面向夜间低光照场景下的轻量型车辆检测算法

张荣辉[1]，邹润宗[1]，刘佶豪[2]

（1. 中山大学 智能工程学院 广东省智能交通系统重点实验室，广州 510275；2. 湖南高速信息科技有限公司，长沙 410001）

摘 要：针对低光照条件下车辆与背景难以区分以及光源干扰导致车辆检测难度较大的问题，提出了一种改进的轻量级车辆检测模型——全局增强-局部注意力YOLO（GALA-YOLO）。现有的深度学习目标检测模型大多针对白天场景设计，忽视了夜间环境中的特殊挑战，而夜间低光照场景下的车辆检测已成为高阶辅助驾驶系统研究中的难点和重要组成部分。该模型基于轻量级的YOLOv8s进行优化，引入全局增强（GA）和局部注意力（LA）模块，分别嵌入不同层次以提升特征表达能力。GA模块使用大卷积核捕捉图像的全局信息（如物体空间分布和光照强度），并针对夜间环境进行自适应调整；LA模块是一个多尺度卷积注意力模块，采用不同大小的卷积核捕捉多尺度局部信息，综合考虑复杂夜间环境中不同层次的细节。进一步地，本文对CIoU损失函数进行了改进，通过采用MPDIoU，以更灵活地捕捉边界框形状特征，有效提升了模型的鲁棒性和收敛速度。实验结果表明，GALA-YOLO在夜间场景数据集上的mAP达到了59.2%，较现有最优目标检测模型提升了0.8%；在KITTI数据集上，车辆检测的mAP_{50}达到了76.7%，相比当前最优检测模型提升了0.5%。研究结果验证了本文所提方法在夜间环境中的有效性和正确性，为高级辅助驾驶系统的智能检测与安全预警研究提供了一种参考。

关键词：计算机视觉；高级辅助驾驶系统；车辆检测；夜间场景；注意力机制；特征增强

中图分类号：TP391 **文献标志码**：A **文章编号**：1671-5497（2026）06-1655-15

DOI：10.13229/j.cnki.jdxbgxb.20241286

Lightweight vehicle detection algorithm for low-light nighttime scenarios

ZHANG Rong-hui[1], ZOU Run-zong[1], LIU Ji-hao[2]

（1. *Guangdong Key Laboratory of Intelligent Transportation System, School of Intelligent Systems Engineering, Sun Yat-sen University, Guangzhou*, 510275, *China*; 2. *Hunan Expressway Information Technology Co., Ltd, Changsha* 410001, *China*）

Abstract: To address the challenge of distinguishing vehicles from the background under low-light conditions and the increased detection difficulty caused by light source interference, this paper proposes an

收稿日期：2024-11-30.

基金项目：国家自然科学基金项目（52172350）；广东省自然科学基金区域联合重点项目（2022B1515120072）；广州市重点研发计划项目（2024B01W0079）

作者简介：张荣辉（1981-），男，副教授，博士. 研究方向：智能车辆与辅助驾驶，交通信息与安全，车路协同与安全控制. E-mail：Zhangrh25@mail.sysu.edu.cn

improved lightweight vehicle detection model—Global Enhancement and Local Attention YOLO (GALA-YOLO). Existing deep learning-based object detection models are primarily designed for daytime scenarios, overlooking the unique challenges of nighttime environments. Vehicle detection in low-light nighttime scenarios has become a critical and challenging aspect of advanced driver assistance system (ADAS) research. The model is optimized based on the lightweight YOLOv8s, incorporating Global Attention (GA) and Local Attention (LA) modules, which are embedded at different layers to enhance feature representation. The GA module uses large convolutional kernels to capture global information, such as object spatial distribution and lighting intensity, with adaptive adjustments for nighttime environments. The LA module is a multi-scale convolutional attention module that employs convolutional kernels of varying sizes to capture multi-scale local information, considering the different levels of detail in complex nighttime environments. Furthermore, the paper improves the CIoU loss function by adopting MPDIoU, which allows for more flexible capture of boundary box shape features, significantly improving the model's robustness and convergence speed. Experimental results demonstrate that GALA-YOLO achieves an mAP of 59.2% on nighttime scene datasets, a 0.8% improvement over the best existing object detection model. On the KITTI dataset, the vehicle detection mAP50 reaches 76.7%, surpassing the current optimal detection model by 0.5%. The results validate the effectiveness and robustness of the proposed method in nighttime environments, providing a reference for intelligent detection and safety warning research inADAS(Advanced Driver Assistance Systems).

Key words: computer vision; advanced driver assistance systems; vehicle detection; nighttime scene; attention mechanism; feature enhancement

0 引 言

近年来,交通事故的发生频率逐渐上升,其中人为错误约占66.8%,成为导致事故的主要因素[1]。特别是在夜间,交通事故发生率较高。其中,追尾碰撞是最常见且致命的事故类型[2]。为了减少道路交通事故的发生,高级辅助驾驶系统(ADAS)[3-5]得到了迅速发展,其中针对道路车辆的检测发挥着至关重要的作用。然而,夜间驾驶中的车辆检测面临诸多挑战。恶劣的光照条件显著降低了道路及周围车辆的可见性,且车辆灯光的不均匀分布进一步增加了检测的难度。因此,如何提升夜间车辆检测的准确性,已成为高级辅助驾驶系统领域亟待解决的关键问题。目前,现有的车辆检测方法多集中于白天场景,而针对夜间的检测技术,尤其在光照不足、远距离检测和环境干扰等方面,仍面临诸多困难。因此,基于深度学习和图像处理技术的夜间车辆检测方法亟待突破这些瓶颈,以确保高级辅助驾驶系统在夜间环境中的安全性与可靠性。

传统的夜间车辆检测方法通常依赖于车灯的特征[6],通过阈值分割、连通域分析等技术提取特征区域,并根据空间位置、形状等信息对车灯进行配对,以实现车辆检测。例如,Bai等[7]提取前大灯和尾灯的特征进行配对检测,而Kavya等[8]则通过捕捉尾灯的颜色特征来识别车辆。然而,这些方法过于依赖单一的车灯特征,容易受到复杂光源的干扰,导致鲁棒性较差,难以适应智能驾驶的复杂场景。与白天相比,夜间车辆图像有效特征(如纹理、颜色和形状)较少,且车辆与背景之间的对比度较低[9],使夜间车辆检测面临更大挑战。此外,交通标志、路灯等其他光源可能会被误识别为车辆,而灯光较弱或被遮挡的车辆则难以检测。因此,基于传统图像处理技术的车灯检测方法在性能和稳定性上存在显著限制,难以满足高级辅助驾驶系统对检测精度和鲁棒性的高要求[10]。

近年来,卷积神经网络(CNN)在车辆检测领域取得了显著进展[11-13]。许多研究者尝试将深度学习技术应用于夜间车辆检测任务,并取得了较好的效果。基于深度学习的检测方法不仅能够提取车灯颜色、轮廓等基本特征,还能捕捉更为丰富的语义信息。Arinaldi等[14]采用Faster R-CNN进行夜间车辆分类与检测,取得了较好的结果,但在低光照场景下,依然存在误判和漏判的情况。Park等[15]结合SSD和VGG架构的规则分类器进

行高速公路夜间车辆检测，但SSD在远距离车辆检测中的表现较差。Chen等[16]通过将检测到的灯光区域与车辆目标提议相结合，生成兴趣区域（ROI），从而提升了检测的鲁棒性。Mo等[10]则通过结合多尺度灯光特征与车辆视觉特征，采用特征聚合机制，增强了检测性能。然而，这些方法的性能在很大程度上依赖于车辆灯光的准确提取，在复杂的灯光环境下，车灯的掩膜难以精确检测，且灯光较弱的车辆同样容易被忽略。此外，经典的一阶段目标检测模型YOLO[17]一经提出，便受到众多研究者的青睐。YOLO系列模型采用多尺度特征融合策略，能够检测不同尺度的目标，使得算法具备强大的适用性和泛化能力，广泛应用于多种目标检测任务[18-20]。其中，YOLO系列模型在特定版本中，通常包括t、s、m和l等不同规模的模型（如YOLOv8t、YOLOv8s、YOLOv8m和YOLOv8l）。它们具有相同的模型架构，但参数量、计算复杂度和精度依次增加。根据具体的应用场景需求，可以在速度、精度和计算资源之间进行平衡选择，以优化性能和效率。然而，现有的YOLO算法在低光照场景中仍存在特征提取能力不足的问题，限制了其在复杂夜间环境下的车辆检测性能与适应性。

针对夜间光照条件差和特征信息匮乏等问题，本文基于轻量级的YOLOv8s模型进行了多方面的改进，旨在保证实时性的前提下，提升夜间车辆检测的鲁棒性和精度。具体贡献总结如下：①全局增强模块（Global augmentation，GA）：为了应对夜间低光照条件下的低对比度问题，本文提出了全局增强模块。该模块通过采用大卷积核来捕捉图像的全局信息，增强上下文特征，并能够自适应地调整光照强度和颜色分布等参数，以有效缓解低亮度和低对比度对目标检测性能的负面影响。②局部注意力模块（Local Attention，LA）：在夜间环境中，由于车辆与背景之间的差异通常较小，局部特征的提取变得尤为关键。为此，本文引入了局部注意力模块，该模块通过多尺度深度卷积进行多分支局部特征的提取与聚合，从而增强了模型对图像中不同尺度的局部信息和细节特征的敏感性。③损失函数改进：针对现有损失函数在夜间车辆检测中的不足，本文引入了MPDIoU（Minimum point distance based IoU），该损失函数专门设计用于优化目标框的定位精度。MPDIoU损失能够有效应对目标物体在图像中发生重叠，或检测框与真实框长宽比相同但大小不同的情况，从而实现更加精确的目标定位。

本文将改进后的模型命名为GALA-YOLO。该模型通过上述创新改进，显著提升了夜间车辆检测的性能。为验证GALA-YOLO的有效性，在一个开源的夜间车辆检测数据集上进行了详细对比实验。实验结果表明，GALA-YOLO在精度、鲁棒性和检测效率方面均优于当前主流的目标检测模型。此外，为确保模型的泛化性和适应性，还在KITTI数据集上进行了额外的验证，进一步验证了GALA-YOLO在白天复杂环境中的检测性能。

1　模型构建

在YOLOv8模型的网络结构中，其骨干网络采用Darknet-53框架，并引入了一个新的C2f模块作为特征提取的主要模块，使用空间金字塔快速池化（Spatial pyramid pooling fast，SPPF）[21]来提取不同感受野的特征，以适应不同尺度的目标。特征融合层主要由FPN＋PAN结构组成，可以将自下而上与自上而下的特征图进行融合。其中，FPN[22]采用自上而下的方式，通过上采样将高层特征信息与低层特征信息融合，从而计算出预测特征图。

然而，在夜间场景中，由于光照不足，图像通常较暗，图像整体的特征信息区分度较低，限制了YOLOv8模型对场景中全局特征的提取。同时，光照不足和摄像头可能出现的模糊，导致车辆局部的细节特征信息匮乏，对C_2f模块的特征提取造成了较大影响。此外，夜间背景中常充满干扰，如道路灯光、车灯等光源造成的反射，这些都会对车辆识别造成不利影响。这些因素导致原始的YOLOv_8模型在复杂夜间环境中的检测性能大打折扣。

针对这些挑战，本文提出了GALA-YOLO模型，旨在解决夜间场景中的车辆检测问题。具体而言，为确保车辆检测的实时性，必须考虑模型的参数量和计算复杂度。如前所述，本文选择YOLOv_8的“s”规模模型，它在车辆检测的精度与实时性之间提供了良好的平衡。GALA-YOLO基于YOLOv8s，在模型的输入端引入了全局增强模块（Global augmentation，GA），并在每个C2f模

块后添加了局部注意力模块(Local attention, LA)。两个模块协同工作,有效应对夜间场景中全局特征区分度低、局部细节特征匮乏和背景干扰等问题。GALA-YOLO模型的整体架构如图1所示,由四部分组成:输入端、主干网络、特征融合网络和输出端。

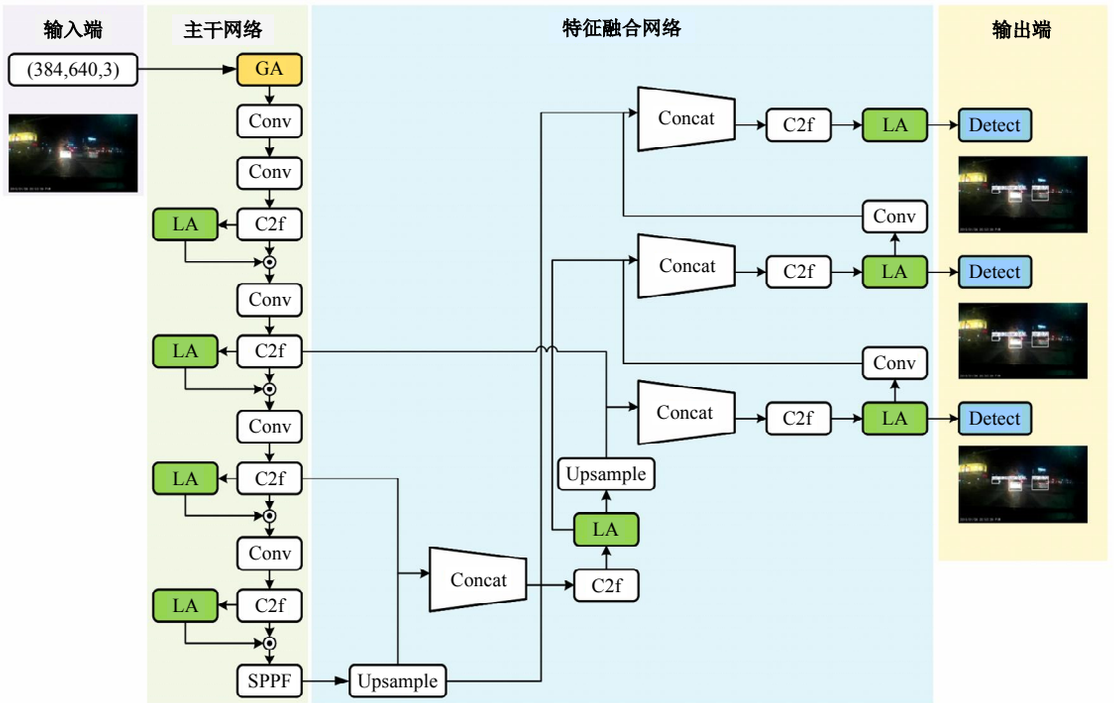


图1 GALA-YOLO整体架构

Fig. 1 Overall architecture of GALA-YOLO

1.1 全局增强模块(GA)

为解决夜间场景光照不足导致图像全局特征信息区分度不足的难题,本文提出一种基于大卷积核的全局增强模块。尽管YOLOv$_8$模型在白天场景的目标检测中表现出色,但面对光照条件恶劣的夜间场景时,其对图像全局特征的建模能力却显得不足。YOLOv8对输入的图片先进行一个卷积核大小为3的卷积操作进行下采样和通道扩充,然后不断叠加小卷积核的卷积操作来增加感受野。然而,这种方法效率低下,甚至可能造成过拟合等问题[23]。为解决这一问题,本文引入了带有大卷积核的全局增强模块,该模块用于捕捉图像的全局信息,并进行有效的上下文建模,从而在夜间场景中更好地理解图像的全局结构。其中,大卷积核能够在较大感受野内进行计算,覆盖更广泛的图像区域,帮助模型识别图像中的全局结构、物体间的空间关系,并有效检测长距离特征。事实上,当前已有研究重新审视大卷积核对CNN性能的影响[24],并已在图像分类、目标检测[25,26]和语义分割等各种计算机视觉任务中实现突破。

如图2所示,GA模块的结构图使用大小为11×11的卷积核,因此具有更大的感受野,能够充分提取特征图的全局信息,在更大的感受野内聚合上下文信息。通过全局特征增强,使得YOLOv$_8$模型能够更好地捕捉到夜间场景中隐藏的信息,提升检测性能。

GA模块主要由两层卷积核大小为11×11的

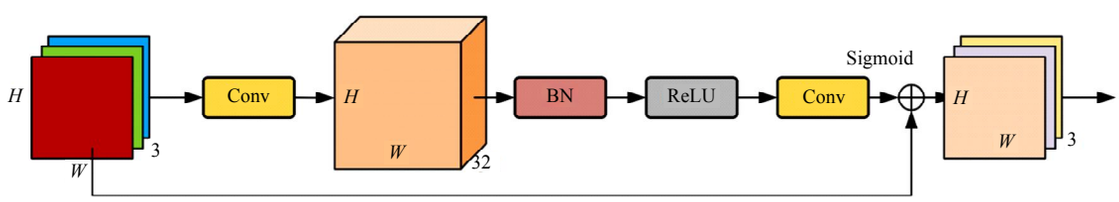


图2 全局增强模块结构图

Fig. 2 Architecture of GA module

卷积层构成，将其插入 YOLOv$_8$模型初始的输入端。给定一张输入图片（大小为 $H\times W\times 3$），首先经过第一个卷积层 Conv1，将输入通道数从 3（RGB 图像的通道数）扩充到 32。此时得到的特征图尺寸为 $H\times W\times 32$，然后依次输入到 BatchNorm 函数和 ReLU 激活函数，目的是帮助模型加速训练，提高稳定性以及非线性能力。然后，再经过第二个卷积层 Conv2，将通道数恢复为 3。最后，通过 Sigmoid 函数生成一个 0 到 1 之间的注意力权重张量，然后与最初的输入图片进行逐元素相加。具体来说，GA 模块可以表达为：

$$X=\mathrm{Conv}_2\left(\mathrm{ReLU}\left(\mathrm{BN}\left(\mathrm{Conv}_1\left(\mathrm{Identity}\right)\right)\right)\right) \tag{1}$$

$$\mathrm{Output}=\mathrm{Identity}\oplus\mathrm{Sigmoid}(X) \tag{2}$$

式中：Identity 代表原始输入图像；⊕ 表示逐元素相加操作。

这种扩充通道再压缩的操作，是一种特征信息处理的引导，使得模型能够在更高维的空间中进行特征表达、编码和选择，从而提取出更为丰富和精确的特征信息。通过这种方式，输入图片的每个像素都会根据对应的权重进行调整，从而实现全局特征的增强。例如，GA 模块能够对图像的整体亮度进行平滑处理，从而改善低光照条件下的图像质量，使物体的轮廓更加清晰。此外，GA 模块还能够在较大范围内帮增强不同物体之间的关联性。这些都有助于在夜间环境中提高对物体低对比度的适应性，从而有效提升车辆检测的性能。

1.2　局部注意力模块（LA）

C2f 模块是卷积神经网络中常用的组件，用于将卷积层的输出转化为全连接层的输入。然而，在传统的 C2f 模块中，特征图信息之间的传递相对有限，这导致了仅使用 C2f 模块作为主干网络的 YOLOv8 模型，其在夜间场景的目标检测准确率较低。而注意力机制的核心思想是根据输入特征图动态计算特征权重，以在特征融合过程中对不同特征进行加权。注意力机制更关注图像的局部信息，细化输入映射。考虑到这些因素，本文构建一种新的局部注意力模块，并将其嵌入 YOLOv8 的网络的不同位置中，以提高局部特征提取性能。

如图 3 所示，LA 是一个多尺度卷积注意力，包含了 3 个部分：用于聚合局部信息的深度卷积，捕获多尺度上下文的多分支深度条形卷积，以及

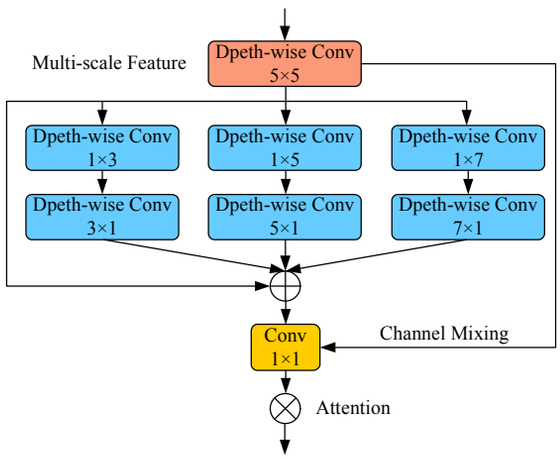


图 3　局部注意力模块结构图

Fig. 3　Architecture of LA module

用于建模不同通道之间关系的 1×1 卷积。1×1 卷积的输出被直接用作注意力权重，用于重新加权 LA 的输入。具体来说，LA 可以表达为：

$$\mathrm{Attn}=\mathrm{Conv}_{1\times 1}\left(\sum_{i=0}^{3}\mathrm{Scale}_i\left(\mathrm{DWConv}(x)\right)\right) \tag{3}$$

$$\mathrm{Output}=\mathrm{Attn}\otimes x \tag{4}$$

式中：x 为输入特征；Att 和 Out 分别为注意力图和 LA 模块的输出；⊗ 表示逐元素矩阵乘法操作；DWConv 表示深度卷积；$\mathrm{Scale}_i(i\in\{0,1,2,3\})$ 表示图 3 中的第 i 个分支，其中，$Scale_0$ 是恒等连接。

根据文献[27]中的做法，在每个分支中，本文使用两次个度条形卷积来近似标准的深度卷积操作，卷积核大小分别为 3、5 和 7。本文选择深度条形卷积主要考虑两点：①条形卷积具有轻量化的特点，为了模拟标准的 7×7 卷积，只需一对 7×1 和 1×7 的卷积；②条形卷积对于一些条状物体（如迎面驶来的汽车、汽车尾部和停在路边的汽车）有很好的表现，因此，条形卷积可以补充网格卷积，并有助于提取条状特征[27,28]。

LA 模块利用不同大小的卷积核提取局部特征，尤其是在车辆边缘细节的增强方面表现突出。通过生成的注意力，模块能够聚焦于车辆的重要区域，从而提高边缘模糊区域的特征提取能力，进一步提升模型的检测精度。LA 能够在图像中聚焦于具有显著特征的区域，抑制无关区域的干扰，从而增强模型在复杂背景下对夜间车辆的检测能力，特别是在对车辆前灯或尾灯的有效识别方面。

1.3　回归损失函数的改进

YOLOv8 采用了 CIoU 回归损失函数[29]，考虑了 3 个关键的几何因素：重叠面积、中心点之间

的距离以及长宽比。CIoU函数利用IoU来评估预测边界框与真实框之间的重叠情况,同时结合了欧几里得距离以及相关的长宽比和角度。具体的损失函数如下所示:

$$R_{\mathrm{CIoU}}=\rho^2\left(b,b^{\mathrm{gt}}\right)/c^2+av \tag{5}$$

$$v=\frac{4}{\pi^2}\left(\arctan\frac{w^{\mathrm{gt}}}{h^{\mathrm{gt}}}-\arctan\frac{w}{h}\right)^2 \tag{6}$$

$$L_{\mathrm{CIoU}}=1-\mathrm{IoU}+\rho^2\left(b,b^{\mathrm{gt}}\right)/c^2+a \tag{7}$$

式中:IoU为真实框与预测框的交并比;

由上述表达式可知,CIoU损失函数通过反正切函数来构造矩形长宽比的惩罚项,但该方法在鲁棒性上较弱,容易受到离群点的影响,导致损失值波动较大。此外,由于反正切函数值域的限制,归一化要求需要引入新的系数,增加了计算复杂度。尽管CIoU考虑了长宽比,但它以相对长宽比而非绝对长宽值的形式使用,这在车辆检测场景中存在问题,因为车辆目标的尺度变化较大,可能导致长宽比一致但实际尺寸不同,从而削弱惩罚效果,难以优化损失函数,增加漏检的风险。为了解决这一问题,本文引入了MPDIoU[30]损失函数来替代CIoU计算边框损失。

具体来说,假定$\left(x_1^{\mathrm{prd}},y_1^{\mathrm{prd}}\right)$与$\left(x_2^{\mathrm{prd}},y_2^{\mathrm{prd}}\right)$分别表示预测框的左上角坐标和右下角坐标,$\left(x_1^{\mathrm{gt}},y_1^{\mathrm{gt}}\right)$与$\left(x_2^{\mathrm{gt}},y_2^{\mathrm{gt}}\right)$分别表示真实框的左上角和右下角坐标,$H$和$W$分别表示输入图像的长和宽,MPDIoU损失函数的示意图如图4所示,具体定义如下:

$$d_1=\left(x_1^{\mathrm{gt}}-x_1^{\mathrm{prd}}\right)^2+\left(y_1^{\mathrm{gt}}-y_1^{\mathrm{prd}}\right)^2 \tag{8}$$

$$d_2=\left(x_2^{\mathrm{gt}}-x_2^{\mathrm{prd}}\right)^2+\left(y_2^{\mathrm{gt}}-y_2^{\mathrm{prd}}\right)^2 \tag{9}$$

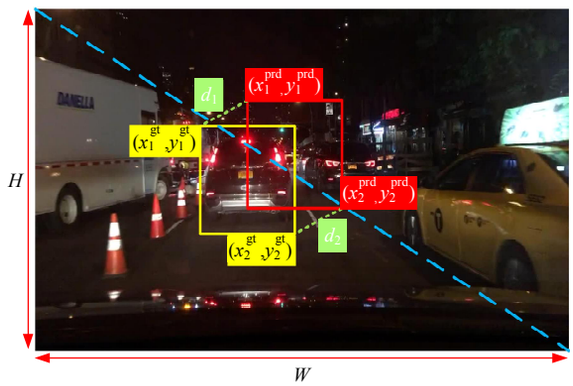


图4 MPDIoU损失函数的各项参数

Fig. 4 Parameters of the MPDIoU loss function

$$\mathrm{MPDIoU}=\mathrm{IoU}-\frac{d_1^2}{h^2+w^2}-\frac{d_2^2}{h^2+w^2} \tag{10}$$

式中:d_1为真实框左上角与预测框左上角的距离;d_2为真实框右下角与预测框右下角的距离。

可以看出,MPDIoU损失函数是一种基于点距离的度量方法,能够有效解决传统CIoU损失函数在处理具有相同长宽比但实际尺寸差异较大的边界框时的不足。同时,MPDIoU损失也能更好地应对多个真实框重叠的情况。因此,该损失函数特别适用于具有显著尺度差异的应用场景,如车辆检测任务,能够显著提升模型在此类问题中的检测准确性。

2 数据集与实验设置

本文将所提出的GALA-YOLO应用于两个数据集:一个开源的夜间车辆检测数据集和KITTI数据集[31]。

2.1 数据集

图5为实验中使用的两个公开可用的数据集,用来评估模型的性能。第一个数据集是开源



(a)夜间车辆检测数据集示例



(b) KITTI数据集示例

图5 数据集示例

Fig. 5 Dataset examples

的夜间车辆检测数据集，包含6 388张来自真实世界夜间场景的图像。该数据集中的图像来源于多种场景类型，包括城市街道、高速公路、住宅区和停车场等。每张图像中的所有车辆类别统一标记为“Car”，因此数据集中只有一个类别标签。在数据预处理阶段，本文首先进行了数据清洗，去除了空样本图像。随后，按照4∶1的比例将数据集划分为训练集（5 110张图像）和验证集（1 278张图像）。

第二个数据集为KITTI检测数据集，主要用于评估本文提出的GALA-YOLO模型在白天场景下进行车辆检测的性能。KITTI检测数据集中的每张图像尺寸为1 242×375，共包含7 481张训练图像和7 518张测试图像。由于测试集的边界框标注未公开，本文将原始训练集同样按4∶1的比例划分为训练集和验证集。此外，本文主要关注车辆行驶场景中的车辆检测，因此将KITTI数据集中的所有车辆类别合并为一个统一的标签“Vehicle”，同时将“Pedestrian”和“Person_sitting”合并为“Pedestrian”，并保留“Tram”（有轨电车）、“Cyclist”（骑行者）和“Misc”（其他难以归类的物体）作为独立标签。最终的训练集包含5 985张图像，测试集包含1 496张图像。

2.2　实验环境配置及参数设置

本文的模型和相关实验均基于Python语言和PyTorch框架实现，运行于Ubuntu 22.04操作系统。实验平台配备IntelXeonGold 6240处理器（主频2.60 GHz）和RTX 3090显卡（显存24 GB）。使用的Python版本为3.8，Pytorch版本为2.0.1，并通过CUDA 11.8实现GPU加速。

在模型评估方面，将GALA-YOLO与多种主流目标检测方法进行了比较，包括YOLOv6[32]、YOLOv8[33]、YOLOv9[34]、YOLOv10[35]和YOLO11[36]。本文聚焦于轻量级车辆检测模型，因此所有对比模型均选用“s”规格版本。模型训练过程中，其损失函数收敛效果及精度指标如图6所示。所有模型均从零开始训练，未使用预训练权重。训练集中的图像通过随机翻转和裁

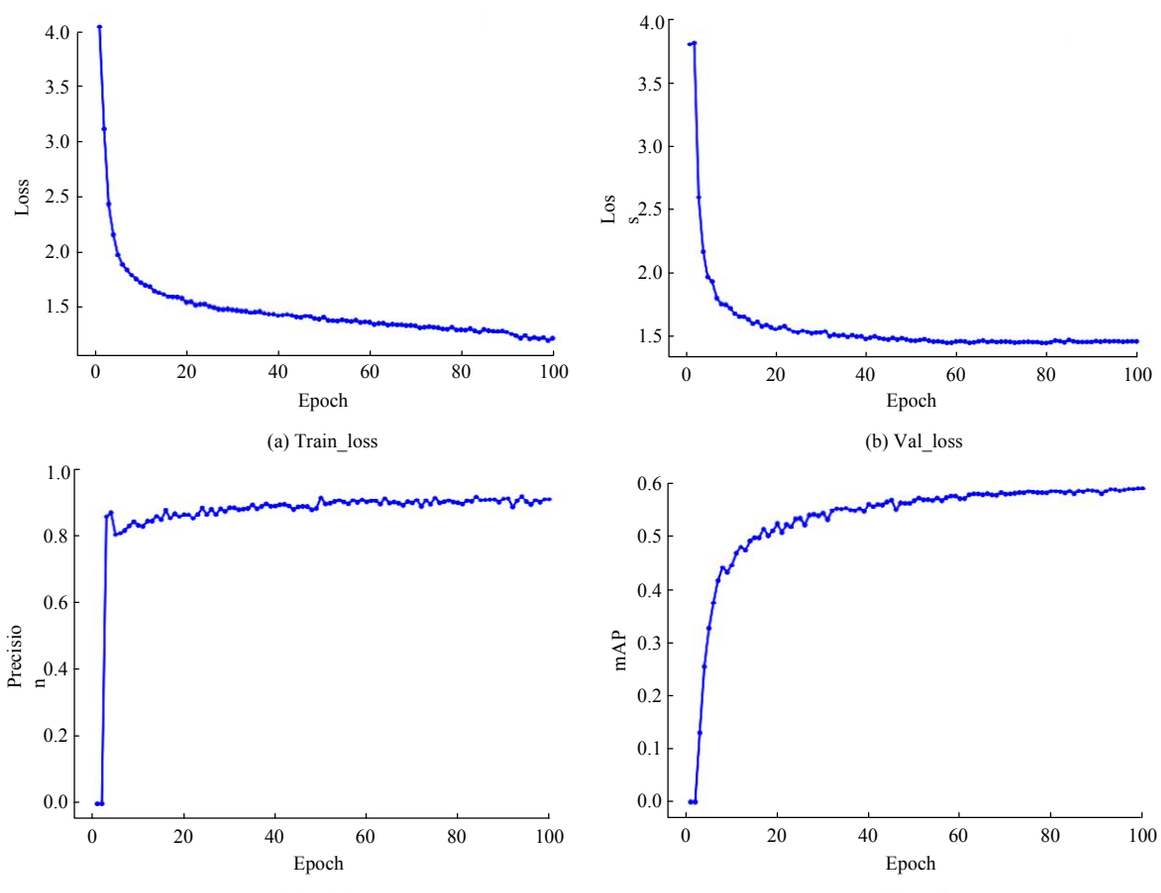


图6　GALA-YOLO模型损失及精度指标曲线

Fig. 6　Loss and accuracy metrics curves of the GALA-YOLO model

剪等数据增强手段进行处理,训练轮次设定为100轮,批次大小为128。优化器选择随机梯度下降(SGD),动量(momentum)设置为0.937,并采用Warm-Up方法对学习率进行预热,初始学习率为0.01,学习率衰减权重设定为5e-4,且使用余弦退火算法自动调整学习率。所有模型的超参数(如锚框大小、损失函数权重等)均保持默认值。检测结果的评估采用IoU阈值为0.7的非极大值抑制(NMS)方法以去除重叠框。

2.3 评价指标

本文通过多种技术指标对模型进行评估,包括精确度(Precision)、召回率(Recall)、平均精确率(mAP)以及模型的参数数量等。这些指标均用于定量分析模型的性能与效率,相关计算方法将在接下来的部分中详细说明。

精确度(Precision)用于评估模型在正类预测中的准确性,表示所有预测为正类的实例中,正确预测为正类的比例。其计算公式为真正例(TP)与真正例和假正例(FP)之和的比值,如公式(11)所示:

$$\mathrm{Precision}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \tag{11}$$

召回率(Recall)衡量模型在所有实际正类实例中,正确预测为正类的比例,计算公式如下所示:

$$\mathrm{Recall}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{12}$$

在公式(11)和(12)中,假正例(FP)指的是在检测结果中被预测为正类,但实际上并不在真实标注中出现的实例数;真正例(TP)指的是在检测结果和真实标注中都存在的实例数;假负例(FN)指的是在真实标注中存在但未能在检测结果中识别出的实例数。

精确率与召回率是机器学习中常用的性能评估指标,但通常二者之间存在一定冲突。为了解决这一问题,引入了平均精确率均值(mAP),它综合考虑了精确度和召回率,提供一个整体评分,反映模型在不同召回率下的性能表现。mAP通过计算在不同召回率水平下的平均精确度,对模型在特定应用场景下的目标识别能力进行全面评估。得分越高,表示模型的性能越好。mAP的计算公式如下:

$$\mathrm{mAP}=\frac{1}{N}\sum_{i=1}^{N}\mathrm{AP}_i \tag{13}$$

式中:N为类别或类别总数;AP_i为第i个类别的平均精确率。

对于特定类别,平均精确率(AP)计算公式为:

$$\mathrm{AP}_i=\frac{1}{R_i}\sum_{k=1}^{R_i}P(k)\cdot precAtK(k) \tag{14}$$

式中:R_i为第i类中的相关实例数;$P(k)$为第k个检索到实例的精确度;$precAtK(k)$为指示函数,当第k个检索到的实例为相关实例时,其值为1,否则为0。

此外,参数数量指的是模型中包含的偏置和权重参数,它反映了模型的复杂性以及其从训练数据中学习的能力。通常,模型中的参数数量越多,表示其复杂度越高。

3 实验设计与结果分析

3.1 消融实验

3.1.1 GA模块卷积核大小设置及有效性实验

首先,本文所提出的全局增强模块(GA)采用了大卷积核,旨在通过更大的感受野覆盖更广泛的图像区域,从而有效地捕获图像的全局信息。通常,为了在图像中获取较大的感受野,同时考虑到计算资源的限制,研究者们通常采用堆叠多个小卷积核来代替单个大卷积核。如图7所示,在Inception-v3[37]论文中,两个3×3卷积核被用来替代一个5×5卷积核的计算。然而,随着计算机视觉技术的发展以及计算资源的不断提高,大卷积核在优化卷积神经网络(CNN)架构中的应用逐渐受到青睐。大卷积核能够直接提供更大的感受野,从而减少小卷积核的堆叠需求,进而降低网络的深度和模型的复杂度。此外,较大的卷积核能够包含更多相邻的图像像素信息,因其对特征图的平滑作用更加显著,能够有效减少局部噪声并增强对全局上下文的捕捉能力。

然而,大卷积核的计算量随着卷积核尺寸的

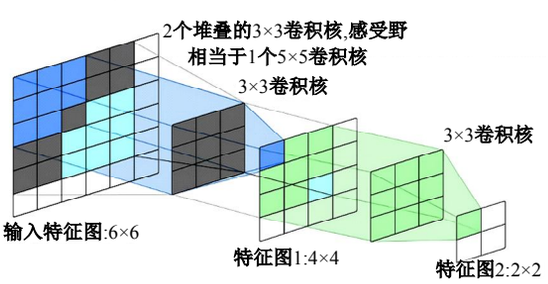


图7 多个3×3卷积替代5×5卷积[37]

Fig. 7 Multiple 3×3 Conv. replacing a 5×5 Conv.[37]

平方增长，且其较为侧重于全局信息的提取，可能会忽略局部细节特征。此外，单个大卷积核通常包含更多的参数，这在训练过程中可能增加过拟合的风险，尤其是在数据集较小或数据多样性不足的情况下。

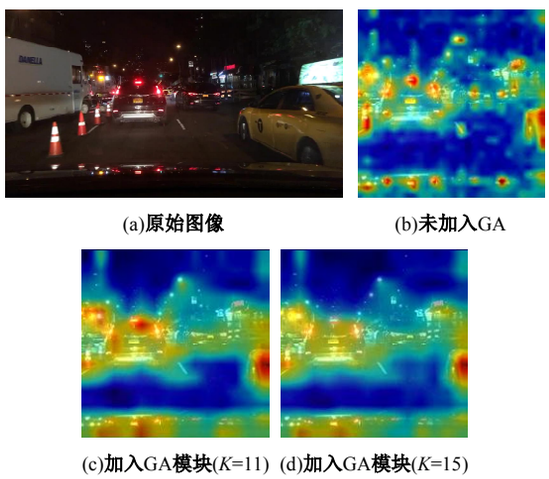
因此，大卷积核的具体尺寸选择需要根据任务需求和实时性要求进行权衡。为此，本文设计了消融实验，旨在确定GA模块中大卷积核的最佳尺寸，以实现模型性能的最优化。首先，本文进行了定量实验，使用的实验数据集为夜间车辆检测数据集。所有实验均在训练集上从零开始训练，并在验证集上进行评估。实验结果如表1所示。结果表明，当卷积核尺寸逐步增大时，模型的精确率和召回率也逐渐提高。当卷积核大小为11时，模型的精确率和召回率达到了最佳值。然而，随着卷积核尺寸进一步增大至13和15时，模型的精确率和召回率出现了下降。这表明，卷积核尺寸为13和15时，模型的性能较差，难以在夜间车辆检测任务中收敛并有效拟合。

表1　GA模块卷积核尺寸对比实验

Table 1　Comparison experiment of convolution kernel size in GA module

Kernel size	Precision /%	$mAP_{0.5}$/%	$mAP_{0.5:0.95}$/%
7	91.2	93.3	58.5
9	90.8	93.4	58.5
11	90.6	93.9	58.8
13	90.6	93.5	58.3
15	91.4	93.0	58.2

为了进一步验证定量实验中的结果，本文开展了定性分析实验，以直观地展示不同尺寸卷积核对模型捕捉图像信息的影响。本文通过Grad-CAM热力图可视化了GA模块输出特征的空间响应。Grad-CAM能够直观展示卷积神经网络在特征提取过程中所关注的区域和决策依据。特征热力图如图8所示，彩图参见电子版，下同。第一列为原始图像，第二列为未引入GA模块时的热力图，第三列为使用卷积核大小为11的GA模块的热力图，最后一列为卷积核大小为15的GA模块的热力图。在热力图中，蓝色区域表示关注度最低，红色区域表示关注度最高。通过对比可以看出，在未使用GA模块时，模型的感受野较小，导致未能提取足够的全局图像信息；而在引入GA模块后，模型的全局特征提取能力得到了有效增强，使得卷积层在特征提取时的关注度分布



(a)原始图像　(b)未加入GA　(c)加入GA模块(K=11)　(d)加入GA模块(K=15)

图8　Grad-CAM热力图可视化

Fig. 8　Grad-CAM heatmap visualization

更加均匀，有助于模型在夜间图像中的全局感知和对光照变化的适应。然而，当卷积核尺寸过大时，卷积操作可能出现不连续，导致卷积层的关注范围过于广泛，缺乏足够的聚焦区域，从而忽视了车辆位置的细节特征，进而影响了车辆检测效果。

3.1.2　其余改进的有效性实验

为了验证本文其余改进的有效性，设计了多组消融实验，实验结果见表2。本文以原始的YOLOv8s模型作为基线进行验证与对比实验，所使用的实验数据集为夜间车辆检测数据集。所有实验均在训练集上从零开始训练，并在验证集上进行评估。在实验2和实验3中，分别引入了全局增强模块(GA)和局部注意力模块(LA)，以评估这两个模块在轻量化YOLOv8s模型中的有效性。实验4将实验2和实验3的结果结合，评估同时引入GA模块和LA模块后模型在夜间场景中进行车辆检测的综合表现，此时得到的模型即为本文提出的GALA-YOLO。实验5在实验4的基础上，对GALA-YOLO模型的训练过程中的损失函数进行了改进，引入了MPDIoU损失函数，并展示了其在提升检测性能方面的重要作用。

上述5组消融实验的结果验证了本文所提出的方法在保持网络复杂度几乎不变的前提下，能够有效提升检测精度。根据表2的实验结果，实验2通过在YOLOv8s模型的输入端插入轻量化的全局增强模块(GA)，增加的模型参数量可忽略不计(仅增加0.2%)，而GA-YOLO模型的$mAP_{0.5}$和$mAP_{0.5:0.95}$分别提高了0.6%和0.5%，

表2 消融实验

Table 2 Ablation experiment

序号	backbone	GA	LA	MPDIoU	Param/M	Precision/%	Recall/%	$mAP_{0.5}$/%	$mAP_{0.5:0.95}$/%
1	YOLOv8s				11.1	91.6	84.7	93.3	58.3
2	YOLOv8s	√			11.1	90.6	86.0	93.9	58.8
3	YOLOv8s		√		12.0	90.3	85.8	94.0	58.8
4	YOLOv8s	√	√		12.0	91.1	86.1	94.3	59.0
5	YOLOv8s	√	√	√	12.0	91.9	86.4	94.3	59.2

这表明GA模块能够显著提升模型的检测性能，同时对模型复杂度的影响微乎其微。实验3通过引入多尺度局部注意力模块(LA)，LA-YOLO模型的$mAP_{0.5}$和$mAP_{0.5:0.95}$分别提高了0.7%和0.5%，同时模型的参数量增加了8.0%。这一结果表明，LA模块在提升检测效果方面具有显著优势。在实验4中，结合使用GA模块和LA模块后，GALA-YOLO模型的$mAP_{0.5}$和$mAP_{0.5:0.95}$分别提高了1.0%和0.7%，且模型的参数量增加了8.3%。这一实验表明，通过GA和LA模块的协同作用，GALA-YOLO能够更好地分离背景噪声和目标特征，从而提升对车辆的检测精度和鲁棒性。

最后，进行了实验5，旨在验证MPDIoU边框损失函数对训练收敛性的影响。实验结果显示，使用MPDIoU作为回归损失函数训练GALA-YOLO模型能够显著提升其性能，模型的$mAP_{0.5:0.95}$提高了0.2%，而$mAP_{0.5}$保持不变，且未引入额外的计算复杂度。如图9所示，本文比较了原始YOLOv8s、未采用MPDIoU的GALA-YOLO以及引入MPDIoU的GALA-YOLO在训练过程中训练集上的损失收敛效果。该结果验证了MPDIoU损失函数在提升模型检测精度方面的有效性。

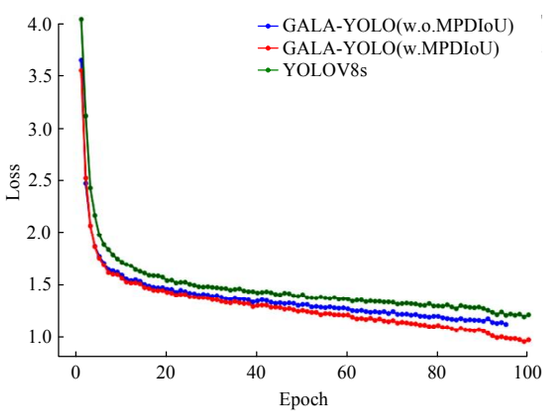


图9 损失函数收敛效果对比

Fig. 9 Comparison of loss function convergence performance

这是因为，在车辆检测任务中，边界框回归对模型的检测精度至关重要。而原始YOLOv8所使用的CIoU损失函数未能充分利用边界框回归的几何特性，因此在预测边界框和真实边界框具有相同长宽比但不同尺寸时，其效果较差，这限制了收敛速度和准确性。相比之下，MPDIoU基于水平矩形的最小点距离进行计算，能够综合考虑重叠区域、中心点距离及宽度和高度的偏差，从而提供更精确的损失度量。

图10展示了两种不同的边界框回归结果。其中，黄色框代表真实的边界框，红色框代表预测



(a) $Loss_{CIoU}$ = 0.75, $LOSS_{MPDIoU}$ = 0.79 (b) $Loss_{CIoU}$ = 0.75, $LOSS_{MPDIoU}$ = 0.76

图10 两种不同边界框回归结果的情况

Fig. 10 Two cases with different bounding box regression results

的边界框。根据文献[30]中的公式推导，在这两种情况下，CIoU 计算出的损失值相同，而使用 MPDIoU 计算的损失值则有所不同。这表明，传统方法在某些特定情况下可能无法有效区分不同的预测结果，而 MPDIoU 能够更准确地反映预测框与真实框之间的差异。上述案例凸显了 MPDIoU 在处理边界框回归问题上的优势，特别是在区分具有相同宽高比但不同尺寸或位置的边界框时。MPDIoU 通过直接计算预测框与真实框之间的关键点距离，提供了更精确的损失度量方法。

3.2　对比实验

为了验证本文所提出的算法在检测精度和网络复杂性方面的优势，本文与当前主流的目标检测算法进行了对比实验。在实验中，本文选取了 YOLOv6s、YOLOv8s、YOLOv9s、YOLOv10s 和 YOLOv11s 五种主流目标检测算法，并额外选择了参数量和计算复杂度较高的 YOLOv_8m、YOLOv_9m 和 YOLO$_{11}$m 模型。实验的目的是验证本文方法在保持轻量化和实时性的前提下，仍能实现较高的夜间车辆检测精度。所有模型均在夜间车辆检测数据集上从零开始训练，未使用任何预训练模型。

根据表 3 中的实验结果，本文提出的 GALA-YOLO 模型在该夜间车辆检测数据集的验证集上，精确度和召回率分别达到了 91.9% 和 86.3%，而 $mAP_{0.5}$ 和 $mAP_{0.5:0.95}$ 分别为 94.3% 和 59.2%。与 YOLOv6s、YOLOv8s、YOLOv9s、YOLOv10s 和 YOLO11s 模型相比，精确度分别提高了 0.7%、0.3%、3.4%、2.4% 和 0.4%；召回率则分别高出 0.5%、1.6%、1.7% 和 1.1%，但比 YOLOv9s 低 1.8%。同时，$mAP_{0.5}$ 较 YOLOv6s、YOLOv8s、YOLOv9s、YOLOv10s 和 YOLO11s 模型分别提升了 1.0%、1.0%、0.7%、1.6% 和 0.9%；$mAP_{0.5:0.95}$ 也分别提高了 1.3%、0.9%、0.8%、1.5% 和 1.0%。

表 3　对比实验

Table 3　Comparative experiment

Model	Param/M	Precision/%	Recall/%	$mAP_{0.5}$/%	$mAP_{0.5:0.95}$/%	FPS
YOLOv6s	16.3	91.2	85.5	93.3	57.9	145
YOLOv8s	11.1	91.6	84.7	93.3	58.3	123
YOLOv9s	7.2	88.5	88.1	93.6	58.4	72
YOLOv10s	8.0	89.5	84.6	92.7	57.7	97
YOLOv11s	9.4	91.5	85.2	93.4	58.2	96
GALA-YOLO (ours)	**12.0**	**91.9(+0.3)**	86.3	**94.3(+0.7)**	**59.2(+0.8)**	**88**
YOLOv8m	25.8	90.3	86.6	94.0	58.9	86
YOLOv9m	20.0	90.8	86.9	94.4	59.5	57
YOLO11m	20.0	89.9	87.1	94.1	59.1	77

注：最优结果用蓝色加粗显示；次优结果用绿色加粗显示；括号中为相较次优结果的精度提升。

值得注意的是，在与 YOLOv8m、YOLOv9m 和 YOLO11m 这 3 个参数量和计算复杂度更高的模型进行对比时，GALA-YOLO 在各项指标上仍表现出显著优势。具体而言，GALA-YOLO 的精确度高于这三者，$mAP_{0.5}$ 仅比 YOLOv9m 低 0.1%，且高于其余两个模型，在 $mAP_{0.5:0.95}$ 上，GALA-YOLO 仅比 YOLOv9m 低 0.3%，也优于其余两个模型。

此外，本文使用单张 RTX 3090 显卡（显存 24GB）测试各模型在推理时的 FPS，从表 3 中可以看出本文的 GALA-YOLO 模型能够很好满足车辆检测的实时性要求。这一系列的实验对比结果证明了提出的 GALA-YOLO 作为一种轻量级的车辆检测模型，能够在保持着较低参数量和计算复杂度的同时，在夜间车辆检测任务中显著优于同级别的模型，并且与参数量更大的模型相比，也仍保持着竞争力。

为了直观展示各模型在夜间车辆检测中的表现，图 11 展示了不同模型在夜间车辆检测数据集上的检测结果对比。检测结果以蓝色方框标注，漏检的结果则用红色方框标出。综合来看，其他模型在夜间车辆检测中或多或少存在漏检和错检的情况。尤其是对于黑色车辆，由于夜间的对比度低，难以与背景区分，且远距离的车辆也给其他模型带来了较大挑战。相比之下，本文提出的 GALA-YOLO 模型能够有效捕捉图像的全局整体特征和多尺度的局部细节特征，从而在夜间车辆检测任务中取得了较高的检测精度。综上所

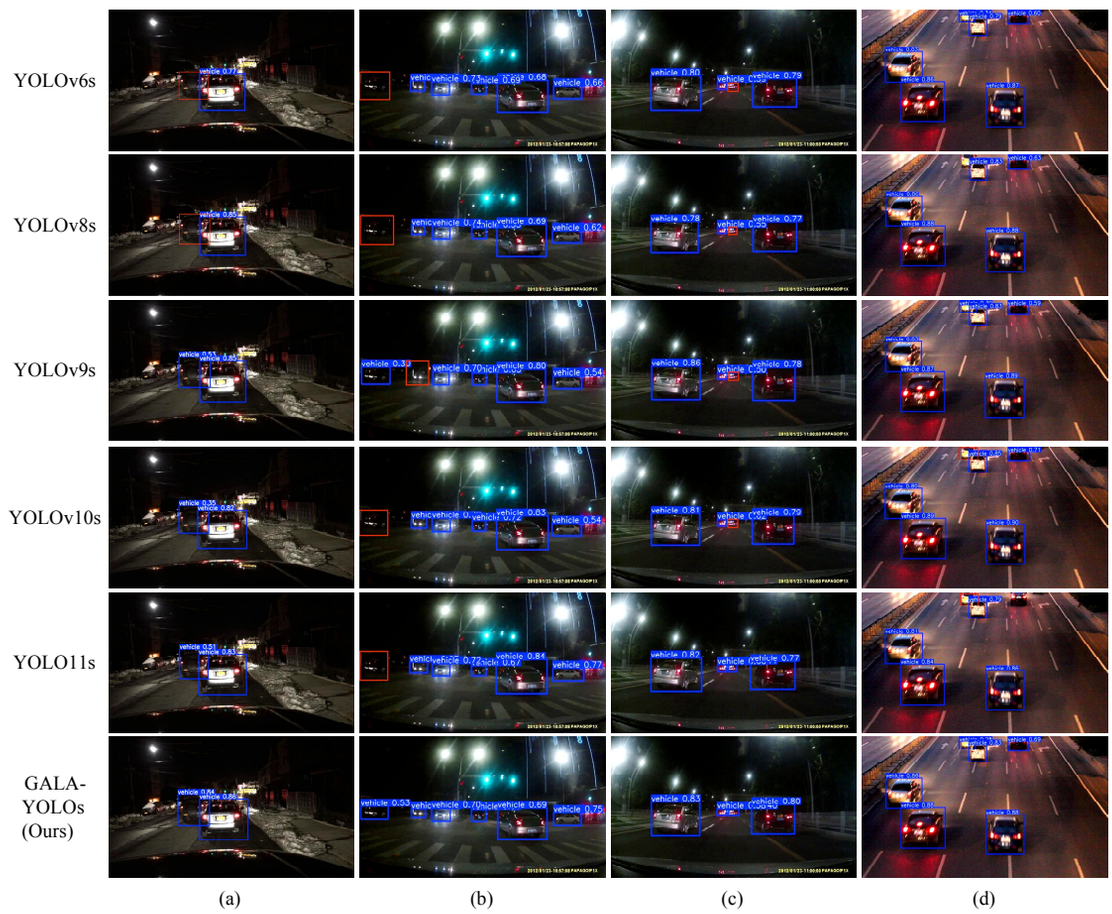


图11 各模型在夜间车辆检测的结果对比

Fig. 11 Comparison of the results of different models in nighttime vehicle detection

述,尽管保持了较低的参数量和计算复杂度,本文方法在性能上优于当前主流的目标检测算法,实验结果验证了其在实际应用中的有效性和优势。

3.3 KITTI上的额外验证

为了验证本文所提出的GALA-YOLO模型在白天场景中的性能,本文选择了KITTI数据集进行车辆检测实验。该数据集的场景为白天,实验结果如表4所示,本文使用单张RTX 3090显卡(显存24GB)来测试各模型在推理时的FPS。在与YOLOv6s、YOLOv8s、YOLOv9s、YOLOv10s和YOLO11s五种主流目标检测模型进行对比后,GALA-YOLO模型在各个类别上均取得了显著的性能提升,整体精确度为89.2%,$mAP_{0.5:0.95}$为58.3%。尤其在"Vehicle"类别的检测中,GALA-YOLO表现尤为突出。精确度达到93.2%,相比YOLOv6s、YOLOv8s、YOLOv9s、YOLOv10s和YOLO11s分别提高了2.9%、0.6%、1.8%、5.2%和5.0%;$AP_{0.5:0.95}$为76.7%,分别比YOLOv6s、YOLOv8s、YOLOv9s、YOLOv10s和YOLO11s提高了3.3%、0.5%、0.7%、1.6%和9.0%。为了直观体现,图12展示了各模型在KITTI数据集上的检测结果对比。这表明,GALA-YOLO模型在白天场景下,特别是在车辆检测任务中,依然表现出显著的优势。具体而言,GA模块通过全局特征提取增强了模型对车辆的检测精度,而LA模块则使模型能够聚焦于含有车辆的区域,优化了细节特征的提取,尤其对于长条形物体(如车辆)具有更好的识别效果。然而,LA模块的条形卷积虽然在车辆检测上有显著提升,但也导致了对行人、骑行者等物体类别检测精度的轻微下降,这反映了该模块的某些局限性。但整体而言,LA模块仍在车辆检测任务中发挥了重要作用。此外,MPDIoU损失函数进一步优化了预测框与真实框之间的匹配,特别是在长宽比相似的框之间,提高了检测精度。综上所述,GALA-YOLO模型通过GA和LA模块的协同作用,在车辆检测任务中实现了精度与效率的平衡。

表 4 KITTI数据集上的检测结果

Table 4　Detection results on the KITTI dataset

Model	Vehicle		Tram		Pedestrian		Cyclist		Misc		Average		FPS
	Precision	AP	P.	AP	P.	AP	P.	AP	P.	AP	P.	mAP	
	/%	/%	/%	/%	/%	/%	/%	/%	/%	/%	/%	/%	
YOLOv6s	90.3	73.4	85.0	59.6	84.0	36.8	76.9	41.5	85.1	47.5	84.3	51.7	149
YOLOv8s	92.6	76.2	87.2	68.2	85.9	40.8	85.6	49.5	88.8	58.5	88.0	58.6	128
YOLOv9s	91.4	76.0	85.6	67.5	80.6	39.8	83.7	48.8	85.3	55.7	85.3	57.6	67
YOLOv10s	88.0	75.1	82.9	63.1	80.0	38.5	82.9	46.9	81.1	51.5	83.0	55.0	100
YOLO11s	88.2	67.7	63.8	52.0	75.0	31.2	80.1	32.7	79.7	39.3	77.3	44.6	102
GALA-YOLO (ours)	93.2	76.7	87.3	68.1	86.5	40.3	87.0	48.6	91.8	58.0	89.2	58.3	105

注:最优结果用蓝色加粗显示,次优结果用绿色加粗显示。

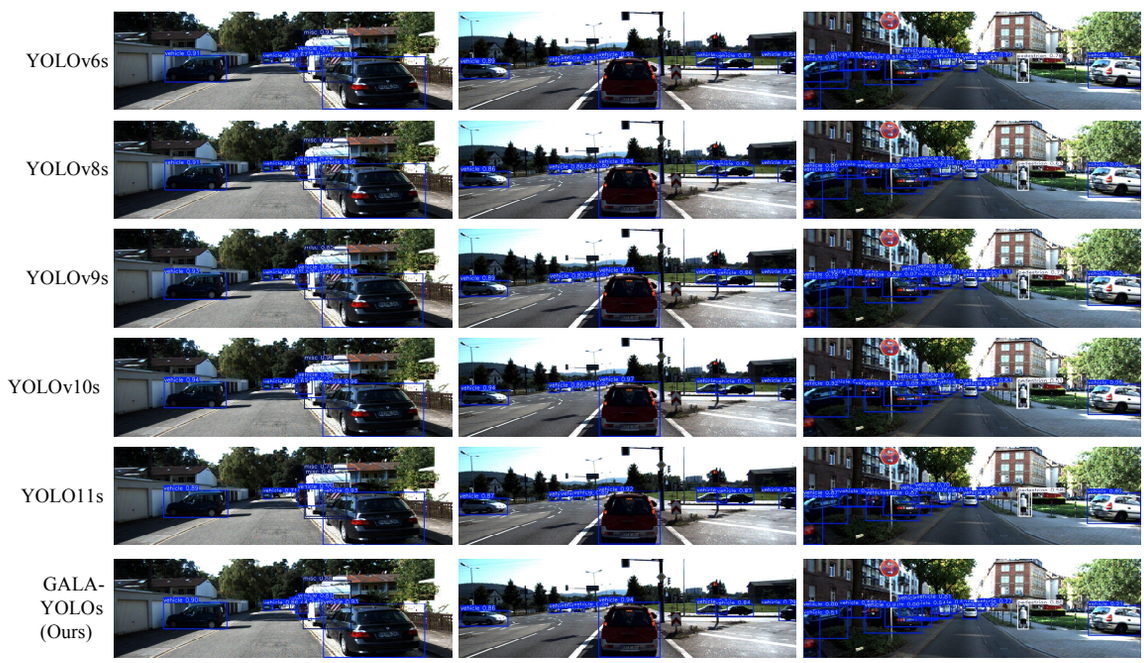


图 12 各模型在KITTI数据集的检测结果对比

Fig. 12　Comparison of detection results of different models on the KITTI dataset

无论是在夜间还是白天场景中,GALA-YOLO均能保持较高的检测性能。

4　结束语

为解决夜间驾驶中车辆检测精度较低的问题,本文提出了一种基于YOLOv8的改进模型——GALA-YOLO。该模型通过全局增强模块(GA)和局部注意力模块(LA)的协同作用,有效分离了背景噪声和车辆特征,显著提升了车辆检测的精度与鲁棒性。全局增强模块改善了模型在低亮度和低对比度环境中的表现,而局部注意力模块则增强了对夜间车辆细节的捕捉能力。同时,MPDIoU损失函数的引入进一步优化了目标的定位精度。实验结果表明,GALA-YOLO不仅在夜间场景下表现优异,还在KITTI数据集的白天场景中展现了卓越的性能,验证了该模型的鲁棒性与广泛适用性。

综上所述,GALA-YOLO通过多模块的协同优化,不仅提升了夜间复杂环境下的检测精度,还在各种环境下保持了较为稳定的表现,进一步证明了该模型的实用性与优势。GALA-YOLO为高级辅助驾驶系统中的车辆检测任务提供了一种高效且精准的解决方案,具有重要的参考价值。

参考文献:

[1] Iqbal A, Rehman Z U, Ali S, et al. Road traffic accident analysis and identification of black spot locations on highway [J]. Civil Engineering Journal, 2020,

6(12): 2448-2456.

[2] Hang J, Yan X, Li X, et al. In-vehicle warnings for work zone and related rear-end collisions: a driving simulator experiment [J]. Accident Analysis & Prevention, 2022, 174: 106768.

[3] Arthurs P, Gillam L, Krause P, et al. A taxonomy and survey of edge cloud computing for intelligent transportation systems and connected vehicles [J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 23(7): 6206-6221.

[4] 黄玲, 郭亨聪, 张荣辉, 等. 人机混驾环境下基于LSTM的无人驾驶车辆换道行为模型[J]. 中国公路学报, 2020, 33(7): 156-166.
Huang Ling, Guo Heng-cong, Zhang Rong-hui, et al. LSTM-based lane-changing behavior model for unmanned vehicle under environment of heterogeneous human-driven and autonomous vehicles [J]. China Journal of Highway and Transport, 2020, 33(7): 156-166.

[5] 张荣辉, 游峰, 初鑫男, 等. 车-车协同下无人驾驶车辆的换道汇入控制方法[J]. 中国公路学报, 2018, 31(4): 180-191.
Zhang Rong-hui, You Feng, Chu Xin-nan, et al. Lane change merging control method for unmanned vehicle under V2V cooperative environment[J]. China Journal of Highway and Transport, 2018, 31(4): 180-191.

[6] Zhang X, Story B, Rajan D. Night time vehicle detection and tracking by fusing vehicle parts from multiple cameras[J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 23(7): 8136-8156.

[7] Bai P F. Nighttime vehicle detection method based on headlight features[J]. Electronic Measurement Technology, 2020, 43(14): 89-95.

[8] Kavya T S, Tsogtbaatar E, Jang Y M, et al. Nighttime vehicle detection based on brake/tail light color [C]//International SoC Design Conference (ISOCC), Daegu, Korea (South), 2018: 206-207.

[9] Kuang H, Chen L, Chan L L H, et al. Feature selection based on tensor decomposition and object proposal for night-time multiclass vehicle detection[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2018, 49(1): 71-80.

[10] Mo Y, Han G, Zhang H, et al. Highlight-assisted nighttime vehicle detection using a multi-level fusion network and label hierarchy[J]. Neurocomputing, 2019, 355: 13-23.

[11] 范佳琦, 李鑫, 霍天娇, 等. 基于单阶段算法的智能汽车跨域检测研究[J]. 中国公路学报, 2022, 35(3): 249-262.
Fan Jia-qi, Li Xin, Huo Tian-jiao, et al. Research on cross domain detection in intelligent vehicles based on one-stage algorithm[J]. China Journal of Highway and Transport, 2022, 35(3): 249-262.

[12] 程鑫, 王宏飞, 周经美, 等. 基于体素柱形的激光雷达点云车辆目标检测算法[J]. 中国公路学报, 2023, 36(3): 247-260.
Cheng Xin, Wang Hong-fei, Zhou Jing-mei, et al. Vehicle detection algorithm based on voxel pillars from LiDAR point clouds[J]. China Journal of Highway and Transport, 2023, 36(3): 247-260.

[13] 张炳力, 潘泽昊, 姜俊昭, 等. 基于交叉注意力机制的多模态感知融合方法[J]. 中国公路学报, 2024, 37(03): 181-193.
Zhang Bing-li, Pan Ze-hao, Jiang Jun-zhao, et al. Multi-modal perception fusion method based on cross attention[J]. China Journal of Highway and Transport, 2024, 37(03): 181-193.

[14] Arinaldi A, Pradana J A, Gurusinga A A. Detection and classification of vehicles for traffic video analytics[J]. Procedia Computer Science, 2018, 144: 259-268.

[15] Park J M, Lee J W. Combination of SSD and a Rule-based approachfor nighttime vehicle detection on roads [J]. Journal of Institute of Control, Robotics and Systems, 2020, 26(6): 493-498.

[16] Chen L, Hu X, Xu T, et al. Turn signal detection during nighttime by CNN detector and perceptual hashing tracking[J]. IEEE Transactions on Intelligent Transportation Systems, 2017, 18(12): 3303-3314.

[17] Redmon J. You only look once: unified, real-time object detection[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, Nevada, USA, 2016: 779-788.

[18] 王艾迪, 彭一川, 郎洪, 等. 基于YOLOX-Transformer两步模型的路面坑槽提取方法[J]. 中国公路学报, 2023, 36(12): 304-317.
Wang Ai-di, Peng Yi-chuan, Lang Hong, et al. Pavement pothole extraction based on YOLOX-transformer two-step model[J]. China Journal of Highway and Transport, 2023, 36(12): 304-317.

[19] 胡晓伟, 闫奕昕, 王大为, 等. 基于YOLOM算法的路面病害轻量化检测方法[J]. 中国公路学报, 2024, 37(12): 381-391.
Hu Xiao-wei , Yan Yi-xin Wang Da-wei, et al. Lightweight pavement disease detection based on YOLOM algorithm[J]. China Journal of Highway and

Transport, 2024, 37(12): 381-391.

[20] 邓实强,丁浩,蒋树屏,等. 基于改进YOLOv5s的公路隧道火灾烟雾智能检测算法[J]. 中国公路学报, 2024, 37(11): 194-209.
Deng Shi-qiang, Ding Hao, Jiang Shu-ping, et al. Intelligent detection algorithm for fire smoke in highway tunnel based on improved YOLOv5s[J]. China Journal of Highway and Transport, 2024, 37(11): 194-209.

[21] He K, Zhang X, Ren S, 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.

[22] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017: 2117-2125.

[23] Dai J, Qi H, Xiong Y, et al. Deformable convolutional networks[C]//2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017: 764-773.

[24] Ding X, Zhang X, Han J, et al. Scaling up your kernels to 31×31: Revisiting large kernel design in CNNs[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 2022: 11963-11975.

[25] 张帆,王景波,朱杨,等. 面向遥感图像舰船识别的轻量化目标检测算法[J]. 吉林大学学报:工学版, 2026, 56(2): 533-542.
Zhang Fan, Wang Jing-bo, Zhu Yang, et al. Lightweight object detection algorithm for ship recognition in remote sensing image[J]. Journal of Jilin University (Engineering and Technology Edition), 2026, 56(2): 533-542.

[26] 高云龙, 任明, 吴川, 等. 基于注意力机制改进的无锚框舰船检测模型[J]. 吉林大学学报:工学版, 2024, 54(5): 1407-1416.
Gao Yun-long, Ren Ming, Wu Chuan, et al. An improved anchor-free model based on attention mechanism for ship detection[J]. Journal of Jilin University (Engineering and Technology Edition), 2024, 54(5): 1407-1416.

[27] Peng C, Zhang X, Yu G, et al. Large kernel matters-improve semantic segmentation by global convolutional network[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017: 4353-4361.

[28] Hou Q, Zhang L, Cheng M M, et al. Strip pooling: Rethinking spatial pooling for scene parsing[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020: 4003-4012.

[29] Zheng Z, Wang P, Liu W, et al. Distance-IoU loss: faster and better learning for bounding box regression [C]//AAAI Conference on Artificial Intelligence, New York, NY, USA, 2020: 12993-13000.

[30] Ma S, Xu Y. MPDIoU: a loss for efficient and accurate bounding box regression[J]. arXiv preprint arXiv: 2307.07662, 2023.

[31] Geiger A, Lenz P, Stiller C, et al. Vision meets robotics: The KITTI dataset[J]. The International Journal of Robotics Research, 2013, 32(11): 1231-1237.

[32] Li C, Li L, Geng Y, et al. YOLOv$_6$ v3.0: a full-scale reloading[J]. arXiv preprint arXiv:2301.05586, 2023.

[33] Ultralytics. YOLOv8 release v8.1.0[EB/OL]. 2024-11-28 (2024-01-10). https://github.com/ultralytics/ultralytics/releases/tag/v8.1.0.

[34] Wang C Y, Yeh I H, Mark Liao H Y. YOLOv9: learning what you want to learn using programmable gradient information[C]//European Conference on Computer Vision (ECCV), Milan, Italy, 2025: 1-21.

[35] Wang A, Chen H, Liu L, et al. YOLOv10: Real-time end-to-end object detection[J]. Advances in Neural Information Processing Systems (NeurIPS), 2024, 37: 107984-108011.

[36] Ultralytics. YOLO11 release v8.3.38[EB/OL]. 2024-11-28(2024-11-26). https://github.com/ultralytics/ultralytics/releases/tag/v8.3.38.

[37] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, Nevada, USA, 2016: 2818-2826.