

DOI:10.19651/j.cnki.emt.2417421

基于 QARep-YOLOv8n 的城市物联网车辆 目标检测方法^{*}

杨沙沙¹ 徐新源²

(1. 西安交通大学中兴通信学院 西安 710300; 2. 陕西航天动力高科股份有限公司 西安 710076)

摘要: 针对现有的方法在密集、遮挡以及小目标检测中存在的问题,本研究提出了一种用于城市道路场景中车辆检测的 QARep-YOLOv8n 算法。首先,本研究采用了一种 Haar 小波下采样模块来缓解传统跨步卷积或池化操作所导致的特征信息丢失的问题;其次,提出了一种批量正则化注意力模块和 QARepC2f 模块来提高 YOLOv8 的特征提取能力;最后,采用了 NWD 边界框损失和 Slide 分类损失来提高对于小目标以及遮挡目标的检出效果。在 4 个主流的车辆检测基准数据集上的广泛消融实验和验证实验表明,QARep-YOLOv8n 相比于 YOLOv8n, mAP 分别提升了 3.3%、3.2%、2.7% 和 1.5%,此外,本研究方法具有更显著的小目标、遮挡目标检测效果。

关键词: 城市物联网; 车辆目标检测; 特征提取; 损失函数

中图分类号: TN929.5 文献标识码: A 国家标准学科分类代码: 520.6040

Urban internet of things vehicle object detection method based on QARep-YOLOv8n

Yang Shasha¹ Xu Xinyuan²

(1. School of Zhong Xing Communication, Xi'an Traffic Engineering Institute, Xi'an 710300, China;

2. Shaanxi Aerospace Power Hi-Tech Co., Ltd., Xi'an 710076, China)

Abstract: Aiming at the problems of existing methods in dense, occluded and small object detection, this paper proposes a QARep-YOLOv8n algorithm for vehicle detection in urban road scenes. First, this paper adopts a Haar wavelet downsampling module to alleviate the problem of feature information loss caused by traditional stepwise convolution or pooling; secondly, this paper proposes a batch-normalization attention module and QARepC2f module to improve the feature extraction ability of YOLOv8; finally, this paper uses NWD bounding box loss and Slide classification loss to improve the detection effect of small and occluded objects. Extensive ablation experiments and validation experiments on four mainstream vehicle detection benchmark datasets shows that QARep-YOLOv8n improves mAP by 3.3%, 3.2%, 2.7% and 1.5%, respectively, compared with YOLOv8n. In addition, the proposed method has more significant detection effects on small and occluded objects.

Keywords: urban internet of things; vehicle object detection; feature extraction; loss function

0 引言

随着物联网技术的发展以及智能设备的广泛应用,如何对物联网数据进行有效地处理和利用成为了学术界关注的重要话题^[1]。随着家用汽车的普及,通过目标检测算法对城市道路的车辆进行自动检测,能够帮助交通执法部门快速地分析出车辆状态、车流量、违法情况等信息,这极大地释放了人工监测的压力^[2]。

基于深度学习的目标检测算法包括一阶段(one-stage)和二阶段(two-stage)两种。尽管 two-stage 目标检测算法能保证不错的车辆检测效果,但其复杂的网络结构,导致这种类型的算法难以保证实时的车辆检测。相反,one-stage 目标检测算法则能够兼顾精度和实时性,实现了检测精度和速度之间的完美权衡。其中,基于 YOLO(you only look once)的目标检测算法是 one-stage 类算法的典型代表,例如: YOLOv5^[3]、YOLOv6^[4]、YOLOX^[5]、YOLOv7^[6]、YOLOv8^[7]等。

收稿日期:2024-11-21

* 基金项目:陕西省教育厅科学研究计划项目(23JK0528)、西安交通大学中青年基金项目(2023KY-34)资助

尽管先前的 YOLO 系列算法取得了不错的车辆检测效果,但是它们一般难以处理日常生活中密集、遮挡的车辆检测场景。与此同时,在城市道路的场景中,往往存在诸多小尺度的车辆目标,由于其外观信息相对有限,因此,提高小目标的检出率对现有的目标检测器是一个极大的挑战^[8]。为了缓解上述问题,原蕾等^[9]提出对瓶颈注意力机制(bottleneck attention module, BAM)的通道和空间注意力分支进行改进,并引入带权重的特征融合和跨层连接,将不同层的特征进行融合,来减少特征信息在传递过程中的丢失。李珣等^[10]提出在 YOLOv7 骨干网络中加入全局注意力机制来增强特征提取性能,并利用全维动态高效聚合网络(ODEANet)重构主干网络,提高算法的鲁棒性与精度。冯晶等^[11]提出在 YOLOv8 的骨干网络中增加注意力机制模块,以增强网络间特征的联系。许德刚等^[12]提出在高分辨的特征图上增加一个小目标预测头来提高 YOLOv8 的小目标检测效果。火久元等^[13]通过引入大内核深度可分离卷积来拓展上下文信息以增强模型对小目标的信息捕获能力。梁超等^[14]提出采用 Adown(anisotropic downsample)下采样模块来提高模型的等效感受野,并引入卷积块注意力机制(convolutional block attention module,CBAM)来提高小目标的检测精度。张利丰等^[15]提出采用双向特征金字塔网络(feature pyramid network, FPN)模块来增强模型的特征提取能力,并通过 Soft-NMS(soft non-maximum suppression)来解决目标框重叠的问题。史涛等^[16]提出在颈部网络采用 CF-FPN 特征融合结构来提升对小目标的检测精度,并将原始 YOLOv8 的头部替换为 Dynamic Head 检测头,通过结合尺度、空间和任务 3 种注意力机制,进一步提高模型的检测性能。高德勇等^[17]提出 PA-AFPN 特征融合方式来提升网络对多尺度特征的融合能力,并在 SPPF(spatial pyramid pooling fast)中引入平均池化分支和三重注意力机制,有效整合多尺度信息,降低背景干扰对检测的影响。综上所述,先前的方法主要通过扩大目标检测网络的等效感受野,或通过设计一些复杂的模块来增强模型的特征提取能力,从而提高模型对于遮挡目标和小目标的检出。然而,这些方法会引入大量的参数量和计算成本,从而增加模型的推理成本。

为了解决上述问题,本文提出了一种 QARes-YOLOv8n 的车辆检测算法。首先,为了缓解传统的检测网络中跨步卷积或池化模块可能导致的特征信息丢失的问题,本文提出采用 Haar 小波下采样^[18](Haar wavelet downsampling, HWD)模块来降低特征图的空间分辨率。HWD 模块通过将空间维度的信息编码到通道维度,并通过一个卷积层将通道映射到指定维度和提取具有判别性的特征,极大程度地保留了小目标的有效信息。

其次,为了增强 YOLOv8 对于小目标的特征提取能力,本文提出了一种批量正则化注意力(batch-normalization attention,BNA)模块。BNA 模块通过对 BN

中仿射变换的尺度因子进行归一化,并将归一化的权重与输入特征进行矩阵乘法来对输入特征进行重校准,有助于模型抑制无用特征,突出有用特征。值得注意的是,与先前的注意力模块不同的是,BNA 模块没有引入任何额外的参数量,因此,其对模型的推理性能不会产生任何负面的影响。此外,为了进一步提高 YOLOv8 对于小目标的信息捕获能力,本文提出了一种 QAResC2f 模块来替换 YOLOv8 中的 C2f 模块。该模块采用重参数化的思想,在训练阶段,利用多分支结构来提高模型的感受野,从而有利于模型捕获小目标的有效信息,在推理阶段,多分支结构通过重参数化合并成一个分支,通过这种方式,QAResC2f 模块相比 C2f 模块,不会引入额外的参数量和计算成本。

最后,为了提高 YOLOv8n 对于密集、遮挡目标的检测效果,本文采用了归一化高斯 Wasserstein 距离(normalized Gaussian Wasserstein distance, NWD)损失函数^[19]和 Slide 损失函数^[20]来学习边界框坐标位置信息和分类信息。NWD 损失将边界框和真实框建模成两个二维的高斯分布,然后,引入 Wasserstein 距离,用于衡量两个分布的相似程度,通过这种方式,减轻了小目标的微小位置偏差会导致明显的交并比(intersection over union, IoU)下降,从而导致不准确的标签分配的问题。Slide 损失则通过对边界困难样本进行强加权,促使模型在训练的过程中,显著地关注密集、遮挡的困难样本。

1 QARes-YOLOv8n 车辆检测方法

1.1 Haar 小波下采样模块

为了降低目标检测模型的计算复杂度,先前的方法一般采用池化或跨步卷积来降低中间特征图的空间分辨率。然而,池化或跨步卷积会导致局部信息丢失,丢失的信息如边界、尺度、纹理等在目标检测中起着至关重要的作用,导致小目标检测效果较差。因此,本文提出了一种 Haar 小波下采样来缓解上述问题,旨在降低特征图的空间分辨率的同时保留小目标的有效信息。

如图 1 所示,HWD 模块包括两个模块:无损特征编码模块和特征表示学习模块。其中,无损特征编码模块负责特征转换和降低空间分辨率。无损特征编码模块首先经过一个 Haar 小波变换(harr wavelet transform, HWT),其中 H_0 和 H_1 分别代表低通和高通分解滤波器,分别用于从输入图像中提取近似信息和高频信息,符号 $\downarrow 2$ 表示应用于近似分量和细节分量的下采样。由图可知,Haar 小波变换会产生 4 个分量:近似分量 A ,水平(H)、垂直(V)和对角线(D)方向上的细节分量,并且,所产生的每个分量的空间分辨率大小为 $\frac{H}{2} \times \frac{W}{2}$ 。值得注意的是,每个分量的分辨率降低到 $\frac{H}{2} \times \frac{W}{2}$,而输入特征的通道数则扩大为原来的四倍,Haar 小波变换可以将空间维度的信息编码到通道维

度,因此,不会造成局部信息丢失。表示学习块由标准卷积层、批处理归一化层和 ReLU 激活层组成,其被用来提取判别特征,并且将通道映射回指定的维度,从而避免引入大量的运算成本。

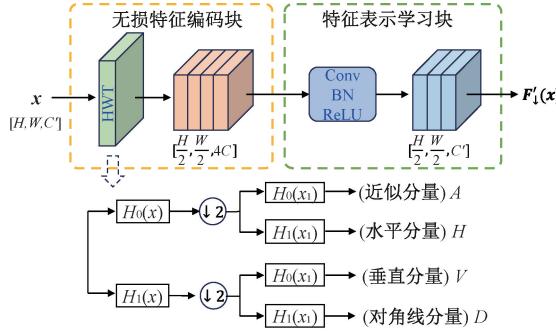


图 1 Haar 小波变换结构

Fig. 1 Haar wavelet transform structure

1.2 BNA 模块

许多先前的工作试图通过抑制网络中无关的权重来改善检测网络的检测性能。其一般是采用空间或通道池化来获取全局信息,并通过多层感知机结构来学习跨空间或跨通道的交互信息,从而学习网络中具有区分性的特征。然而,这些工作一般忽略了来自训练中调整权重信息的重要性。为了解决上述问题,本文提出了一种轻量化的批量正则化注意力模块(BNA)模块,该模块的核心思想是通过训练归一化模型权重的方差度量来突出显著特征,抑制无用特征,BNA 模块的网络结构如图 2 所示。

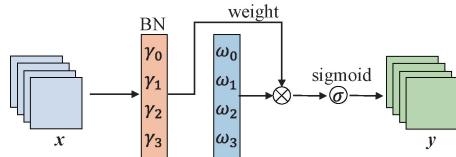


图 2 BNA 模块的结构

Fig. 2 Structure of the BNA module

首先,BNA 模块对从训练中学习到的输入特征的每个通道尺度因子 γ 通过一个正则化操作来计算归一化的权重,如式(1)所示。然后,归一化的权重与通道尺度因子进行点乘,并通过一个 sigmoid 运算来计算注意力权重,注意力权重测量通道的方差并衡量通道特征的重要性。最后,注意力权重与 BN 中的尺度因子进行相乘,来获取激活后的特征,完成的计算流程如式(2)所示。值得注意的是,BNA 模块相比于先前的注意力模块,没有采用池化来获取全局信息,而是在训练中主动去调整特征的权重,因此,其不会导致局部信息丢失。此外,BNA 模块也没有引入任何额外的参数量,因此,BNA 模块是一个部署友好型的注意力模块,其不会影响模型的推理速度。

$$\omega_i = \frac{\gamma_i}{\sum_{j=0}^c \gamma_j} \quad (1)$$

$$B_{out} = \text{sigmoid}(\omega \otimes \gamma) \frac{B_{in} - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} + \beta \quad (2)$$

其中, μ_B 为均值, σ_B 为标准差; γ 和 β 是可训练的仿射变换参数(尺度和位移)。

1.3 QARepC2f 模块

为了进一步提高 YOLOv8 的特征提取能力,本文提出将 YOLOv8 的 C2f 模块中的 Bottleneck 模块替换成 QARepBottleneck 模块,进而构建了 QARepC2f 模块,C2f 模块和 QARepC2f 模块的结构图分别如图 3(a)和图 3(b)所示。其中,图 3(b)中的 QARepC2f 模块的核心子模块为 QARepBottleneck 模块,该模块由一个 QARepCBS 模块和一个卷积核大小为 1×1 的 CBS 模块组成。

如图 3(d)所示,QARepCBS 采用了重参数化的思想,在训练阶段,采用多分支结构,即一个包含 BN 层的 3×3 卷积、一个包含偏置的 1×1 卷积和一个恒等映射层,QARepCBS 训练阶段的计算流程如式(3)所示。通过这种方式,网络在训练阶段能够利用多分支的梯度流来缓解梯度消失的问题,并且,多分支结构有利于增大模型的等效感受野,从而提高模型对于多尺度信息的捕获能力。在推理阶段,多分支将合并为一个单分支结构,整个过程不会引入额外的推理成本。

$$\mathbf{Y} = \mathbf{Y}_1 + \mathbf{Y}_2 + \mathbf{Y}_3 = BN(\mathbf{W}_{(3)} * \mathbf{X}) + (\mathbf{W}_{(1)} * \mathbf{X} + \mathbf{B}_{(1)}) + \mathbf{X} \quad (3)$$

其中, $\mathbf{W}_{(3)}$ 和 $\mathbf{W}_{(1)}$ 分别表示 3×3 卷积和 1×1 卷积的权重, $\mathbf{B}_{(1)}$ 表示 1×1 卷积的偏置,* 代表卷积。

融合后的卷积权重和偏置的计算过程分别如式(4)和(5)所示,其中, μ 和 σ 分别表示融合前 3×3 卷积分支 BN 的均值和标准差, γ 和 β 表示可训练的仿射变换参数, $\mathbf{W}'_{(1)}$ 则表示 1×1 卷积填充为 3×3 卷积的权重, $\mathbf{W}'_{(ex)}$ 表示恒等映射层的权重(全为 1),* 代表卷积。值得注意的是,与先前的 QARepConv 不同的是,QARepConv 在训练阶段对 3×3 卷积、 1×1 卷积和恒等映射的融合特征进行 BN 操作,而不是分别对 3×3 卷积和 1×1 卷积进行 BN 操作,有助于提升训练的稳定性。

$$\mathbf{W}'_{(3)} = \mathbf{W}_{(3)} * \frac{\gamma}{\sqrt{\sigma + \epsilon}} + \mathbf{W}'_{(1)} + \mathbf{W}'_{(ex)} \quad (4)$$

$$\mathbf{B}'_{(3)} = \beta - \gamma * \frac{\mu}{\sigma} + \mathbf{B}_{(1)} \quad (5)$$

1.4 NWD 损失函数

小目标指的是绝对尺度小于 32×32 的物体,由于其外观信息相对有限,因此增大了网络学习其特征的难度。先前的目标检测器一般采用基于 IoU 的边界框回归损失函数,然而,IoU 对不同尺度物体的灵敏度差异很大,对于小目标而言,微小的位置偏差会导致明显的 IoU 下降从而导致不准确的标签分配。为了解决上述问题,本文通过采用 NWD 损失函数来为模型提供小目标的高质量正负样本。

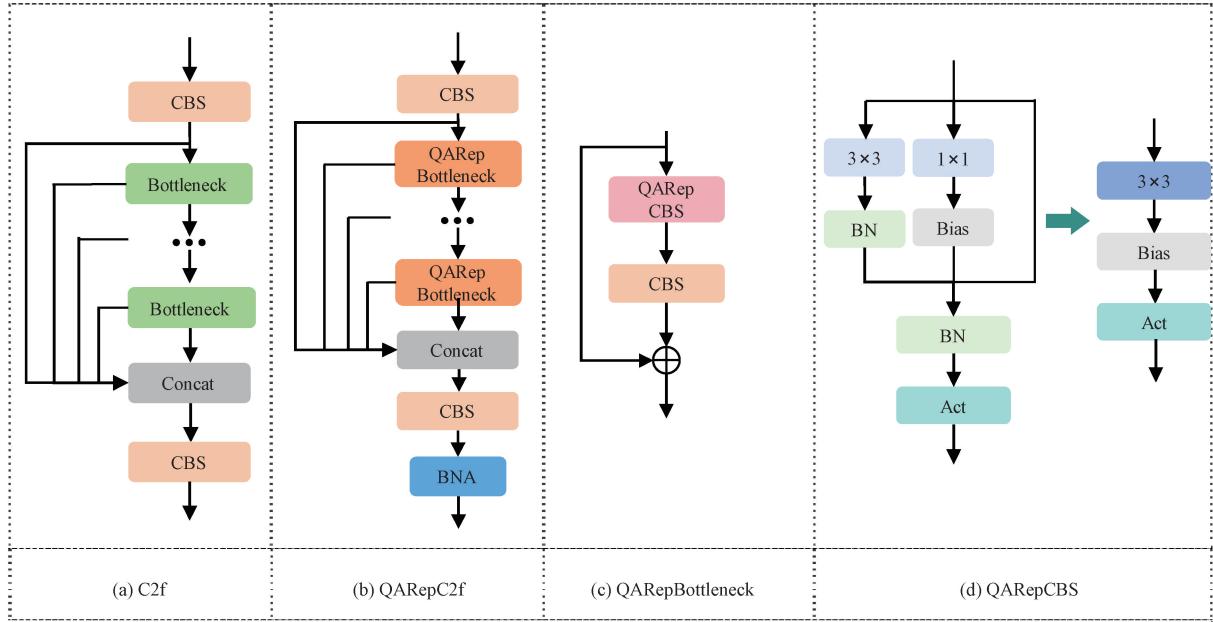


图 3 QARepC2f 模块的结构

Fig. 3 Structure of the QARepC2f module

首先, NWD 利用边界框的内接椭圆将网络预测的边界框和标签的真实框建模成两个高斯分布 N_a 和 N_b ; 然后, 引入了 Wasserstein 距离, 该距离可以用来衡量两个分布的相似程度, 并计算 N_a 和 N_b 的二阶 Wasserstein 距离, 如式(6)所示, 其中 x, y, w, h 分别表示边界框的中心点和宽高的坐标; 最后, 将二阶 Wasserstein 距离进行归一化, 并计算最终的损失, NWD 损失的计算过程如式(7)所示。总体而言, NWD 损失采用高斯分布的 Wasserstein 距离来表征预测框和真实框的距离, 这极大地缓解了 IoU 类型的损失对小目标敏感的问题, 提高了小目标的召回率。

$$W_2^2(N_a, N_b) = \left\| \left(\begin{bmatrix} x_a, y_a, \frac{w_a}{2}, \frac{h_a}{2} \end{bmatrix}^T, \begin{bmatrix} x_b, y_b, \frac{w_b}{2}, \frac{h_b}{2} \end{bmatrix}^T \right)^T \right\|_2^2 \quad (6)$$

$$L_{NWD} = 1 - \exp\left(-\frac{\sqrt{W_2^2(N_a, N_b)}}{2}\right) \quad (7)$$

为了充分地保证模型的检测性能, 本文采用了 NWD 损失和 CIoU 损失的线性组合来最小化预测边界框和真实框的位置距离, 并且, 采用 DFL 损失用于回归边界框中心点到四条边的距离, 这进一步提高了回归的精度, 因此, 完整的边界框损失函数如式(8)所示。

$$L_{bbox} = (\alpha L_{NWD} + \beta L_{CIoU}) + L_{DFL} \quad (8)$$

1.5 Slide 损失函数

先前的方法一般采用的是二元交叉熵来优化模型的分类结果, 但是其不能有效地处理正负样本分配环节中的样本不平衡问题。在目标检测任务中, 由于简单样本的数量要远远多于困难样本的数量, 因此, 简单样本的损失将主导分类损失的优化, 最终导致模型不能有效地更新困难

样本的梯度, 从而导致密集、遮挡的目标检出率差。尽管之前的一些方法在一定程度上优化了样本不平衡问题, 但其超参数较多, 针对不同的训练数据, 需要对其超参数进行特定的设置, 优化相对困难。

为了解决上述问题, 本文提出了一种 Slide 损失函数来优化样本不平衡问题。考虑到简单样本和困难样本的是基于预测框和真实框的 IOU 来区分的, 由于分类的不确定性, 位于边界的样本(即 $IoU=0.5$)往往具有较大的损失, 因此, Slide 损失的核心思想是为困难样本分配更高的损失权重, 从而让模型去着重优化此类样本, Slide 损失的函数表达式如式(9)所示。

$$f(x) = \begin{cases} 1, & x \leqslant \mu - 0.1 \\ e^{1-\mu}, & \mu - 0.1 < x < \mu \\ e^{1-x}, & x \geqslant \mu \end{cases} \quad (9)$$

其中, μ 表示所有预测框和真实框之间的平均 IoU。

当预测框与对应的真实框的 IoU 小于平均 IoU 时, Slide 损失将其作为负样本, 并且其损失具有较小的权重 1; 当 IoU 位于边界时, 表明该样本是难分样本, 具有最高的权重 $e^{1-\mu}$; 当大于平均 IoU 时, 权重呈指数衰减。这表明约靠近边界时, 权重越大, 从而促使模型更好地学习困难样本。值得注意的是, Slide 损失在计算时, 不需要设置任何超参数, 因此, 其优化过程相对简单。

1.6 QARes-YOLOv8n 网络的整体结构

本文所提出的 QARes-YOLOv8n 的整体网络结构如图 4 所示。具体而言, 首先, 将 QAResC2f 模块替换了 YOLOv8n 的 backbone 和 neck 中的 C2f 模块; 然后, 将所提出的 HWD 模块替换了 backbone 中的后 3 个下采样跨

步卷积层以及 neck 中的 3 个下采样跨步卷积层;最后,采用 CIOU、NWD、DFL 损失进行边界框回归,采用 Slide 损

失进行分类分数预测。

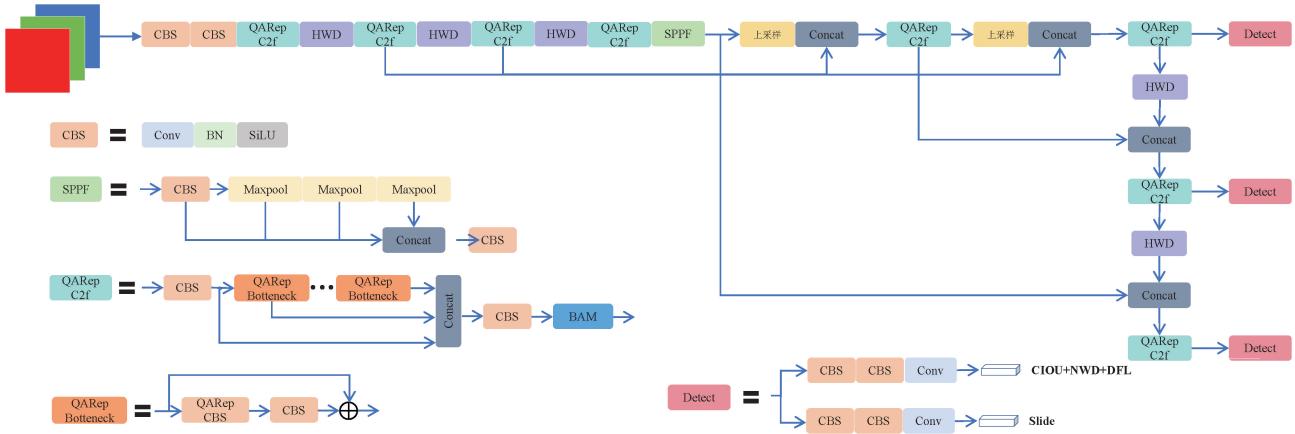


图 4 QARep-YOLOv8n 的整体网络结构

Fig. 4 Overall network structure of QARep-YOLOv8n

2 实验数据与处理

2.1 数据集

本文在 COCO^[21]、BIT-Vehicle^[22]、UA-DETRAC^[23]、VisDone2019^[24]4 个数据上进行广泛的消融实验或对比实验来验证所提出的 QARep-YOLOv8n 的有效性。

本文从 COCO 数据集的 80 个类别中选取了和城市道路场景相关的 bicycle、car、motorcycle、bus、truck 车辆类别,并从训练集和验证集中随机抽取 7 000 张图片,然后按照 8 : 2 的比例随机划分训练集和验证集。

BIT-Vehicle 数据集主要为公路视角下的车辆行驶场景,车辆主要是大像素目标,该数据总共包含 6 类:公共汽车(bus)、微型客车(microbus)、小型货车(minivan)、轿车(sedan)、SUV 和卡车(truck),总共 9 850 张样本图片,按照 8 : 1 : 1 的比例划分为训练集、测试集和验证集。

UA-DETRAC 数据集是一个极具挑战性的现实世界目标检测基准,该数据集的类别包含轿车(car)、公共汽车(bus)、货车(van)、卡车(truck)和其他车辆(others),总共 8 250 辆车和 121 万个手动标注框,训练集和验证集分别包含 8 539 和 2 231 张图片。

VisDrone2019 数据集涵盖 10 种类别,本文去除了其中和车辆无关的行人(pedestrian)和人群(people),该数据具有目标密集,小目标数量较多的特点,比较适用于验证本文算法对小目标检测的有效性。其中,训练集 6 471 张图片,验证集 548 张,测试集 1 580 张图片。

本文在上述数据集的训练集上训练网络,在验证集上评估模型的性能。

2.2 实现细节

本文在 pytorch 框架上实现了所提出的 QARep-YOLOv8n 模型。所有的实验均在两块 Tesla v100 GPU 上进行,总批量大小设为 128。与 YOLOv8 算法相同,采

用随机梯度下降算法用于优化模型,其中动量参数设置为 0.937。在学习率更新方面,本文采用了线性 warm-up 和余弦退火的学习率更新方式;在数据增强方面,额外添加一个 10% 概率的旋转数据增强来进一步提升模型的在 VisDone2019 数据集上的泛化性能。在训练策略方面,本文首先在 obj365v1 数据集上进行预训练,模型总共训练 40 轮,学习率初始化为 0.01;然后,在 COCO、BIT-Vehicle、UA-DETRAC 和 VisDone2019 数据集上进行微调,微调的轮数为 150 轮,并且将学习率初始化为 0.001,来进一步提升模型在小数据集上的表现。

2.3 评价指标

为了和先前的方法进行公平的比较,本文采用了召回率(R)和 mAP 进行模型的指标评价,召回率和 mAP 的计算公式如式(10)和(11)所示。

$$R = \frac{TP}{TP + FN} \quad (10)$$

$$mAP = \frac{1}{c} \sum_{i=1}^c AP_i \quad (11)$$

其中,TP 表示真正例,即模型预测为正例的数量,FN 表示假负例,即模型错误预测为负例的数量,召回率衡量的是模型能够正确识别的正类实例占所有实际正类实例的比例。在 mAP 的计算公式中,c 表示数据集中的类别数量,那么其衡量的则是模型预测的各类别平均精度,用于总体衡量模型的预测准确率。

此外,本研究也采用了参数量和浮点运算总数(GFLOPs)来衡量模型的复杂度,这两项指标用来表示模型整体的计算性能。

3 实验结果与分析

3.1 消融实验

为了验证所提出的方法对城市道路场景车辆检测的

优化效果,本文以 YOLOv8n 为基准进行消融实验,所有的实验均在 COCO 数据集上进行,并采用了相同的训练策略的超参数设置。

1) 不同的下采样模块对模型性能的影响

为了验证所提出的 HWD 模块的有效性,本实验将 HWD 模块与几种常见的下采样模块进行了比较,实验结果如表 1 所示。由实验结果可知,最大值池化、平均池化模块均能显著减少基准模型的参数量,但是其 mAP 指标相对较低,这表明了池化模块会丢失车辆目标的局部特征信息,从而影响模型的检测性能。本文的方法取得了 67.6% 的 mAP50 和 71.3% 的召回率,比 ADown 和 MP 模块分别高出 1.0% 和 0.6%,并且,召回率分别提升了 0.8% 和 0.4%,这进一步表明了 HWD 模块通过将空间信息编码到通道维度,有效地缓解了上述模块因池化所导致的局部空间特征信息丢失的问题。

表 1 不同的下采样模块对模型性能的影响

Table 1 Effect of different downsampling modules on model performance

| 模块 | mAP50/ % | Recall/ % | 参数量/ 10^6 | GFLOPs |
|---------------|-------------|--------------|----------------|--------|
| CBS(stride=2) | 66.1 | 70.0 | 3.0 | 8.1 |
| Max-Pooling | 65.0 | 68.5 | 2.6 | 7.4 |
| Avg-Pooling | 65.2 | 68.9 | 2.6 | 7.4 |
| ADown | 66.6 | 70.5 | 3.2 | 8.4 |
| MP | 67.0 | 70.9 | 3.2 | 8.4 |
| HWD(Ours) | 67.6 | 71.3 | 2.8 | 7.7 |

2) 不同的 C2f 模块对模型性能的影响

为了验证所提出的 QAResC2f 模块的有效性,本实验将 QAResC2f 和几种改进的 C2f 模块进行比较,包括 Ghost-C2f、GAM-C2f、ML-C2f,实验结果如表 2 所示。由实验结果可知,与 Ghost-C2f 模块相比,本文所提出的方法 mAP50 和召回率分别提升了 1.0% 和 0.6%,这表明 QAResConv 具有更强的特征提取能力。同时,与 ML-C2f 模块相比,本文所提出的方法仍实现了 0.4% 的指标提升,这表明了 QAResC2f 的多分支相比多分支 Ghost 卷积,能够更加有效地捕获多尺度的上下文信息。此外, QAResC2f 模块中的 BNA 模块有利于促进模型提取有用信息,进一步提高了模型的特征提取能力。

3) 不同的损失函数对模型性能的影响

为了验证所采用的 NWD 损失和 Slide 损失的有效性,本实验对所采用的分类损失和回归损失进行了消融研究,实验结果如表 3 所示。由表 3 可知,当分类损失为 Slide 损失时,相比于 Focal 损失和 VFL 损失,mAP 提升了 0.4%,这验证了 Slide 损失能够更加有效地处理样本不平衡问题。同时,当回归损失采用 NWD 损失和 ClIoU 损失的组

表 2 不同的 C2f 模块对模型性能的影响

Table 2 Effect of different C2f modules on model performance

| 模型 | mAP50/ % | Recall/ % | 参数量/ 10^6 | GFLOPs |
|-----------|-------------|--------------|----------------|--------|
| C2f | 66.1 | 70.0 | 3.0 | 8.1 |
| Ghost-C2f | 67.0 | 71.2 | 2.8 | 7.7 |
| GAM-C2f | 67.2 | 71.2 | 2.8 | 7.7 |
| ML-C2f | 67.6 | 71.1 | 2.9 | 8.0 |
| QAResC2f | 68.0 | 71.8 | 3.0 | 8.1 |

合时,mAP 和召回率比单独采用 ClIoU 损失分别高 0.3% 和 0.4%,NWD 采用高斯分布的 Wasserstein 距离来表征预测框和真实框的距离,缓解 IoU 损失对小目标尺度敏感的问题,提高了小目标的召回率。

表 3 不同的损失函数对模型性能的影响

Table 3 Effect of different loss functions on model performance

| 分类损失 | 回归损失 | mAP@50/ | |
|-------|---------------|-------------|-------------|
| | | % | % |
| BCE | ClIoU+DFL | 68.7 | 72.4 |
| Focal | ClIoU+DFL | 68.8 | 72.6 |
| VFL | ClIoU+DFL | 68.8 | 72.7 |
| Slide | ClIoU+DFL | 69.1 | 72.6 |
| Slide | NWD+ClIoU+DFL | 69.4 | 73.0 |

3.2 对比实验

为了进一步验证本文所提出方法的有效性和鲁棒性,本文在 UA-DETRAC 数据集将 QARes-YOLOv8n 与几种主流的 one-stage 检测算法进行了对比,实验结果见表 4 所示。由表 4 可知,QARes-YOLOv8n 相比于 YOLOv8n,mAP50 提升了 2.6%,同时,并没有额外引入其他参数量和计算成本,这充分体现了本文对 YOLOv8n 所提出的几点改进的有效性。此外,与几种其他的改进 YOLOv8n 算法,本文所提出的方法同样实现了具有竞争力的表现,这主要得益于本文对下采样模型、特征提取模块以及损失函数进行特殊设计,缓解了局部特征损失的问题,提高了模型对于小目标以及遮挡目标的检出率,使得模型能够更好地适应交通场景中的检测任务,提升模型的鲁棒性。

此外,本文还对比了 YOLOv8n 改进前后在 BIT-Vehicle 和 VisDrone2019 数据集的检测精度,实验结果如表 5 所示。由表 5 可知,改进后的 QARes-YOLOv8n 算法相比于 YOLOv8n,在 BIT-Vehicle 和 VisDrone2019 数据集上分别提升了 2.7% 和 1.5% 的 mAP50,并且,在 BIT-Vehicle 数据集上的召回率提升了 6.1%,这充分体现了改进后的算法具有更大的目标以及小目标的检测能力。

表 4 本文算法与主流算法在 UA-DETRAC 数据集上的性能对比

Table 4 Performance comparison between this paper's algorithm and mainstream algorithms on UA-DETRAC dataset

| 模型 | mAP50/% | mAP50-90/% | Recall/% | 参数量/ 10^6 | GFLOPs |
|-------------------------------|-------------|-------------|-------------|-------------|--------|
| YOLOv5s ^[3] | 55.0 | 41.0 | 56.1 | 2.7 | 14.1 |
| YOLOv6n ^[4] | 56.0 | 41.4 | 56.9 | 11.0 | 15.2 |
| YOLOXs ^[5] | 56.7 | 41.8 | 57.4 | 8.9 | 26.8 |
| YOLOv7tiny ^[6] | 57.4 | 42.1 | 57.7 | 6.0 | 13.1 |
| YOLOv8n ^[7] | 58.0 | 42.5 | 58.2 | 3.0 | 8.1 |
| OD-YOLOv8n ^[25] | 58.2 | 43.0 | 58.8 | 2.8 | 7.7 |
| GS-YOLOv8n ^[26] | 58.7 | 43.7 | 59.4 | 2.8 | 7.7 |
| Ghost-YOLOv8n ^[27] | 59.7 | 44.2 | 60.2 | 2.5 | 11.7 |
| ML-YOLOv8n ^[12] | 60.1 | 44.4 | 60.7 | 2.8 | 12.5 |
| GAM-YOLOv8n ^[15] | 60.4 | 44.6 | 61.2 | 4.1 | 12.0 |
| QARes-YOLOv8n(Ours) | 60.6 | 44.7 | 61.5 | 3.0 | 8.1 |

表 5 YOLOv8n 改进前后在 BIT-Vehicle 和 VisDrone2019 数据集上的性能对比

Table 5 Performance comparison of YOLOv8n before and after improvement on BIT-Vehicle and VisDrone2019 datasets %

| 数据集 | 模型 | mAP50 | mAP50-90 | Recall |
|--------------|---------------------|-------------------------|-------------------------|-------------------------|
| BIT-Vehicle | YOLOv8n | 58.0 | 42.5 | 54.4 |
| | QARes-YOLOv8n(Ours) | 60.7 _(+2.7%) | 44.8 _(+2.3%) | 61.5 _(+6.1%) |
| VisDrone2019 | YOLOv8n | 35.2 | 20.9 | 34.6 |
| | QARes-YOLOv8n(Ours) | 36.7 _(+1.5%) | 21.5 _(+0.6%) | 36.4 _(+1.8%) |

同时,本文比较了改进前后模型在 VisDrone 数据集上的边界框损失和分类损失的大小,实验结果如图 5 所示,其中,绿线表示 YOLOv8n 算法,红线表示本文提出的 QARes-YOLOv8n 算法。由图 5 可知,QARes-YOLOv8n 在网络训练的前期,box 损失下降得更快,这表明通过引入

NWD 损失,有助于模型对于小样本的学习,从而提升了模型在小目标数据集上的收敛速度。此外,QARes-YOLOv8n 通过引入 Slide 损失,进一步促进了模型对于困难样本的分类。

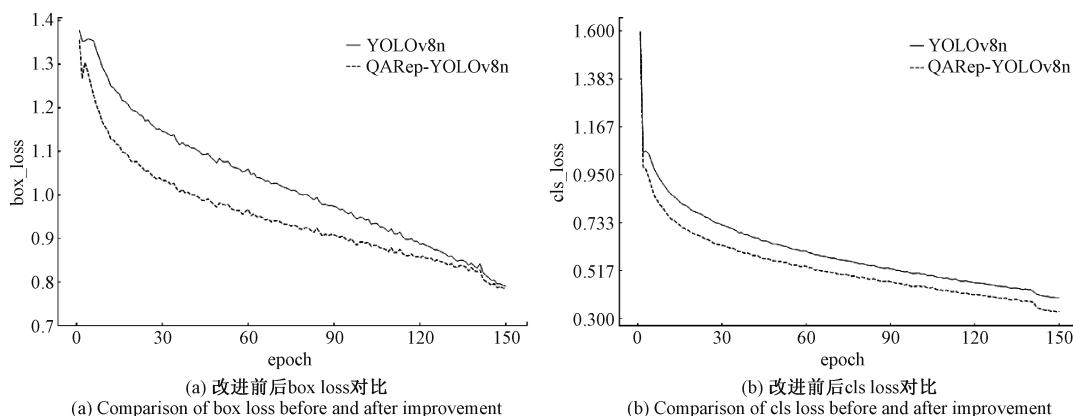


图 5 模型改进前后的损失函数对比

Fig. 5 Comparison of loss functions before and after model improvement

最后,本文对比了 YOLOv8n 改进前后在 COCO、VisDrone 数据集上的检测效果,实验结果分别如图 6 前两行和后两行所示。如图 6 所示,由第 1 行结果图可知,改进后的 QARes-YOLOv8n 算法能够检出被遮挡严重的卡车

(truck),同时也能检出远离镜头的小目标汽车(car);同时,由第 2 行和第 3 行结果图可知,YOLOv8n 将靠近镜头的汽车、水箱和铁皮建筑误检成了卡车(truck),这表明 YOLOv8n 不能有效地区分外观结构相似的目标,这进一

步体现了 QARes-YOLOv8n 出色的特征提取能力;最后,由第 4 行结果图可知,本文方法能够检测出被树木严重遮

挡的小目标车辆(右上角),这进一步体现了改进后的 QARes-YOLOv8n 强大的小目标、遮挡目标检测能力。



图 6 模型改进前后的检测结果对比

Fig. 6 Comparison of detection results before and after model improvement

4 结 论

本文基于 YOLOv8n 提出了一种用于小目标、遮挡车辆目标检测的 QARes-YOLOv8n 算法。首先,采用了 HWD 模块替换跨步下采样卷积,用于减少局部信息丢失。然后,提出了一种 QAResC2f 模块和 BNA 模块来提高 YOLOv8n 的特征提取能力。最后,采用了 NWD 损失和 Slide 损失来优化小目标检测和遮挡检测问题。广泛的消融实验和对比实验结果表明,本文所提出的算法具有更显著的检出能力。

参考文献

- [1] 吴佳桓,刘晨阳,陈天森,等.基于物联网的电动自行车充电过程监测系统[J].电子测量技术,2023,46(24):127-132.
WU J H, LIU CH Y, CHEN T M, et al. Electric bicycle charging process monitoring system based on

IOT[J]. Electronic Measurement Technology, 2023, 46(24):127-132.

[2] 梁继然,陈壮,董国军,等.结合注意力机制和密集连接网络的车辆检测方法[J].电子测量与仪器学报,2022,36(3):210-216.

LIANG J R, CHEN ZH, DONG G J, et al. Vehicle detection method combining attention mechanism and dense connection network [J]. Journal of Electronic Measurement and Instrumentation, 2022, 36 (3): 210-216.

[3] JOCHER G. YOLOv5 [EB/OL]. [2022-02-20]. <https://github.com/ultralytics/YOLOv5>.

[4] LI CH Y, LI L L, JIANG H L, et al. YOLOv6: A single-stage object detection framework for industrial applications[J]. ArXiv preprint arXiv:2209.02976, 2022.

[5] ZHANG G, LIU S T, WANG F, et al. YOLOX: Exceeding YOLO series in 2021[J]. ArXiv preprint arXiv: 2107.08430, 2021.

[6] 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]. IEEE Conference on Computer Vision and Pattern Recognition, 2023:7464-7475.
- [7] REIS D, KUPEC J, HONG J, et al. Real-time flying object detection with YOLOv8 [EB/OL]. ArXiv preprint arXiv: 2305.09972, 2023.
- [8] 刘罡, 闫曙光, 刘钰, 等. 基于感受野增强的复杂道路场景目标检测研究[J]. 电子测量技术, 2024, 47(13): 157-166.
- LIU G, YAN SH G, LIU Y, et al. Research on target detection in complex road scenes based on receptive field enhancement [J]. Electronic Measurement Technology, 2024, 47(13): 157-166.
- [9] 原蕾, 王科俊. 基于注意力机制与特征融合的改进 YOLOv7 车辆检测方法[J]. 国外电子测量技术, 2023, 42(9): 49-57.
- YUAN L, WANG K J. Vehicle detection based on YOLOv7 improved by attention mechanism and feature fusion [J]. Foreign Electronic Measurement Technology, 2023, 42(9): 49-57.
- [10] 李珣, 伍荣兴, 周慧龙, 等. 基于 YOLOv7-R 的多车辆目标识别[J]. 东南大学学报(自然科学版), 2024, 54(5): 1260-1270.
- LI X, WU R X, ZHOU H L, et al. Multi-vehicle object recognition based on YOLOv7-R [J]. Journal of Southeast University(Natural Science Edition), 2024, 54(5): 1260-1270.
- [11] 冯晶, 符丹丹, 刘强, 等. 基于 YOLOv8 的航拍车辆检测技术研究[J]. 计算机科学与应用, 2023, 13(12): 2399-2407.
- FENG J, FU D D, LIU Q, et al. Research on aerial vehicle detection technology based on YOLOv8 [J]. Computer Science and Application, 2023, 13(12): 2399-2407.
- [12] 许德刚, 王双臣, 尹柯栋, 等. 改进 YOLOv8 的城市车辆目标检测算法[J/OL]. 计算机工程, 1-8 [2024-07-26]. <https://doi.org/10.19678/j.issn.1000-3428.0069125>.
- XU D G, WANG SH CH, YIN K D, et al. Improved YOLOv8 urban vehicle target detection algorithm [J/OL]. Computer Engineering, 1-8 [2024-07-26]. <https://doi.org/10.19678/j.issn.1000-3428.0069125>.
- [13] 火久元, 苏泓瑞, 武泽宇, 等. 基于改进 YOLOv8 的道路交通小目标车辆检测算法[J]. 计算机工程, 2025, 51(1), 246-257.
- HUO J Y, SU H R, WU Z Y, et al. Road traffic small target vehicle detection based on improved YOLOv8 [J]. Computer Engineering, 2025, 51(1), 246-257.
- [14] 梁超, 来跃深, 常宏. 改进 YOLOv8 的道路小目标检测算法[J]. 西安工业大学学报, 2024, 44(4): 514-524.
- LIANG CH, LAI Y SH, CHANG H. Small road object detection algorithm based on improved YOLOv8 [J]. Journal of Xi'an Technological University, 2024, 44(4): 514-524.
- [15] 张利丰, 田莹. 改进 YOLOv8 的多尺度轻量型车辆目标检测算法[J]. 计算机工程与应用, 2024, 60(3): 129-137.
- ZHANG L F, TIAN Y. Improved YOLOv8 multi-scale
- and lightweight vehicle object detection algorithm [J]. Computer Engineering and Applications, 2024, 60(3): 129-137.
- [16] 史涛, 崔杰, 李松. 优化改进 YOLOv8 实现实时无人机车辆检测的算法[J]. 计算机工程与应用, 2024, 60(9): 79-89.
- SHI T, CUI J, LI S. Algorithm for real-time vehicle detection from UAVs based on optimizing and improving YOLOv8 [J]. Computer Engineering and Applications, 2024, 60(9): 79-89.
- [17] 高德勇, 陈泰达, 缪兰. 改进 YOLOv8n 的道路目标检测算法[J]. 计算机工程与应用, 2024, 60(16): 186-197.
- GAO D Y, CHEN T D, MIAO L. Improved road object detection algorithm for YOLOv8n [J]. Computer Engineering and Applications, 2024, 60(16): 186-197.
- [18] XU G P, LIAO W T, ZHANG X, et al. Haar wavelet downsampling: A simple but effective downsampling module for semantic segmentation [J]. Pattern Recognition, 2023, 143: 109819.
- [19] WANG J W, XU CH, YANG W, et al. A normalized gaussian wasserstein distance for tiny object detection [J]. ArXiv preprint arXiv: 2110.13389, 2021.
- [20] YU Z P, HUANG H B, CHEN W J, et al. YOLO-FaceV2: A scale and occlusion aware face detector [J]. ArXiv preprint arXiv: 2208.02019, 2022.
- [21] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: Common objects in context [C]. 13th European Conference on Computer Vision, 2014: 740-755.
- [22] DONG ZH, WU Y W, PEI M T, et al. Vehicle type classification using a semi supervised convolutional neural network [J]. IEEE Transactions on Intelligent Transportation Systems, 2015, 16(4): 2247-2256.
- [23] WEN L, DU D, CAI Z, et al. UA-DETRAC: A new benchmark and protocol for multi-object detection and tracking [J]. Computer Vision and Image Understanding, 2020, 193: 102907.
- [24] DU D W, ZHU P F, WEN L Y, et al. VisDrone-DET2019: The vision meets drone object detection in image challenge results [C]. 2019 IEEE/CVF International Conference on Computer Vision Workshop, 2019: 213-226.
- [25] LI CH, ZHOU AO J, YAO AN B. Omni-dimensional dynamic convolution [J]. ArXiv preprint arXiv: 2209.07947, 2022.
- [26] LI H L, LI J, WEI H B, et al. Slim-neck by GSConv: A better design paradigm of detector architectures for autonomous vehicles [J]. ArXiv preprint arXiv: 2206.02424, 2022.
- [27] HAN K, WANG Y H, TIAN Q, et al. GhostNet: More features from cheap operations [J]. ArXiv preprint arXiv: 1911.11907, 2019.

作者简介

杨沙沙(通信作者), 硕士, 讲师, 主要研究方向为图像处理、物联网技术、嵌入式。

E-mail: 1073606001@qq.com

徐新源, 本科, 高级工程师, 主要研究方向为嵌入式、物联网技术。