DOI:10.11918/202406004

复杂场景下无人驾驶障碍检测算法

程铄棋[1,2],伊力哈木·亚尔买买提[1,2],谢丽蓉[1,2],侯雪扬[1],马 颖[1]

(1.新疆大学 电气工程学院,乌鲁木齐 830017;2.新疆露天矿智能生产与管控重点实验室(新疆大学),乌鲁木齐 830017)

摘 要:为解决复杂路况下因目标遮挡及小目标信息缺失导致现有无人驾驶目标检测算法准确率低的问题,提出了基于改进YOLOv8的无人驾驶障碍检测算法(YOLOv8 effectual accurate,YOLOv8-EA)。该算法首先引入快速神经网络作为主干网络,利用部分卷积提取空间特征,保证特征的完整性;其次,利用大内核深度卷积层重构快速金字塔池化层,采用并行多尺度连接的方式融合不同分辨率的自注意力特征,增强模型在复杂环境中的特征提取能力;然后,采用多分支结构和重参数化抑制信息干扰,并通过不断堆叠梯度流的方式提升特征融合能力;最后,基于部分卷积设计小目标检测头以处理小目标像素级特征信息。对比实验结果表明,相较于原模型,上述改进后,模型在性能上均有明显提升,并在检测精度上显著优于其他改进方式。消融实验结果表明,YOLOv8-EA在障碍检测精度方面取得显著提升,在KITTI数据集下,mAP50和mAP50-95分别提升了2.4%和4.7%;采用SODA10M数据集进行二次验证,mAP50和mAP50-95分别提升了1.4%和1.1%,证明YOLOv8-EA算法具有很好的泛化能力。所提算法在处理遮挡目标及小目标时,展现了出色的性能,为无人驾驶系统中的后续决策任务提供了更加可靠的支持。

关键词:目标检测;无人驾驶;复杂道路场景;部分卷积;大内核深度卷积层

中图分类号:TP399 **文献标志码**:A **文章编号**:0367-6234(2025)06-0160-11

Obstacle detection algorithm for unmanned driving in complex scenarios

CHENG Shuoqi[1,2], YILIHAMU Yaermaimaiti[1,2], XIE Lirong[1,2], HOU Xueyang[1], MA Ying[1]

(1. School of Electrical Engineering, Xinjiang University, Urumqi 830017, China;
2. Key Laboratory of Intelligent Production and Control of Open-Pit Mines (Xinjiang University), Urumqi 830017, China)

Abstract: To solve the problem of low accuracy of existing unmanned target detection algorithm caused by object occlusion and small-object information loss in complex road scenarios, this paper proposes an enhanced obstacle detection algorithm based on YOLOv8, named YOLOv8-EA (effectual accurate). The algorithm incorporates a lightweight backbone using partial convolution to preserve spatial feature integrity. A large-kernel depthwise convolutional layer is introduced to reconstruct the pyramid pooling structure, and the multi-scaled self-attention features are fused through parallel connections, which Enhances the feature extraction ability of the model in complex scenarios. Additionally, a multi-branch architecture with reparameterization suppresses noise interference and enhances feature fusion via stacked gradient flow. A small-object detection head based on partial convolution is also designed to improve pixel-level feature extraction for small targets. Experimental results show that YOLOv8-EA achieves notable improvement in detection accuracy compared to the original YOLOv8. On the KITTI dataset, mAP50 and mAP50-95 increased by 2.4% and 4.7%, respectively, while on the SODA10M dataset, gains of 1.4% and 1.1% were observed, which demonstrate the strong generalization ability of YOLOv8-EA. The proposed algorithm shows superior capability in handling occlusion and small-object detection, offering more reliable perception support for unmanned driving systems.

Keywords: object detection; unmanned driving; complex road scenarios; partial convolution; large-kernel depthwise convolution layers

目标检测作为环境感知中的一部分,在无人驾驶中起到“眼睛”的作用,是智驾系统中不可或缺的一环。无人驾驶汽车在行驶过程中,若车速过快或遇到颠簸路段,可能会产生运动模糊,导致传感器采集的图像出现信息缺失,从而增加小目标检测的难度。此外,车辆行驶环境复杂多变,如城市道路和高速公路中常见的建筑物、树木及其他车辆等遮挡物,可能影响目标的可见性,进而降低检测的准确性。

收稿日期:2024-06-03;**录用日期**:2024-07-08;**网络首发日期**:2025-04-28
网络首发地址:https://link.cnki.net/urlid/23.1235.T.20250428.1129.003
基金项目:国家自然科学基金(62362063);新疆维吾尔自治区自然科学基金(2023B01006)
作者简介:程铄棋(1999—),女,硕士研究生;伊力哈木·亚尔买买提(1978—),男,教授,硕士生导师;谢丽蓉(1969—),女,教授,博士生导师
通信作者:伊力哈木·亚尔买买提,65891080@qq.com

深度学习模型凭借其强大的学习能力，可以从大量的驾驶数据中学习驾驶规则和模式，为无人驾驶的实际应用提供重要技术支撑。近年来，目标检测算法迅速发展。Girshick 等[1]提出基于区域的卷积神经网络（region-based convolutional neural network，RCNN）的两阶段目标检测算法，第 1 阶段生成候选区域，第 2 阶段对候选区域进行精细化处理。Ren 等[2]在两阶段的基础上添加了区域候选网络，提出了快速区域卷积神经网络（faster region-based convolutional neural network，Faster R-CNN），但提取到的特征图分辨率较低。赵锟[3]在 Faster R-CNN 的基础上，加入了人类视觉系统机制，利用物体的边缘信息构造目标候选区域，提升了检测准确率。Redmon 等[4]提出的 YOLO 是一种集目标分类、检测与分割于一体的高效算法，因性能卓越而受研究者的青睐。目前，YOLO 已更新至第 8 代[5-9]。李经宇等[10]在 YOLOv3 的基础上加入了空间金字塔池化模块，将多尺度局部特征进行融合和拼接，有效地提高了性能，使模型更适用于复杂场景下的检测任务。Wu 等[11]对 YOLOv5 的主干网络进行了简化，同时优化了残差结构，通过跨层及多次连接残差组件提升特征融合能力，并将其应用于交通信号灯识别。田鹏等[12]在 YOLOv8 的基础上引入了平衡区域注意力，以增强网络对小目标的感知能力。同时，采用可变形卷积改善因不规则特征导致的特征提取能力不足问题，从而提升了算法对重叠遮挡目标的检测性能。尽管这些改进推动了无人驾驶感知技术的发展，但要实现全面普及，需要进一步提升目标识别的精度。准确识别是决策的前提，也是安全的保障。

针对上述问题，本文提出了复杂场景下的无人驾驶障碍检测算法（YOLOv8 effectual accurate，YOLOv8-EA），对小目标和遮挡目标进行优化。该算法结合图像修复技术补全缺失的特征信息，采用大内核卷积层增强特征提取能力，并引入多分支结构丰富信息流。最后，通过实验验证 YOLOv8-EA 的有效性，旨在为无人驾驶目标检测提供新的思路。

1　YOLOv8-EA 网络

YOLOv8 是一种单阶段目标检测算法，无需生成额外的候选区域，可以直接从输入图像预测目标类别和位置。YOLOv8 的应用范围非常广泛，不仅能用于目标检测领域，还能用于跟踪、分割、分类及姿态估计等任务中。YOLOv8 的网络结构如图 1 所示。总体分为 3 部分：1）主干网络（backbone）。主要用于特征提取，即从输入图像中提取多尺度的特征表示，以提供丰富的特征信息。backbone 的尾部添加了金字塔池化层（spatial pyramid pooling faster，SPPF），目的是增加感受野以便于捕获场景中更多层次的特征信息。2）颈部网络（neck）。该部分将 backbone 提取到的特征进行融合，帮助浅层特征向深层特征聚合。3）检测头（head）。该部分采用解耦头结构，分为分类端和定位预测端，以避免两种任务之间存在冲突。

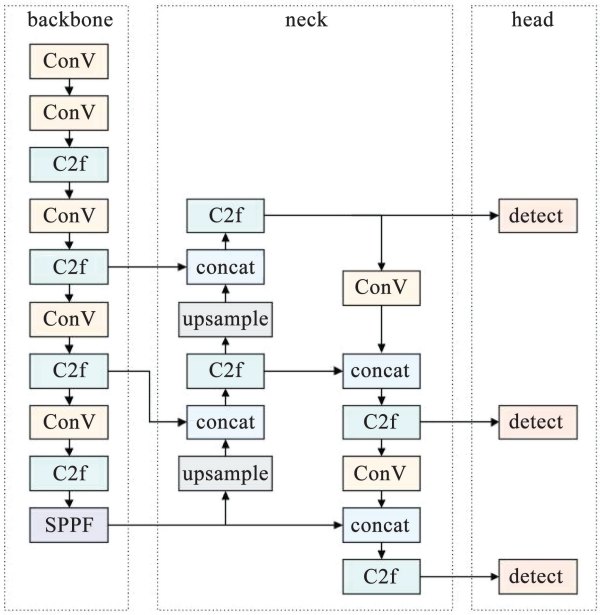


图 1　YOLOv8 网络结构

Fig. 1　YOLOv8 network architecture

本文选择 YOLOv8 作为基础网路，针对现有问题，对 YOLOv8 的 backbone、neck 和 head 进行相应改进，构建了 YOLOv8-EA，其网络结构如图 2 所示。该算法引入快速神经网络（faster neural network，FasterNet）作为主干网络，通过部分卷积（partial convolution，PConV）提取空间特征，仅对非缺失数据进行卷积运算，并补充缺失像素。在 SPPF 模块中融合大内核深度卷积层（large separable kernel attention，LSKA），并对 SPPF 结构进行重构，提出基于大内核深度卷积的快速空间金字塔池化层（faster spatial pyramid pooling with large separable kernel attention，SPPLA），用于处理不同尺度图像特征，以融合更大尺度的全局信息。在 neck 部分引入多分支结构和重参数化技术（rcs-one-shot aggregation，RCSOSA），通过不同通道提取多尺度特征，并采用重复堆叠的方式增强信息流动。在 head 部分，利用 PConV 构造小目标检测头（partial convolution head，Phead），处理小目标像素级特征信息，以提升模型的鲁棒性。

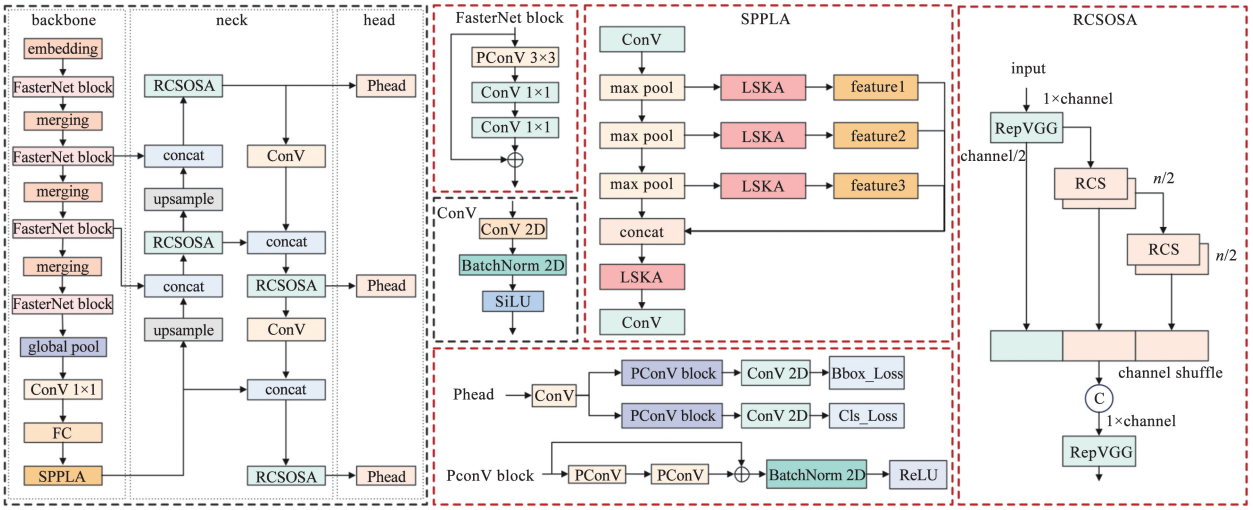


图 2　YOLOv8-EA 网络结构

Fig. 2　YOLOv8-EA network architecture

2　优化策略

2.1　FasterNet 融入主干网络

FasterNet 是一种基于 PConV 的高精度快速神经主干网络[13]，能够有效提取空间特征和不同尺度目标特征。FasterNet 的结构如图 3 所示，包含 4 个层次级，每个层次的起始阶段设有 1 个步长为 4 的常规 4 ×4 卷积层作为嵌入层（embedding）负责空间下采样，或者是 1 个步长为 2 的常规 2 ×2 卷积层作为合并层（merging）用于扩展通道数量。每个嵌入层或合并层后接上一个基础网络模块（FasterNet block）以进行特征提取。经过 4 个层级之后，通过全局平均池化（global pool）操作，对整个特征图的所有像素值进行平均，从而实现特征的降维和转换。最终，通过 1 个 1 ×1 卷积（convolution，ConV）调整通道数，并加入 1 个 1 ×1 的全连接层（fully connected layer，FC）以进行分类任务处理。

FasterNet block 包含 1 个 PConV 和两个 ConV 层，用于提取图像特征和非线性激活，并利用残差结构加深模型，提升特征信息的传递效率。受汽车颠簸或光线条件不佳的影响，小目标的信息通常会丢失。PConV 是一种专为处理图像中缺失或损坏区域而设计的卷积[14]，其通过二进制掩码识别图像中有效数据的区域，仅对这些有效区域进行卷积操作，同时智能补充缺失像素，恢复图像特征的完整性。图 3 中 w、h、c_p 分别代表张量的宽度、高度和每一层的通道数，k 代表卷积核的大小。PConV 结合了恒等映射（identity）技术和卷积核（filters），其中，identity 允许部分输入特征直接传递至输出，避免信息丢失，而 filters 则对有效的区域进行卷积操作。通过这种方式，PConV 不仅能够保留原始特征中的重要信息，还能有效应对图像中缺失或损坏区域，提升图像特征的完整性和准确性。

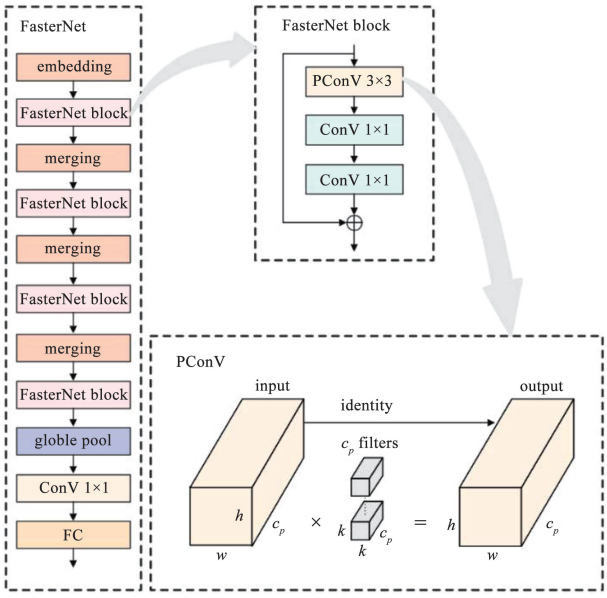


图 3　FasterNet 网络结构

Fig. 3　FasterNet network architecture

2.2　SPPLA 模块设计

由于汽车行驶过程中场景复杂多变，对目标的特征信息有一定的干扰。针对这一问题，本文将 LSKA[15] 与 SPPF 相结合，提出 SPPLA，其设计思路如图 4 所示。SPPLA 通过结合 LSKA 的局部特征捕捉能力和 SPPF 的多尺度特征提取优势，提高目标检测的准确性和鲁棒性。

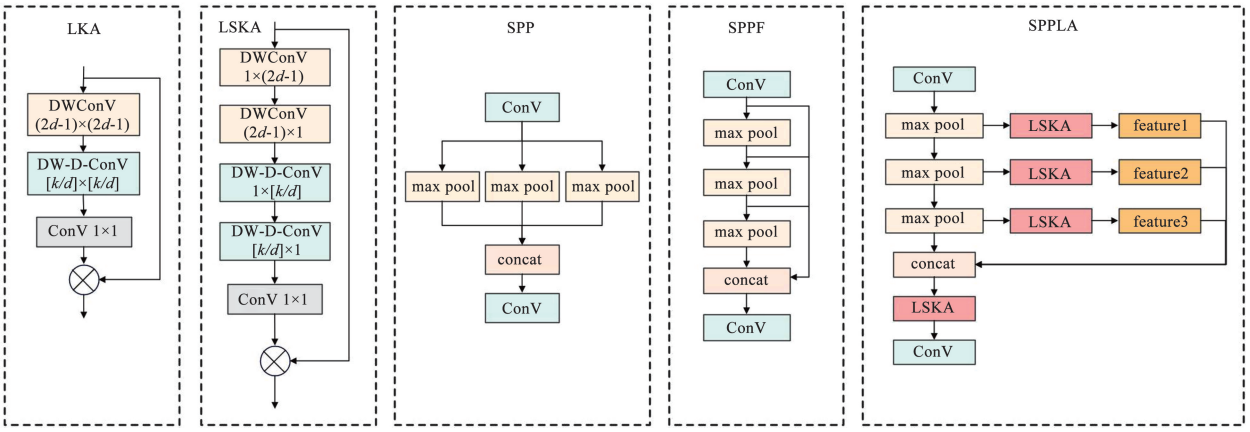


图 4　SPPLA 设计思路

Fig. 4　SPPLA design approach

局部关键点注意力模块(large kernel attention, LKA)包含深度卷积(depthwise convolution, DWConV)和深度空洞卷积(depthwise dilation convolution, DW-D-ConV),其中 DWConV 用于捕获局部特征,DW-D-ConV 用于建模长距离特征。DWConV 和 DW-D-ConV 中引入了膨胀率 d,通过在卷积核内部插入空洞,使感受野在保持参数量增长可控的情况下有效扩大,从而能够感知更大范围的上下文信息。这种 DWConV 和 DW-D-ConV 的组合相当于一个大规模的卷积神经网络(convolutional neural network, CNN)[16] 卷积核,DW-D-ConV 的输出作为 1×1 卷积的输入,生成注意力图,然后输入特征与生成的注意力图进行逐元素相乘,得到自适应细化的特征。LSKA 模块是 LKA 模块的可分离版本,通过将 DWConV 中的二维卷积核分解为垂直和水平的一维卷积核,使得大卷积核能够在注意力机制模块中更高效地使用。

空间金字塔池化网络层(spatial pyramid pooling, SPP)[17] 通过对不同区域进行最大池化(max pool)操作,成功解除 CNN 对输入大小的限制,使得即便输入图像的尺寸不同,也能获得固定的输出尺寸。SPPF 在 SPP 的基础上进一步提升了网络在处理不同尺寸输入时的精度和速度,通过更快的区域池化操作,其可以在输入图像尺寸不同的情况下,仍然获得相同尺寸的输出图像,并保持较高的精度。然而,SPPF 在进行连续 max pool 时会忽略某些细节信息。为了解决这一问题,SPPLA 采用并行多尺度连接(concat),通过多次 max pool,生成不同分辨率的特征图(feature),捕捉高分辨率的局部细节和低分辨率的全局信息。每个尺度的特征图随后会经过 LSKA 模块处理,从而保留更多细节信息,得到更丰富的局部自注意力特征图。将特征图进行拼接后,能够获得更加多样化的特征表示,增强模型的特征表达能力。最后,SPPLA 通过串联一个 LSKA 模块,进一步获取更大尺度的全局信息,从而学习不同特征的重要性,聚焦于关键信息,帮助网络忽略背景干扰,提取更多有效特征信息。SPPLA 不仅可以处理不同尺寸的输入图像,还能在复杂多变的环境下有效提取容易被忽略的有用特征信息,从而增强网络性能,进一步提升模型的泛化能力。

2.3　RCSOSA 模块优化颈部网络

为了进一步抑制无人驾驶汽车行驶过程中背景对目标的干扰,引入 RCSOSA[18] 优化颈部网络的 C2f 模块,提升颈部网络的特征融合能力。在训练阶段,RCSOSA 模块采用多分支结构丰富特征表达,推理阶段采用结构重参数化(reparameterization)将训练时结构较大但优秀的某种性质(例如更好的精度)保留至结构更小的推理阶段,并通过堆叠加强不同通道之间的信息流动,丰富梯度流信息。RCSOSA 的结构框架如图 5 所示。

基于通道重排的重参数卷积(reparameterized convolution based on channel shuffle, RCS)是一种将视觉几何群网络(RepVGG/RepConV)[19] 结合了通道混洗(channel shuffle)[20] 的 reparameterization 卷积,可以增强网络的特征提取能力,提供更多的特征。RCS 结构的左侧是训练阶段的 RepVGG,在给定输入的特征维度为 $c \times h \times w$ 后,经过通道分裂(channel split)运算形成两个尺寸均为 $c \times h \times w$ 通道的张量,其中一个张量由恒等分支、1 个 1×1 卷积和 3×3 卷积构成,另一个张量直接进行融合。右侧是推理阶段,将训练阶段的 RepVGG 通过重参数化转换为 3×3 的 RepConV。这种多分支的拓扑结构有助于学习训练中更多的特征信息,经过多支训练的张量通过通道方式进行连接,不仅能丰富特征表达,还能增强不同特征之间的信息融合。

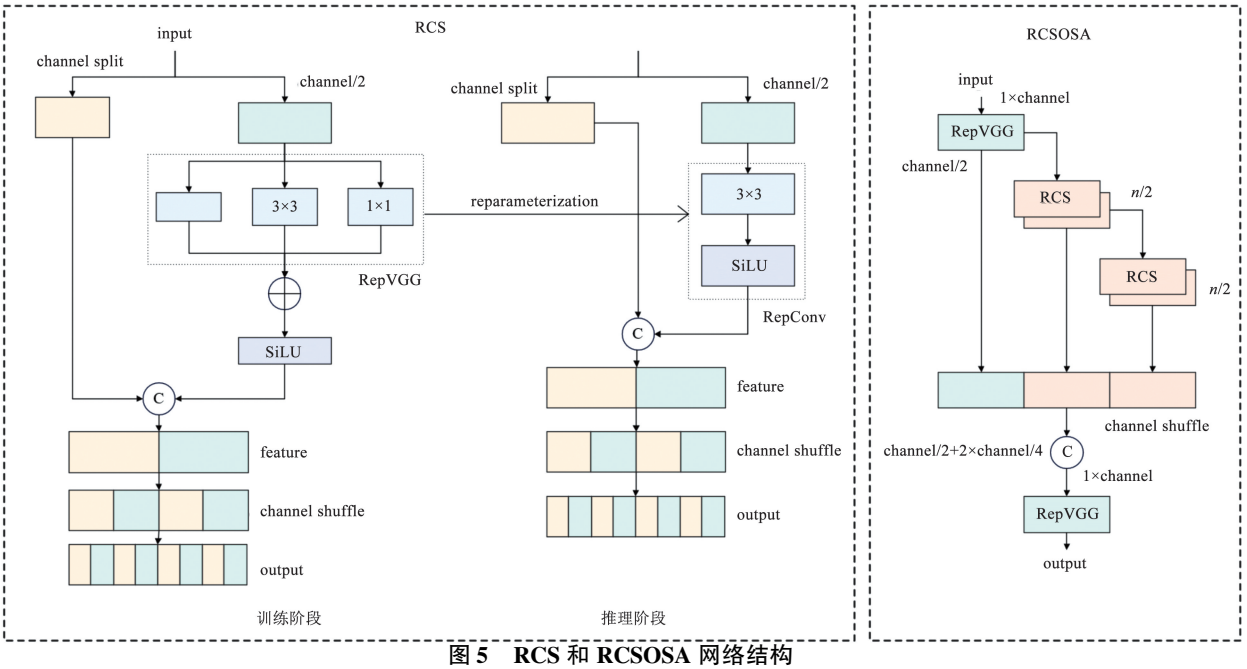


图 5 RCS 和 RCSOSA 网络结构

Fig. 5 RCS and RCSOSA network architectures

RCSOSA 模块是在 RCS 的基础上加入一次性聚合(one-shot aggregation, OSA)技术,这样可以增强网络中的密集连接,并通过多感受野得到多样化特征,最后一次性融合所有特征,提升模型的效率。为了使特征可以被充分复用,采用重复堆叠 RCS 模块的方式,使得相邻特征层中不同通道间的信息更好地流动。在 RCSOSA 模块中,输入被分流成两部分,一部分进入堆叠的 RCS 模块进行处理,另一部分则直接与经过处理的部分在 channel shuffle 处进行融合。YOLOv8 中原有的 C2f 模块的特征融合方式相对简单,主要通过特征的分割及拼接实现,在处理复杂场景下的多尺度目标时,表现有所局限。将 C2f 替换为 RCSOSA 模块,可以更有效地传递有用的特征信息并进行融合,抑制复杂场景下冗余信息多样性对模型的影响。

2.4 Phead 检测头设计

小目标的检测能力不仅受到汽车抖动或不良光照等外部因素的影响,同样也受网络结构设计的制约,许多检测网络通过堆叠卷积层的方式获取更丰富的语义信息特征,导致细节特征被忽略。YOLOv8 检测头(YOLOv8_head)与 Phead 的结构框架如图 6 所示。YOLOv8_head 主要依赖传统的 3 × 3 卷积层,在处理小目标特征时,由于连续的卷积和池化操作,细节信息会逐渐丢失,导致小目标检测能力下降。为了改善这一问题,基于 PConV 模块的设计理念,提出了一种新的结构(PConV block),并重构了 YOLOv8 原有的检测头,提出了小目标解耦头 Phead。Phead 采用分支解耦结构,初始卷积层将提取到的特征共享给两个并行路径,每个路径均包含 1 个 PConV block 模块和 1 个 ConV 2D 卷积层,最终分别输出,用于边界框回归损失(bounding box loss, Bbox_Loss)和分类损失(classification loss, Cls_Loss)。

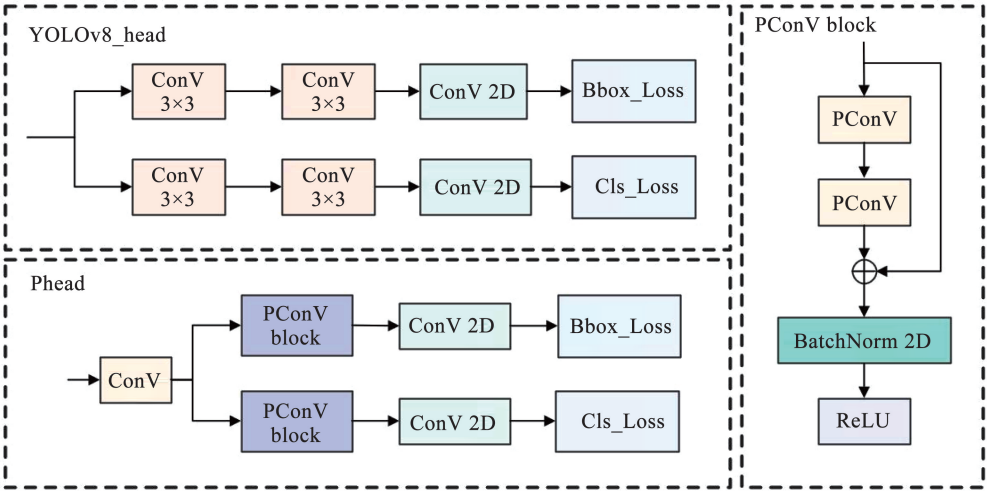


图 6 YOLOv8_head 和 Phead 网络结构

Fig. 6 YOLOv8_head and Phead network architectures

PConV block 是由 PConV 构成的，通过串联两个 PConV 处理有缺失部分和不规则形状的输入，避免传统卷积在处理这些区域时引入噪声。并且通过自适应学习，只对非缺失的数据进行卷积操作，维持图像上下文信息，生成更一致的特征。在此基础上，加入残差结构，缓解梯度消失及网络退化问题，加速收敛。并且，残差结构使得网络在面对不同的数据分布时更具鲁棒性，能够更好地适应训练集和测试集之间的差异，提升网络性能。加入二维批量归一化（batch normalization for 2-dimensional inputs，BatchNorm 2D）缓解过拟合问题，使训练过程更加稳定。最后，通过修正线性激活函数（rectified linear unit，ReLU）提高网络特征表达能力，从而更有效地捕捉图像中的细节特征。

3　实验结果与分析

3.1　数据集及实验环境

采用德国卡尔斯鲁厄理工学院和丰田美国技术研究院联合制作的 KITTI[21] 数据集进行实验，并选用华为与中山大学联合创建的 SODA10M 无人驾驶数据集[22] 进行验证。KITTI 是一种用于计算机视觉研究的无人驾驶场景数据集，包含了市区、乡村及高速公路等真实环境数据，每张图像包含最多 15 辆车和 30 名行人。SODA10M 数据集包含了从 32 个城市采集的一千多万张道路场景车辆行驶视角图像。将原始的 KITTI 和 SODA10M 数据集经过筛选和归类，最终划分为“行人”、“自行车”和“汽车”3 种类别，其中 KITTI 数据集共计 7 481 张，SODA10M 数据集共计 10 000 张。

实验环境为：AutoDL 算力云服务器，NVIDIA GeForce RTX3090 显卡，24 GB 运存，12 核 Xeon(R) Platinum 8255C CPU，43 GB 内存；1.11.0 版本的 PyTorch 框架以及 11.3 版本的 CUDA，对应 3.8 版本的 python。

3.2　评价指标

在目标检测任务中，准确率（precision）是指在模型所预测为正类别（即目标存在）的所有样本中，实际为正类别的样本所占的比例；召回率（recall）是指在所有实际为正类别（即目标确实存在）的样本中，被模型正确预测为正类别的样本所占的比例。准确率和召回率的曲线即为 PR 曲线，以召回率为横坐标，准确率为纵坐标，PR 曲线所围成的面积即为平均准确率（average precision，AP），曲线面积越大，说明模型性能越好。

通常，目标检测任务会涉及多个类别的目标，每个类别绘制一条 PR 曲线，并计算该类别的 AP 值。最终，所有类别的 AP 值取平均，即为平均精度值（mean average precision，mAP）。当交并比（intersection over union，IoU）设置为 0.5 时，得到的就是 mAP50。如果 IoU 阈值设得较高，模型的精度通常会较低。mAP50-95 指在 IoU 从 0.50 到 0.95 之间的多个不同值下计算得到的平均 mAP，用于全面评估模型在不同 IoU 条件下的检测性能。

参数量是神经网络中的可训练参数总数，其大小直接影响模型的复杂度和学习能力，更多的参数通常意味着模型有更大的能力去捕捉数据中的复杂特征，但同时也会增加训练难度并存在拟合风险。计算量通常指执行一次前向传播所需的浮点运算次数，反映了模型在处理数据时的计算成本。

3.3　对比实验及消融实验

3.3.1　主干网络横向对比

为了验证最佳主干网络，将 efficient vision transformer（EfficientViT）及可逆神经网络（RevCol）、FasterNet 作为主干网络引入 YOLOv8 中，并与原 YOLOv8 进行对比，结果见表 1。由表 1 可以看出：在相同实验条件下，EfficientViT 的准确率达到 90.6%，但召回率和 mAP 较低，且参数量及计算量相对于 YOLOv8 均有所上升；RevCol 作为主干时，虽使模型参数量压缩 29.0%，计算量下降 21.0%，但模型各项性能指标均明显下降；FasterNet 作为主干时，虽然引发了额外的计算开销，模型变得更加复杂，但召回率、mAP50 和 mAP50-95 对比 YOLOv8 分别提高了 3.3%、0.6% 和 2.6%，在所有主干网络中表现最好。

表 1　主干网络横向对比

Tab. 1　Horizontal comparison of backbone networks

主干网络	准确率/%	召回率/%	mAP50/%	mAP50-95/%	计算量/10^9	参数量/10^6
YOLOv8	89.8	75.9	86.7	58.9	8.2	3.1
EfficientViT	90.6	76.7	86.6	58.6	10.1	4.1
RevCol	90.0	75.1	84.3	56.6	6.4	2.2
FasterNet	90.1	79.2	87.3	61.5	10.7	4.1

3.3.2 RCSOSA 模块横向对比

为了选择最契合的模块,将高效多尺度注意模块(EMA_Faster)、可变形卷积 V2(DCNV2)、可变形卷积 V3(DCNV3)和 RCSOSA 模块分别引入 YOLOv8 中,并与 YOLOv8 原有的 C2f 模块进行横向对比,结果见表 2。由表 2 可知:在相同实验条件下,引入 EMA_Faster 后,模型更加轻量,但各项性能指标对比 C2f 模块均有所下降;加入 DCNV2 和 DCNV3 后,相比 C2f 模块,准确率和召回率有小幅提升,DCNV3 的 mAP50 提升 0.9%,DCNV2 和 DCNV3 的 mAP50-95 分别提升了 0.8% 和 1.9%,在性能上虽有提升,但考虑模型复杂度和计算量均有上涨的前提下,性能提升幅度较小;在引入 RCSOSA 后,相较于未改进前,虽引发了大量的计算开销,模型也较为复杂,但准确率、召回率、mAP50 和 mAP50-95 分别提升 0.9%、5.3%、2.2% 和 4.2%,提升效果十分可观。

3.3.3 SPPLA 模块横向对比

为了验证模块与网络是否契合,将焦点调制网络(FocalModule)、基于注意力的内部尺度特征交互模块(AIFI)和本文提出的 SPPLA 模块分别引入 YOLOv8 中,并与 YOLOv8 中原有的 SPPF 进行横向对比实验,结果见表 3。由表 3 可知:在相同实验条件下,引入 FocalModul 后,准确率较未改进前提升了 1.4%,mAP50-95 上涨了 0.9%,性能表现良好,产生的额外计算开销也较小,参数量基本维持;引入 AIFI 模块后,相较于 YOLOv8,计算量上涨且各项性能指标均下降,与网络不契合;采用本文设计的 SPPLA 模块后,提升最为显著,召回率、mAP50 和 mAP50-95 相较于 SPPF 分别提升了 3.3%、1.3% 和 1.9%,虽然产生额外 11.0% 的计算开销,并增加了 9.7% 的参数量,但性能领先于其他模块,与 YOLOv8 模型契合度最高,证明了 SPPLA 模块的有效性。

表 2 RCSOSA 模块横向对比

Tab. 2 Horizontal comparison of RCSOSA modules

模块	准确率/%	召回率/%	mAP50/%	mAP50-95/%	计算量/10^9	参数量/10^6
C2f	89.8	75.9	86.7	58.9	8.2	3.1
EMA_Faster	86.2	79.5	86.5	58.0	7.3	2.4
DCNV2	89.4	78.0	86.5	59.7	8.8	3.1
DCNV3	92.3	78.4	87.6	60.8	8.1	2.9
RCSOSA	90.7	81.2	88.9	63.1	19.5	6.8

表 3 SPPLA 模块横向对比

Tab. 3 Horizontal comparison of SPPLA modules

模块	准确率/%	召回率/%	mAP50/%	mAP50-95/%	计算量/10^9	参数量/10^6
SPPF	89.8	75.9	86.7	58.9	8.2	3.1
FocalModule	91.2	78.2	86.5	59.8	8.9	3.2
AIFI	88.2	78.3	86.1	59.0	8.7	3.0
SPPLA	89.0	79.2	88.0	60.8	9.1	3.4

3.3.4 检测头横向对比

为了选择最佳的网络检测头,将 YOLOv7 的检测头、渐近特征金字塔网络(AFPN)、Phead 分别引入 YOLOv8 中,与 YOLOv8 原有的检测头进行对比,结果见表 4。由表 4 可知:在相同条件下,YOLOv7 的检测头在召回率上领先其他检测头,较 YOLOv8 检测头提升了 2.0%,mAP50-95 也有明显提升,表现良好,但参数量上涨了 29.0%,也产生了 52.4% 的额外计算量;引入 AFPN 检测头后,相较于 YOLOv8,不但增加了 13.4% 的额外计算量,而且各项性能指标均有所下降,整体表现较差;引入本文提出的 Phead 后,相较于原模型,mAP50 提升了1.3%,mAP50-95 提升了 1.9%,没有产生额外的计算量,比 YOLOv7 检测头表现更好,与 YOLOv8 网络契合度最高,整体表现最优。

3.3.5 KITTI 数据集消融实验

为了验证所选模块的有效性及其与网络的契合度,在相同实验条件下,将不同模块加入 YOLOv8 网络中进行消融实验,结果见表 5。由表 5 可知:分别单独加入 4 个优化模块后,相较于 YOLOv8 均有所提升,但参数量及计算量也均有上涨;在颈部网络加入 RCSOSA 模块时,模型变得最为复杂,但同时也获得最好的 mAP,检测效果最好;将 FasterNet 作为主干网络、Phead 作为检测头后,相较于未改进时的

YOLOv8，mAP50 和 mAP50-95 分别提升了 1.0% 和 2.6%，表明两个模块与模型契合度较高，但造成了额外的计算负担；在此基础上加入 SPPLA 模块后，性能进一步提升了 0.4% 和 0.1%，且基本未增加额外参数量，显示出了 SPPLA 模块的优越性；最后，加入 RCSOSA 模块后，构成完整的 YOLOv8-EA，相较 YOLOv8，mAP50 和 mAP50-95 分别提升了2.4% 和 4.7%，提升明显，说明检测效果很好。

表 4　检测头横向对比

Tab. 4　Horizontal comparison of detection heads

检测头	准确率/%	召回率/%	mAP50/%	mAP50-95/%	计算量/10^9	参数量/10^6
YOLOv8	89.8	75.9	86.7	58.9	8.2	3.1
YOLOv7	87.3	77.9	85.6	59.2	12.5	4.0
AFPN	88.7	74.6	84.1	57.7	9.3	2.7
Phead	88.9	77.5	86.9	59.5	8.2	3.8

表 5　KITTI 数据集消融实验

Tab. 5　Ablation experiments on the KITTI dataset

FasterNet	SPPLA	RCSOSA	Phead	准确率/%	召回率/%	mAP50/%	mAP50-95/%	计算量/10^9	参数量/10^6
×	×	×	×	89.8	75.9	86.7	58.9	8.2	3.1
√	×	×	×	90.1	79.2	87.3	61.5	10.7	4.1
×	√	×	×	89.0	79.2	88.0	60.8	9.1	3.4
×	×	√	×	90.7	81.2	88.9	63.1	19.5	6.8
×	×	×	√	88.9	77.5	86.9	59.5	8.2	3.8
√	×	×	√	89.0	80.5	87.7	61.5	10.7	5.0
√	√	×	√	89.5	80.7	88.1	61.6	11.1	5.0
√	√	√	√	93.6	81.0	89.1	63.6	23.0	9.4

注：√表示在 YOLOv8 原模型的基础上加入该模块，×表示未加入。

为了更直观地展现消融实验的结果，绘制了每一轮的 mAP50 和 mAP50-0.95 的变化曲线，如图 7 和图 8 所示。可以看出，YOLOv8-EA 模型在 mAP 上的效果更好。

3.4　目标检测算法对比分析

为了验证 YOLOv8-EA 的优越性，选择目前主流的目标检测算法（SSD、Faster R-CNN 及 YOLO 系列）与 YOLOv8-EA 进行对比，结果见表 6。由表 6 可知：在相同实验条件下，SSD、Faster R-CNN 以及 YOLOv3 算法参数量十分庞大，且计算量较多，各项指标也非常低；YOLOv5 和 YOLOv6 在性能上相较于 YOLOv3 有明显的提升，且更加轻量，但对比 YOLOv8 仍有不足，各项性能指标均低于 YOLOv8，鲁棒性差；优化后的 YOLOv8-EA 各项指标表现更加优秀，mAP50 和 mAP50-95 分别达到了 89.1% 和 63.6%，领先于其他模型；对比 YOLOv8，虽然 YOLOv8-EA 的参数量更大，也引起了很多额外的计算开销，但与参数量、计算量更大，且与拥有更好性能的 YOLOv8-L（YOLOv8 large）相比，YOLOv8-EA 的各项性能指标更好，且计算复杂度更低，实用性和鲁棒性更强。

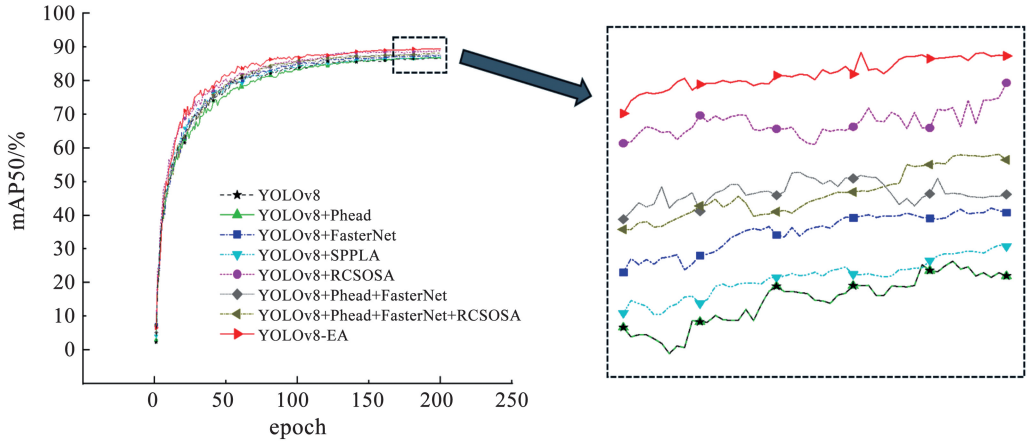


图 7　消融实验 mAP50 结果可视化对比

Fig. 7　Visualization comparison of mAP50 results in the ablation experiment

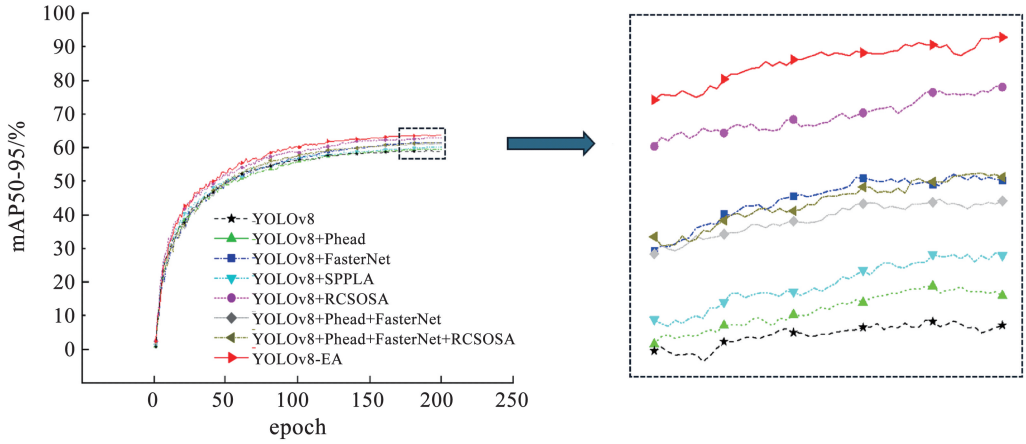


图 8 消融实验 mAP50-95 结果可视化对比

Fig. 8 Visualization comparison of mAP50-95 results in the ablation experiment

表 6 目标检测算法对比实验

Tab. 6 Comparative experiments of object detection algorithms

模型	准确率/%	召回率/%	mAP50/%	mAP50-95/%	计算量/10^9	参数量/10^6
SSD	89.3	29.0	58.0	24.1	269.3	23.9
Faster R-CNN	46.7	85.8	81.0	57.4	369.7	29.2
YOLOv3	61.8	54.8	73.1	33.5	152.8	61.5
YOLOv5	89.1	76.2	86.2	57.5	15.8	7.0
YOLOv6	85.4	74.1	81.5	55.3	13.1	4.5
YOLOv8	89.8	75.9	86.7	58.9	8.2	3.1
YOLOv8-L	91.0	80.5	88.3	63.3	28.7	11.0
YOLOv8-EA	93.6	81.0	89.1	63.6	23.0	9.4

3.5 SODA10M 数据集消融实验

从 SODA10M 数据集中选取 10 000 张包含小目标及遮挡目标的样本进行效果验证，结果见表 7。由表 7 可知：分别加入 4 个改进点后，相较于未改进前，模型在各方面均有小幅提升；在 FasterNet 的基础上分别加入 SPPLA 和 RCSOSA 后，模型准确率进一步提升；相较于 YOLOv8，YOLOv8-EA 的准确率和召回率分别提升了 2.1% 和 1.5%，mAP50 和 mAP50-95 分别提升了 1.4% 和 1.1%，再次证明了 YOLOv8-EA 的优越性。

表 7 SODA10M 数据集消融实验

Tab. 7 Ablation experiments on the SODA10M dataset

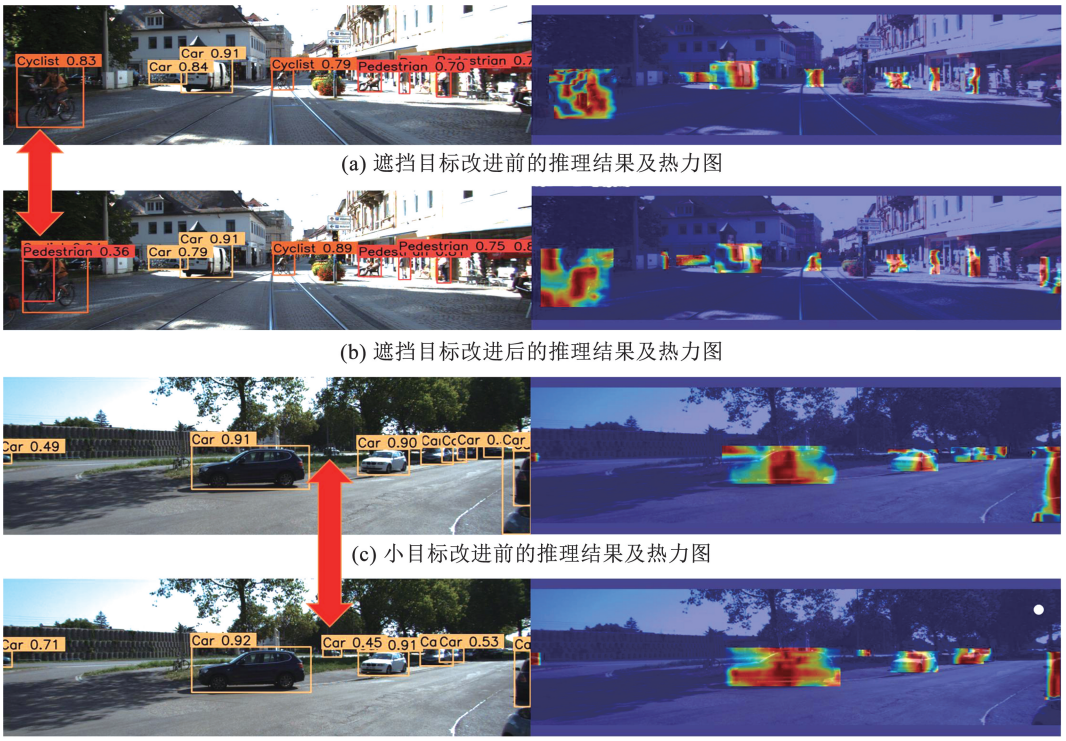
FasterNet	SPPLA	RCSOSA	Phead	准确率/%	召回率/%	mAP50/%	mAP50-95/%	计算量/10^9	参数量/10^6
×	×	×	×	70.0	55.6	61.5	38.3	8.2	3.1
√	×	×	×	71.3	55.4	62.0	38.7	10.7	4.1
×	√	×	×	71.7	54.4	62.2	38.2	9.1	3.4
×	×	√	×	70.8	56.1	62.3	38.8	19.5	6.8
×	×	×	√	70.3	55.9	61.9	38.0	8.2	3.8
√	√	√	×	73.7	55.4	62.7	38.9	11.1	5.0
√	√	√	√	72.1	57.1	62.9	39.4	23.0	9.4

注：√表示在 YOLOv8 原模型的基础上加入该模块，×表示不加入。

3.6 效果验证

为了更直观地体现网络模型在实际场景中的应用效果，选取了部分场景进行可视化展示，如图 9 所示。图 9(a)和图 9(b)分别为遮挡目标改进前后的推理结果及其对应的热力图，可以看到，箭头标记的区域，改进前存在明显的漏检现象，部分遮挡目标未能检测到；而改进后，遮挡目标能够被模型准确检测，并且特征图中的关注区域更加集中。图 9(c)和

图 9(d)分别为小目标改进前后的推理结果及其对应的热力图，可以观察到，箭头标记的区域，改进前算法对小目标检测存在漏检现象，而改进后算法能够有效检测出小目标，显著提高了检测的准确性。通过对比推理结果及热力图，进一步验证了改进算法在处理复杂场景，特别是在遮挡目标和小目标检测方面性能的优越性。



(a) 遮挡目标改进前的推理结果及热力图

(b) 遮挡目标改进后的推理结果及热力图

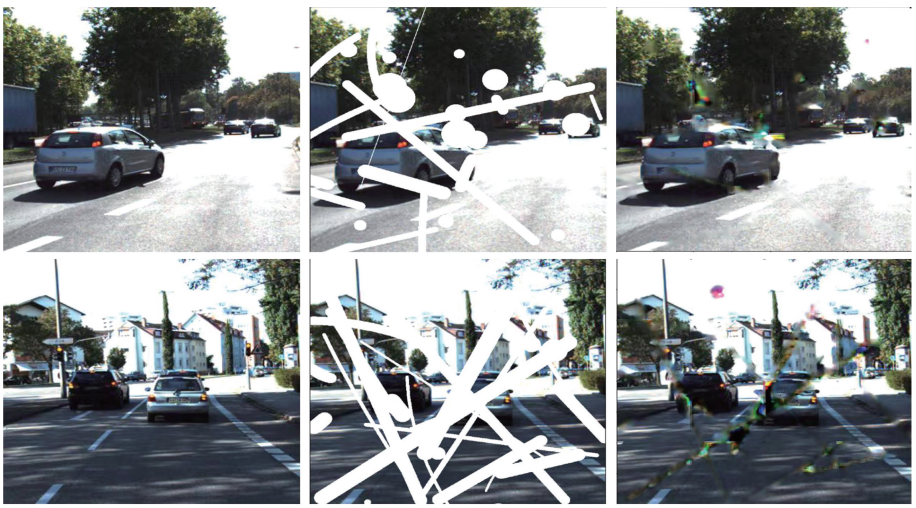
(c) 小目标改进前的推理结果及热力图

(d) 小目标改进后的推理结果及热力图

图 9　模型改进前后效果对比

Fig. 9　Comparison of model performance before and after improvement

为了展示 PConV 在修复图像特征方面的能力，选取 KITTI 数据集中的部分图像进行修复效果展示，如图 10 所示。图 10(a)为原图，图 10(b)为随机生成的 mask，图 10(c)为修复后结果。可以看出，经过 PConV 处理后，图像恢复了缺失的细节和结构，体现出了 PConV 的有效性。



(a) 正常原图　　(b) 加入mask　　(c) 修复后图像

图 10　PConV 图像修复效果展示

Fig. 10　PConV restoration effect display

4 结 论

1)剖析了 YOLOv8-EA 模型的特征提取机制及优化策略,建立了基于 FasterNet、PConV 和 LSKA 的目标检测框架。通过引入 PConV,解决小目标特征缺失问题,提高特征完整度,并构建 SPPLA 实现多尺度特征融合,提升模型对复杂环境的适应能力。

2)颈部网络采用一种基于多分支结构和 reparameterization 的信息抑制方法,通过梯度流的优化提高密集区域遮挡目标的检测能力。同时,基于 PConV 构建小目标检测头,增强模型对小目标像素级特征的捕捉能力。

3)实验表明:在 KITTI 数据集上,YOLOv8-EA 的 mAP50 和 mAP50-95 分别达到了 89.1% 和 63.6%;在 SODA10M 数据集上,mAP50 和 mAP50-95 分别达到了 62.9% 和 39.4%;相较于原始的 YOLOv8 模型,指标提升明显,验证了优化算法在复杂路况下的有效性。未来研究将进一步探索实时检测优化策略,并结合更多复杂场景进行泛化能力评估,以提升模型在多变环境下的目标检测性能。

参考文献

[1] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]//Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014: 580. DOI:10.1109/CVPR.2014.81

[2] REN Shaoqing, HE Kaiming, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137. DOI:10.1109/TPAMI.2016.2577031

[3] 赵锟. 基于深度卷积神经网络的智能车辆目标检测方法研究[D]. 长沙: 国防科学技术大学, 2015
ZHAO Kun. Deep convolutional neural network-based object detection methods with applications to autonomous vehicle [D]. Changsha: National University of Defense Technology, 2015

[4] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C]//Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 779. DOI:10.1109/CVPR.2016.91

[5] REDMON J, FARHADI A. YOLOv3: an incremental improvement [EB/OL]. (2018-04-08) [2024-05-10]. https://arxiv.org/abs/1804.02767

[6] BOCHKOVSKIY A, WANG Chienyao, LIAO Hongyuanmark. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. (2020-04-23) [2024-05-10]. https://arxiv.org/abs/2004.10934

[7] GE Zheng, LIU Songtao, WANG Feng, et al. YOLOX: exceeding YOLO series in 2021 [EB/OL]. (2021-07-18) [2024-05-10]. https://doi.org/10.48550/arXiv.2107.08430

[8] ZHU X, LYU S, WANG X, et al. TPH-YOLOv5: improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios [C]//18th IEEE/CVF International Conference on Computer Vision Workshops, ICCVW 2021. Montreal: IEEE, 2021: 2778. DOI:10.1109/ICCVW54120.2021.00312

[9] WANG Chienyao, BOCHKOVSKIY A, LIAO Hongyuanmark. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023: 7464. DOI:10.1109/CVPR52729.2023.00721

[10] 李经宇, 杨静, 孔斌, 等. 基于注意力机制的多尺度车辆行人检测算法[J]. 光学精密工程, 2021, 29(6): 1448
LI Jingyu, YANG Jing, KONG Bin, et al. Multi-scale vehicle and pedestrian detection algorithm based on attention mechanism[J]. Optics and Precision Engineering, 2021, 29(6): 1448. DOI:10.37188/OPE.20212906.1448

[11] WU Jie, ZHANG Zhian. Research on enhanced multi-task traffic scene object detection and classification method based on improved YOLOv5[C]//4th International Seminar on Artificial Intelligence, Networking and Information Technology, AINIT 2023. Nanjing: IEEE, 2023: 414. DOI:10.1109/AINIT59027.2023.10212696

[12] 田鹏, 毛力. 改进 YOLOv8 的道路交通标志目标检测算法[J]. 计算机工程与应用, 2024, 60(8): 202
TIAN Peng, MAO Li. Improved YOLOv8 object detection algorithm for traffic sign target[J]. Computer Engineering and Applications, 2024, 60(8): 202. DOI:10.3778/j.issn.1002-8331.2309-0415

[13] CHEN Jierun, KAO Shiuhong, HE Hao, et al. Run, don't walk: chasing higher FLOPS for faster neural networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023: 12026. DOI:10.1109/CVPR52729.2023.01157

[14] LIU Guilin, REDA F A, SHIH K J, et al. Image inpainting for irregular holes using partial convolutions[C]//Proceedings of the 15th European Conference on Computer Vision. Munich: Springer, 2018: 89. DOI:10.1007/978-3-030-01252-6_6

[15] LAU K W, PO Laiman, REHMAN Y A U. Large separable kernel attention: rethinking the large kernel attention design in CNN[J]. Expert Systems with Applications, 2024, 236: 121352. DOI:10.1016/j.eswa.2023.121352

[16] KIM Y. Convolutional neural networks for sentence classification [C]//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. Doha: ACL, 2014: 1746. DOI:10.3115/v1/d14-1181

[17] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, 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. DOI:10.1109/TPAMI.2015.2389824

[18] KANG Ming, TING Cheeming, TING Fungfung, et al. RCS-YOLO: a fast and high-accuracy object detector for brain tumor detection[C]//Proceedings of the 26th International Conference on Medical Image Computing and Computer-Assisted Intervention. Vancouver: Springer, 2023: 600. DOI:10.1007/978-3-031-43901-8_57

[19] DING Xiaohan, ZHANG Xiangyu, MA Ningning, et al. RepVGG: making VGG-style ConvNets great again [C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021: 13733. DOI:10.1109/CVPR46437.2021.01352

[20] ZHANG Xiangyu, ZHOU Xinye, LIN Mengxiao, et al. ShuffleNet: an extremely efficient convolutional neural network for mobile devices [C]//31st Meeting of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 6848. DOI:10.1109/CVPR.2018.00716

[21] GEIGER A, LENZ P, STILLER C, et al. Vision meets robotics: the KITTI dataset[J]. International Journal of Robotics Research, 2013, 32(11): 1231. DOI:10.1177/0278364913491297

[22] HAN Jianhua, LIANG Xiwen, XU Hang, et al. SODA10M: a large-scale 2D self/semi-supervised object detection dataset for autonomous driving [EB/OL]. (2021-06-21) [2024-05-10]. https://doi.org/10.48550/arXiv.2106.11118

(编辑 李 美)