

重庆理工大学学报(自然科学)

Journal of Chongqing University of Technology(Natural Science)

ISSN 1674-8425,CN 50-1205/T

《重庆理工大学学报(自然科学)》网络首发论文

题目: 改进 YOLOv7 的道路多目标检测算法
作者: 张琦, 张赛军, 周广生, 谢豪
网络首发日期: 2024-05-28
引用格式: 张琦, 张赛军, 周广生, 谢豪. 改进 YOLOv7 的道路多目标检测算法[J/OL]. 重庆理工大学学报(自然科学).
<https://link.cnki.net/urlid/50.1205.t.20240527.1643.002>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

改进 YOLOv7 的道路多目标检测算法

张琦, 张赛军, 周广生, 谢豪

(华南理工大学 机械与汽车工程学院, 广州 510640)

摘要: 道路环境的感知是自动驾驶任务中的重要组成部分, 为解决道路环境感知中小目标检测困难、检测目标尺寸不一致以及检测目标的遮挡给检测任务带来的困难, 提出一种深度学习增强方法以提高目标检测性能。首先, 设计了 Bottleneck-ELAN(Bottleneck-Efficient Layer Aggregation Networks)模块作为主干, 加强了模型的特征提取能力。同时, 使用 Gather-and-Distribute(GD)机制实现了特征图之间跨尺度的直接融合, 解决了颈部网络的信息丢失问题。此外, 采用 Complete-IOU(CIOU)和 Normalized Wasserstein Distance(NWD)相结合的损失函数组, 克服了单一 IOU 损失函数对不同尺度物体位移敏感性不一致和平滑性差的问题。实验结果表明, 改进后的模型在 BDD100K 数据集上的平均精度均值达到了 43.4%, 相较于原始的 YOLOv7 算法提高了 3.1%, 并且在小目标检测中精度提升更为明显, 提升达到了 10%。

关键词: 计算机视觉; 目标检测; 深度学习; YOLOv7 算法

中图分类号: TP391.4

Improved YOLOv7 object detection algorithm for the target on the road

ZHANG Qi, ZHANG Saijun, ZHOU Guangsheng, XIE Hao

(School of Mechanical and Automotive Engineering, South China University of Technology, Guangzhou 510640, China)

Abstract: The perception of road environment is an important component of autonomous driving tasks. In order to solve the difficulties of detecting small targets, inconsistent detection target sizes, and occlusion of detection targets in road environment perception, a deep learning enhancement method is proposed to improve target detection performance. Initially, the Bottleneck-ELAN module is developed as the backbone to enhance the model's feature extraction capability. The Gather-and-Distribute (GD) mechanism is also employed to achieve cross-scale fusion between feature maps, addressing the issue of information loss during feature fusion across different scales. Additionally, a combination of the Complete-IOU (CIOU) and Normalized Wasserstein Distance (NWD) loss functions is employed to address the inconsistency in sensitivity to object displacement and the smoothness disparity present in the single IOU loss function. The experiment shows that the average accuracy of the improved model on the BDD100K dataset reaches 43.4%, which is 3.1% higher than the original YOLOv7 algorithm. Moreover, the accuracy improvement is more significant in small object detection, with an increase of 10%.

Keywords: computer vision; object detection; deep learning; YOLOv7 algorithm

0 引言

基于计算机视觉的道路多目标检测技术是自动驾驶领域的研究基础之一, 并被广泛应用于辅助驾驶、车辆行人跟踪等智能驾驶任务中。道路目标检测通过带有类别标签的预测框来检测行驶道路上的不同目标, 并为车辆控制决策提供关键信息, 算法的准确性和实时性直接影响到自动驾驶车辆的安全性能和反应速度, 因此道路场景下的目标检测仍然是无人驾驶的研究热点之一。

早期的目标检测方法主要基于传统的计算机视觉算法, 如 Haar 特征分类器^[1]、HOG 特征分类器^[2]和 SIFT 特征分类器^[3]等, 这些方法需要手动选择特征并对特征提取和分类进行调试, 难以适应复杂场景的要求, 鲁棒性和泛化性相对较差。随着计算机性能的不断提高, 基于深度学习的目标检测算法逐渐成为了研究的主流方向^[4]。

近年来, 道路目标检测任务的研究大多是在现有目标检测算法的基础上, 结合算法的优缺点和道路目标检测场景的特点设计有效的模块, 以获得更高的检测精度和更快的检测速度。如针对道路中的行人检测问题, 樊嵘等^[14]以 DETR 为基础模型, 引入注意力编码器来有效利用小目标的特征信息, 并在训练时采用 Smooth-L1 与 GIOU^[15]结合的损失函数来加速模型收敛, 在 Wider-Person 数据集上取得了较好的检测效果。针对道路中的交通标志检测问题, 韦强等^[16]以 YOLOv5 为基础

投稿日期: 2024-01-16

基金项目: 广东省重点领域研发项目(2020B010184002)

作者简介: 张琦, 男, 硕士研究生, 主要从事计算机视觉与目标检测研究, E-mail: 598423240@qq.com; 通信作者张赛军, 男, 博士, 副教授, 主要从事数字图像处理研究, E-mail: mesjzhang@scut.edu.cn。

模型, 引入递归门控卷积^[17]用于骨干网络, 并采用 F-Elou^[18]损失函数和 SOCA^[19]注意力机制提高小目标的检测精度, 在 TT100K 数据集上得到了较大的检测精度提升。在道路行驶过程中的信号灯检测问题, Müller 等^[20]改进了 SSD 算法的结构, 提出了 TL-SSD, 该算法结合了上下文和本地信息的特点, 具有较高的检测准确率, 但检测速度较慢, 实时性能较差。对于复杂道路场景中的目标检测问题, 盛博莹等^[21]提出基于反馈机制的特征提取网络(RFP-PANet), 并引入级联注意力机制(SECA)用来提高杂交通场景下的小目标检测精度; Wang 等^[22]基于 YOLOv5 增加浅层检测层来提高道路小目标的检测精度, 但导致模型提取的有效特征被过度分散而不利于常规目标的检测, 最终整体精度提升并不明显; 刘辉等^[23]在 YOLOv5s 算法的基础上提出了改进模型 DPE-YOLO, 在复杂道路场景中的小目标和遮挡目标检测任务上获得了性能的提升。

目前出于对实时性的需求，大多数的道路目标检测算法的研究选用了一阶段目标检测算法为研究基础。然而，一阶段目标检测算法在道路目标检测任务中仍然存在一些问题。首先，道路目标检测任务中存在着大量的小尺寸目标，而一阶段检测算法在主干网络对图像进行特征提取时，需要压缩原始图像和特征图像的尺寸，这会导致小目标的特征丢失，进而降低小目标的检测精度。其次，在颈部网络的特征融合中，跨层级之间的特征融合是通过中间层级的间接融合，这导致不同层级之间对相同种类检测目标的特征信息不一致。在检测多尺度检测目标时，这种特征的不一致可能会导致只有部分尺度大小的检测目标得到检测。另外，道路目标检测任务中会存在大量的目标遮挡问题，这同样给检测任务带来了极大的困难。

本文中，针对道路目标检测任务中检测目标易受到遮挡、同类目标尺度较多且跨度较大以及一阶段目标检测算法对小目标检测精度较差等问题，基于 YOLOv7 算法进行改进研究：(1)针对道路目标环境检测任务中检测目标被遮挡问题，结合 ELAN(Efficient Layer Aggregation Networks)和 Bottleneck 网络，在骨干网络中设计多梯度流残差结构的特征增强模块 Bottleneck-ELAN，并融入了 CBAM 注意力机制，更好地提取图形特征，增强检测能力，降低目标检测的误检率和漏检率；(2)通过使用基于 Gather-and-Distribute(GD)机制的特

征融合网络替换原模型中的 FPN-PANet 特征融合结构,使不同尺度的特征实现跨层级的直接融合,获得更好的特征融合效果;(3)使用复合损失函数组,即 CIOU 和 NWD 的加权求和作为模型训练的约束,缓解检测中 IOU 损失对位移的敏感性问题,并增加损失上升的平滑性,有效改善了 IOU 系列损失函数的缺陷。

1 YOLOv7 算法介绍

YOLOv7^[13]是一款优秀的目标检测模型，它是在 YOLOv5 的基础上进行改进和优化得到的，结合了众多优秀的设计思想，保证了在边缘设备推理时的精度与速度平衡。因此，本文中选择在 YOLOv7 算法基础上进行改进，在保证推理实时性的前提下获得更高的检测精度。

YOLOv7 算法由输入端(Input), 主干网络(Backbone), 颈部网络(Neck), 检测头(Head)组成, 网络结构图如图 1 所示。

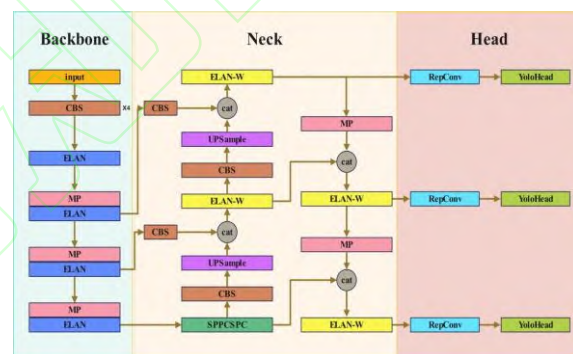


图 1 YOLOv7 网络结构图

1.1 主干网络

YOLOv7 的主干网络由多个 ELAN 模块和 MP 模块组成, 通过 32 倍下采样处理输入图像。CBS 模块由 Conv、BN、SiLu 组成, ELAN 由多个 CBS 模块堆叠组成, 蕴含丰富的梯度流信息, 可以高效地利用网络参数并加速网络的推理, MP 模块由 CBS 模块和最大池化层组成, 主要用于下采样操作。各个模块结构图如图 2 所示:

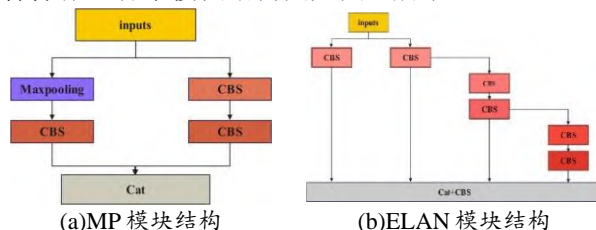


图 2 YOLOv7 主干网络模块结构

1.2 颈部网络

YOLOv7 的颈部网络是通过 ELAN-w 模块和

MP 模块构建的自上而下的 FPN^[24](Feature Pyramid Network)和自下而上的 PANet^[25](Path Aggregation Network)。FPN 对含有高维强语义信息的特征进行上采样,增强多个尺度上的语义表达, PANet 将浅层的定位和细节信息传递到深层,增加多个尺度上的定位能力。特征在颈部网络中实现不同层次之间的高效融合,实现了不同尺度特征之间的互补,为检测头的回归提供更加完善的特征。

1.3 多尺度检测头

YOLOv7 的多尺度检测头采用了 RepConv^[26]结构,通过增加训练成本来换取推理阶段的高效性。在检测头中,利用颈部网络输出的不同尺度特征进行目标类别和目标位置的回归,从而实现目标的检测。

2 模型改进

尽管 YOLOv7 算法可以适应许多目标检测场景,但在道路目标检测任务中,由于存在检测目标尺寸跨度较大和检测目标易受遮挡的问题且检测算法在小目标检测任务中的不足,目标检测效果往往较差。为了应对这些问题,对 YOLOv7 的模型网络和损失函数进行改进以获得更好的性能。

2.1 主干网络改进

YOLOv7 算法通过 ELAN 结构对图像特征进行提取。ELAN 模块通过并行更多的梯度流分支以获得更丰富的梯度信息,进而获得更高的精度和更合理的延迟。本文中,主干网络的设计沿用并行梯度流分支的做法,并在模块中融合 Bottleneck^[27]模块和注意力机制,设计了一种新的 Bottleneck-ELAN 模块,模块结构如图 3 所示。

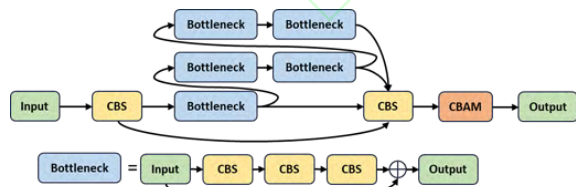


图 3 Bottleneck-ELAN 模块结构

Bottleneck-ELAN 模块首先使用恒等卷积将输入特征图的通道数量压缩,以提升待聚合特征图的信息密度并减少后续模块的运算量。然后,通过不同深度的 Bottleneck 结构提取特征,生成具有不同感受野和层次语义信息的特征图并将不同深度和浅度的特征图进行拼接融合,得到包含复杂梯度流特征的特征图。最后,通过 CBAM 模块提

取注意力,筛选重要特征,并将筛选后的特征图与输入特征图相加,形成残差结构,以防止网络退化。

CBAM^[28](Convolutional Block Attention Module)是一种注意力机制模块,用于增强卷积神经网络在特征表示和视觉表达能力方面的性能。CBAM 模块通过自适应地学习特征通道维度和空间维度的注意力权重,使网络更加关注重要特征并抑制噪音。具体模块结构如图 4 所示。

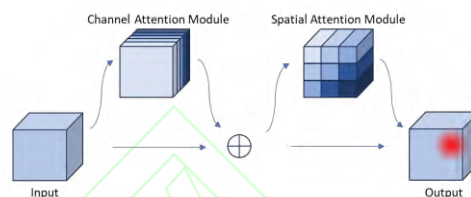


图 4 CBAM 注意力机制

道路多目标检测任务中多数目标存在着相对明确的位置信息,即不同目标出现在图像中的位置相对固定, CBAM 机制能够提取这种位置信息并对后续的检测任务提供依据,因此采用该模块在 Bottleneck-ELAN 模块对特征图进行注意力的提取,进一步增强特征提取的效果。通过 Bottleneck-ELAN 和 CBAM 模块的结合,关键细节信息可以通过多梯度流结构和残差连接有选择性地传递到深层,并强化检测目标的位置特征和语义特征,最终为目标检测提供更为有效的特征信息。

2.2 颈部网络改进

YOLOv7 的颈部网络采用了 FPN-PANet 结构,实现了语义信息和细节信息的双向传播,如图 5 所示。

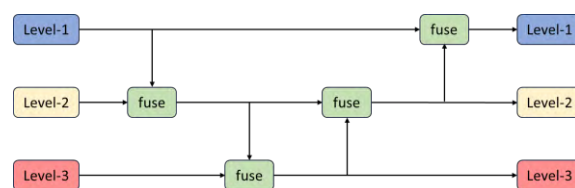


图 5 FAN-PANet 网络通道间的信息传输

图 5 表明,在特征融合的过程中,FPN-PANet 结构间隔层级之间的信息融合只能通过中间层级的间接融合来实现,无法直接融合不同层级之间的信息。这种传输模式导致间隔层级之间传输的信息是经过中间层级筛选后的残缺信息,从而造成某些信息的丢失,因此限制了信息融合的整体有效性。在道路目标检测任务中,存在着大量的多尺度目标,这种整体有效性的破坏,会导致在检测多尺度目标时仅能实现对特定尺寸的对象进行检测,并漏检其他尺寸的对象。

为了避免传统 FPN-PANet 结构在传输过程中出现信息丢失的问题, 本文中的颈部网络采用了基于 Gather-and-Distribute(GD)^[29] 机制的颈部网络。使用统一的模块来收集和融合各个层次的信息, 然后将融合后的特征分配到不同层级的特征图中, 不仅避免了传统 FPN 结构中固有的信息丢失问题, 并且在不显著增加延迟的情况下增强了颈部网络的部分信息融合能力。网络结构如图 6 所示。

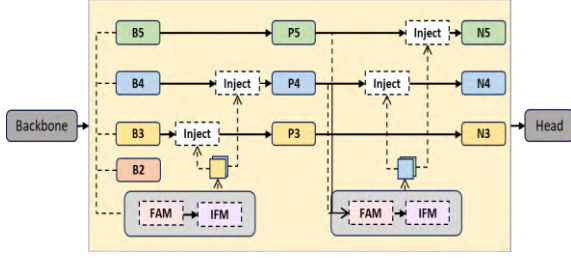


图 6 GD 机制的信息传输

图 6 显示, GD 机制用卷积和自注意力机制分别实现了三个模块: Feature Alignment Module (FAM), Information Fusion Module (IFM), 信息注入模块 Information Injection Module (Inject)。FAM 模块收集主干网络中不同尺度的特征图, 并通过上采样、下采样的方式对齐, IFM 模块将收集的特征生成全局特征并切分, Inject 模块将切分后的全局特征分发到各个层级, 最终实现不同层级特征之间信息的直接融合, 有效解决了目标检测中跨层级特征图之间检测特征信息不一致的问题。

2.3 损失函数改进

YOLO 系列算法的损失函数包括置信度损失、分类损失和定位损失三个部分。其中定位损失采用的是 IOU 系列损失函数。IOU 损失函数是通过计算检测框和真实框之间的交并比以确定检测框定位精度的一种损失函数, 这类损失函数对小目标的位移非常敏感, 并且函数变化的平滑性受检测目标尺寸影响严重, 如图 7 所示:

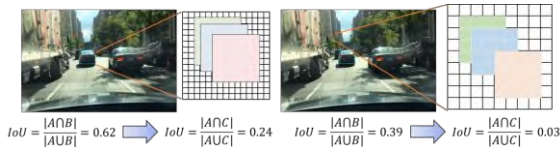


图 7 IOU 对小目标的位移敏感性更强且更容易失效

为了解决这一问题, 随后的研究提出了

GIOW^[15]、DIOU^[30]和 CIOU^[30], 这些方法考虑了检测框面积、中心点距离和长宽比等各种几何性质, 以克服 IOU 计算的局限性。CIOU 公式如下

$$L_{\text{CIOU}} = \text{IOU} + \frac{\rho^2(b, b^{\text{gt}})}{c^2} + \alpha v \quad (1)$$

式中: b, b^{gt} 分别表示检测框和真实框的中心点, ρ 表示计算两个中心点之间的欧氏距离, c 是同时包含预测框和真实框的最小闭合区域的对角线距离, α 是用于进行权衡的参数, v 是用于测量长宽比一致性的参数。然而, 在小目标检测任务中, 基于 IOU 的损失仍然存在平滑性较差且易失效的问题, 这导致模型在训练过程中受到的监督信息不足, 从而使网络的收敛变得困难。

为弥补 IOU 系列损失函数的缺陷, 引入 NWD^[31] 损失函数作为补充, 将 CIOU 损失和 NWD 损失进行加权求和作为最后的损失函数, 其中 NWD 的相关公式如下:

$$W_2^2(N_a, N_b) = \left\| \left[\begin{matrix} cx_a, cy_a, \frac{\omega_a}{2}, \frac{\omega_a}{2} \end{matrix} \right]^T, \left[\begin{matrix} cx_b, cy_b, \frac{\omega_b}{2}, \frac{\omega_b}{2} \end{matrix} \right]^T \right\|_2^2 \quad (2)$$

$$L_{\text{NWD}} = \exp \left(-\frac{\sqrt{W_2^2(N_a, N_b)}}{C} \right) \quad (3)$$

式中: N_a 和 N_b 代表两个不同的检测框,

$cx, cy, \frac{\omega}{2}, \frac{\omega}{2}$ 分别表示边界框的中心坐标和尺寸。

NWD 将检测框建模成二维正态分布, 并计算二维正态分布之间的 Wasserstein 距离, 以解决 IOU 系列函数平滑性较差和位移敏感性不一致的问题。最终的损失函数公式为:

$$L_{\text{com}} = 1 - \lambda L_{\text{CIOU}} - (1 - \lambda) L_{\text{NWD}} \quad (4)$$

其中 λ 为两种损失函数的权重。NWD 损失函数在小目标检测任务中也能保证有效性和较好的平滑性, 能够有效弥补 CIOU 损失函数中 IOU 系数快速失效导致的性能缺陷。

表 1 改进算法和 YOLOv7 算法各检测目标精度对比

Algorithm	$mAP_{50}/\%$	Precision /%	Recall /%	FPS	$AP_{50}/\%$								
					car	bus	person	bike	truck	motor	rider	sign	light
YOLOv7	40.3	70.6	39.2	98	62.1	52.8	42.1	30.2	52.4	27.2	27.7	41.9	26.1
Improved-YOLO	43.4	73.4	41.3	79	65.8	50.5	45.9	29.1	50.4	30.9	28.2	52.7	37.9

3 实验过程及结果分析

实验硬件平台为 NVIDIA RTX 4080 16G，软件环境为 PyTorch 2.0。所有模型使用 Adam 优化器进行训练。在所有实验中，最大训练期设置为 500，初始学习率为 0.01，最终学习率为 0.001。学习速率衰减采用余弦函数衰减方法。所有实验均未使用预先训练过的模型。

3.1 数据集选择与预处理

本次实验使用 BDD100K 数据集，该数据集由 Berkeley Deep Drive 团队开源，涵盖了来自加州和纽约市不同地区的约 10 万张图像，图像像素均为 1280×780 。BDD100K 数据集包含了丰富的驾驶场景，如城市街道、高速公路、乡村道路等，以及不同天气光照条件。数据集中有多个类别的目标，包括 Bus、Light、Sign、Person、Bike、Truck、Motor、Car、Rider。实验选取数据集中完成标注的 80000 张图片作为实验数据并按照 7:2:1 进行划分，其中 56000 张图片作为训练集，16000 张作为验证集，8000 张作为测试集。为提高模型的鲁棒性，实验过程中采用随机剪裁拼接的方

式，对数据集进行数据增强。

3.2 实验结果

3.2.1 改进算法检测结果

在处理过的数据集上使用训练集分别训练 YOLOv7 算法和改进后的算法，然后使用测试集对训练好的模型进行测试验证。实验选用目标检测通用指标 mAP_{50} 作为判断检测效果的标准。YOLOv7 算法和改进后的算法在测试集上的检测结果如表 1 所示。根据表 1 可知，改进算法在所有检测类别的平均 mAP_{50} 指标提高了 3.1%，虽然检测速度相较于 YOLOv7 算法略有下降，但仍然满足实时性的需求。在一些尺度较多的目标检测任务中，例如 car 和 person，精度提升达 3.5% 以上，这表明改进算法能够有效缓解尺度跨度过大带来的检测困难。在一些小目标的检测任务中，例如 sign 和 light，精度的提升达 10% 以上。这表明改进后的算法能够有效提高小目标的检测精度。为验证模型的泛化性能，使用训练完成后的 YOLOv7 算法以及改进算法在测试集上进行测试，并选取具有代表性的三组图片进行分析，检测结果如图 8 所示。

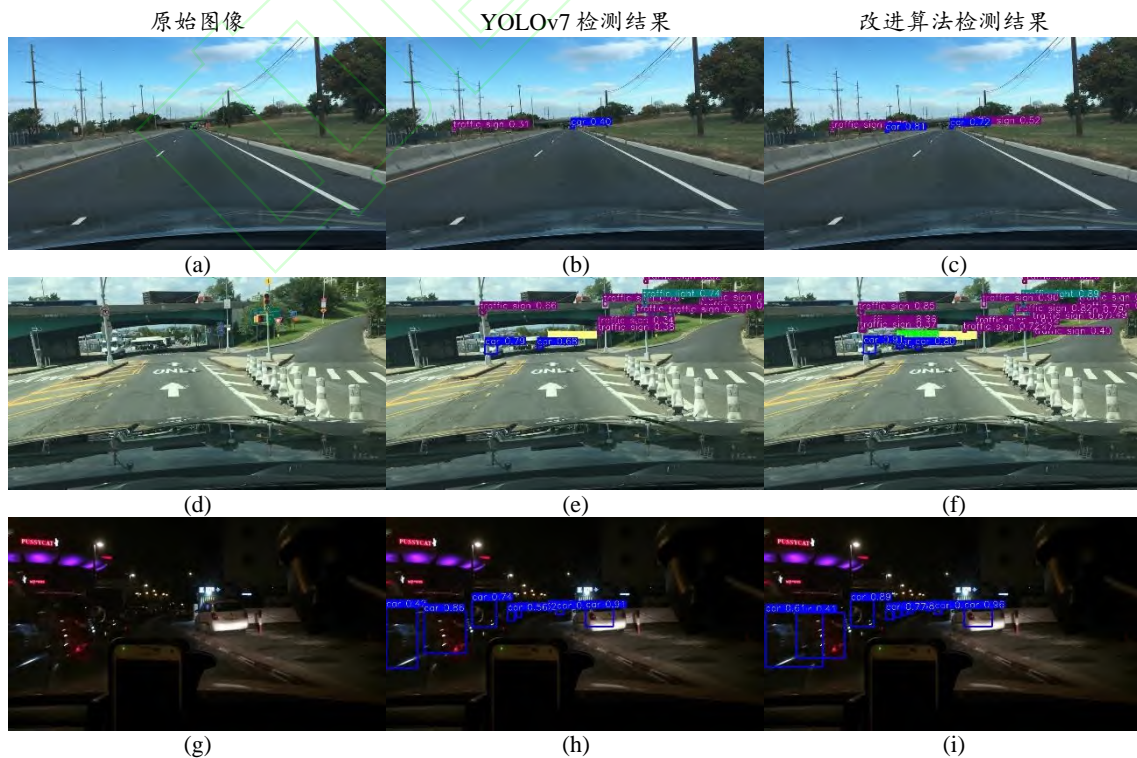


图 8 改进模型检测结果对比

图 8 可以观察到, 在对图 8(a)的检测中, YOLOv7 算法对图像中尺寸较小的目标发生了漏检, 而改进的算法避免了这些对小目标的漏检, 证明改进算法在小目标检测任务中拥有更高的准确率。而在对图 8(d)的检测中, 由于图像中检测对象数量增多并且受到了不同程度的遮挡, 使得检测任务变得更加困难, YOLOv7 算法在检测过程中漏检了大量受遮挡目标, 而改进算法在检测的过程中避免了这些漏检, 证明改进模型具备更好的鲁棒性。在图 8(g)所示的图像中, 检测对象随着距离的变化而逐渐缩小, 并且天气导致的图像昏暗也会影响检测的结果。YOLOv7 算法在检测时, 随着对象尺寸的缩小逐渐出现了漏检, 而改进算法避免了尺寸变化导致的检测缺失, 表明改进算法在检测多尺度对象中能够保持良好的稳定性并准确检测出不同尺寸的同类目标。

综上, 改进的算法不仅能够有效缓解 YOLOv7 算法检测小目标和多尺度目标困难的缺陷, 并且能够更好得适应恶劣的检测环境, 拥有更好的鲁棒性。

3.2.2 消融实验对比

为了验证 Bottleneck-ELAN 模块、GD 机制和组合损失函数对模型性能的提升效果以及不同模块之间的耦合性能, 对上述模块进行消融实验研究。实验在 BDD100K 数据集上进行, 实验结果如表 2 所示。

表 2 消融实验结果

Algorithm	$mAP_{50}/\%$
YOLOv7	40.3
YOLOv7-Bottleneck-ELAN	42.1
YOLOv7-GD	41.8
YOLOv7-Loss	40.9
All-Improved	43.4

表 2 显示, Bottleneck-ELAN、GD 机制和联合损失函数单独作用时, 使 mAP_{50} 指标分别提高了 1.8%、1.5%和 0.6%, 说明了各个模块对道路目标检测任务的有效性。并且在上述模块共同作用时, 检测精度提升了 3.1%, 相较于各个模块单独作用, 上述模块共同作用时模型获得了更高的精度, 说明上述模块的耦合性能较好。

3.2.3 不同模型对比

为进一步对比实验, 将本文算法与先进的目标检测算法作比较, 包括一阶段目标检测算法 YOLOv5n、YOLOv3-spp、YOLOr-csp、YOLOv7, 两阶段目标检测算法 Fast R-CNN、Faster R-CNN。不同模型之间的性能对比如表 3 所

示:

表 3 不同模型的性能对比

Algorithm	$mAP_{50}/\%$	Precision/ /%	Recall/ /%	FPS
Fast R-CNN	37.5	67.1	38.2	27
Faster R-CNN	48.9	75.4	51.3	39
YOLOv5n	27.5	56.5	28.4	232
YOLOv3-spp	36.6	57.3	35.6	120
YOLOr-csp	36.9	63.1	34.0	92
YOLOv7	40.3	70.6	39.2	98
Improved-YOLO	43.4	73.4	41.3	79

根据表 3 可以观察到, 在主流的道路场景检测算法中, yolov5n 具有最好的实时性, 检测速度达到了 232fps, 但是检测效果也是主流算法中较差的。Faster R-CNN 的精度和召回率达到了最高, 但是检测速度相较于其他算法相比有着明显的不足。本文改进的算法取得了检测精度和检测速度的平衡, 不仅 FPS 达到了 79, 能够满足实时检测的需要, 并且检测精度也仅次于 Faster R-CNN, 达到了 43.4%。综合对比表中算法, 本文所提出的算法在具有较好的检测速度的同时, 具有更高的检测精度, 因此更能准确地进行道路目标检测。

3.2.4 损失函数的平衡验证

本文中, 训练选用 CIOU 和 NWD 两种损失函数作为约束, 并用权重参数 λ 进行平衡。为验证不同的权重参数对检测结果的影响, 以及选取最优的权重参数, 选用不同的权重参数 λ , 并进行实验对比:

表 4 不同权重参数对模型 λ 的影响

λ	$mAP_{50}/\%$
0	42.7
0.2	43.1
0.4	43.4
0.6	42.5
0.8	42.8
1	42.4

根据表 4 可以观察到, 随着权重参数 λ 的改变, 模型的检测精度也发生了变化, 当权重参数 λ 为 0 或 1 时, 即损失函数的构成由单一的 CIOU 和 NWD 损失函数构成时, 目标检测的精度要低于两种损失函数共同约束时的精度。当选取权重参数 λ 为 0.4 时, 模型的检测精度达到了最高, 达到了 43.4%。

4 结论

针对道路目标检测任务中存在大量小型检测目标、目标尺寸跨度大且目标易受遮挡等问题,选用 YOLOv7 目标检测模型为基础进行模型改进以获得更好的性能并通过实验验证其有效性:

1)综合 ELAN 模块多梯度流分支的思想和 Bottleneck 模块压缩通道减少运算量的思想,设计 Bottleneck-ELAN 模块作为主干网络,加强模型的特征提取能力并控制模型运算量的增加。

2)为解决颈部网络在特征跨尺度融合时的信息丢失问题,使用 GD 机制实现了不同尺度特征图之间的直接融合。

3)采用 CIOU 和 NWD 相结合的损失函数,克服了单一 IOU 损失函数对不同尺度物体位移敏感性不一致和平滑性差的缺陷。

4)实验证明上述模块能够有效提高 YOLOv7 算法在小目标检测上的精度,并有效解决算法在检测尺寸跨度较大的目标时部分尺寸目标漏检的问题。上述改进模块分别使模型的 mAP_{50} 指标提高了 1.8%、1.5% 和 0.6%,且模块之间的耦合性能良好,模块的叠加能够有效提升模型的检测效果,最终相较于原 YOLOv7 算法, mAP_{50} 指标提升了 3.1%。

参考文献:

- [1] VIOLA P, JONES M. Rapid object detection using a boosted cascade of simple features[C]//Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. Cambridge, MA: IEEE 2001.
- [2] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]//2005 IEEE computer society conference on computer vision and pattern recognition, Cambridge, MA: IEEE, 2005.
- [3] LOWE D G. Distinctive image features from scale-invariant key points[J]. International journal of computer vision, 2004, 60: 91-110.
- [4] WU X, SAHO D, HOI S C H. Recent advances in deep learning for object detection[J]. Neurocomputing, 2020, 396: 39-64.
- [5] HE K, GKIOXARI G, DOLLAR P, et al. Mask r-cnn[C]//Proceedings of the IEEE international conference on computer vision. Venice, ITALY: IEEE 2017.
- [6] GIRSHICK R. Fast r-cnn[C]//Proceedings of the IEEE international conference on computer vision. Santiago, Chile: IEEE, 2015.
- [7] REN Shaoqing, HE K, GIRSHICK R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[J]. Advances in neural information processing systems, 2017, 39(6): 1137-1149.
- [8] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector[C]//Computer Vision-ECCV 2016: 14th European Conference, Amsterdam, The Netherlands: Springer International Publishing, 2016.
- [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, 2016.
- [10] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Seattle, WA: IEEE 2017.
- [11] REDMON J, FARHADI A. YOLOv3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.
- [12] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv:2004.10934, 2020.
- [13] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, BC, Canada: IEEE 2023.
- [14] 樊嵘, 马小陆. 面向拥挤行人检测的改进 DETR 算法[J]. 计算机工程与应用, 2023, 59(19): 159-165.
FAN Rong, MA Xiaolu. An improved DETR algorithm for crowded pedestrian detection[J]. Computer Engineering and Applications, 2023, 59(19): 159-165.
- [15] REZATOOFIHI 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, CA, USA: IEEE, 2019.
- [16] 韦强, 胡晓阳, 赵虹鑫. 改进 YOLOv5 的交通标志检测方法[J]. 计算机工程与应用, 2023, 59(13): 229-237.
WEI Qiang, HU Xiaoyang, ZHAO Hongxin. Improved traffic sign detection method for YOLOv5[J]. Computer Engineering and Application, 2023, 59(13): 229-237.
- [17] RAO Y, ZHAO W, TANG Y, et al. Hornet: Efficient high-order spatial interactions with recursive gated convolution s[J]. Advances in Neural Information Processing Systems, 2022, 35: 10353-10366.

-
- [18] ZHANG Y F, REN W, ZHANG Z, et al. Focal and efficient IOU loss for accurate bounding box regression[J]. *Neuro computing*, 2022,506: 146-157.
- [19] FRANK E, HALL M. A simple approach to ordinal classification[C]//European conference on machine learning. Springer, Berlin, Heidelberg, Hamilton, New Zealand: ECML, 2001.
- [20] MÜLLER J, DIETMAYER K. Detecting traffic lights by single shot detection[C]//2018 21st International Conference on Intelligent Transportation Systems (ITSC). Maui, HI: IEEE, 2018: 266-273.
- [21] 盛博莹,侯进,李嘉新等.面向复杂交通场景的道路目标检测方法[J]. *计算机工程与应用*,2023,29(15): 87-96.
SHENG Boying, HOU Jin, LI Jiaxin, et al. A road target detection method for complex traffic scenarios[J]. *Computer Engineering and Applications*, 2023,29(15): 87-96.
- [22] WANG P, FU S, CAO X. Improved Lightweight Target Detection Algorithm for Complex Roads with YOLOv5[C]//2022 International Conference on Machine Learning and Intelligent Systems Engineering (MLISE). Guangzhou, China: IEEE, 2022: 275-283.
- [23] 刘辉, 刘鑫满, 刘大东. 面向复杂道路目标检测的 YOLOv5 算法优化研究[J], *Journal of Computer Engineering & Applications*, 2023, 59(18): 207-217
LIU Hui, LIU Xinman, LIU Dadong. Optimization of YOLOv5 algorithm for complex road target detection[J]. *Journal of Computer Engineering & Applications*, 2023, 59(18): 207-217
- [24] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Honolulu, HI :IEEE, 2017.
- [25] 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. Salt Lake City, UT: IEEE, 2018.
- [26] DING X, ZHANG X, MA N, et al. Repvgg: Making vgg-style convnets great again[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. ELECTRONIC NETWORK: IEEE, 2021.
- [27] 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. Las Vegas, NV, USA: IEEE, 2016: 770-778.
- [28] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). Munich, GERMANY: Lecture Notes in Computer Science 2018: 3-19.
- [29] WANG C, HE W, NIE Y, et al. Gold-YOLO: Efficient Object Detector via Gather-and-Distribute Mechanism[J]. *arXiv preprint arXiv:2309.11331*, 2023.
- [30] 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, NY: aai 2020.
- [31] WANG J, XU C, YANG W, et al. A normalized Gaussian Wasserstein distance for tiny object detection[J]. *arXiv preprint arXiv:2110.13389*, 2021.