhinav G, *et al.* Non-local neural networks [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. [S. l.]: IEEE, 2018.
- [8] Ning C, Zhou H, Yan S, *et al.* Inception single shot multibox detector for object detection [C]//Proceedings of the 2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW). [S. l.]: IEEE, 2017.
- [9] Redmon J, Divvala S, Girshick R, *et al.* You only look once: unified, real-time object detection [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. [S. l.]: IEEE, 2016.
- [10] Yun S, Han D, Chun S, *et al.* CutMix: regularization strategy to train strong classifiers with localizable features [C]//Proceedings of the IEEE/CVF international conference on computer vision. [S. l.]: IEEE, 2019.
- [11] Wang C Y, Liao H, Wu Y H, *et al.* CSPNet: a new backbone that can enhance learning capability of CNN [C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Seattle: IEEE, 2020.
- [12] Lin T Y, Dollar P, Girshick R, *et al.* Feature pyramid networks for object detection [C]//Proceedings of the IEEE conference on computer vision and pat-

tern recognition. [S. l.]: IEEE, 2017.

[13] Liu S, Qi L, Qin H, *et al.* Path aggregation network for instance segmentation [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. [S. l.]: IEEE, 2018.

[14] Huang G, Liu Z, Laurens V, *et al.* Densely connected convolutional networks [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. [S. l.]: IEEE, 2017.

[15] He K, Zhang X, Ren S, *et al.* Deep residual learning for image recognition [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. [S. l.]: IEEE, 2016.

[16] Yu F, Chen H, Wang X, *et al.* BDD100K: a diverse driving dataset for heterogeneous multitask learning [C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2020.

引用本文格式:

中 文: 章程军, 胡晓兵, 牛洪超. 基于改进 YOLOv5 的车辆目标检测研究[J]. 四川大学学报: 自然科学版, 2022, 59: 053001.

英 文: Zhang C J, Hu X B, Niu H C. Vehicle object detection based on improved YOLOv5 method [J]. J Sichuan Univ; Nat Sci Ed, 2022, 59: 053001.