

# 基于感受野增强的复杂道路场景目标检测研究<sup>\*</sup>

刘罡<sup>1,2</sup> 闫曙光<sup>2</sup> 刘钰<sup>1</sup> 侯恩翔<sup>2</sup> 黄应征<sup>2</sup>

(1. 无锡学院江苏省集成电路可靠性技术及检测系统工程研究中心 无锡 214105;

2. 南京信息工程大学电子与信息工程学院 南京 210044)

**摘要:** 针对当前自动驾驶场景下道路目标检测算法对远距离小目标、遮挡目标容易出现漏检和误检的问题,提出一种基于改进 YOLOv8n 的道路目标检测算法。在特征提取方面,对感受野注意力卷积进行轻量化改进,重新构造 C2f 模块,以解决卷积计算中参数无法共享问题,使网络有效捕捉关键信息;然后引入轻量化点采样算子,以减少上采样过程中特征细节损失,更好保留图像的细节信息;在特征融合方面,设计多尺度特征融合网络,以增强小目标特征信息,丰富不同尺度特征的双向融合;同时使用归一化注意力机制,以抑制无关背景信息干扰,提高模型抗干扰能力。实验结果表明,提出的改进算法在 KITTI 数据集和 Udacity 数据集上的检测精度分别达到了 92.8% 和 78.7%,相比原始算法分别提高了 2.2% 和 1.6%,模型依然满足轻量化要求,在一定程度上提高了对复杂道路场景的适应能力。

**关键词:** 自动驾驶;目标检测;YOLOv8n;上采样;注意力机制

中图分类号: TP391.4; TN914 文献标识码: A 国家标准学科分类代码: 520.6040

## Research on target detection in complex road scenes based on receptive field enhancement

Liu Gang<sup>1,2</sup> Yan Shuguang<sup>2</sup> Liu Yu<sup>1</sup> Hou Enxiang<sup>2</sup> Huang Yingzheng<sup>2</sup>

(1. Jiangsu Province Engineering Research Center of Integrated Circuit Reliability Technology and Testing System, Wuxi University, Wuxi 214105, China; 2. School of Electronic and Information Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China)

**Abstract:** To address the issue of missed and false detections for distant small objects and occluded objects in current road target detection algorithms in autonomous driving scenarios, a road target detection algorithm based on an improved YOLOv8n is proposed. In terms of feature extraction, the Receptive-Field Attention Convolution is lightweightly improved, and the C2f module is reconstructed to solve the problem of non-shared parameters in convolution calculations, enabling the network to effectively capture critical information. Then, a lightweight point sampling operator is introduced to reduce the loss of feature details during the upsampling process, better preserving image detail information. In terms of feature fusion, a multi-scale feature fusion network is designed to enhance small target feature information and enrich the bidirectional fusion of features at different scales. Simultaneously, a normalization attention mechanism is used to suppress irrelevant background information interference, improving the model's anti-interference capability. Experimental results show that the proposed improved algorithm achieves detection accuracies of 92.6% and 78.7% on the KITTI dataset and the Udacity dataset, respectively, representing improvements of 2.1% and 1.6% compared to the original algorithm. The model still meets lightweight requirements and enhances adaptability to complex road scenes to a certain extent.

**Keywords:** autonomous driving; object detection; YOLOv8n; upsampling; attention mechanism

## 0 引言

伴随着人工智能以及通信技术的不断发展,自动驾驶

以及辅助驾驶技术已经变得越来越成熟。自动驾驶汽车在道路上行驶过程中,需要持续地通过安装在车身周围的摄像头和雷达来获取周围瞬息万变的环境信息,并根据获取

到的信息进行处理采取相应的措施<sup>[1]</sup>。在实际复杂交通环境中,快速准确的检测目标对于自动驾驶的行车安全至关重要。但是实际道路场景情况复杂多变,比如车辆与行人障碍物的互相遮挡、夜间光照条件不佳、恶劣天气干扰等也给当前的目标检测算法提出了更多的挑战<sup>[2]</sup>。

近年来,得益于计算机硬件的快速发展,基于深度学习的卷积神经网络目标检测算法成为主流,其在图像分类和目标追踪等领域得到了广泛应用,通过神经网络可以学习目标的高级语义特征,具备优异的检测性能。基于卷积神经网络的目标检测算法主要分为两类:一是以区域卷积神经网络(region convolutional neural networks, RCNN)系列<sup>[3]</sup>为代表的两阶段检测算法,如 Fast R-CNN<sup>[4]</sup>, Faster R-CNN<sup>[5]</sup>等;二是以 SSD (single shot multibox detector)<sup>[6]</sup>、YOLO(you only look once)系列<sup>[7]</sup>为代表的单阶段算法,如 YOLOv5<sup>[8]</sup>, YOLOv7<sup>[9]</sup>等。两阶段算法虽然检测精度较高,但网络参数复杂,检测速度慢,难以满足实时检测要求。单阶段检测算法直接利用网络检测目标位置和类别等信息,网络结构简单检测速度快,能实现实时检测。

YOLO 系列作为单阶段算法的主流,拥有较高的检测速度,但在精度等方面还存在不足,较多研究人员提出多种改进策略。其中,史涛等<sup>[10]</sup>针对小目标难检测问题,通过增加小目标检测头,设计下采样模块,重新构建金字塔网络,提高了对小目标的检测精度,但检测速度变慢;原蕾等<sup>[11]</sup>在 YOLOv7 模型基础上改进特征融合过程,以及调整瓶颈注意力机制的分支结构,减少了特征信息的损失,但模型参数量过大,检测帧率较低。魏陈浩等<sup>[12]</sup>针对道路场景多尺度目标,在 YOLOv8n 模型中通过增加小目标检测层,在特征提取中引入可变形卷积,同时增加双层路由注意力机制,更好地适应了道路目标复杂多变的场景,有效降低了背景信息干扰,但检测的准确率不高;高德勇等<sup>[13]</sup>提出一种改进的 YOLOv8n 道路目标检测算法,通过引入多样化分支块,改进特征金字塔融合方式,设计三重注意力机制模块,提升了对多尺度目标的识别能力,但检测精度有待提高。

在现阶段上述算法虽然在一些方面作了改进,但在面对复杂道路背景以及目标尺度多变的道路环境中,还存在较大改进空间,性能亟待提升。YOLOv8 是目前 YOLO 系列较为成熟的算法,在前几代版本的基础上,引入了较多新的策略和架构创新。YOLOv8n 作为 YOLOv8 的最小模型,检测速度快,能很好应对常规检测目标,但在应对实际道路场景下目标尺度多变、目标互相遮挡以及背景干扰等挑战,存在漏检和误检率较高等问题。因此为应对上述挑战,本文以 YOLOv8n 轻量化模型为框架进行改进,提出一种轻量化的改进 YOLOv8n 检测算法。主要贡献如下:通过改进感受野注意力卷积(receptive-field attention conv, RFACConv)<sup>[14]</sup>,使其更为轻量化高效,保留目标的突出特征

信息,提升不同目标特征提取的重要性;同时在颈部(neck)网络部分加入新的特征融合层,加强高层次特征融合,增强特征提取能力;此外,引入轻量化点采样算子 Dysample<sup>[15]</sup>,增强特征提取的感知范围;最后,为消除负样本干扰,加入归一化注意力机制(normalization-based attention module, NAM)<sup>[16]</sup>,对特征权重进行分配,降低负面影响。

## 1 基于改进 YOLOv8n 的目标检测算法

### 1.1 YOLOv8n 网络模型

YOLO 系列模型因为其准确度高和模型参数量少在计算机视觉领域得到了广泛应用。YOLOv8 网络主要由骨干(backbone)、颈部(neck)、检测头(head)3 部分构成,按照模型的深度和宽度从小到大分为 n、s、m、l、x 五个版本。骨干网络采用 Darknet-53 框架,将 C3 模块替换为 C2f, C2f 模块通过更多的分支跨层连接,进一步丰富了模型的梯度流。此外骨干网络末端还保留了 YOLOv5 网络中使用的快速空间金字塔池化(spatial pyramid pooling fast, SPPF)模块<sup>[17]</sup>,能够有效捕捉不同尺度的特征。在头部网络,采用解耦头结构,将分类和检测分离,使网络的训练和推理更加高效,同时在检测中采用效果更好的无锚框机制<sup>[18]</sup>。此外,框匹配策略由静态匹配改为了正负样本动态分配策略<sup>[19]</sup>,可以根据训练损失或者其他指标动态调整权重,更好地适应不同的数据集。本文采用 YOLOv8n 作为基准模型,虽然 YOLOv8n 模型体积足够小,检测速度快,但在面临一些复杂环境时容易出现检测精度不高,小目标漏检和误检严重等情况,本文对 YOLOv8n 模型进行优化,在速度和精度之间达到一个不错的平衡。

### 1.2 改进的 YOLOv8n 目标检测模型

当前道路环境较为复杂,车多人多且互相遮挡,光照条件多变,存在较多背景干扰,为了解决在复杂道路场景下远距离小目标和被遮挡目标检测精度低,漏检率和误检率较高等问题,本文对 YOLOv8n 模型进行以下方面改进:针对复杂道路场景下存在较多远距离小目标,导致特征融合不完全、感受野缺失等问题,改进轻量化感受野注意力卷积 SRFAConv,引入 C2f 模块,对不同大小目标使用不同的感受野,提高不同目标特征提取的重要性;同时对加权双向特征金字塔 (bi-directional feature pyramid network, BiFPN)<sup>[20]</sup> 进行改进,加入 P2 目标层进行特征融合,保留更多小目标细节信息,使其对于小目标更加敏感。针对道路场景背景复杂,模型易受负面信息干扰的问题,引入一种轻量化点采样算法 Dysample,改进特征提取过程,最大程度保留原始特征信息;与此同时加入 NAM 注意力机制,通过降低不重要特征信息的权重,消除背景干扰,改进后的 YOLOv8n 网络结构如图 1 所示。

#### 1) 轻量化感受野注意力卷积

YOLOv8 在进行特征提取时通常采用  $3 \times 3$  卷积,虽然运算简单,但感受野较小,对于小目标特征信息关注不

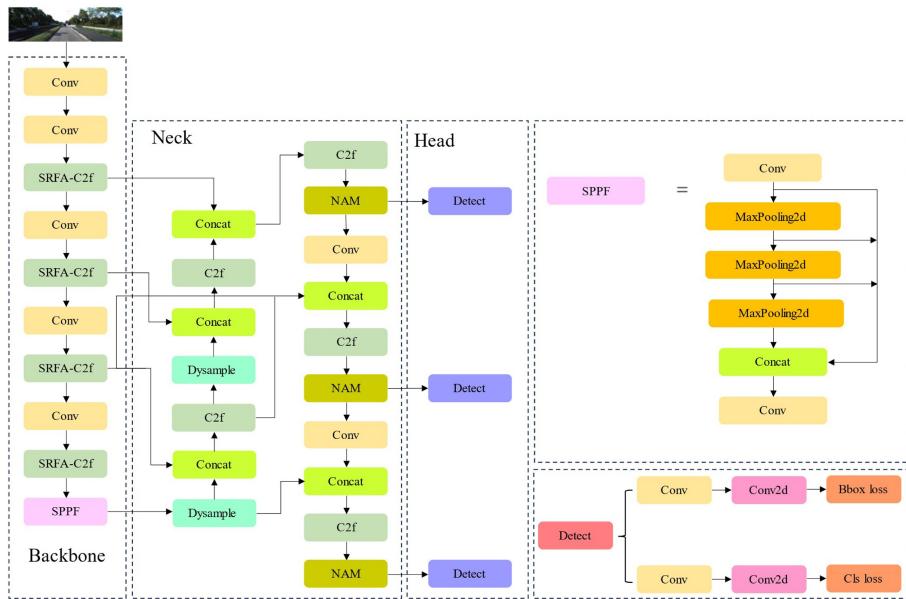


图 1 改进的 YOLOv8n 网络结构

Fig. 1 Improved YOLOv8n network architecture

够,同时缺乏相关空间特征信息。坐标注意力机制 (coordinate attention, CA)<sup>[21]</sup> 和卷积块注意力机制 (convolutional block attention module, CBAM)<sup>[22]</sup> 等空间注意力机制证明了关注空间特征信息的有效性,能显著提升模型性能。然而,它们只关注于空间特征信息,却没有考虑来自不同位置特征信息之间的差异,同时也存在卷积核参数无法共享的问题。而 RFACConv 从本质上解决了卷积核的参数共享问题,使卷积核能够平等地处理每个像素,加入的感受野注意力使感受野内的空间特征更加集中,通过调整感受野内特征图的权重分布,突出不同位置重要的特征信息。受 RFACConv 的启发,本文对 RFACConv 进行改进,使其更为高效和轻量化,并命名为 SRFAConv,提升模型在复杂环境下的识别能力。

SRFAConv 的结构如图 2 所示,主要分为两部分:提取感受野空间特征和生成注意力特征图。首先进行感受野空间特征的提取,感受野空间特征根据卷积核的大小动态生成,通过滑动窗口对原始特征映射进行变换,之后利用分组卷积动态生成展开特征,将原始特征映射为新特征,突出感受野滑块内不同特征的重要程度。每个滑动窗口都代表一个感受野滑块,可以生成不同大小的感受野空间特征。最后使用 SoftMax 函数调整感受野内每个特征图的权重,最终生成感受野特征图。

在生成注意力特征图时,会对输入的特征图进行平均池化操作,保留全局信息的同时聚集所有感受野的信息,但这样会导致缺少目标的关键特征信息,尤其是在复杂道路环境中存在不同大小的目标,受到背景信息干扰,需要保留目标的突出特征,以便提取有效信息。本文采用平均池化+最大池化相结合的方式,首先进行平均池化更好地

捕捉特征图中的整体信息,并降低噪声;接着进行最大池化,可以在保留尺寸变换的同时,强调目标的最突出特征,更好地提取目标的边缘以及纹理信息。之后通过将组卷积替换为更为轻量化的深度可分离卷积 (depthwise separable convolution, DWConv)<sup>[23]</sup>, 实现对 RFACConv 的轻量化改造。

将 C2f 模块的 Bottleneck 替换为更为轻量化高效的 SRFAConv,将传统的标准卷积与空间注意机制结合,提高网络捕捉关键信息的能力。利用分组卷积提取感受野特征,通过使用平均池化和最大池化对每个感受野特征的全局信息进行聚合,同时保留突出特征信息,使不同滑块内的信息能够互相交互,进一步改善网络特征融合。

## 2) 多尺度特征融合模块

传统的目标检测网络特征金字塔 (feature pyramid networks, FPN) 结构采用自顶向下的单向信息流进行特征融合<sup>[24]</sup>, YOLOv8 系列采用 FPN+PAN 的方式进行特征融合<sup>[25]</sup>, 加入自底向上的路径增强融合信息流,同时通过横向连接融合各个尺度信息,这种方式有效增强了信息流的传输,但原始特征在经过不断的下采样和上采样后会丢失一些关键特征信息,不利于小目标检测, BiFPN 网络则在 PANet 的基础上进一步细化了特征融合的过程,改善了存在的问题。

BiFPN 的网络结构如图 3(a) 所示, BiFPN 采用双向融合策略,删除了 PANet 中的冗余节点和连接,增加跳跃连接进行跨尺度特征融合,使网络更多保留原始信息。同时重新配置自上而下和自下而上的路径,简化了网络路径,并确保不同尺度和层级的信息能够融合,提高网络处理多尺度目标的能力。此外,引入可以学习的权重,通过自适

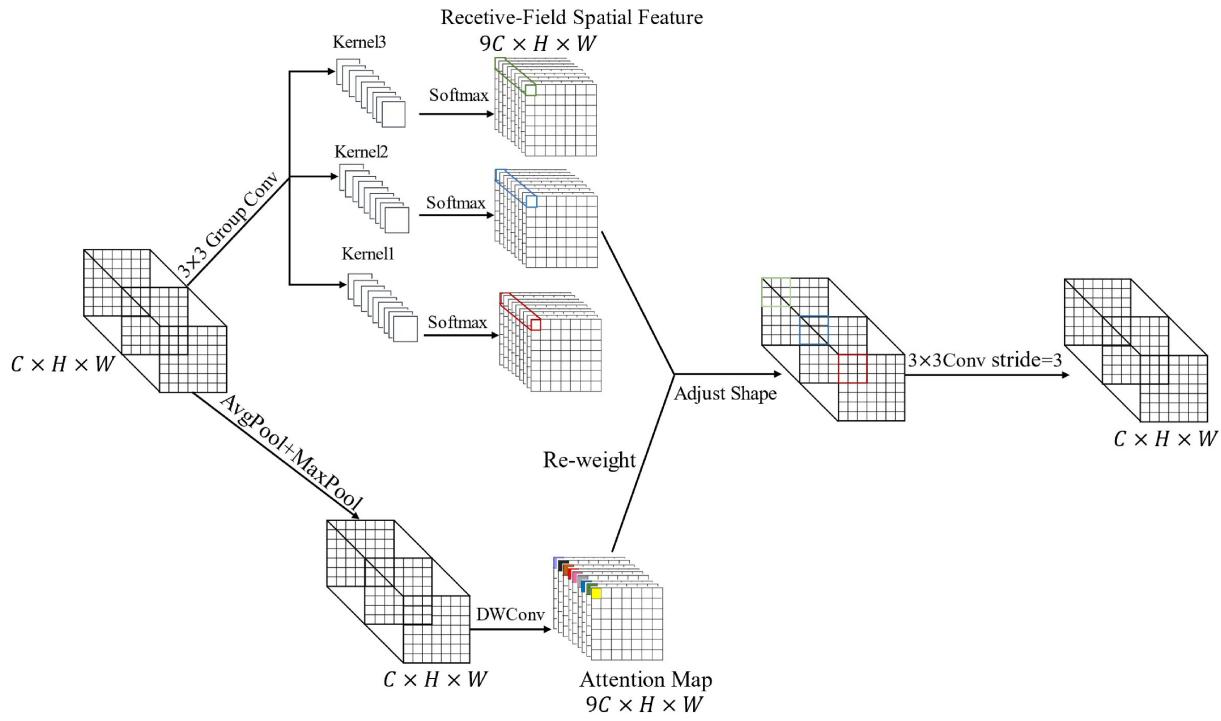


图 2 SRFAConv 结构

Fig. 2 SRFAConv structure

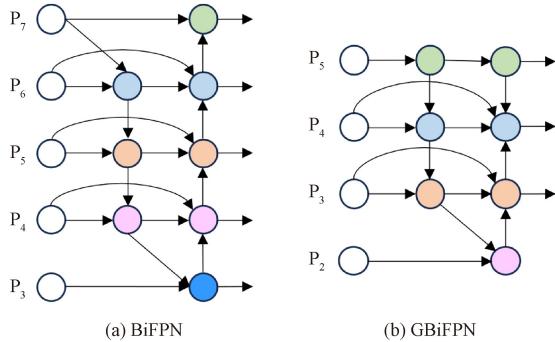


图 3 BiFPN 和 GBiFPN 结构

Fig. 3 BiFPN and GBiFPN structures

应权值调整优化特征融合过程,学习不同输入特征的重要性,以调整不同层级的特征更好地匹配不同任务的需求。

为了进一步提高模型检测不同尺度交通目标的能力,本文在 BiFPN 基础上进一步优化,通过加入跨尺度连接增强 BiFPN 特征融合能力,减少特征信息丢失,同时不会增加较多计算成本。提出的 GBiFPN 如图 3(b)所示,通过增加 P2 浅特征层进行特征融合,浅层特征图包含较多的细节信息,分辨率更高,更适合小目标检测,能够更敏锐地定位和检测小目标,提高小目标检测的准确性。同时在 P3 层引入跳跃连接,使特征信息在不同层级之间直接传递,更好的保留不同层级的语义信息。特征之间的连接方式采用 Conact 进行拼接,最大程度保留各层次的特征信息,提高网络的整体感知能力,改善小目标检测效果。

### 3) 轻量化上采样算子 DySample

特征上采样算法对于物体检测效果方面有着举足轻重的作用。上采样操作通过将小分辨率特征图映射到大分辨率特征图上,将图像恢复到原来的尺寸,以便更好的捕捉细节和特征信息,常见的上采样方法有最近邻插值和双线性插值等。YOLOv8 模型采用最近邻插值算法,这种算法虽然所需计算资源较少,速度快,但它只利用了像素的空间位置来确定上采样核,无法利用特征图的语义信息,导致感知范围有限,只有  $1 \times 1$  个像素。

为了改进存在的问题,采用一种新的超轻量级 DySample 上采样模块。与最近的一些基于内核的动态上采样器相比,DySample 采用一种点采样的方法,可以在较低的计算资源下实现高效的上采样效果。网格采样过程如图 4 所示,其中  $x$  为输入特征,  $x'$  为上采样特征,采样点生成器生成采样集,输入特征通过网格采样函数生成上采样特征。图 5 是通过网格采样设计的静态点采样器和动态点采样器,其中采样集是生成的偏移量与原始的网格位置之和,  $o$  代表偏移量,通过线性投影生成。图 5(a)为静态范围因子版本,图 5(b)为动态范围因子版本,静态范围因子在采样过程中不会发生变化,动态范围因子在采样过程中会不断调整,  $\sigma$  为 sigmoid 函数,可以用于调制偏移量的大小。

采样过程如下:首先利用双线性初始化方法确定初始采样点位置,之后通过线性投影+像素洗牌生成偏移量进行内容感知偏移构造新的采样点,内容感知偏移可以看作

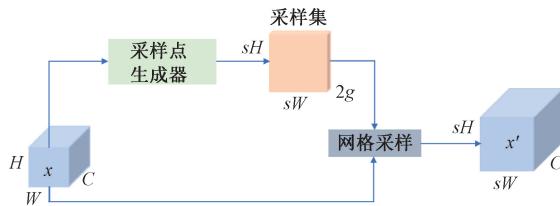


图4 网格采样

Fig. 4 Grid sampling

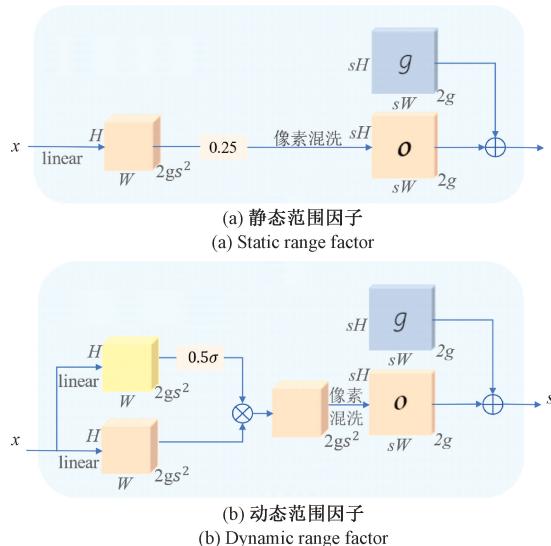


图5 静态和动态采样点生成器

Fig. 5 Static and dynamic sampling point generators

是为每个点在其周围寻找一个语义上差异较大的区域,将偏移量乘以静态范围因子 0.25 以限制采样点位置,防止发生重叠。之后将特征图按照通道维度分组,分组进行上采样操作,最后每个采样点用双线性插值对输入特征映射进行重采样。利用全新的点采样过程,能够对恢复特征信息的过程更加细腻,特别是对于一些密集场景或者存在部分遮挡场景更加有效,能够使图像边缘细节信息更加清晰,使模型更容易检测和定位。

#### 4) NAM 注意力机制

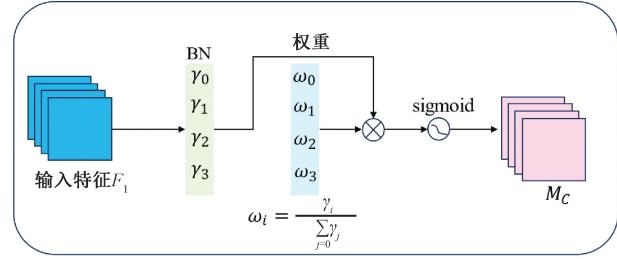
道路场景下背景复杂,模型在进行特征提取和特征融合时很容易受到背景的干扰,为了降低干扰,提升模型准确性,在网络中引入轻量级 NAM 注意力机制。NAM 注意力机制是归一化的注意力机制,通过在注意力模块上应用稀疏的权重惩罚,可以降低不太显著的特征权重,这使得模型在计算时更加高效。

轻量化 NAM 注意力采用了 CBAM 的模块集成方式,并重新设计了通道注意力模块和空间注意力模块,如图 6 所示,通道注意力模块如图 6(a)所示,空间注意力模块如图 6(b)所示。就通道注意力模块而言,通过使用批量归一化(batch normalization, BN)的缩放因子得到通道方差,可以表示出每个通道的重要程度,方差越大代表该通

道变化越明显,包含的信息也会越丰富,批归一化缩放因子的计算如式(1)所示。

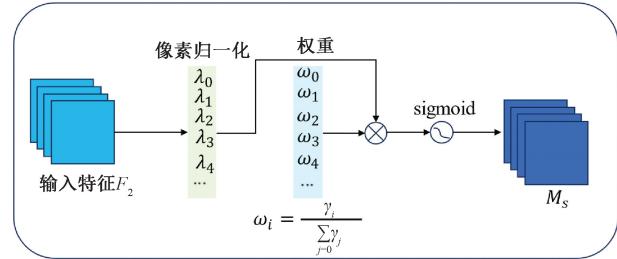
$$B_{\text{out}} = BN(B_{\text{in}}) \gamma \frac{B_{\text{in}} - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} + \beta \quad (1)$$

式中:  $B_{\text{in}}$  和  $B_{\text{out}}$  分别表示输入的小批量  $B$  和经归一化后输出的结果,  $\gamma$  和  $\beta$  为可训练的缩放变换参数,  $\mu_B$  和  $\sigma_B$  分别为小批量  $B$  的均值和标准差,  $\epsilon$  为常数,用于保证数值稳定。



(a) 通道注意力子模块

(a) Channel attention sub-module



(b) 空间注意力子模块

(b) Spatial attention sub-module

图6 NAM 注意力机制

Fig. 6 NAM attention mechanism

通道注意力子模块的输出特征为式(2)所示,其中,  $M_c$  表示输出特征,  $F_1$  为输入特征,  $W_\gamma$  代表通道权重, 同时权重的计算公式如式(3)所示,  $\gamma_i$  为比例因子。

$$M_c = \text{sigmoid}(W_\gamma(BN(F_1))) \quad (2)$$

$$W_\gamma = \frac{\gamma_i}{\sum_{j=0}^n \gamma_j} \quad (3)$$

同理,空间注意力子模块的输出特征如式(4)所示,其中,  $M_s$  为对应输出特征,  $W_\lambda$  为空间权重,权重的计算公式如式(5)所示,  $\lambda_i$  为比例因子。

$$M_s = \text{sigmoid}(W_\lambda(BN(F_2))) \quad (4)$$

$$W_\lambda = \frac{\lambda_i}{\sum_{j=0}^n \lambda_j} \quad (5)$$

## 2 仿真实验与结果分析

### 2.1 数据集

本节提供了各种数值仿真来验证所提算法的有效性。实验部分主要采用公开自动驾驶数据集 KITTI 进行算法验证。同时,为进一步验证模型的泛化与抗干扰能力,在

另一公开自动驾驶数据集 Udacity 上进行辅助验证。

### 1) KITTI 数据集

KITTI 数据集由卡尔斯鲁厄理工学院 (KIT) 和丰田工业大学芝加哥分校 (TTIC) 共同创立, 是目前国际上较大的自动驾驶场景下的交通数据集<sup>[26]</sup>。KITTI 训练集包含真实图像数据共 7 481 张图片, 还有未完成标注的图像 7 518 张, 将其作为测试集使用, 图像分辨率大部分为  $1242 \times 375$  像素, 共分为 8 个类别, 考虑到真实自动驾驶场景的情况, 将其中的 8 个类别合并删除为 3 个类别, 3 个类别分别是: 汽车, 行人和骑车的人。本文将数据集按照 8 : 2 (5 985 : 1 496) 的比例随机划分为训练集和验证集。

### 2) Udacity 数据集

Udacity 数据集<sup>[27]</sup>是专门为自动驾驶算法比赛所准备的数据集, 对连续的视频图片进行了标注, 共包含 15 000 张城市道路照片, 与 KITTI 数据集相比, 采集环境光线变化更加强烈, 背景干扰更为复杂。将数据集的 11 个类别保留为 4 个类别, 删除 Trafficlight 类标签, 按照 8 : 1 : 1 的比例划分训练集、验证集和测试集。

## 2.2 实验环境和实验参数设置

本次实验采用的实验平台配置为 Windows10 操作系统, 显卡为 RTX3090、CPU 为 Xeon(R) Platinum 8352V、CUDA 版本为 11.3, 以 Pytorch1.11.0 为网络框架。采用随机梯度下降优化器 (SGD), 总训练轮数为 300, 批次大小设置为 32, 其他参数采用默认值。

## 2.3 评价指标

本文采用均值平均精度 (mean average precision, mAP) 和每秒传输帧数 (frames per second, FPS) 衡量改进模型的检测性能, mAP 为检测所有类别的平均精度, FPS 为每秒检测图片的数量。精确率 (Precision) 和召回率 (Recall) 的计算如式(6)和(7)所示。

$$P = \frac{TP}{TP + FP} \quad (6)$$

$$R = \frac{TP}{TP + FN} \quad (7)$$

式中:  $P$  为精确率,  $R$  为召回率,  $TP$  为被正确检测到的目标数量,  $FP$  为被误检的目标数量,  $FN$  为没有被检测到的目标数量。

召回率和精确率两个曲线与坐标轴围成的面积为检测精度 (average precision, AP), 如式(8)所示。

$$AP = \int_0^1 P(R) dR \quad (8)$$

对有  $N$  个类别的检测目标, 各类别平均精度  $mAP$  可以表示为:

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (9)$$

式中:  $N$  为目标的类别数量,  $AP_i$  为第  $i$  个类别的检测精度。

FPS 是衡量模型检测速度的重要指标, 计算公式如

式(10)所示。

$$FPS = \frac{1000}{P + I + N} \quad (10)$$

式中:  $P$  为图像前处理时间,  $I$  为模型推理时间,  $N$  为后续处理时间。

## 2.4 注意力机制对比实验

为了验证 NAM 注意力模块的优越性, 将其与 CA 注意力模块、CBAM 注意力模块和全局注意力模块 (global attention mechanism, GAM)<sup>[28]</sup> 在 KITTI 数据集上进行对比实验, 结果如表 1 所示。

表 1 注意力机制对比

Table 1 Comparison of attention mechanisms

模型	参数量	mAP@0.5/%
YOLOv8n	3 011 433	90.6
YOLOv8n+CA	3 023 153	90.3
YOLOv8n+CBAM	3 114 606	91.0
YOLOv8n+GAM	3 446 889	90.7
YOLOv8n+NAM	3 084 713	91.1

实验结果表明, 添加 NAM 注意力模块精度提升最高, mAP@0.5 有效提升了 0.5 个百分点, 但参数量却只有少量增加; 虽然添加 CBAM 注意力模块精度也有 0.4 个百分点的提升, 但参数引入量却比 NAM 注意力大; 而 GAM 注意力模块引入参数最多, 精度却只提升了 0.1 个百分点。因此综合考虑, YOLOv8n+NAM 的效果最好, 通过利用稀疏权重抑制无关紧要的特征图权重, 减少了复杂背景信息的干扰。

## 2.5 上采样模块对比实验

为了进一步验证轻量级点采样算子 DySample 是否能够改进特征上采样过程, 将小分辨率特征图恢复到大分辨率特征图, 并且最大程度保留局部细节信息, 提升模型的特征提取能力, 将改进后的模型与原始模型和同样是轻量化上采样算子的内容感知特征重组 (content-aware reassembly of features, CARAFE)<sup>[29]</sup> 改进的模型作进一步对比实验, 如表 2 所示。

表 2 上采样模块对比

Table 2 Comparison of upsampling modules

模型	参数量	mAP@0.5/%	FPS
YOLOv8n	3 011 433	90.6	120
YOLOv8n+CARAFE	3 151 865	90.9	111
YOLOv8n+DySample	3 023 785	91.0	116

从表 2 可以看出, 加入 DySample 上采样算子后, 参数量仅有小幅增加, 说明引入计算量较小, 推理速度方面有所下降, 且检测精度上升有限。因此, 从检测精度和速度方面考虑, DySample 比 CARAFE 更具优势, 更加契合改

进模型。

## 2.6 消融实验

为了验证 DySample、RFAConv、GBiFPN 和 NAM 注

意力机制对模型性能的提升效果,采用控制变量法对 KITTI 数据集进行消融实验,进一步探究各个模块的有效性,实验结果如表 3 所示。

表 3 消融实验

Table 3 Ablation study

SRFACConv	DySample	GBiFPN	NAM	mAP@0.5/%	mAP@0.5:0.9/%	GFLOPS
×	×	×	×	90.6	64.8	8.2
√	×	×	×	91.3	66.3	8.4
×	√	×	×	91.0	65.7	8.2
×	×	√	×	91.5	67.2	8.7
×	×	×	√	91.1	66.9	8.3
√	√	×	×	91.8	68.3	8.5
×	×	√	√	92.0	69.1	8.8
√	√	√	√	92.8	70.8	9.2

表 3 中“√”表示模型加入了对应的模块,“×”表示模型没有使用该模块。由表可知,加入轻量级 SRFACConv 后,参数量上升有限,通过对不同大小的目标使用不同的感受野,以及对一些关键特征进行重加权,提高了网络特征提取的效率和整体性能,使 mAP@0.5 提升了 0.7 个百分点;在原模型基础上使用轻量化上采样 DySample 算子,模型大小没有发生变化,但检测精度小幅提升 0.4%,在不增加算法复杂性的同时,进一步提升了精确度;将改进的 GBiFPN 加入后,提升了网络的特征融合能力,增强了对于小目标的识别和检测,使检测精度值提升了 0.9 个百分点;NAM 注意力的加入抑制了背景噪声的干扰,检测精度提高了 0.5 个百分点。在将所有改进模块加入后,各模块之间耦合性能良好,检测精度均得到不同程度提升,证明了本文所添加模块的有效性。

## 2.7 对比实验

为了进一步验证本文算法的优越性,将改进后的算法

与其他目标检测算法进行对比,实验结果如表 4 所示。从实验结果可以看出,改进后的 YOLOv8n 模型与基线模型相比,在增加少量参数量的基础上精度获得了较大提升,参数量只增加了 1 GFLOPS,检测精度却提高了 2.2 个百分点,同时检测速度依然处于前列。与同类型的 YOLOv4, YOLOv5s, YOLOv7-tiny 等模型相比,参数量降低的同时,检测精度也有着显著优势。与最新改进的 YOLOv8n 模型 SES-YOLOV8n 相比,在实验环境相同的情况下, mAP@0.5 基本相同, mAP@0.5:0.95 有 1% 的领先,但本文在检测速度方面有着明显的优势,更加轻量化,说明 SES-YOLOV8n 模型参数量较大,本文以较低的参数量实现了相同的检测精度,更加具有轻量化,证明了本文方法的先进性。从整体上看,本文提出的算法在拥有最高的检测精度的同时,依然保持着轻量化的优势,检测速度仍然满足实时检测要求,更适合部署在一些嵌入式设备,证明了本文所改进算法的优越性。

表 4 对比实验

Table 4 Comparative experiments

模型	精确率/%	召回率/%	mAP@0.5/%	mAP@0.5:0.95/%	GFLOPS	FPS
Faster R-CNN	32.3	65.7	54.5	28.4	220.2	19
SSD	89.7	64.4	72.4	48.0	62.3	54
YOLOv4	63.2	72.0	86.9	—	20.6	46
YOLOv5n <sup>[30]</sup>	84.9	75.4	83.3	55.0	4.5	80
YOLOv5s	92.5	84.1	91.6	68.4	16.5	97
YOLOv7-tiny <sup>[31]</sup>	88.3	80.8	86.9	55.5	13.2	86
YOLOv8n	91.7	82.7	90.6	65.3	8.2	120
SES-YOLOV8n	92.5	86.5	92.7	69.2	—	68
本文	92.4	86.7	92.8	70.2	9.2	104

## 2.8 检测效果分析

为了更直观的体现所改进模型的性能,在测试集中随机抽取一定比例图片进行测试,测试结果如图 7 所示,包含光照不均,远距离小目标,被遮挡目标等不同场景。在图 7(a)光线较强存在较多远距离小目标情况下,原始模型将图中左下部分小目标广告牌误检为行人,同时较远处的行人小目标也存在漏检一人,改进网络通过增加 P2 层进行特征融合,改善了识别小目标的能力,对于小目标的敏感程度更高,检测能力更强;在图 7(b)光照不均以及存在大量密集遮挡的情况下,检测出了更多的被遮挡的小目标,以及与背景类似被漏检的行人,精确度也有显著提高,说明融入改进的 SRFAConv 后,网络聚焦关键位置特征的能力更强;在图 7(c)阴影背光路口环境中,检测到了远处过曝的小目标,以及改善了错检的问题,说明加入归一化注意力机制后,降低了环境带来的干扰,使网络更加专注于正样本目标信息。

为了进一步验证模型的泛化性,将改进后的 YOLOv8n 模型继续在 Udacity 数据集上进行训练与测试,结果如表 5 所示。从表 5 可以看出,改进后的模型对比原始模型各参数均获得不同程度提升,其中精确度和召回率分别提升了 0.3 个百分点和 2.3 个百分点,召回率提升明显,平均精度提升了 1.6 个百分点。

表 5 Udacity 数据集模型性能对比

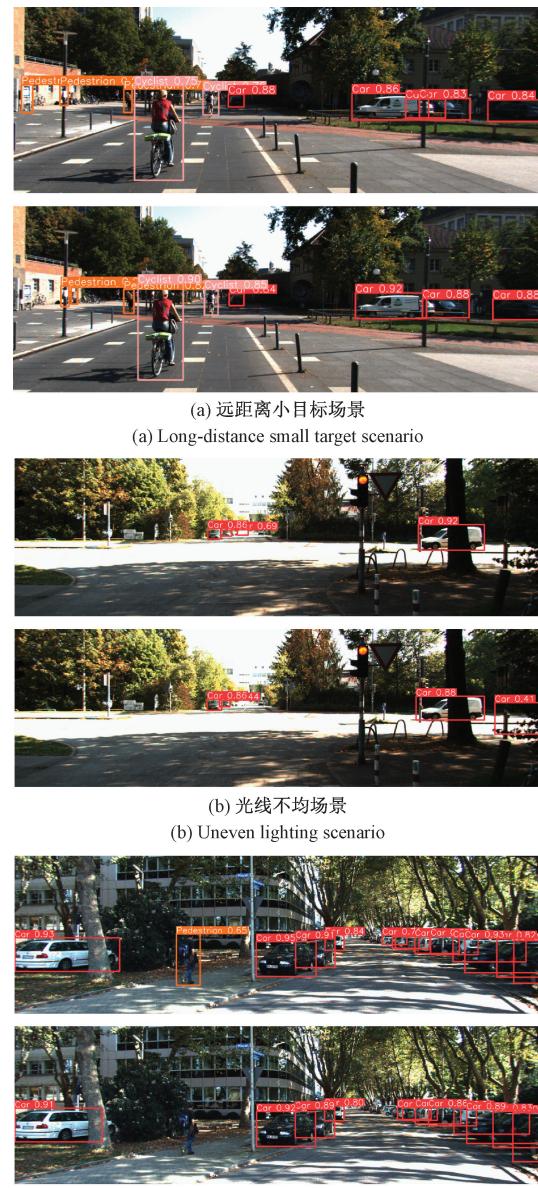
Table 5 Performance comparison of models on the Udacity dataset

模型	精确率	召回率	mAP@0.5	%
YOLOv8n	83.0	70.1	77.1	
改进模型	83.3	72.4	78.7	

Udacity 数据集的检测效果如图 8 所示,Udacity 数据集的第 1 张图远处阴影下的小目标原始模型没有检测到;第 2 张图为密集遮挡场景,原始模型将远处的广告牌识别为卡车,同时被遮挡的汽车也存在漏检,而改进 YOLOv8n 模型很好解决了漏检和误检现象,检测精度也有一定提升。可以看出,改进 YOLOv8n 模型降低了远距离小目标和被遮挡目标的漏检率和误检率,在复杂道路场景中有着更好的鲁棒性,更卓越的检测性能。

## 3 结 论

针对现实道路环境下背景复杂,存在遮挡和光照不均等情况导致的精度低、漏检和误检等情况,提出改进的 YOLOv8n 道路目标检测模型。在保持模型轻量化的基础上,融入改进的轻量级感受野注意力卷积 SRFAConv,使网络聚焦关键特征能力更强,提升对于不同大小目标的识别率;采用轻量化上采样算子 Dysample,使模型更好理解特征图中的语义信息,减少上采样过程的特征损失。此



(c) Shadow backlit obscures a large area of the scene

图 7 KITTI 数据集检测效果对比

Fig. 7 Detection performance comparison on the KITTI dataset

外,将改进后的 BiFPN 网络进行特征融合,通过引入小目标融合层,提高了对于小目标的敏感度,更容易检测到远距离小目标;归一化注意力机制 NAM 的引入通过稀疏权重惩罚不重要的特征权重,有效抑制了无关背景信息的干扰。

实验结果表明,改进算法在自动驾驶数据集 KITTI 和 Udacity 上表现良好,检测精度分别提升了 2.2 个百分点、1.6 个百分点。与原模型及其他算法相比,在控制模型大小的基础上,实现了更高的检测精度和较快的检测速度,有效降低了小目标和遮挡目标的漏检和误检率,更易于在一些嵌入式终端部署。在后续工作中,还需要提升模型的

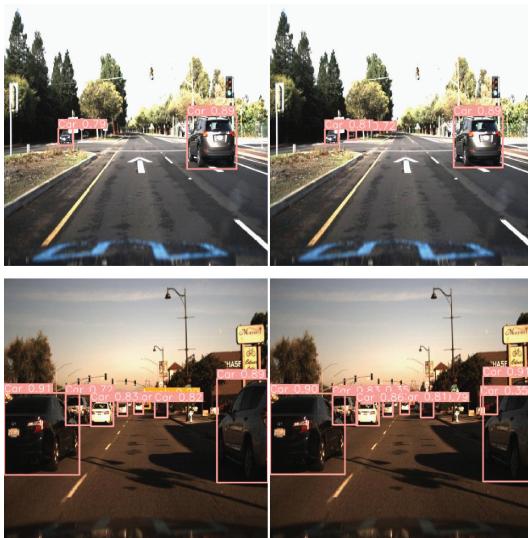


图 8 Udacity 数据集检测效果对比

Fig. 8 Detection performance comparison on the Udacity dataset

鲁棒性,增强抗干扰能力,以满足在更加复杂多变的环境中稳定工作。

## 参考文献

- [1] LIU L K, LU S D, ZHONG R, et al. Computing systems for autonomous driving: State of the art and challenges[J]. IEEE Internet of Things Journal, 2020, 8(8): 6469-6486.
- [2] GRIGORESCU S, TRANSNEA B, COCIAS T, et al. A survey of deep learning techniques for autonomous driving[J]. Journal of Field Robotics, 2020, 37(3): 362-386.
- [3] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.
- [4] GIRSHICK R. Fast R-CNN[C]. IEEE International Conference on Computer Vision, 2015: 1440-1448.
- [5] REN SH Q, HE K M, 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, 2016, 39(6): 1137-1149.
- [6] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]. Computer Vision-ECCV 2016, 2016: 21-37.
- [7] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.
- [8] GLENNJ, ALEXS, JIRKAB. YOLOv5 [EB/OL]. [2021-04-12]. <https://github.com/ultralytics/yolov5>.
- [9] WANG CH 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/CVF Conference on Computer Vision and Pattern Recognition, 2023: 7464-7475.
- [10] 史涛, 刘祖林, 朱文旭, 等. 基于改进 YOLOv5s 的车辆行人检测[J]. 国外电子测量技术, 2023, 42(12): 195-200.
- SHI T, LIU Z L, ZHU W X, et al. Vehicle and pedestrian detection based on improved YOLOv5s[J]. Foreign Electronic Measurement Technology, 2023, 42(12): 195-200.
- [11] 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. Improved YOLOv7 vehicle detection method based on attention mechanism and feature fusion [J]. Foreign Electronic Measurement Technology, 2023, 42(9): 49-57.
- [12] 魏陈浩, 杨睿, 刘振丙, 等. 具有双层路由注意力的 YOLOv8 道路场景目标检测方法[J]. 图学学报, 2023, 44(6): 1104-1111.
- WEI CH H, YANG R, LIU ZH B, et al. YOLOv8 road scene object detection method with dual-route attention[J]. Journal of Image and Graphics, 2023, 44(6): 1104-1111.
- [13] 高德勇, 陈泰达, 缪兰. 改进 YOLOv8n 的道路目标检测算法[J/OL]. 计算机工程与应用, 2024: 1-14 [2024-07-15]. <http://kns.cnki.net/kcms/detail/11.2127.TP.20240528.1330.002.html>.
- GAO D Y, CHEN T D, MIAO L. Improved road object detection algorithm based on YOLOv8n [J/OL]. Computer Engineering and Applications, 2024: 1-14 [2024-07-15]. <http://kns.cnki.net/kcms/detail/11.2127.TP.20240528.1330.002.html>.
- [14] ZHANG X, LIU CH, YANG D G, et al. Rfaconv: Innovating spatial attention and standard convolutional operation [J]. ArXiv preprint arXiv: 2304.03198, 2023.
- [15] LIU W Z, LU H, FU H T, et al. Learning to upsample by learning to sample [C]. IEEE/CVF International Conference on Computer Vision, 2023: 6027-6037.
- [16] LIU Y CH, SHAO Z R, TENG Y Y, et al. NAM: Normalization-based attention module [J]. Arxiv preprint arxiv:2111.12419, 2021.
- [17] HE K M, ZHANG X Y, REN SH Q, et al. Spatial

- pyramid pooling in deep convolutional networks for visual recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37 (9): 1904-1916.
- [18] TIAN ZH, SHEN CH H, CHEN H, et al. FCOS: A simple and strong anchor-free object detector [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 44(4): 1922-1933.
- [19] FENG CH J, ZHONG Y J, GAO Y, et al. Tood: Task-aligned one-stage object detection [C]. 2021 IEEE/CVF International Conference on Computer Vision(ICCV), 2021: 3490-3499.
- [20] TAN M X, PANG R M, LE Q V. Efficientdet: Scalable and efficient object detection [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 10781-10790.
- [21] HOU Q R, ZHOU D Q, FENG J SH. Coordinate attention for efficient mobile network design [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13713-13722.
- [22] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module [C]. European Conference on Computer Vision(ECCV), 2018: 3-19.
- [23] HOWARD A G, ZHU M, CHEN B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications [J]. Arxiv preprint arxiv:1704.04861, 2017.
- [24] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117-2125.
- [25] LIU S, QI L, QIN H F, et al. Path aggregation network for instance segmentation [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8759-8768.
- [26] GEIGER A, LENZ P, URTASUN R. Are we ready for autonomous driving the kitti vision benchmark suite [C]. 2012 IEEE Conference on Computer Vision and Pattern Recognition, 2012: 3354-3361.
- [27] BUYVAL A, GABDULLIN A, MUSTAFIN R, et al. Realtime vehicle and pedestrian tracking for didi udacity self-driving car challenge [C]. 2018 IEEE International Conference on Robotics and Automation (ICRA), 2018: 2064-2069.
- [28] LIU Y CH, SHAO Z R, HOFFMANN N. Global attention mechanism: Retain information to enhance channel-spatial interactions [J]. Arxiv preprint arxiv: 2112.05561, 2021.
- [29] WANG J Q, CHEN K, XU R, et al. Carafe: Content-aware reassembly of features [C]. IEEE/CVF International Conference on Computer Vision, 2019: 3007-3016.
- [30] SUN Y, ZHANG Y H, WANG H Y, et al. SES-YOLOv8n: Automatic driving object detection algorithm based on improved YOLOv8 [J]. Signal, Image and Video Processing, 2024, 18 (5): 3983-3992.
- [31] 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/CVF Conference on Computer Vision and Pattern Recognition, 2023: 7464-7475.

## 作者简介

**刘罡**,副教授,硕士生导师,主要研究方向为移动通信、深度学习等。

E-mail:mrliug@163.com

**闫曙光**,硕士研究生,主要研究方向为深度学习、目标检测。

E-mail:1394934824@qq.com

**刘钰**(通信作者),博士,讲师,主要研究方向为智能信号处理、通感一体化和干扰消除等。

E-mail:860420@cwxu.edu.cn

**侯恩翔**,硕士研究生,主要研究方向为深度学习、目标检测。

**黄应征**,硕士研究生,主要研究方向为深度学习、目标检测。