

基于 YOLO-NPDL 的复杂交通场景检测方法

张浩晨,张竹林*,史瑞岩,曹士杰,王文翰,雷镇诺

山东交通学院汽车工程学院,山东 济南 250357

摘要:为提高复杂交通场景下车辆目标检测模型的检测精度,以 YOLOv8n(you only look once version 8 nano)为基准模型,设计具有复合主干的 Neck-ARW(包括辅助检测分支、RepBlock 模块、加权跳跃特征连接)颈部结构,减少信息瓶颈造成沿网络深度方向的信息丢失;引入 RepBlock 结构重参数化模块,在训练过程中采用多分支结构提高模型特征提取性能;添加 P2 检测层捕捉更多小目标细节特征,丰富网络内小目标的特征信息流;采用 Dynamic Head 自注意力机制检测头,将尺度感知、空间感知和任务感知自注意力机制融合到统一框架中,提高检测性能;采用基于层自适应幅度的剪枝(layer-adaptive magnitude based pruning, LAMP)算法,移除模型的冗余参数,构建 YOLO-NPDL(Neck-ARW, P2, Dynamic Head, LAMP)车辆目标检测模型。以 UA-DETRAC(university at Albany detection and tracking)数据集为试验数据集,分别进行 RepBlock 模块嵌入位置试验、不同颈部结构对比试验、剪枝试验、消融试验、模型性能对比试验,验证 YOLO-NPDL 模型的平均精度均值。试验结果表明:RepBlock 模块同时嵌入辅助检测分支和颈部主干结构时对多尺度目标的特征提取能力更优,在训练过程中可保留更多的细节信息,但参数量和计算量均增大;采用 Neck-ARW 颈部结构后模型的平均精度均值 E_{mAP50} 、 $E_{mAP50-95}$ 分别提高 1.1%、1.7%,参数量减小约 17.9%,结构较优;剪枝率为 1.3 时,模型参数量、计算量分别减小约 38.0%、24.0%,冗余通道占比较少,结构较紧凑;与 YOLOv8n 模型相比,YOLO-NPDL 模型在参数量基本相同的基础上,召回率增大 2.7%, E_{mAP50} 增大 2.7%,达到 94.7%, $E_{mAP50-95}$ 增大 6.4%,达到 79.7%;与目前广泛使用的 YOLO 系列模型相比,YOLO-NPDL 模型在较少参数量的基础上,检测精度较高。YOLO-NPDL 模型在检测远端目标、雨天及夜景等实际复杂交通情景中无明显误检、漏检情况,可检测到更多的远端小目标车辆,检测效果更优。

关键词:目标检测;复杂交通场景;YOLOv8n;Neck-ARW;RepBlock;LAMP 算法

中图分类号:U495;TP183

文献标志码:A

文章编号:1672-0032(2025)02-0034-14

引用格式:张浩晨,张竹林,史瑞岩,等.基于 YOLO-NPDL 的复杂交通场景检测方法[J].山东交通学院学报,2025,33(2):34-47.

ZHANG Haochen, ZHANG Zhulin, SHI Ruiyan, et al. Complex traffic scene detection method based on YOLO-NPDL[J]. Journal of Shandong Jiaotong University, 2025, 33(2): 34-47.

0 引言

车辆和行人检测技术是计算机视觉领域的经典技术,在自动驾驶、交通流量监控、交通规划等方面具有广泛的应用价值^[1]。车辆行驶过程中,车辆目标检测模型准确、快速地检测周围车辆与行人,对辅助车辆规划路径、保障交通安全至关重要。检测设备位置相对固定,提取图像中近端车辆目标像素占比远大于远端车辆,检测模型对远端目标检测不灵敏、不精确。远端车辆目标在被检测到之前,往往已出现在检

收稿日期:2024-09-03

基金项目:山东交通学院研究生科技创新项目(2024YK001)

第一作者简介:张浩晨(2000—),男,山东临沂人,硕士研究生,主要研究方向为深度学习与目标检测,E-mail:1270602501@qq.com。

***通信作者简介:**张竹林(1979—),男,山东青岛人,副教授,硕士研究生导师,工学博士,主要研究方向为智能车辆与深度学习,E-mail:qcxzhang@126.com。

测设备视野中,并运动了一段时间,提高检测模型对小目标车辆的检测能力,有助于早期发现并精准定位远端车辆目标。

在物体检测算法中,YOLO(you only look once)模型的检测速度和检测精度具有显著的均衡性,已广泛应用于工业场景,可根据不同任务需求改进模型结构^[2]。YOLOv8(you only look once version 8)模型是目标检测领域广泛应用的模型之一,由主干网络、颈部网络、检测头组成:采用主干网络提取目标特征信息,将输入图像转换为具有语义信息的特征图,用于后续识别任务;颈部网络采用路径融合网络-特征金字塔网络(path aggregation network-feature pyramid networks,PAN-FPN)结构,有自上而下和自下而上两条特征融合路径,可进行特征提取和信息传播;检测头具有解耦结构,能将分类和检测任务分离。相比前代YOLO系列模型,YOLOv8模型的检测精度较高,但对远端小目标的检测精度仍较低,具体表现为:一是对远端车辆目标边界框定位难度大,在预测过程中边界框仅偏移几个像素点也会对小目标的定位产生较大影响;二是对远端、密集、遮挡车辆目标的漏检率较高,导致模型召回率较低。YOLOv8模型对远端车辆目标检测精度较低的内部原因为:信息传播过程中,逐层压缩特征信息,沿网络深度方向可能丢失大量信息,小目标特征信息易淹没在信息噪声中或被当作无关信息丢弃;检测模型对远端车辆目标的特征提取性能较差,远端车辆目标在检测图像中像素占比较少,模型难以提取有足够鉴别特征的目标,网络中传播的小目标检测有效特征信息较稀疏。

提高模型对远端车辆目标的检测精度是实际应用中亟待解决的重要问题。随着深度学习算法的发展,许多学者将浅层网络的强位置特征信息和高层网络的强特征语义信息进行特征融合,减少特征传播过程中的信息丢失。Lin等^[3]提出特征金字塔网络(feature pyramid networks,FPN)框架,将低分辨率强语义特征与高分辨率弱语义特征结合,提高检测模型提取特征信息的能力。Tan等^[4]提出重复加权双向特征金字塔网络(bidirectional feature pyramid network,BiFPN),构建跨层连接,引入可学习的权重,提高特征融合效率。孙庆^[5]将深度可分离卷积模块与BiFPN结合,提高特征融合效率。Xu等^[6]从网络结构、损失函数、推理速度和鲁棒性等方面改进YOLO模型,提高在实时检测任务中的检测精度。以上改进通过路径增强和改进网络结构等方式融合浅层位置特征信息与深层网络信息,降低信息传播过程中的损失,提高模型提取信息的能力。

通过结合上下文信息可提高复杂背景下模型对小目标的感知能力^[7]。深层网络中车辆目标聚集场景下,聚集区域相邻的远端车辆目标在深层特征图上聚合成一个模糊的点,检测难度较大。预测的边界框可能会因后处理的非极大值抑制而过滤大量正确预测的边界框,出现漏检情况^[8]。Feng等^[9]提出三重上下文感知网络,采用全局与局部上下文信息提高检测性能。Guan等^[10]提出语义上下文感知网络,通过融合多级信息增强对小目标的特征感知能力。以上改进能提高模型特征提取性能,并通过结合局部和全局信息的相互关系,增强模型提取特征的表征能力,在一定程度上提高了车辆目标检测模型的性能,但随着汽车保有量的增大,交通场景的复杂性增大,对检测模型的检测精度提出了更严苛的要求。

考虑到实际交通场景的应用需求,以YOLOv8n(you only look once version 8 nano)模型为基准模型,设计具有复合主干的Neck-ARW(包括辅助检测分支(auxiliary detection branch)、RepBlock模块、加权跳跃特征连接(weighted jump feature fusion))颈部结构,引入RepBlock结构重参数化模块,添加P2检测层,采用Dynamic Head动态自注意力机制检测头,采用基于层自适应幅度的剪枝(layer-adaptive magnitude based pruning,LAMP)算法,提出YOLO-NPDL模型,以期提高复杂交通场景下模型的检测精度。

1 模型改进

为提高复杂交通场景下车辆目标检测模型的检测精度,从以下5方面改进YOLOv8n模型。

1)设计具有复合主干的Neck-ARW颈部结构。针对沿网络深度方向的信息丢失,精简颈部主干网络的部分上采样路径,降低模型深度,针对精简后的颈部结构特征信息提取性能不足,设计基于三重尺度特征融合(triple scale feature fusion,TSFF)的辅助检测分支;将辅助检测分支与主干结构、颈部结构进行加

权跳跃特征连接 (weighted jump feature fusion, WJFF), 提高模型对辅助检测分支提取的多尺度特征信息的关注度。

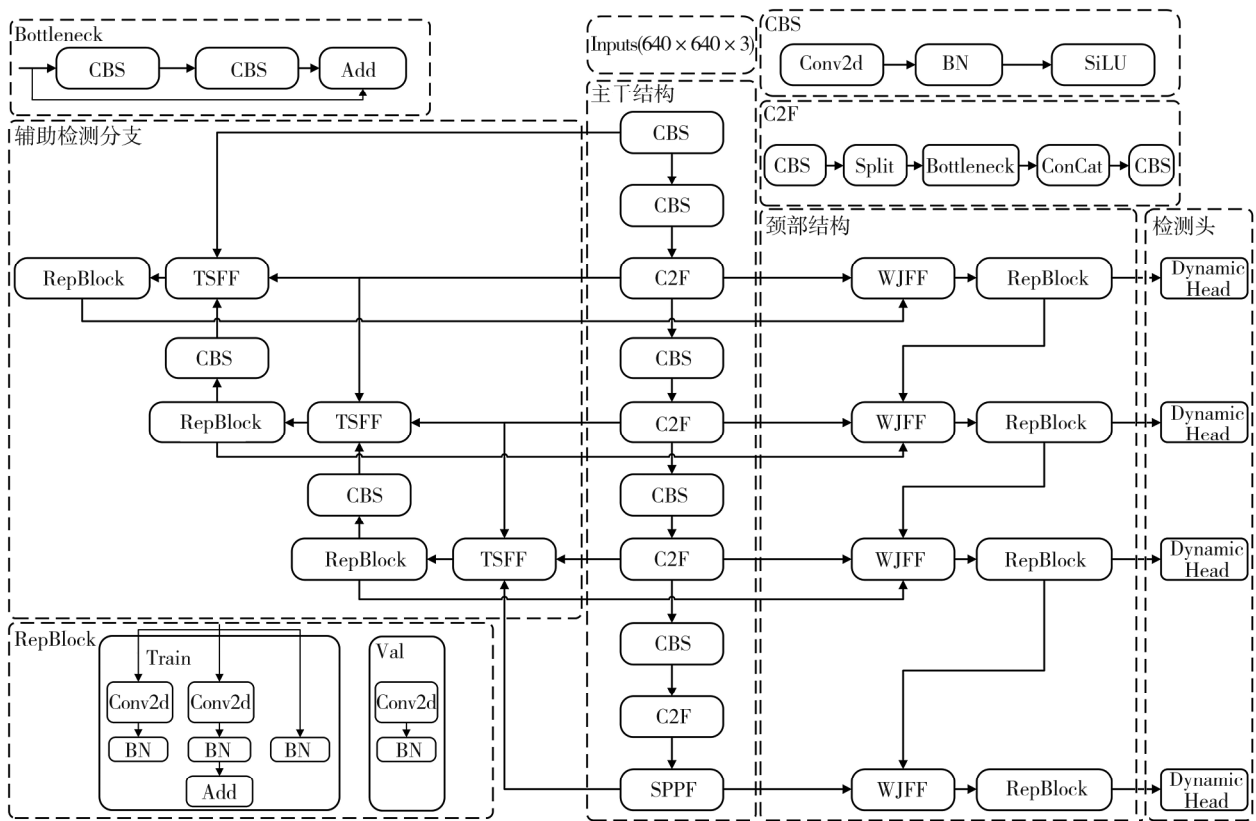
2) 引入结构重参数化模块 RepBlock, 在训练过程中采用多分支结构提高模型特征提取性能, 在推理阶段重参数化为单一的线性层, 不引入额外的推理成本, 提高推理速度。

3) 添加 P2 检测层捕捉更多小目标细节特征, 丰富网络中的小目标特征信息流。

4) 采用 Dynamic Head 动态自注意力机制检测头, 结合尺度感知、空间感知和任务感知自注意力模块, 提高检测模型对多尺度目标的感知能力, 将注意力应用于目标空间结构, 自适应地聚合多个特征层级, 学习更具辨别力的特征表达。

5) 采用 LAMP 算法, 剪除模型的冗余参数, 降低模型运算量、参数量, 减小权重文件大小, 提高模型实时检测性能。

经过以上改进提出 YOLO-NPDL 模型, 其网络结构如图 1 所示。



注: Bottleneck 为瓶颈层; CBS 为卷积层 (convolutional layer, Conv)-批归一化层 (batch normalization, BN)-激活函数 SiLU 的连续组合模块; Add 为相加操作; Train 为训练阶段模型结构; Val 为验证阶段模型结构; Inputs (640×640×3) 为输入特征图数据 (行数×列数×张数); C2F (cross stage partial fusion with 2 convolutions) 为改进的特征融合模块, 通过两个卷积操作实现跨阶段的部分特征融合; SPPF 为快速空间金字塔池化模块 (spatial pyramid pooling fast); Conv2d 为二维卷积操作; Split 为拆分操作; ConCat 为拼接操作。

图 1 YOLO-NPDL 模型网络结构

1.1 Neck-ARW 颈部结构

YOLOv8n 模型采用 PAN-FPN 颈部结构, 输入信息经逐层压缩过程中丢失大量原始特征^[11], 这种丢失信息的现象称为信息瓶颈, 导致偏倚的梯度流用于模型更新, 深度网络建立目标与输入信息间出现错误关联, 产生错误检测或定位^[12]。信息传递过程可表示为:

$$I(Y, X) \geq I(Y, f_{\theta}(X)) \geq I(Y, g_{\phi}(f_{\theta}(X))) \geq \dots \geq I(Y, \hat{X}),$$

式中: I 为互信息, X 为输入信息, Y 为期望输出信息, θ 为变换函数 f 的参数, ϕ 为变换函数 g 的参数, $\hat{X}$ 为隐藏层中期望输出信息的最小充分统计量。

根据信息瓶颈原理,神经网络训练过程中最理想的情况是学习过程中最大化地压缩输入信息 X ,最大化地保留期望输出信息 Y ,通常在训练过程中通过隐式最小化实现,如随机梯度下降^[13]。通过计算损失函数生成的梯度更新深度神经网络的参数,使网络输出与给定目标间的差异最小化。神经网络在训练过程中因信息丢失只能采用不完整信息,造成梯度不可靠、收敛困难,且产生误差积累,若浅层监督在训练过程中丢失信息,后续层将无法检索到所需信息,由于无法完整保留预测目标的信息,层级越深,所受影响越大。在深度神经网络中缓解信息瓶颈现象的主要方法有:1)采用可逆架构重复输入数据;2)采用掩码建模,通过重构损失,采用隐式最大化提取特征并保留输入信息。引入复合主干会增大网络宽度,提高模型表达能力,通过多分支结构可在训练过程中减少梯度消失^[14]、精简深度结构及提高颈部结构特征提取性能。

通过精简 Neck-ARW 颈部结构的部分上采样路径,缩短梯度传递路径,减少 PAN-FPN 颈部结构逐层信息瓶颈导致的原始特征丢失;采用辅助检测分支加颈部主干的多分支结构增大网络基础信息量,辅助检测分支与颈部主干共享来自主干网络中的特征输入。辅助检测分支输出通过 WJFF 模块与颈部结构输出连接,获取浅层的空间和细节信息,融合后兼顾底层高清位置信息与高层语义信息。Neck-ARW 颈部结构节点如图2所示。

1.1.1 TSFF 模块

多尺度特征集成是目标检测的关键部分。对主干网络输出的3个近邻层特征进行集成,结合深度语义特征,保留更精准的目标定位信息。

基于自适应空间特征融合 YOLO (adaptive spatial feature fusion YOLO, ASF-YOLO) 框架对大尺度特征图的处理方式^[15],采用自适应最大池化 (adaptive_max_pool) 和自适应平均池化 (adaptive_avg_pool) 组合模块改进 ASF-YOLO 框架的上采样模块。辅助检测分支采用 TSFF 模块进行特征融合,自适应池化操作输出尺寸无需显式地指定固定池化窗口的大小,通过指定输出特征尺寸计算相应的步长 S 和卷积核尺寸 K ,在多次、多尺度特征融合的情况下具有更强的模块泛用性,公式为:

$$\begin{cases} S = \lfloor P_a / P_a' \rfloor \\ K = P_a - S(P_a' - 1) \end{cases}$$

式中: P_a 为特征图输入尺寸, P_a' 为特征图输出尺寸。

下采样过程中采用步长 S 、卷积核尺寸 K 均为2、填充 p 为1的转置卷积模块调整小尺度特征图尺寸与通道数。转置卷积输出尺寸

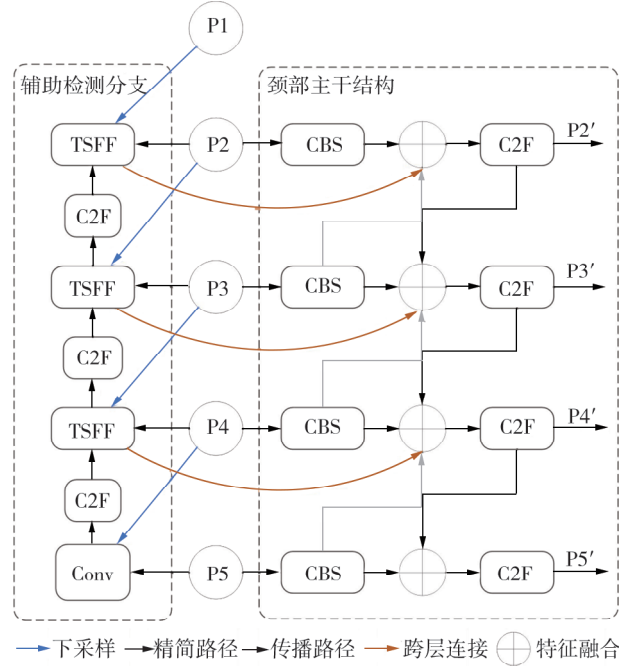
$$P_c' = S(P_c - 1) - 2p + K,$$

式中 P_c 为小尺度特征图尺寸。

假设 P_a 、 P_b 、 P_c 为3个相邻的特征图的尺寸。TSFF 模块结构如图3所示。图3中3个特征图的融合过程为:1) P_a 经卷积调整通道数后输入到自适应最大池化和自适应平均池化的组合模块中,并将输出相加,再通过卷积调整通道数;2) P_b 经卷积模块调整通道数;3) 将 P_c 输入特征通道维度调整为原来的1/2,经转置卷积模块进行上采样,使特征图尺寸为 $2S$;4) 将上述模块输出特征拼接,得到特征图。

1.1.2 WJFF 模块

通过跳跃连接可连接不同层级的网络单元,其先后应用于残差网络 (residual neural network,



注: $P_1 \sim P_5$ 为主干结构输出特征; $P_2' \sim P_5'$ 为颈部结构输出特征。

图2 Neck-ARW 颈部结构节点

ResNet)^[16]、密集连接卷积网络(densely connected convolutional networks, DenseNet)^[17]等。通过重复引入浅层信息,跳跃连接可缓解信息丢失造成的精度降低。不同的特征对网络学习的贡献程度不同,部分特征的特征能力相对较弱^[18],小目标特征作为一种弱特征,在信息传播过程中易被当作背景或噪声丢弃。

设计 WJFF. A、WJFF. B 两种 WJFF 模块连接辅助检测分支、模型主干结构和颈部结构,通过权重调整和动态特征选择,优化融合结果,确保重要特征得到有效利用。

WJFF. A 模块连接模型主干结构浅层信息和辅助检测分支信息,有 2 个输入接口,通过归一化模块后的初始权重为 0.5;WJFF. B 模块连接主干结构浅层信息、辅助检测分支信息和颈部主干结构中的特征信息,有 3 个输入接口,各结构输入特征的初始权重相等。

WJFF 模块结构如图 4 所示。

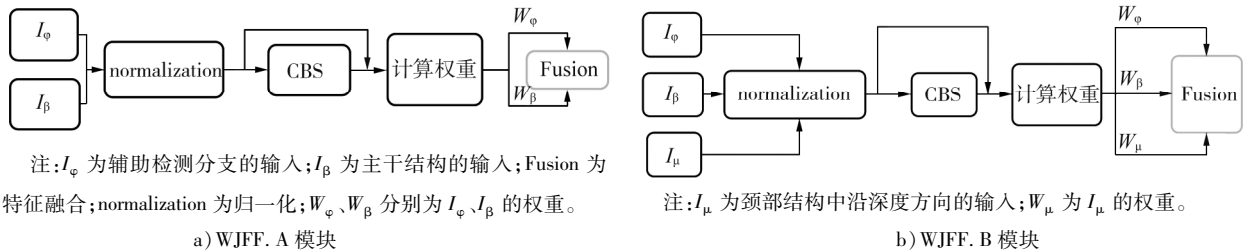


图 4 WJFF 模块结构

通过调整输入的权重确定输入特征的重要性,输出特征

$$O = \sum_i \frac{W_i I_i}{\varepsilon + \sum_j W_j},$$

式中: ε 为避免权重不稳定的参数, $\varepsilon = 0.0001$; W_i 、 W_j 分别为第 i 、 j 个输入特征的权重; I_i 为第 i 个输入特征。

Neck-ARW 颈部结构的 P3 层训练过程加权融合模块输入特征权重的可视化结果如图 5 所示。由图 5 可知: WJFF 模块提高了辅助检测分支输入权重,降低了颈部主干结构深层信息权重,可缓解深层信息丢失造成的检测精度降低。

1.2 RepBlock 模块

采用多分支结构替代颈部结构中的 C2F 模块,提高网络对多尺度特征的提取能力。多分支结构作为网络结构的隐式集成,能减少训练过程中梯度消失带来的影响,引入 RepBlock 结构重参数化模块^[19],从检测精度与推理速度两方面优化模型,多分支结构在训练阶段与推理阶段的结构不同,多分支并行结构在训练过程中可得到更优质的权重参数,捕获多尺度范围的目标特征,丰富原模型难以感知的细节特征。训练后重参数化成为单一的线性层,不引入额外的推理成本。

训练过程中,RepBlock 模块保持多分支并行结构,分支 1 采用 1×1 卷积与批归一化层连接,分支 2 采用 3×3 卷积与批归一化层连接,分支 3 采用批归一化层连接。在推理过程中,RepBlock 模块重参数化为单

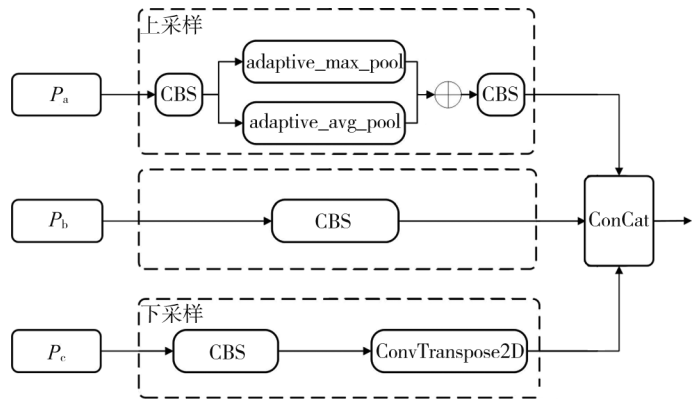


图 3 TSFF 模块结构

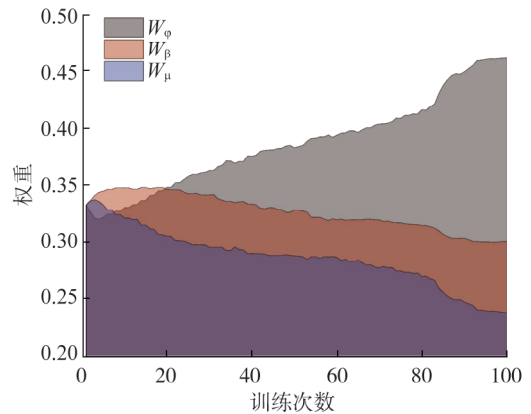


图 5 输入特征权重可视化结果

一的线性层,分支1中 1×1 卷积层填充为 3×3 卷积模块,然后与批归一化层融合,分支2中的 3×3 卷积层与批归一化层融合,分支3中批归一化层分支转化成 1×1 卷积层,再通过填充转化为 3×3 卷积层。最后将多分支转换为单一 3×3 卷积层。

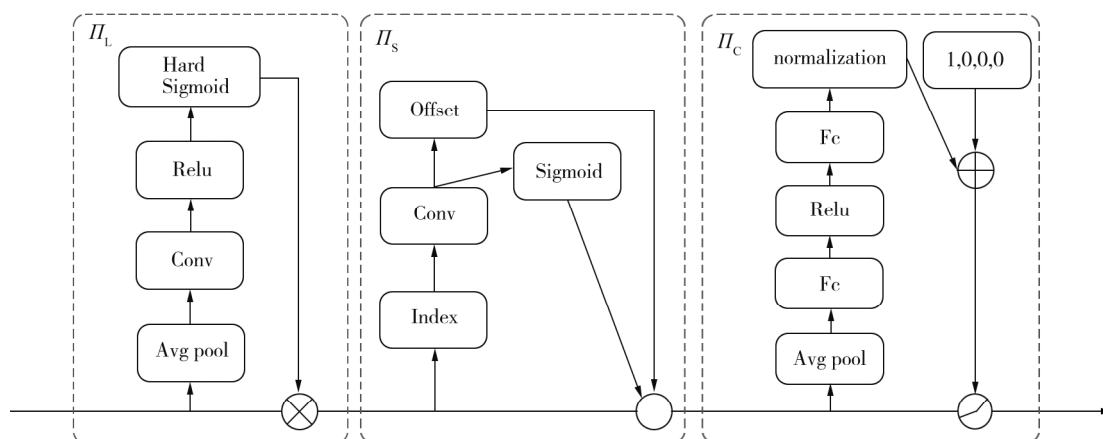
1.3 P2 检测层

通常需更多的特征信息检测小目标。YOLOv8n 模型采用比原图大的下采样倍率,在信息提取过程中易丢失对小目标检测极为重要的纹理和边缘等细节特征信息,检测小目标处于有效特征信息匮乏的状态。

P2 检测层提取到的特征图为 160 像素 $\times$ 160 像素,包含对小目标定位精度重要的位置和空间特征信息。实际驾驶场景中包含大量小目标车辆、车辆遮挡场景、密集车辆场景,这些场景中存在大量尺寸较小且密集的车辆样本。在原网络中增加一个具有更多原始图像的细节信息的检测层,丰富网络中的小目标特征信息流。

1.4 Dynamic Head 检测模块

YOLOv8n 模型检测头的参数量约占总参数量的 1/5,此模型仅从单独的特征图尺度上进行预测,对多尺度目标的检测能力有限,缺乏对检测目标空间结构的感知能力。Dynamic Head 检测模块将尺度感知、空间感知和任务感知融合到统一的框架中,在特征的每个特定维度上部署不同的自注意力机制,能提高模型检测性能和效率^[20],其结构如图 6 所示。



注: Π_L 为任务感知注意力模块; Π_S 为空间感知注意力模块; Π_C 为尺度感知注意力模块; Hard Sigmoid、Relu 为激活函数; Avg pool 为平均池化; Offset 为偏移算子; Index 为输入特征索引; Sigmoid 为激活函数; Fc (fully connected layers) 为全连接层。

图 6 Dynamic Head 检测模块结构

通过 3 种自注意力机制序列排布,构建嵌套形的模块,检测模块的输出特征

$$W(F) = \Pi_C(\Pi_S(\Pi_L(F)F)F)F,$$

式中 F 为检测模块的输入特征。

1.5 模型剪枝

提高模型检测精度会增大参数量和计算量,采用神经网络模型剪枝算法修剪网络中对提高模型性能无帮助的冗余参数,减少模型对算力资源的消耗。

在神经网络中,每个连接都有权重,权重决定输入信号的重要性^[21]。在剪枝过程中,根据连接的评分决定是否保留该通道结构,剪除评分低的连接^[22]。Han 等^[23]提出基于权重剪枝的方法,剪除权重小于某阈值的连接后,微调直至达到预期准确率。Gale 等^[24]评估基于幅值的剪枝 (magnitude-based pruning, MP) 算法,发现该算法的剪枝性能比权重表征方法好。Lee 等^[25]在 MP 算法的基础上提出 LAMP 算法,能避免过度剪枝造成的层崩溃,模型准确率、自适应稀疏分配、训练过程的稳定性等性能优

于 MP 算法。

本文采用 LAMP 算法对 YOLO-NPDL 模型进行剪枝。权重评分是对连接权重的表征方法,以向量表示连接权重, u 、 v 为连接权重的索引, $W(u)$ 、 $W(v)$ 分别为索引 u 、 v 映射权重的幅值,以索引为序将权重幅值按大小排列,即 $u < v$ 时, $W(u) < W(v)$ 。LAMP 剪枝算法中权重评分

$$s(u; W) = \frac{[W(u)]^2}{\sum_{v>u} [W(v)]^2}。 \quad (1)$$

将所有的权重幅值按从大到小的顺序排序,式(1)中分母表示从当前目标连接权重索引 u 开始该层中所有更大的连接权重幅值的平方和,索引不大于 u 的权重项已被剪除。设置目标修剪比例后,LAMP 算法从最小权重幅值处开始剪枝,直到模型达到设定的目标剪枝比例,且每一层至少有一个评分为 1 的通道,避免发生层崩溃。

2 数据集与评价指标

2.1 数据集介绍

本文采用 UA-DETRAC(university at Albany detection and tracking)数据集^[26]作为试验数据集,UA-DETRAC 是检测难度较大的多目标、多场景交通视频数据集,包含大量小目标车辆、车辆遮挡场景、密集车辆场景,且有单张图片中超过 30 辆车辆的场景,有夜晚、雨天等不同天气的图片,包括十字路口、三岔路口等道路场景,能体现现实中的复杂交通情景^[27]。

UA-DETRAC 数据集图片来自视频,相邻帧间图片相似度较高,训练价值较低。对视频采取每隔 20 帧取 1 张图片的方式,共截取 5 435 张图片,将图片分辨率转换为 640 像素×640 像素,将图片数量按 7:1:2 划分为训练集、测试集、验证集。删除与车辆类别无关的标注信息,将重复的类别标签统一合并。

2.2 评价指标

采用准确率 P 、召回率 R 、平均精度均值(mean average precision, mAP) E_{mAP} 作为模型检测性能的评价指标。采用浮点运算次数(floating point operations, FLOPs) n_f 、参数量 n_p 、权重文件大小 s_w 作为模型复杂度的评价指标^[28]。

平均精度(average precision, AP)反映模型对单个类别检测结果的准确性,mAP 表示对多个类别的平均精度求平均值,具体使用中多通过交并比(intersection over union, IoU)阈值固定为 0.50 时的平均精度均值(mean average precision at IoU threshold 0.50, mAP50) $E_{\text{mAP}50}$ 及 IoU 阈值从 0.50 增至 0.95(步长为 0.05)时的平均精度均值(mean average precision at IoU thresholds from 0.50 to 0.95, mAP50-95) $E_{\text{mAP}50-95}$ 评估模型整体性能, $E_{\text{mAP}50}$ 和 $E_{\text{mAP}50-95}$ 越大,模型的整体预测精度越高。

3 试验结果与分析

在 Windows10 系统上进行本文所有试验,CPU 为 Intel(R) Core(TM) i7-10700, GPU 为 NVIDIA GeForce RTX 3070,显存为 8 GB,深度学习框架为 Pytorch1.7.1。参数环境为: CUDA 版本为 12.4,编译语言为 Python-3.8.16,训练批尺寸为 16;迭代次数为 100;输入图片统一裁剪为 640 像素×640 像素;采用 AdamW 优化器,初始学习率为 0.002,动量为 0.9。在剪枝训练中,采用的 torch-pruning 版本为 1.3.7,迭代次数为 250。

3.1 RepBlock 模块嵌入位置试验

分析 RepBlock 模块嵌入位置对颈部结构的影响,设计 4 组对比试验确定 RepBlock 模块的最佳嵌入位置。Rep0 为无 RepBlock 模块嵌入的模型,Rep1、Rep2、Rep3 分别为 RepBlock 模块嵌入辅助检测分支、嵌入颈部主干结构、同时嵌入辅助检测分支和颈部主干结构的模型,试验结果如表 1 所示。

由表 1 可知:检测模型嵌入 RepBlock 模块后, n_p 、 n_f 均增大;相较于采用 C2F 模块的 Rep0 模型,Rep1 模型倾向于将样本标记为正例,以确保捕捉到更多的真正例,同时增加了假正例,导致 P 减小, R 增大;Rep2 模型则相反,Rep1、Rep2 模型的 $E_{mAP50-95}$ 分别增大 0.4%、0.6%;相较于 Rep0 模型,Rep3 模型的 P 、 R 、 $E_{mAP50-95}$ 分

别增大 0.1%、0.5%、0.6%,表明 RepBlock 模块同时嵌入辅助检测分支和颈部主干结构的检测模型对多尺度目标的特征提取能力更优,在训练过程中可保留更多的细节信息。因此,YOLO-NPDL 模型将 RepBlock 模块同时嵌入辅助检测分支和颈部主干结构。

3.2 不同颈部结构对比试验

为验证 Neck-ARW 颈部结构的有效性,选取不同颈部结构的模型进行性能对比试验,结果见表 2。

表 2 不同颈部结构的模型性能对比试验结果

颈部结构	$E_{mAP50}/\%$	$E_{mAP50-95}/\%$	n_p	$n_f/10^9$	s_w/MB
PAN-FPN ^①	92.0	73.3	3 005 843	8.1	6.2
GFPN ^②	92.0	73.3	3 290 659	8.5	6.8
GD Neck ^③	92.9	75.1	5 976 019	12.0	12.4
RepBiPAN ^④	92.8	74.5	3 732 387	12.3	7.8
BiFPN ^⑤	92.2	73.8	2 788 207	8.2	5.8
Neck-ARW	93.1	75.0	2 467 674	8.8	5.2

①Ultralytics 公司发布的 YOLOv8n 模型的颈部结构。

②阿里巴巴达摩院发布的 GIRAFFEDET 模型的颈部结构 GFPN(giraffe feature pyramid network)^[29]。

③华为诺亚方舟实验室提出的 Gold-YOLO 模型的颈部结构^[30]。

④美团基于硬件感知神经网络思想设计的特征融合结构^[31]。

⑤谷歌团队提出的经典加权双向特征金字塔网络^[4]。

由表 2 可知:与采用 PAN-FPN 颈部结构的 YOLOv8n 模型相比,GFPN 颈部结构的适配度较低,模型检测精度持平的情况下 n_p 和 n_f 均增大;与 YOLOv8n 模型相比,Gold-YOLO 模型的 n_p 、 n_f 分别增大 98%、48%,计算负担较重,难以满足实时检测需求;与 YOLOv8n 模型相比,采用 RepBiPAN 颈部结构时,模型的 E_{mAP50} 、 $E_{mAP50-95}$ 分别增大 0.8%、1.2%,但增大幅度不及 Gold-YOLO 模型;与 YOLOv8n 模型相比,采用 BiFPN 颈部结构时,模型具有轻量化优势, n_p 减小约 7.0%, s_w 减小约 16.0%, E_{mAP50} 、 $E_{mAP50-95}$ 分别增大 0.2%、0.5%,检测精度增大有限,在对比试验中缺乏竞争优势;采用 Neck-ARW 颈部结构的 YOLO-NPDL 模型的检测精度与 Gold-YOLO 模型相近,但 n_p 、 n_f 仅有 Gold-YOLO 模型的 41%、73%,在参数利用率和计算量方面具有明显的优势;与 YOLOv8n 模型相比,YOLO-NPDL 模型的 E_{mAP50} 、 $E_{mAP50-95}$ 分别增大 1.1%、1.7%, n_p 减小约 17.9%,具有更好的检测性能,结构较优。

不同颈部结构的模型训练输出特征图可视化结果如图 7 所示。由图 7 可知:采用 PAN-FPN 颈部结构时,YOLOv8n 模型的输出特征图丢失大部分目标位置信息;采用 GFPN 颈部结构时,模型输出特征图中检测目标和背景的位置产生混淆;采用 GD Neck 颈部结构的 Gold-YOLO 模型对图像色彩变化较大区域敏感,缺乏对于检测目标的特征理解;采用 RepBiPAN 颈部结构时,输出特征忽略图像右侧部分目标的位置信息;采用 BiFPN 颈部结构时,输出特征对左侧检测背景混淆;采用 Neck-ARW 颈部结构时,输出特征

更大比例地保留目标位置信息,且输出信息准确率较高。

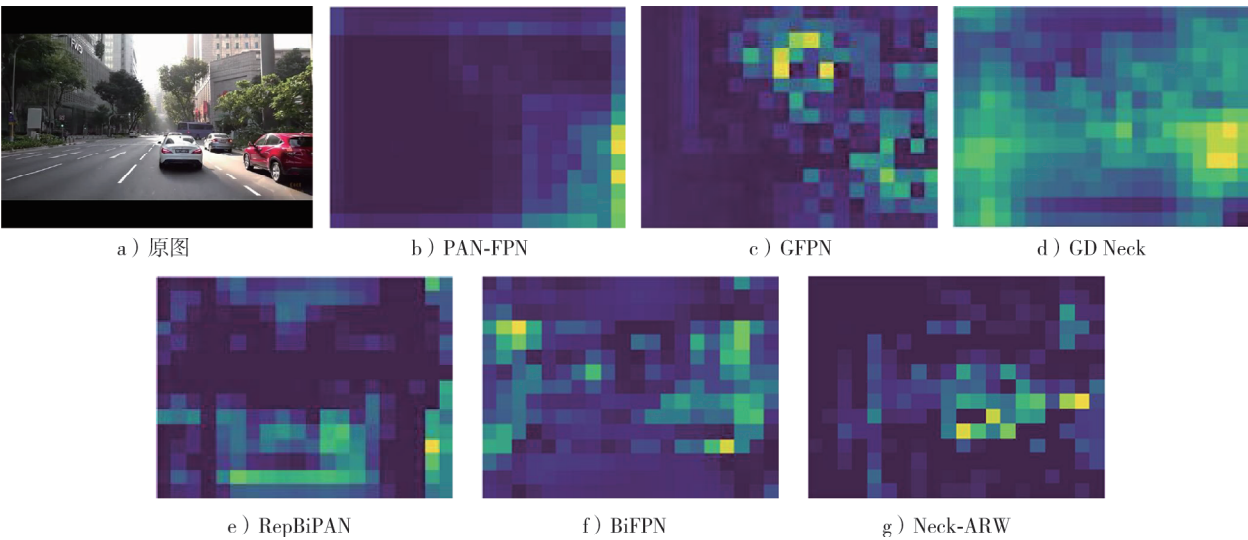


图 7 不同颈部结构的模型输出特征图可视化结果

3.3 剪枝试验

采用 LAMP 剪枝算法对 YOLO-NPDL 模型进行轻量化处理,不同剪枝率下模型性能对比如表 3 所示。由表 3 可知:剪枝率从 1.1 增至 1.3 时,剪枝算法剪除了模型的冗余结构,产生冗余结构的原因可能是同一层的卷积核学习到较多相似特征,产生参数和计算冗余,删除低评分的连接不会对检测精度有明显影响;当剪枝率增至 1.5 时,模型的检测精度下降较大,说明模型剩余连接对检测精度较重要;剪枝率为 2.0 时,模型的 n_p 和 n_f 与剪枝率为 1.7 时相差不大,说明模型在 LAMP 算法下的最大剪枝率为 1.7,大幅提高剪枝率会对模型造成严重的性能损伤。

表 3 不同剪枝率下 YOLO-NPDL 模型性能对比

剪枝率	$E_{mAP50}/\%$	$E_{mAP50-95}/\%$	n_p	$n_f/10^9$	s_w/MB
1.1	95.4	81.2	3 936 766	42.8	8.9
1.3	94.7	79.7	3 013 595	36.9	7.1
1.5	93.2	76.6	2 330 246	31.5	5.7
1.7	91.7	73.5	2 278 623	31.1	5.6
2.0	91.8	73.7	2 278 623	31.1	5.6

剪枝率为 1.3 时,模型的 n_p 、 n_f 比剪枝前分别减小约 38.0%、24.0%,检测精度降低处于可接受的范围内,本文选择剪枝率为 1.3,此时剪枝前、后模型通道数如图 8 所示。由图 8 可知:冗余通道主要分布在 YOLOv8n 主干结构,Neck-ARW 颈部结构冗余通道占比较少,整体结构较紧凑。

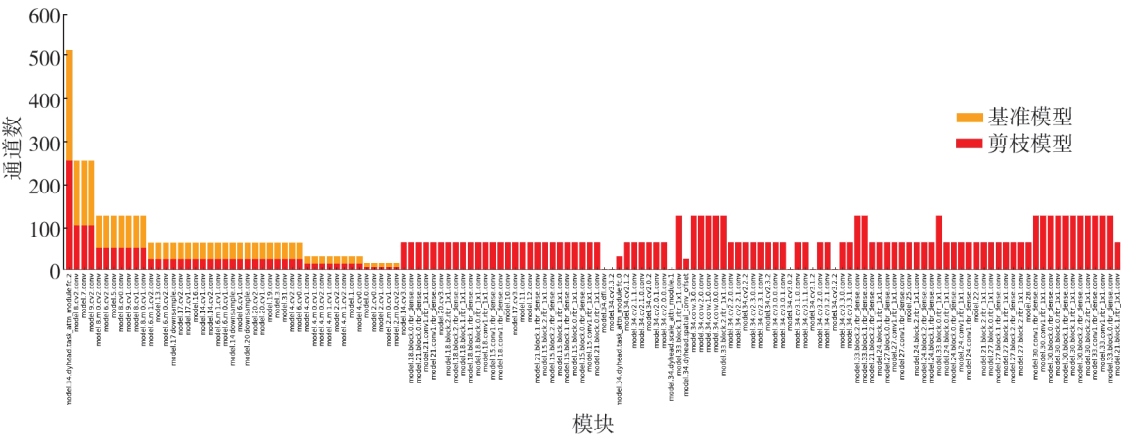


图 8 剪枝前、后通道数对比

3.4 消融试验

为验证YOLO-NPDL模型的改进效果及检测精度,共进行A~F 6组消融试验,试验结果如表4所示。

表4 消融试验结果

试验	Neck-ARW	Repblock	P2	Dynamic Head	LAMP	$P/\%$	$R/\%$	$E_{mAP50}/\%$	$E_{mAP50-95}/\%$	n_p	$n_f/10^9$	s_w/MB
A	×	×	×	×	×	91.2	83.2	92.0	73.3	3 005 843	8.1	6.2
B	✓	×	×	×	×	91.4	84.4	93.1	75.0	2 467 674	8.8	5.2
C	✓	✓	×	×	×	91.4	84.8	93.2	75.8	3 662 171	14.1	8.0
D	✓	✓	✓	×	×	91.4	86.3	94.8	79.4	4 194 015	41.3	9.4
E	✓	✓	✓	✓	×	91.8	86.7	95.1	80.7	4 886 555	48.7	10.8
F	✓	✓	✓	✓	✓	91.9	85.9	94.7	79.7	3 013 595	36.9	7.1

注:✓为采用该结构或算法,×为未采用该结构或算法。

由表4可知:相较于试验A,采用Neck-ARW颈部结构的试验B模型检测精度略提高, n_p 减小17.9%, s_w 减小1.0 MB,Neck-ARW颈部结构优化了信息传播路径,训练过程中信息损失更少且参数利用率更高;相较于试验B,在颈部结构引入RepBlock模块的试验C的 $E_{mAP50-95}$ 增大0.8%,模型检测性能提高;相较于试验C,引入P2检测头的试验D的 $E_{mAP50-95}$ 增大3.6%,表明模型能提取到更多小目标的有效信息,P2检测头丰富了网络中小目标的特征信息流;相较于试验D,引用Dynamic Head模块的试验E的 R 、 E_{mAP50} 、 $E_{mAP50-95}$ 分别增大3.5%、3.1%、7.4%,但 n_p 、 n_f 分别增大16.5%、17.9%,计算负担增幅明显;试验F对模型进行轻量化剪枝,剪枝后模型 n_p 、 n_f 分别减小38.3%、24.2%,轻量化效果显著。

对比YOLOv8n模型,YOLO-NPDL模型在参数量基本相同的基础上, R 、 E_{mAP50} 、 $E_{mAP50-95}$ 分别增大2.7%、2.7%、6.4%。

3.5 模型性能对比试验

为验证YOLO-NPDL模型改进有效性,与目前广泛使用的YOLOv5、YOLOv8及YOLOv9模型进行性能对比,试验结果如表5所示。

表5 不同模型的性能对比试验

模型	$P/\%$	$R/\%$	$E_{mAP50}/\%$	$E_{mAP50-95}/\%$	n_p	$n_f/10^9$	s_w/MB
YOLOv9t	90.8	81.4	91.2	71.8	2 616 950	10.7	6.1
YOLOv9s	90.6	85.3	93.2	76.0	9 743 366	39.6	19.3
YOLOv8n	91.2	83.2	92.0	73.3	3 005 843	8.1	6.2
YOLOv8s	91.4	85.8	93.7	77.2	11 125 971	28.4	22.5
YOLOv5n	91.9	80.1	90.4	66.1	1 760 518	4.1	3.9
YOLOv5s	92.9	85.5	93.9	75.4	7 012 822	15.8	14.4
YOLO-NPDL	91.9	85.9	94.7	79.7	3 013 595	36.9	7.1

由表5可知:与轻量化模型相比,YOLO-NPDL模型的 E_{mAP50} 、 $E_{mAP50-95}$ 比YOLOv5n模型分别增大4.3%、13.6%,比YOLOv9t模型分别增大3.5%、7.9%;与同参数量模型相比,YOLO-NPDL模型的 E_{mAP50} 、 $E_{mAP50-95}$ 比YOLOv8n模型分别增大2.7%、3.7%;与更高参数量的模型相比,YOLO-NPDL模型的 n_p 约为YOLOv5s模型的42.0%,约为YOLOv8s模型的27.0%,但 $E_{mAP50-95}$ 比后二者分别增大4.3%、2.5%。YOLO-NPDL模型在更小参数量的基础上,检测精度较高。

3.6 实际检测效果对比

选取实际交通场景中远端目标、雨天及夜间场景对比YOLOv5s、YOLOv9s、YOLOv8s、YOLO-NPDL模型的检测效果,检测结果如图9~11所示。



图 9 不同模型对远端目标的检测对比

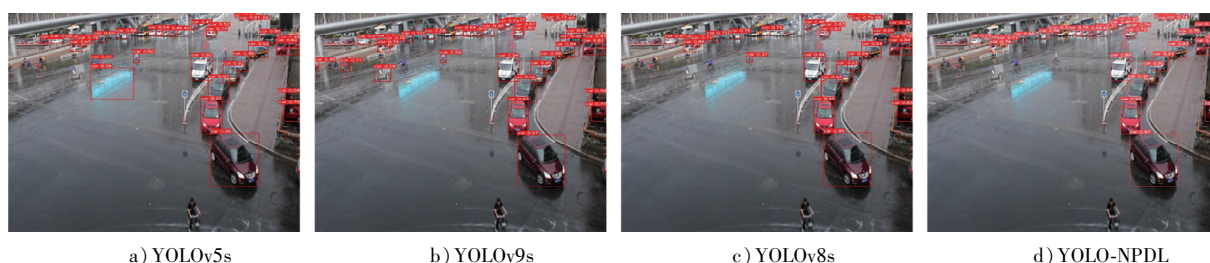


图 10 雨天场景下不同模型的检测对比



图 11 夜间场景下不同模型对的检测对比

由图 9~11 可知:YOLOv5s、YOLOv9s 模型对远端目标均存在不同程度的漏检,且 YOLOv5s 模型存在明显误检,将交通围栏误检为车辆目标,YOLOv8s、YOLO-NPDL 模型在检测远端目标方面更有优势;雨天场景下,受路面积水折射、图像噪声等因素干扰,模型区分检测前景与背景的难度较大,YOLOv5s、YOLOv9s、YOLOv8s 模型均存在将行人归类为车辆目标的误检,且 YOLOv5s 模型将路面积水折射的灯光误检为车辆目标,YOLO-NPDL 模型无明显误检;夜间场景下,受光照条件、对比度变化等因素影响,模型对形状边缘模糊目标的分类难度较大,YOLOv5s、YOLOv9s、YOLOv8s 模型均将路边报亭顶棚识别为车辆目标,对路侧目标存在不同程度的漏检,YOLO-NPDL 模型无明显误检、漏检情况。

YOLOv8、YOLO-NPDL 模型检测结果热力图如图 12 所示。由图 12 可知:YOLOv8 模型对目标边缘检测能力较弱,注意力并不分布于模型的几何边缘,仅关注目标的明显特征,无法有效区分目标和背景;YOLO-NPDL 模型注意力分布更精细,可将注意力分布在检测目标的形状边缘,检测目标的热力分布界限与背景的对比更清晰,说明 YOLO-NPDL 模型具有更好的空间结构感知能力;且 YOLO-NPDL 模型在识别距离方面更有优势,可检测到更多的远端目标(黄色椭圆形虚线框内)。

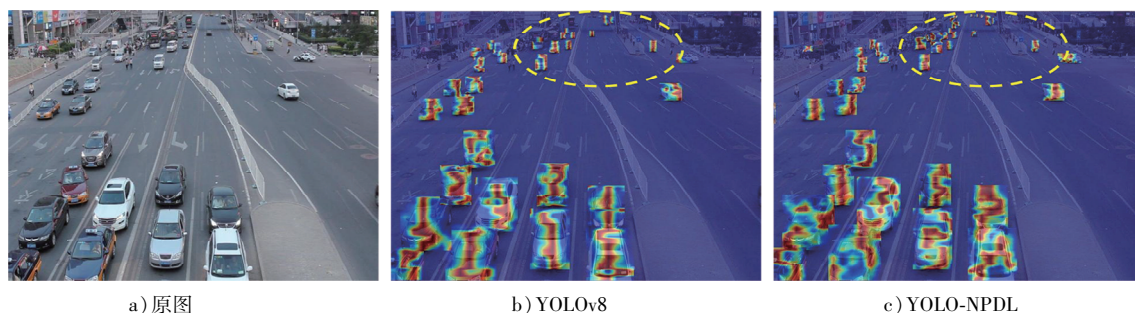


图 12 YOLOv8、YOLO-NPDL 模型检测结果热力图

4 结束语

为提高复杂交通场景下模型对车辆目标的检测精度及识别距离,基于 YOLOv8n 模型,提出改进的目标检测模型 YOLO-NPDL。具体改进包括以下方面:1)基于多分支拓扑结构模型隐式集成可在训练过程中减少梯度消失,设计具有复合主干的 Neck-ARW 颈部结构,通过复合颈部结构精简颈部网络深度,减少信息瓶颈带来的特征信息损失;设计三重尺度特征融合模块,提高模型的特征信息提取能力;通过加权跳跃连接将辅助检测分支、主干结构浅层特征与颈部主干提取特征进行融合,增强对小目标检测重要的浅层特征和辅助检测分支提取的多尺度特征在网络中的表征能力。2)引入 RepBlock 模块提高模型的特征提取性能,并探究其嵌入位置对网络性能的影响。3)增加 P2 小目标检测层,丰富网络中的小尺寸目标的特征信息流。4)采用将尺度感知、空间感知和任务感知融合至统一框架中的动态检测头,全面提高网络的检测性能。5)采用基于层自适应幅度的剪枝算法对模型进行轻量化剪枝,剪枝后模型参数量、计算量分别减小约 38.0%、24.0%,提高了模型的实时检测性能。模型对比试验结果表明:YOLO-NPDL 模型的 $E_{\text{mAP}50}$ 及 $E_{\text{mAP}50-95}$ 分别为 94.7%、79.7%,相较于 YOLOv8n 模型,在几乎同参数量的情况下分别增大 2.7%、6.4%;在复杂交通情景下,YOLO-NPDL 模型无明显误检、漏检情况,可检测到更多的远端目标,检测性能更稳定。

参考文献:

- [1] 李伟东,黄振柱,何精武,等. 改进行为克隆与 DDPG 的无人驾驶决策模型[J]. 计算机工程与应用, 2024, 60(14): 86-95.
- [2] TERVEN J, CORDOVA-ESPARZA D M, ROMERO-GONZÁLEZ J A. A comprehensive review of YOLO architectures in computer vision: from YOLOv1 to YOLOv8 and YOLO-NAS[J]. Machine Learning and Knowledge Extraction, 2023, 5(4): 1680-1716.
- [3] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE, 2017: 936-944.
- [4] TAN M X, PANG R M, LE Q V. EfficientDet: scalable and efficient object detection[C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA: IEEE, 2020: 10778-10787.
- [5] 孙庆. 基于 Transformer 和 BiFPN 的轻量化车辆检测算法研究[D]. 西安: 长安大学, 2023.
- [6] XU X Z, JIANG Y Q, CHEN W H, et al. DAMO-YOLO: a report on real-time object detection design[EB/OL]. (2023-04-24) [2024-06-15]. <https://arxiv.org/abs/2211.15444v4>.
- [7] 廖龙杰, 吕文涛, 叶冬, 等. 基于深度学习的小目标检测算法研究进展[J]. 浙江理工大学学报(自然科学), 2023, 48(3): 331-343.
- [8] 潘晓英, 贾凝心, 穆元震, 等. 小目标检测研究综述[J]. 中国图象图形学报, 2023, 28(9): 2587-2615.
- [9] FENG X X, HAN J W, YAO X W, et al. TCANet: triple context-aware network for weakly supervised object detection in remote sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(8): 6946-6955.
- [10] GUAN L T, WU Y, ZHAO J Q. SCAN: semantic context aware network for accurate small object detection[J]. International Journal of Computational Intelligence Systems, 2018, 11(1): 951-961.
- [11] TISHBY N, ZASLAVSKY N. Deep learning and the information bottleneck principle[C]//Proceedings of IEEE Information Theory Workshop. Jerusalem, Israel: IEEE, 2015: 1-5.
- [12] WANG C Y, YEH I H, MARK LIAO H Y. YOLOv9: learning what you want to learn using programmable gradient information[C]//Computer Vision-ECCV 2024: 18th European Conference. Cham, Switzerland: Springer Nature Switzerland, 2025, 15089: 1-21.
- [13] SHWARTZ-ZIV R, TISHBY N. Opening the black box of deep neural networks via information[EB/OL]. (2017-04-29) [2024-06-15]. <https://arxiv.org/abs/1703.00810>.
- [14] VEIT A, WILBER M, BELONGIE S, et al. Residual networks behave like ensembles of relatively shallow networks[C]//

- Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona, Spain; ACM, 2016: 550–558.
- [15] LI C Y, LI L L, GENG Y F, et al. YOLOv6 v3.0: a full-scale reloading[EB/OL]. (2023-01-13)[2024-06-15]. <https://arxiv.org/abs/2301.05586>.
- [16] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA; IEEE, 2016: 770–778.
- [17] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA; IEEE, 2017: 2261–2269.
- [18] 郭佩林,张德,王怀秀.基于特征可视化探究跳跃连接结构对深度神经网络特征提取的影响[J/OL].计算机工程.(2024-04-29)[2024-07-20].<https://link.cnki.net/doi/10.19678/j.issn.1000-3428.0068885>.
- [19] DING X H, ZHANG X Y, MA N N, et al. RepVGG: making VGG-style ConvNets great again[C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, TN, USA; IEEE, 2021: 13728–13737.
- [20] DAI X Y, CHEN Y P, XIAO B, et al. Dynamic head: unifying object detection heads with attentions[C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, TN, USA; IEEE, 2021: 7369–7378.
- [21] 孙阳, 李佳. 基于通道剪枝的 YOLOv7-tiny 输电线路异物检测算法[J]. 计算机工程与应用, 2024, 60(14): 319–328.
- [22] BLALOCK D, ORTIZ J J G, FRANKLE J, et al. What is the state of neural network pruning? [EB/OL]. (2020-03-26)[2024-07-20]. <https://arxiv.org/abs/2003.03033v1>.
- [23] HAN S, POOL J, TRAN J, et al. Learning both weights and connections for efficient neural networks[C]//Proceedings of the 29th International Conference on Neural Information Processing Systems; Volume 1. Montreal, Canada; ACM, 2015: 1135–1143.
- [24] GALE T, ELSEN E, HOOKER S. The state of sparsity in deep neural networks[EB/OL]. (2019-02-25)[2024-07-20]. <https://arxiv.org/abs/1902.09574>.
- [25] LEE J, PARK S, MO S, et al. Layer-adaptive sparsity for the magnitude-based pruning[EB/OL]. (2021-05-09)[2024-08-19]. <https://arxiv.org/abs/2010.07611>.
- [26] 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.
- [27] 皇甫俊逸, 孟乔, 孟令辰, 等. 基于 GhostNet 与注意力机制的 YOLOv5 交通目标检测[J]. 计算机系统应用, 2023, 32(4): 149–160.
- [28] 姚若禹, 郑世玲, 史怡璇, 等. 基于改进 YOLOv8n 的钢材表面缺陷检测算法[J]. 聊城大学学报(自然科学版), 2025, 38(2): 177–189.
- [29] JIANG Y Q, TAN Z Y, WANG J Y, et al. GiraffeDet: a heavy-neck paradigm for object detection[EB/OL]. (2022-06-22)[2024-08-19]. <https://arxiv.org/abs/2202.04256>.
- [30] WANG C C, HE W, NIE Y, et al. Gold-YOLO: efficient object detector via gather-and-distribute mechanism[EB/OL]. (2023-10-23)[2024-08-19]. <https://arxiv.org/abs/2309.11331>.
- [31] LI C, LI L, JIANG H, et al. YOLOv6: a single-stage object detection framework for industrial applications[EB/OL]. (2022-09-07)[2024-08-19]. <https://arxiv.org/abs/2209.02976>.

Complex traffic scene detection method based on YOLO-NPDL

ZHANG Haochen, ZHANG Zhulin^{*}, SHI Ruiyan, CAO Shijie,
WANG Wenhan, LEI Zhennuo

School of Automotive Engineering, Shandong Jiaotong University, Jinan 250357, China

Abstract: In order to improve the detection accuracy of the vehicle object detection model in complex traffic scenes, using YOLOv8n (you only look once version 8 nano) as the benchmark model, a Neck-ARW

(including auxiliary detection branch, RepBlock module, and weighted jump feature fusion) neck structure with a composite backbone is designed to reduce information loss caused by information bottlenecks along the network depth direction; the RepBlock structure heavy parameterization module is introduced, and the multi-branch structure is used in the training process to improve the model feature extraction performance; the P2 detection layer is added to capture more small target detail features and enrich the feature information flow of small targets in the network; the Dynamic Head self-attention mechanism detection head is used, which integrates scale perception, spatial perception, and task perception self-attention mechanism into a unified framework to improve detection performance. The layer-adaptive magnitude based pruning(LAMP) algorithm is used to remove redundant parameters of the model and construct the YOLO-NPDL(Neck-ARW, P2, Dynamic Head, LAMP) vehicle object detection model. Using the UA-DETRAC(university at Albany detection and tracking) dataset as the experimental dataset, RepBlock module embedding position test, different neck structure comparison test, pruning test, ablation test, and model performance comparison test are conducted to verify the detection accuracy of the YOLO-NPDL model. The experimental results show that: RepBlock module has better feature extraction ability for multi-scale targets when embedding auxiliary detection branches and neck trunk structures at the same time, and can retain more detailed information during the training process, but the amount of parameters and computation increases; after using the Neck-ARW neck structure, the detection accuracy indicators E_{mAP50} and $E_{mAP50-95}$ of the model are increased by 1.1% and 1.7%, respectively, and the number of parameters is reduced by about 17.9%, and the structure is better; when the pruning rate is 1.3, the model parameters and computation are reduced by about 38.0% and 24.0%, respectively, and the redundant channel accounts for less, and the structure is more compact; compared with the YOLOv8n model, the YOLO-NPDL model has a 2.7% increase in recall rate, a 2.7% increase in E_{mAP50} , reaching 94.7%, a 6.4% increase in $E_{mAP50-95}$, reaching 79.7%; compared with the widely used YOLO series models, the YOLO-NPDL model has higher detection accuracy on the basis of fewer parameters. The YOLO-NPDL model has no obvious false detection or omission in detecting remote targets, rainy days and night scenes, and can detect more remote small target vehicles, with better detection effect.

Keywords: object detection; complex traffic scene; YOLOv8n; Neck-ARW; RepBlock; LAMP algorithm

(责任编辑:赵玉真)