

基于改进 YOLOv10n 的泊车图像旋转目标检测算法

梁列金¹ 李 想¹ 何永华² 周 璇³

(1. 广东财经大学信息学院 广州 510320; 2. 广东财经大学统计与数学学院 广州 510320;

3. 华南理工大学机械与汽车工程学院 广州 510640)

摘 要: 目标检测是无人驾驶时代自动泊车智能感知的关键技术之一。鱼眼相机感知过程存在环境因素复杂、障碍物类型多样、鱼眼镜头下检测对象图像失真等问题,常规算法难以保证自动泊车复杂场景下各类对象的检测精度。为此,本文提出了一种基于改进 YOLOv10n 的旋转目标检测方法,在主干网络引入 SPPELAN 模块,并利用 DSConv 改进 C2f 中部分卷积融合 iRMB 模块,以提高鱼眼镜头下的特征提取能力,增强小目标对象的定位能力;然后采用 ATFL 函数,增强模型在检测目标特征上的聚焦能力。实验结果表明,改进后的算法在鱼眼相机泊车数据集上的 mAP@0.5 达到了 89.89%,mAP@0.5:0.95 达到了 69.36%,比基准模型分别提高了 0.62% 和 0.6%,为泊车感知技术提供了新的思路。

关键词: 旋转目标检测;YOLOv10n;泊车图像;动态蛇形卷积;深度学习

中图分类号: TN911.73 **文献标识码:** A **国家标准学科分类代码:** 520.6040

Rotating target detection algorithm in parking images
based on improved YOLOv10nLiang Liequan¹ Li Xiang¹ He Yonghua² Zhou Xuan³

(1. School of Information Science, Guangdong University of Finance and Economics, Guangzhou 510320, China;

2. School of Statistics and Mathematics, Guangdong University of Finance and Economics, Guangzhou 510320, China;

3. School of Mechanical and Automotive Engineering, South China University of Technology, Guangzhou 510640, China)

Abstract: Object detection is one of the key technologies for intelligent perception in autonomous parking systems in the era of autonomous driving. The perception process of fisheye cameras faces several challenges, including complex environmental factors, a diverse range of obstacles, and image distortion of detection targets under fisheye lenses. Conventional algorithms struggle to maintain high detection accuracy for various objects in complex parking scenarios. To address this, this paper proposes a rotation-based object detection method using an improved YOLOv10n model. The approach introduces the SPPELAN module into the backbone network, and utilizes DSConv to enhance the C2f module by improving the convolution fusion of the iRMB. This improves feature extraction capability under fisheye lenses and enhances the localization ability of small objects. Additionally, an ATFL function is employed to strengthen the model's focus on target features. Experimental results demonstrate that the improved algorithm achieves a mAP@0.5 of 89.89% and a mAP@0.5:0.95 of 69.36% on the fisheye camera parking dataset, outperforming the baseline model by 0.62% and 0.6%, respectively. This provides new insights into the development of parking perception technologies.

Keywords: rotating target detection; YOLOv10n; parking image; dynamic snake convolution; deep learning

0 引 言

自动泊车技术作为智能驾驶系统的重要组成部分,在促进智能交通发展和提高停车效率等方面发挥着重要作用^[1]。目标检测是自动泊车环境感知的重要组成部分,车

位与障碍物的准确识别是提高自动泊车准确性的前提。实际泊车场景中,由于泊车图像大多为俯视视角,任意的停车位方向使得水平目标检测算法容易丢失泊车目标方向特性,掺杂更多无用的背景信息^[2],且车位线划分方式多样、车位线的磨损与遮挡、光照条件的变化以及成像畸变等复

杂场景性均对车位识别的准确性与鲁棒性构成了挑战^[3]。

鱼眼相机能够覆盖更广的视野,具有优秀的环境感知能力,成为自动泊车系统中常用的视觉传感器。然而,由于鱼眼相机图像存在严重的径向畸变,标准算法难以直接应用于环视感知任务中^[4],同时,由于畸变,鱼镜头头下的障碍物目标会呈现不规则形状,与常规相机相比,进一步增加了鱼眼相机目标检测的复杂性,引起广大学者的普遍关注。

近年来,许多基于深度学习的泊车目标检测方法被提出。基于深度学习的目标检测算法按照是否产生候选区域,可分为两阶段目标检测算法和单阶段目标检测算法。两阶段目标检测算法的基本思想是先产生候选区域,然后对这些区域进行分类和边界框回归,此类方法的代表是区域卷积神经网络(region-based convolutional neural network, R-CNN)系列算法,例如快速区域卷积神经网络(faster region-based convolutional neural network, Faster R-CNN)^[5]和掩模区域的卷积神经网络(mask region-based convolutional neural network, Mask R-CNN)^[6]等。单阶段目标检测算法直接在图片上预测类别和边界框,不需要先生成候选区域,因此速度通常比两阶段算法快,单阶段目标检测代表算法有 YOLO(you only look once, YOLO)系列^[7-9]、单次多边框检测(single shot multiBox detector, SSD)^[10]和 Retina-Net^[11]等。

针对自动泊车领域存在的诸多挑战,Grbic 等^[12]通过 YOLOv5 算法与基于密度的聚类方法(density-based spatial clustering of applications with noise, DBSCAN)相结合的方式,实现了自动泊车位检测,在复杂场景中表现较高的适应性和准确性。Suhr 等^[13]提出了一种端到端可训练的单阶段车位检测方法,通过卷积神经网络(convolutional neural network, CNN)同时获取停车泊位的全局和局部信息,实现了泊车位的位置、方向、类型和占用状态的精确检测;Yu 等^[14]提出了一种名为选择性剪枝全卷积网络(select and prune the fully convolutional networks, SPFCN)的实时停车泊位检测模型,通过引入选择-剪枝模块,在每个训练阶段自动选择最佳感受野并剪枝冗余通道,从而实现了更高的推理速度和较小的模型尺寸。董宝鑫等^[15]基于快速空间金字塔池化(spatial pyramid pooling, SPPF),进行了目标感受野的扩展,增强了对旋转目标的检测精度。邝先验等^[16]通过引入焦点与高效交并比(focal and efficient intersection over union, Focal-EIoU)损失函数,提升了模型泛化能力。童小钟等^[17]通过多尺度特征与注意力机制的融合,提高了对小目标检测的性能。

Xu 等^[18]为解决城市停车位实时检测与需求预测难题,提出多阶段车位检测框架与卷积神经网络-长短期记忆网络(convolutional neural network-long short-term memory, CNN-LSTM)混合模型,通过预处理增强抗干扰性,结合空间特征提取与时间序列预测,在 1 158 组数据中实现 98.4% 车位识别准确率,预测误差较传统长短期记忆

网络(long short-term memory network, LSTM)降低 47%,但模型对空间特征优化不足。宋绍京等^[19]针对复杂驾驶场景下小目标车辆漏检及遮挡问题,使用 YOLOP 构建多任务环境感知网络模型(multi-task environment perception network, MEPNet),采用 YOLOv7 主干增强实时性,通过四分支感受野模块优化多尺度特征提取,扩展小目标检测层并结合上采样算子模块(content-aware reassembly of features, CARAFE)进行上采样,提升了模型在复杂驾驶场景下目标检测精度。

闫旭东等^[20]为解决自主泊车系统中相邻停车位交叉及复杂场景检测精度不足问题,提出的基于全景环视图像的停车位检测方法,通过结合改进的 Transformer 模块与卷积神经网络,实现了高精度的车位角点检测与定位。杨飞帆等^[21]为解决自动泊车中环视图像车位检测的轻量化与复杂场景精度不足问题,提出改进的 YOLOv8 算法,采用融合部分卷积的跨阶段三卷积快速模块(fast cross-stage partial convolution module with three partial convolutions, C3-Faster)模块和 CoordAtt 注意力机制优化主干网络,结合高效交并比(efficient intersection over union, EIoU)损失函数强化边界定位,在降低模型计算量的同时,显著提升其自建 AVM 数据集的车位检测精度与效率。然而这两项研究未考虑自动泊车场景中行人和障碍物的动态交互关系,限制了复杂场景适应性。

上述方法大部分采用水平目标检测方法来进行停车位检测,针对倾斜车位检测效果有限。传统的水平检测框冗余信息增多、在密集场景下容易出现重叠、遮挡的情况,难以准确识别具有方向性的泊车目标。为解决此问题,本文引入旋转目标检测算法,旋转目标检测通过引入旋转边界框,能够更加紧凑地包围目标,减少背景干扰,从而提高检测精度和定位准确性。

考虑到车载芯片存储空间的限制以及驾驶安全性的需求,泊车场景中的目标检测模型不仅需要具备轻量化的设计,同时也要在检测效率和精度方面有所保障。YOLOv10n 作为实时的端到端检测模型,有着较高的检测精度和速度。因此,本文选用 YOLOv10n 作为基准模型,以满足泊车目标检测任务对精确性和高效性的双重需求。针对泊车图像车位及障碍物并行检测任务的特点,本文以 YOLOv10n 算法框架为基础进行改进,提出一种基于改进 YOLOv10n 的泊车图像旋转目标检测算法,本文主要工作有以下 3 点:

1) 针对泊车图像中细长不规则,畸变的目标定位困难的问题,使用动态蛇形卷积(dynamic snake convolution, DSConv)^[22]替换 C2f 部分普通卷积,提高网络对细长目标特征的感知能力,同时融合倒置残差移动模块(inverted residual mobile block, iRMB)^[23],增强对重要目标的特征提取能力。

2) 考虑到实际泊车场景背景的复杂性,应用自适应阈

值焦点损失 (adaptive threshold focal loss, ATFL) 函数^[24]替换二元交叉熵 (binary cross entropy, BCE) 损失函数, 使模型聚焦于目标特征, 减少对背景特征的关注。

3)面向泊车场景8种典型目标识别过程中,锥筒、限位杆等小目标检测精度不高的问题,在YOLOv10n主干网络中引入高效局部聚合网络的空间金字塔池化(spatial pyramid pooling efficient local aggregation network, SPPELAN)^[25]模块,通过多尺度池化,实现多尺度特征融合,增强模型对泊车小目标的识别能力。

1 YOLOv10n 算法

YOLOv10n 模型于 2024 年由清华大学研究团队提出,其创新亮点在于引入了一致性双重分配策略,实现了无非极大值抑制(non-maximum suppression, NMS)的高效训练与推理^[26]。这一改进大幅减少了推理时的计算开销,同时避免了由于 NMS 参数调整而带来的性能波动,使得模型在端到端推理中保持一致性和高效性。YOLOv10 在提升物体检测精度的同时,维持了优异的实时检测能力,适用于对精度与效率要求严格的应用场景。

YOLOv10n 算法模型主要由骨干网络(Backbone)、颈部网络(Neck)和检测头(Head)三部分组成,如图 1 所示。Backbone 主要负责图片特征的提取,保留了 YOLOv8 中的跨阶段部分模块(cross stage partial network, c2f)和 SPPF 结构,并新增了空间-通道解耦下采样(spatial-channel decoupled downsampling, SCDown)、基于紧凑倒置块的跨阶段部分网络(cross stage partial network with compact inverted blocks, C2fCIB)和部分自注意力(partial self-attention, PSA)模块。Neck 采用特征图金字塔网络(feature pyramid network, FPN)^[27]和路径聚合网络(path aggregation network, PAN)^[28]结构,通过对 Backbone 提取的不同尺度特征进行融合,从而增强网络特征聚合能力。Head 采用 One-to-one Head 和 One-to-many Head 轻量级的检测头设计,分别处理分类、回归及置信度任务,并整合输出结果。最终采用一致性的双重分配策略,消除了对 NMS 的需求,获得检测结果。

YOLOv10n-OBb 算法是在 YOLOv10n 的基础上对检测头进行了修改,在解耦头增添了对角度的预测输出头,即用定向边界框(oriented bounding box, OBb)取代原先的水平边届框(horizontal bounding box, HBb)。并替换了原有的完全交并比(complete intersection over union, CIoU)损失函数,采用概率交并比(probabilistic intersection-over-union, ProbIoU)^[29] 损失函数将真实框和预测框转为高斯分布来对旋转矩形框进行预测,以达到对泊车目标进行旋转检测的目的。

尽管 YOLOv10n 模型具备较高的精度和实时性,但在泊车场景中的应用仍存在一定局限性:泊车场景中复杂背景和密集目标会干扰模型的检测性能,导致漏检或误检

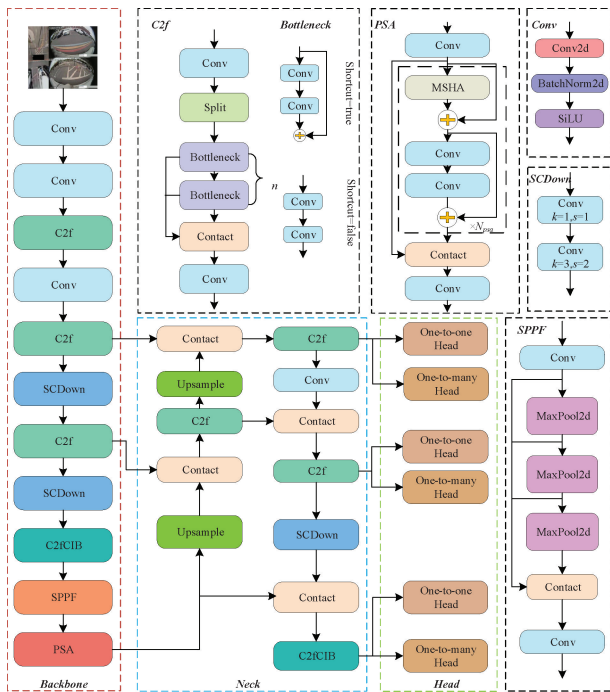


图 1 YOLOv10n 网络结构

Fig.1 YOLOv10n network structure

情况增加。对于小目标的检测, YOLOv10n 的标准结构可能难以准确识别微小特征, 降低了小目标检测的效果; 图 2 为不同泊车图像采用水平框标注和旋转框标注的对比图。从图 2 中可以看出, 泊车位的角度和方向在鱼眼俯视图图像中可能各不相同。在使用水平检测框时, 往往会包含大量冗余信息, 影响了模型对真实目标位置和形状的有效表达。因此需要对 YOLOv10n 进行进一步改进, 以更好完成泊车场景下的检测任务。

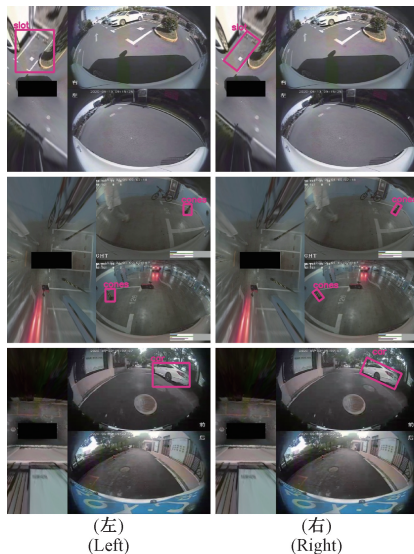


图 2 同一图像使用水平框(左)和旋转框(右)标注对比
Fig.2 Comparison of annotation effects of the same image
using horizontal box (left) and rotation box (right)

2 本文算法

2.1 DSConv 模块

DSConv 是一种专门设计用于捕捉细长和扭曲局部特征的卷积模块。DSConv 通过动态调整卷积核的形状和位置,增强了对细长目标特征的感知能力。不同于传统的固定卷积核大小的卷积操作,DSConv 允许卷积核根据特征间的空间关系自适应调整,从而更好地捕捉特征之间的相关性和局部细节,提高特征建模能力。

DSConv 的设计中,考虑到卷积核在 x 轴和 y 轴方向上的位置变化,以 9×9 大小的卷积核为例,每个网格的具体位置可以用 $\mathbf{K}_{i \pm c} = (x_{i \pm c}, y_{i \pm c})$ 表示,其中 $c = \{1, 2, 3, 4\}$ 代表从中心网格的水平距离。卷积核中每个网格位置的选择是一个累积过程,从中心位置 \mathbf{K}_i 开始,远离中心网格的位置取决于前一个网格的位置; \mathbf{K}_{i+1} 相比于 \mathbf{K}_i 增加了一个偏移量 $\Delta = \{\delta \mid \delta \in [-1, 1]\}$,确保卷积核符合线性形态。其在 x 轴和 y 轴的执行过程如式(1)和(2)所示。

$$\mathbf{K}_{i \pm c} = \begin{cases} (x_{i+c}, y_{i+c}) = (x_i + c, y_i + \sum_{i=1}^{i+c} \Delta y) \\ (x_{i-c}, y_{i-c}) = (x_i - c, y_i + \sum_{i=1}^{i-c} \Delta y) \end{cases} \quad (1)$$

$$\mathbf{K}_{j \pm c} = \begin{cases} (x_{j+c}, y_{j+c}) = (x_i + \sum_{j=1}^{j+c} \Delta x, y_i + c) \\ (x_{j-c}, y_{j-c}) = (x_i + \sum_{j=1}^{j-c} \Delta x, y_i - c) \end{cases} \quad (2)$$

由于偏移量 Δ 通常是小数,因此采用双线性插值将位置以整数形式表达,其实现如式(3)所示。

$$\mathbf{K} = \sum_{\mathbf{K}'} \mathbf{B}(\mathbf{K}', \mathbf{K}) \cdot \mathbf{K}' \quad (3)$$

式中: \mathbf{K} 为式(1)和(2)的小数的位置, \mathbf{K}' 枚举了所有可能的整数的位置, \mathbf{B} 为双线性插值核,可以分解为两个一维核,如式(4)所示。

$$\mathbf{B} = (\mathbf{K}, \mathbf{K}') = \mathbf{b}(\mathbf{K}_x, \mathbf{K}'_x) \cdot \mathbf{b}(\mathbf{K}_y, \mathbf{K}'_y) \quad (4)$$

通过这种方式,DSConv 在保持感知范围灵活性的同时,确保了对细长目标的关注和捕捉更加准确和稳定。图 3 为 DSConv 感受野变化图,展示了 DSConv 的感受野变化情况,表明 DSConv 通过动态调整其卷积核的大小和形状,提高了模型对细长目标的捕捉能力,减少了由于大的变形偏移而导致的感知范围失效问题,进一步增强了模型的鲁棒性和泛化能力。

图 4 为鱼镜头畸变下的骑行者图。泊车图像中,由于鱼镜头造成的透视畸变,骑行者可能呈现为细长且不规则的形态,如图 4 所示。传统的固定大小卷积核难以有效捕捉这类细长目标的完整特征信息,当目标具有较高的纵横比时,传统的卷积操作容易忽略目标沿长度方向上的细节。

DSConv 通过自适应调整卷积核的形状和位置,能够针

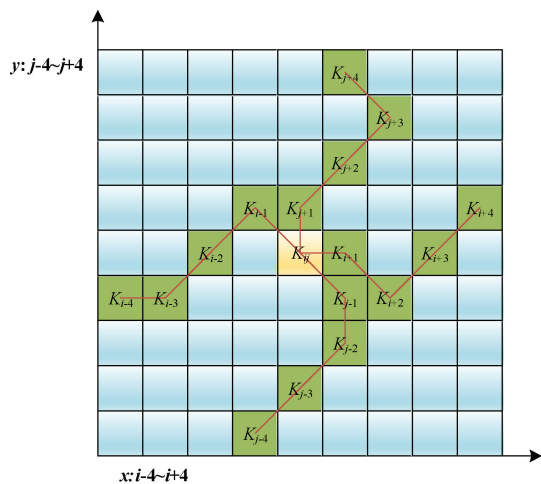


图 3 DSConv 感受野变化图

Fig. 3 DSConv receptive field change map



图 4 鱼镜头畸变下的骑行者图像

Fig. 4 The rider image under fisheye lens distortion

对性地捕捉细长的车辆的局部特征,避免了传统固定大小卷积核在处理此类目标时的局限性。将 DSConv 融入 C2f 模块中,融合后的结构如图 5 所示。C2f 与 DSConv 结合不仅可以保持原有的多尺度特征融合优势,还能进一步增强对细长目标的特征捕捉能力。实验结果表明,模型在处理骑行者这类目标时,融合 DSConv 的 C2f 模块能够提供更高的检测精度,相较于原始 YOLOv10n OBB, P 值提升了 1.43%, R 值提高了 1.32%, mAP@0.5 提高了 0.73%。

2.2 ATFL 损失函数

传统的交叉熵损失函数定义如式(5)所示。尽管交叉熵损失函数简单有效,但它无法解决样本之间的不平衡问题。

$$\mathcal{L}_{BCE} = -(y \log p + (1 - y) \log(1 - p)) \quad (5)$$

式中: p 表示模型输出预测概率, y 表示真实标签。其简化形式如式(6)和(7)所示。

$$\mathcal{L}_{BCE} = -\log(p_i) \quad (6)$$

式中:

$$p_i = \begin{cases} p, & y = 1 \\ 1 - p, & \text{其他} \end{cases} \quad (7)$$

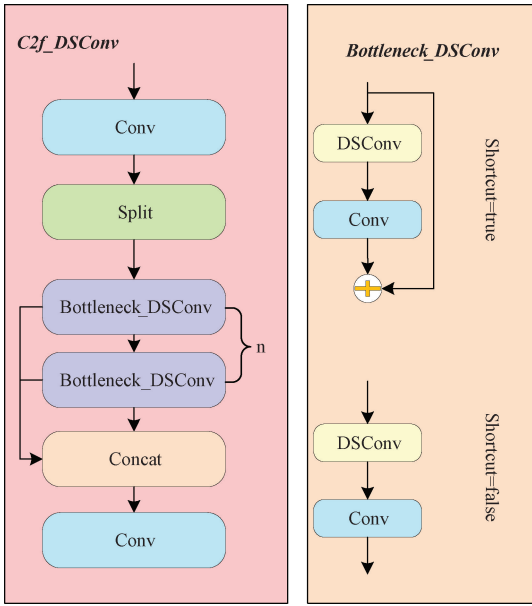


图5 C2f_DSConv 结构

Fig. 5 Structure of C2f_DSConv

针对样本间的不平衡问题,焦点损失(Focal Loss)被提出,并引入了一个调制因子 $(1 - p_t)^\gamma$ 来减少容易分类样本的损失贡献^[11]。焦点损失函数定义如式(8)所示。

$$FL(p_t) = (1 - p_t)^\gamma \log p_t \quad (8)$$

式中: γ 是一个可调参数,用来控制调制因子的作用强度。

焦点损失虽然减少了容易样本的损失权重,同时也降低了困难样本的损失值,这不利于困难样本的学习。为解决此问题,ATFL 引入了一个阈值设置来分离易识别的背景和难以识别的目标。通过增强与目标相关的损失并减少与背景相关联的损失,ATFL 迫使模型更加关注目标特征,从而缓解目标与背景之间的不平衡。此外,ATFL 还采用了自适应设计来减少因调整超参数而导致的时间消耗。ATFL 的定义如式(9)所示。

$$ATFL = \begin{cases} -(\lambda - p_t)^{-\ln(\hat{p}_t)} \log p_t, & p_t \leq 0.5 \\ -(1 - p_t)^{-\ln(\hat{p}_c)} \log p_t, & p_t > 0.5 \end{cases} \quad (9)$$

式中: λ 是一个常数,用于调整损失权重, \hat{p}_c 表示下一个 epoch 的预测概率值。

本文采用的泊车图像旋转目标检测数据集包含了从强光到弱光等多种光照条件下的复杂场景,挑选白天强光与夜晚弱光条件下的两张对比图,如图6所示。这些光照条件导致背景环境多变,增加了旋转目标检测的难度。在这种情况下,ATFL 通过解耦目标与背景,利用自适应机制调整损失权重,确保模型更多地关注目标特征而非背景特征,从而提高模型在复杂背景下的鲁棒性,确保模型在不同光照条件下均能保持良好表现。

2.3 SPPELAN 模块

传统的 SPPF 模块中,通过结合大卷积核和非膨胀卷

(a) 强光
(a) Strong light(b) 弱光
(b) Weak light

图6 不同光线条件下的复杂场景对比图

Fig. 6 Comparison of complex scenes under different lighting conditions

积来扩大感受野,这种方式虽然能够处理不同尺度的目标,但在处理小目标时可能因特征融合过程中丢失信息而导致检测精度下降。图7为锥形桶小目标图,在泊车数据集中,鱼镜头下的锥形桶通常小于 $32 \text{ pixel} \times 32 \text{ pixel}$,这符合 MS-COCO 对小目标的定义。



图7 锥形桶小目标图像

Fig. 7 Small object image of cones

针对此问题,本文引入 SPPELAN 模块,如图8所示,该模块结合了空间金字塔池化(spatial pyramid pooling, SPP)和高效局部聚合网络(efficient local aggregation network, ELAN)的优点,通过引入多尺度的最大池化操作,对输入特征图进行不同尺度的最大池化操作,再通过 ELAN 对局部特征进行聚合。这一过程在处理小目标时,能够更好地捕捉其特征。

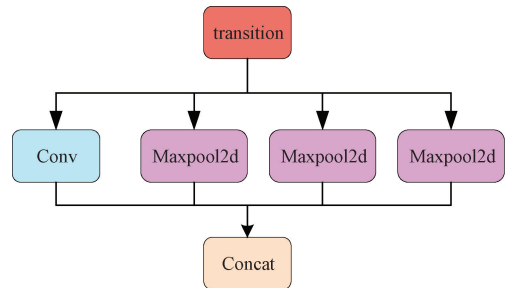


图8 SPPELAN 结构

Fig. 8 Structure of SPPELAN

2.4 iRMB 模块

iRMB 是基于倒置残差块(inverted residual block,

IRB)发展而来的轻量化模块,iRMB有效结合了深度可分离卷积(depthwise separable convolution,DSC)轻量级特性和高效的多头自注意力机制(efficient window-based multi-head self attention,EW-MHSA)动态捕捉能力,在保留计算轻量化的同时,增强了特征提取能力。图 9 为 iRMB 结构范式图。

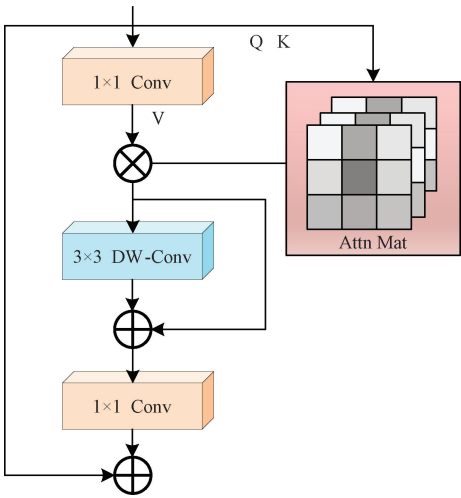


图 9 iRMB 结构范式
Fig. 9 Paradigm of the iRMB

泊车图像中的目标在鱼镜头下通常会因不同的视角和角度而呈现多样化的形态。iRMB 模块的深度可分离卷积与高效注意力机制能够适应不同尺度和方向的目标,有助于精确捕捉旋转目标特征。通过将 iRMB 的长距离捕捉能力融入 C2f 多尺度特征融合过程,能进一步增强了模型在复杂背景下的目标识别能力。具体融合过程如图 10 所示,通过将动态蛇形卷积、倒置残差移动模块和多尺度融合结构有机结合,C2f_DSConv_iRMB 模块具备高度灵活性、轻量化和高效融合的特点,能够在复杂场景和小目标检测中表现更佳。

2.5 改进 YOLOv10n 网络结构

本文所提出的改进 YOLOv10n OBB 网络框架图如图 11 所示。实验中将主干网络中 SPPF 模块替换为 SPPELAN 模块,并在颈部网络中引入 iRMB 模块和 DSConv 模块对 C2f 进行改造,采用 ATFL 损失函数提高预测精度。

3 实验结果及分析

3.1 数据集

本文泊车数据集来源于采集车上安装的全景相机和左右 CMS 相机拍摄的视频。通过截取并拼接各相机拍摄的图像,最终生成包含鱼眼视角的综合图像,共计 17 724 张图像,包含 slot、car、cones、limit_lane、bicycle、person、rider 和 lock 八类典型目标。图 12(a)~(d)所示数据集涵盖了从白天到黑夜,地下停车场到室外停车等各种情况停车场

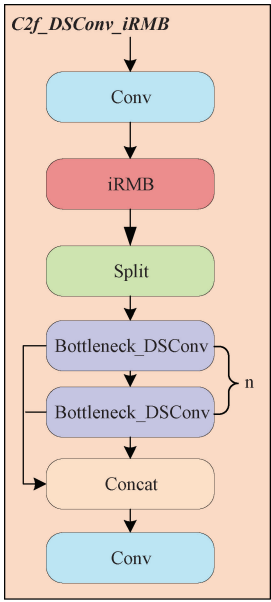


图 10 C2f_DSConv_iRMB 结构
Fig. 10 Structure of C2f_DSConv_iRMB

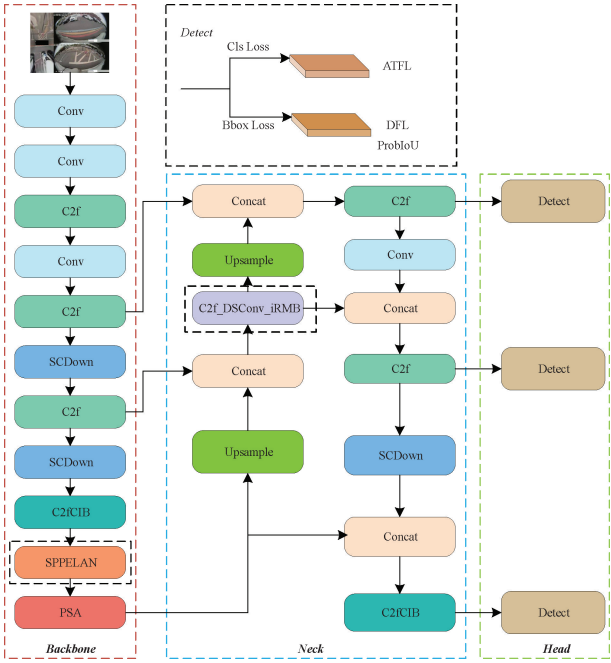


图 11 改进 YOLOv10n OBB 网络结构
Fig. 11 Improved YOLOv10n OBB network structure

景,数据集中每张图像的像素大小为 512×416 。在实验中,数据集以 8:1:1 的比例被划分为训练集 14 179 张、验证集 1 773 张和测试集 1 772 张,部分图像如图 12 所示。

3.2 实验环境及参数配置

实验的软硬件配置如表 1 所示。在训练阶段,批量大小设置为 32,imsgsz 设置为 512,初始学习率为 0.01,动量为 0.937,epoch 设为 200,patience 设为 100,CPU 线程数为 4,采用 SGD 优化器对模型进行优化。

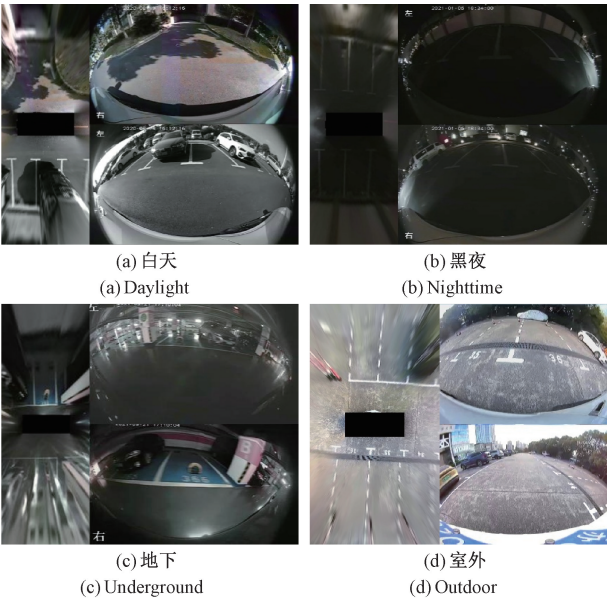


图 12 数据集样例
Fig.12 Dataset samples

表 1 实验环境配置

Table 1 Experimental setup configuration	
配置环境	版本
操作系统	Windows11
CPU	Intel(R) Core(TM) i7-9700K
Pytorch	1.12.0
CUDA	11.4
Python	3.9.19
显卡	NVIDIA GeForce GTX 1080 Ti
内存	96 GB

表 2 消融实验结果

Table 2 Ablation experimental results									
DSCnv	SPPELAN	ATFL	iRMB	Precision/ %	Recall/ %	Params/ M	GFLOPs	mAP@0.5/ %	mAP@0.5:0.95/ %
				84.32	83.41	2.66	8.0	89.27	68.76
✓				84.46	83.42	2.73	8.2	89.46	69.03
✓	✓			84.26	84.02	3.22	8.6	89.61	68.71
✓	✓	✓		82.88	84.54	3.22	8.6	89.72	69.65
✓	✓	✓	✓	83.85	84.24	3.24	9.1	89.89	69.36

注:✓为添加模块,加粗为最优值。

由表 2 可知,相较于原始的 YOLOv10n OBB 模型,将 C2f 中部分卷积替换为 DSCnv 后,该模型在提取细长且不规则目标特征方面表现更佳,虽然模型大小提高了 0.2 M,但 mAP@0.5 由 89.27% 提升到了 89.46%,提高了 0.19%,同时 Precision、mAP@50:95 和 Recall 分别提升了 0.14%、0.27% 和 0.01%;在主干部分,将 SPPF 替换为 SPPELAN 后,增强了对小目标的检测能力,模型的 mAP

3.3 评价指标

本文采用准确率(precision,P)、召回率(recall,R)和平均精度(mean average precision,mAP)作为旋转目标检测模型的性能评价指标。同时,选取参数量(Params)和每秒十亿次浮点运算次数(giga floating point operations per second, GFLOPs),每秒传输帧数(frames per second, FPS)作为模型检测的效率指标。相关计算公式如式(10)~(13)所示。

$$P = \frac{TP}{TP + FP}$$
 (10)

$$R = \frac{TP}{TP + FN}$$
 (11)

$$AP = \int_0^1 P(R) dR$$
 (12)

$$mAP = \frac{\sum_{i=1}^N AP_i}{N} AP$$
 (13)

式中: TP 表示泊车目标预测正确的正样本数量, FP 表示泊车目标预测错误的正样本数量, FN 表示泊车目标预测错误的负样本数量, N 表示类别数量。mAP@0.5 表示 IoU 阈值为 0.5 时所有类别平均检测精度, mAP@0.5:0.95 代表 IoU 阈值自 0.5~0.95 之间,步长为 0.05 的整体平均检测精度。

3.4 消融实验

为研究 4 种改进方法对原始 YOLOv10n OBB 模型的影响,并验证算法改进的有效性,本文在保持训练方法和训练环境一致的前提下,使用泊车数据集进行了消融实验。实验依次将 DSCnv、SPPELAN 和 ATFL 和 iRMB 加入原始 YOLOv10n OBB 算法中,消融实验结果如表 2 所示。

增加了 0.15%,且 Recall 相较于原始模型提升了 0.61%;更换 ATFL 损失函数后,提升了模型在复杂场景下的检测能力,在不增加模型参数量和计算量的情况下,依然可以获得 0.11%的 mAP@0.5 提升, mAP@0.5:0.95 由原来的 68.76%提升到了 69.65%。提升了 0.89%,Recall 也得到了 0.52%的提升;在颈部网络中引入 iRMB, mAP@0.5 提升了 0.17%,但 mAP@0.5:0.95 降低了 0.29%,可能

是 iRMB 模块降低了高分辨率特征的提取能力。

为进一步分析 3 种改进方法对各类检测目标的影响，

本文将原始 YOLOv10n OBB 算法的各类别检测结果与改进后的算法进行了对比，具体结果如表 3 所示。

表 3 各类别消融实验结果

Table 3 Per class results of ablation experiments

DSCnv	SPPELAN	ATFL	iRMB	AP/%								mAP@0.5/
				slot	car	cones	limit_lane	bicycle	person	rider	lock	%
				98.03	96.34	95.70	77.36	82.62	90.89	88.42	84.82	89.27
✓				98.01	96.42	96.00	75.49	82.87	91.79	89.15	85.91	89.46
✓	✓			98.16	96.35	96.26	77.30	83.69	90.77	89.65	84.70	89.61
✓	✓	✓		98.19	96.37	95.68	76.92	84.43	90.91	88.75	86.50	89.72
✓	✓	✓	✓	98.28	96.44	95.52	77.01	84.34	92.11	90.00	85.4	89.89

注:✓为添加模块,加粗为最优值。

由表 3 可知,将 C2f 中部分卷积替换为 DSCnv 后,针对鱼眼图像下畸变的 rider 类的 AP 值提升了 0.9%,证明了 DSCnv 的有效性;替换了 SPPELAN 之后,cones 小目标的 AP 值较原始模型提升了 0.56%,验证了 SPPELAN 模块对小目标检测能力的提升;引入 ATFL 之后,slot、bicylce 和 lock 类的 AP 值,相较于原始模型,分别提升了 0.16%、1.8%和 1.68%;将颈部网络的 C2f 替换为 C2f_DSCnv_iRMB 之后,slot、car、person 和 rider 的 AP 值达到最高,分别较原模型提升了 0.25%、0.1%、1.22%和 1.58%。综上所述,与原始模型对比,改进后的模型在泊车目标检测任务中表现出更优的性能。

3.5 注意力机制对比实验

为验证 iRMB 注意力模块的性能,本文与卷积块注意力模块(convolutional block attention module, CBAM)^[30]、压缩和激励网络(squeeze-and-excitation networks, SE)^[31]、坐标注意力(coordinate attention, CA)^[32]、可变形大核注意力(deformable large kernel attention, D-LKA)^[33]、上下文锚点注意力(context anchor attention, CAA)^[34]和高效多尺度注意力(efficient multi-scale attention, EMA)^[35]注意力模块进行了对比实验,所有实验均在先添加 DSCnv、SPPELAN 和 ATFL 模块的基础上,再次引入不同的注意力模块进行比较,实验结果如表 4 所示。

表 4 注意力模块对比实验结果

Table 4 Comparison experiment results of attention module

注意力模块	Precision/%	Recall/%	Params/M	GFLOPs	mAP@0.5/%	mAP@0.5;0.95/%
CBAM	82.91	84.39	3.23	8.6	89.43	68.50
SE	83.26	83.64	3.22	8.6	89.44	68.95
CA	84.02	83.56	3.22	8.6	89.81	69.29
D-LKA	82.98	84.49	3.62	9.8	89.65	68.78
CAA	83.33	84.18	3.23	8.6	89.19	69.07
EMA	84.21	83.35	3.22	8.6	89.17	68.94
iRMB	83.85	84.24	3.24	9.1	89.89	69.36

注:加粗为最优值。

从表 4 可以看出,与其他注意力模块相比,iRMB 注意力模块在参数量和计算量仅略微增加的情况下,达到了最高的 mAP@0.5(89.89)和 mAP@0.5;0.95(69.36)值,表明 iRMB 能够有效提升模型性能。

3.6 iRMB 模块测试集验证

为进一步验证 iRMB 模块的性能,本文将改进后的 YOLOv10n 模型在 1 772 张测试集上进行验证。模型 1 为改进后的模型去除 iRMB 模块(YOLOv10n+SPPELAN+DSCnv+ATFL),模型 2 为改进后的模型,模型 3 为改进后的模型 Neck 处添加 iRMB 模块。图 13 为模型 3 具体的

网络结构,表 5 为模型 1、2 和 3 在不同类别上性能对比表,表 6 为模型 1、2 和 3 的性能对比表。

由表 5 和表 6 的实验结果可知,在相同测试集条件下,模型 2(添加 1 个 iRMB 模块)和模型 3(添加 2 个 iRMB 模块)相较于模型 1 性能都有所提升。在综合检测指标 mAP@0.5 方面,模型 2 达到 91.59%(+0.44%),模型 3 达到 91.85%(+0.70%);在更严格的 mAP@0.5;0.95 评估标准下,模型 2 和模型 3 分别提升 1.36%和 1.71%,验证了 iRMB 模块对模型鲁棒性的增强作用。

从具体类别检测性能来看,模型 3 在 slot(98.17%)、

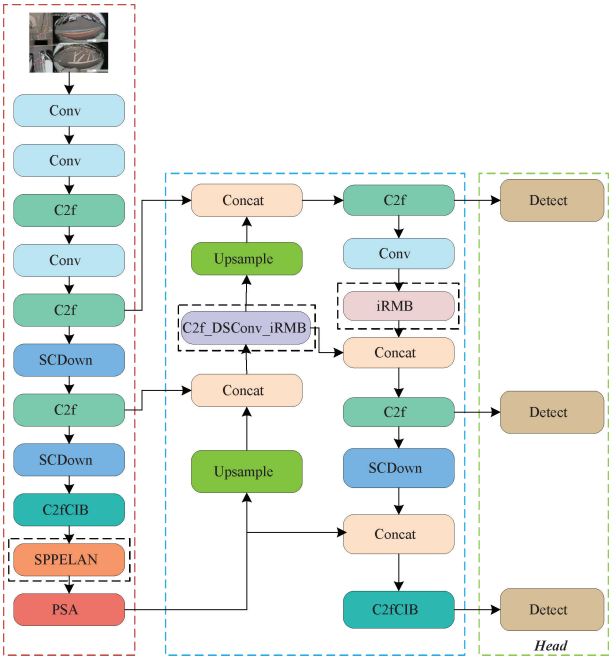


图 13 模型 3 结构
Fig. 13 Structure of model 3

cones(95.99%)、limit_lane(81.72%)、person(87.48%)和 lock(72.10%)五个类别上取得最优 AP 值,针对 limit_lane 检测任务上,模型 3 相较基线模型提升 1.87%,表明 iRMB 模块提升了对细长目标检测性能。

在精度-召回平衡性方面,模型 2 的召回率提升 2.02%(85.77%→87.79%)但精度下降 0.57%,而模型 3 通过增加模块数量使召回率保持 87.48%的同时精度回升 0.16%,说明双重 iRMB 模块可能通过层级特征筛选机制改善了检测稳定性。

尽管 iRMB 模块的引入导致模型复杂度呈可控增长(模型 3 相较基线模型参数量增加 8.7%,推理速度仅下降 5.4%(185→175 FPS),计算复杂度提升 16.3%),并伴随微幅的精度下降(模型 2 和模型 3 的 Precision 分别降低 0.57%/0.41%),但其在 mAP@0.5:0.95 指标上取得的显著提升(+1.71%)验证了该模块在特征表征增强方面的有效性。

3.7 对比实验

为验证本文提出的改进算法在泊车目标检测中的优越性,实验在泊车数据集上将该算法与 YOLOv5n、YOLOv8n、高斯 Wasserstein 距离(gaussian wasserstein

表 5 改进 YOLOv10n 模型变体在不同目标类别上的性能表现对比表
Table 5 Performance comparison of enhanced YOLOv10n variants across different object classes

模型	AP/%								mAP@0.5/	mAP@0.5:0.95/
	slot	car	cones	limit_lane	bicycle	person	rider	lock	%	%
模型 1	98.16	96.65	95.74	79.85	88.86	93.79	93.71	82.47	91.15	70.39
模型 2	98.10	96.69	95.93	81.36	88.26	94.07	92.95	85.34	91.59	71.35
模型 3	98.17	96.63	95.99	81.72	86.69	95.11	93.01	87.48	91.85	72.10

注:加粗为最优值。

表 6 改进 YOLOv10n 模型变体综合检测性能指标对比表
Table 6 Comparative analysis of comprehensive detection metrics for enhanced YOLOv10n variants

模型	Precision/%	Recall/%	Params/M	GFLOPs	mAP@0.5/%	mAP@0.5:0.95/%	FPS
模型 1	86.00	85.77	3.22	8.6	91.15	70.39	185
模型 2	85.43	87.79	3.24	9.1	91.59	71.35	177
模型 3	85.59	87.48	3.50	10.0	91.85	72.10	175

注:加粗为最优值。

distance, GWD)^[36]、YOLOv11n OBB 及 KL 散度(kullback-leibler divergence, KLD)^[37]算法模型进行对比,实验结果如表 7 所示。

由表 7 可知,改进后的算法相较于其他算法,对 slot、car、bicycle、person 和 rider 的检测取得最佳成绩,分别为 98.28%、96.44%、84.32%、92.11%和 90%,平均精度达到了 89.89%,高于其他算法。尽管改进后的算法在计算量和参数量上略有增加,分别达到 3.24 M 和 9.1 GFLOPs,但在泊车检测任务中仍保持在可接受范围内。

综上所述,本文改进的 YOLOv10n OBB 模型,在泊车任务中具有一定的实际应用价值。

3.8 泊车检测结果可视化

为了更直观评价改进 YOLOv10n OBB 模型的性能,本文将 YOLOv10n OBB 基线模型与改进的模型在测试集上进行检测效果对比,对比结果如图 14 所示。可以看出,原始模型针对小目标 cones 存在误检和漏检的情况,在复杂黑夜场景下,未能识别出 limit_lane。相比之下,改进后的模型无论是对小目标的检测亦或是复杂场景下的检测

表 7 对比实验结果

Table 7 Comparative experimental results

算法模型	AP/%								mAP@0.5/	Params/	GFLOPs
	slot	car	cones	limit_lane	bicycle	person	rider	lock	%	M	
YOLOv5n OBB	98.08	96.13	95.65	74.71	81.31	89.62	85.82	79.69	87.63	2.58	7.3
YOLOv8n OBB	98.17	96.29	95.68	76.74	81.95	90.82	88.33	86.95	89.37	3.08	8.3
GWD	90.15	88.03	84.59	55.76	64.07	76.66	74.76	56.82	73.85	36.27	43.27
KLD	90.07	88.02	82.14	53.44	65.04	77.24	77.06	59.56	74.07	36.27	43.27
YOLOv9t OBB	98.08	96.28	95.34	76.62	82.71	90.17	87.53	83.00	88.72	2.02	7.8
YOLOv10n OBB	98.03	96.34	95.70	77.36	82.62	90.89	88.42	84.82	89.27	2.66	8.0
YOLOv11n OBB	98.17	96.25	95.63	77.01	79.46	89.46	88.27	83.86	88.51	2.66	6.6
本文	98.28	96.44	95.52	77.01	84.34	92.11	90.00	85.4	89.89	3.24	9.1

注:加粗为最优值。

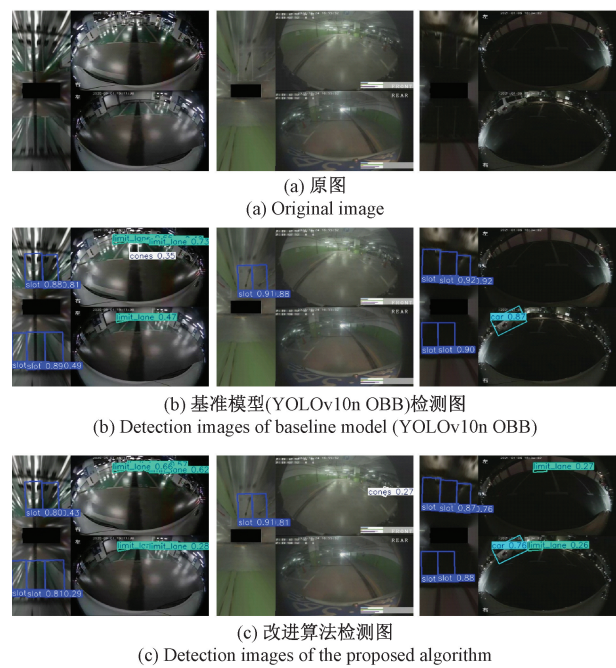


图 14 泊车目标检测结果可视化

Fig. 14 Visualization of parking target detection results

都优于原始模型。

4 结 论

本文针对泊车场景中的复杂背景干扰,小目标定位困难,鱼镜头头下检测目标径向畸变等难题,提出了改进的YOLOv10n 旋转目标检测方法。在主干网络中引入SPPELAN 模块以增强小目标的特征提取能力;在颈部网络中,采用 DSConv 替代 C2f 中的普通卷积,提升了对鱼镜头头下细长目标的检测效果,并进一步集成 iRMB 模块,以增强模型在复杂场景下的鲁棒性;采用 ATFL 损失函数,缓解了目标与背景不平衡。实验结果表明,与其他旋转目标检测模型相比,改进的 YOLOv10n 旋转目标检测方法在检测精度上表现更优,但在参数量和计算量方面仍高

于其他 YOLO 模型。未来研究将聚焦于进一步优化模型的轻量化设计,以更好地满足泊车场景中的实时应用需求。

参考文献

[1] 田杰,叶青. 自动泊车发展现状及运动规划研究进展[J]. 科学技术与工程,2024,24(21):8825-8836.
TIAN J, YE Q. Review of automatic parking and research progress of motion planning [J]. Science Technology and Engineering, 2024, 24 (21): 8825-8836.

[2] 成浪,敬超. 基于改进 YOLOv7 的 X 线图像旋转目标检测[J]. 图学学报, 2023, 44(2): 324-334.
CHENG L, JING CH. X-ray image rotating object detection based on improved YOLOv7[J]. Journal of Graphics, 2023, 44(2): 324-334.

[3] 王玉龙,翁茂楠,黄辉,等. 基于 CNN-Transformer 的自动泊车车位感知算法[J]. 汽车技术, 2024(8): 1-6.
WANG Y L, WENG M N, HUANG H, et al. Algorithm for parking space detection in automatic parking system based on CNN-Transformer [J]. Automobile Technology, 2024(8): 1-6.

[4] KUMAR V R, EISING C, WITT C, et al. Surround-view fisheye camera perception for automated driving: overview, survey & challenges [J]. IEEE Transactions on Intelligent Transportation Systems, 2023, 24(4): 3638-3659.

[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, 2017, 39(6): 1137-1149.

[6] HE K M, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]. 2017 IEEE International Conference on

- Computer Vision(ICCV), IEEE, 2017: 2980-2988.
- [7] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), IEEE, 2016: 779-788.
- [8] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger [C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2017: 6517-6525.
- [9] REDMON J, FARHADI A. YOLOv3: An incremental improvement[J]. ArXiv preprint arXiv: 1804.02767, 2018.
- [10] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]. Computer Vision-ECCV 2016, Springer, 2016: 21-37.
- [11] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 318-327.
- [12] GRBIC R, KOCH B. Automatic vision-based parking slot detection and occupancy classification[J]. Expert Systems with Applications, 2023, 225: 120147.
- [13] SUHR J K, JUNG H G. End-to-end trainable one-stage parking slot detection integrating global and local information [J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(5): 4570-4582.
- [14] YU ZH P, GAO ZH, CHEN H S, et al. SPFCN: Select and prune the fully convolutional networks for real-time parking slot detection [C]. 2020 IEEE Intelligent Vehicles Symposium (IV), IEEE, 2020: 445-450.
- [15] 董宝鑫, 王江涛. 基于重参数化广义金字塔与扩张残差的遥感图像旋转框算法[J]. 电子测量与仪器学报, 2024, 38(12): 54-61.
- DONG B X, WANG J T. Remote sensing image rotation box algorithm leveraging reparameterized generalized pyramid and dilated residual[J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(12): 54-61.
- [16] 邝先验, 程福军, 吴翠琴. 基于改进 YOLOv7-tiny 的高效轻量遥感图像目标检测方法[J]. 电子测量与仪器学报, 2024, 38(7): 22-33.
- KUANG X Y, CHENG F J, WU C Q. Efficient and lightweight target detection method for remote sensing images based on improved YOLOv7-tiny[J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(7): 22-33.
- [17] 童小钟, 魏俊宇, 苏绍璟, 等. 融合注意力和多尺度特征的典型水面小目标检测[J]. 仪器仪表学报, 2023, 44(1): 212-222.
- TONG X ZH, WEI J Y, SU SH J, et al. Typical small target detection on water surfaces fusing attention and multi-scale features[J]. Chinese Journal of Scientific Instrument, 2023, 44(1): 212-222.
- [18] XU ZH Y, TANG X, MA CH X, et al. Research on parking space detection and prediction model based on CNN-LSTM [J]. IEEE Access, 2024, 12: 30085-30100.
- [19] 宋绍京, 陆婷婷, 孙翔, 等. 面向自动驾驶的多任务环境感知算法[J]. 电子测量技术, 2023, 46(24): 157-163.
- SONG SH J, LU T T, SUN X, et al. Multi-task environment perception algorithm for autonomous driving [J]. Electronic Measurement Technology, 2023, 46(24): 157-163.
- [20] 闫旭东, 钱莉. 基于全景环视图像的停车位检测算法[J]. 华中师范大学学报(自然科学版), 2024, 58(5): 526-532.
- YAN X D, QIAN L. Parking slot detection based on surround-view image [J]. Journal of Central China Normal University(Natural Sciences), 2024, 58(5): 526-532.
- [21] 杨飞帆, 李军, 王耀弘. 改进 YOLOv8 的环视车位检测算法研究[J]. 重庆理工大学学报(自然科学), 2024, 38(9): 22-29.
- YANG F F, LI J, WANG Y H. Research on the improved YOLOv8 algorithm for the detection of surrounding parking spaces[J]. Journal of Chongqing University of Technology (Natural Science), 2024, 38(9): 22-29.
- [22] QI Y L, HE Y T, QI X M, et al. Dynamic snake convolution based on topological geometric constraints for tubular structure segmentation[C]. 2023 IEEE/CVF International Conference on Computer Vision, IEEE, 2023: 6047-6056.
- [23] ZHANG J N, LI X T, LI J, et al. Rethinking mobile block for efficient attention-based models[C]. 2023 IEEE/CVF International Conference on Computer Vision(ICCV), IEEE, 2023: 1389-1400.
- [24] YANG B, ZHANG X Y, ZHANG J, et al. EFLNet: Enhancing feature learning network for infrared small target detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62: 1-11.
- [25] 王舒梦, 徐慧英, 朱信忠, 等. 基于改进 YOLOv8n 航拍轻量化小目标检测算法: PECS-YOLO[J/OL]. 计算机

- 工程:1-16[2025-03-28]. <https://doi.org/10.19678/j.issn.1000-3428.0069353>.
- WANG SH M, XU H Y, ZHU X ZH, et al. Lightweight small object detection algorithm based on improved YOLOv8n aerial photography: PECS-YOLO [J/OL]. Computer Engineering; 1-16 [2025-03-28]. <https://doi.org/10.19678/j.issn.1000-3428.0069353>.
- [26] WANG A, CHEN H, LIU L H, et al. YOLOv10: Real-time end-to-end object detection [J]. ArXiv preprint arXiv:2405.14458, 2024.
- [27] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), IEEE, 2017; 2117-2125.
- [28] LIU SH, QI L, QIN H F, et al. Path aggregation network for instance segmentation[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE, 2018; 8759-8768.
- [29] LLERENA J M, ZENI L F, KRISTEN L N, et al. Gaussian bounding boxes and probabilistic intersection-over-union for object detection[J]. ArXiv preprint arXiv:2106.06072, 2021.
- [30] WOO S H, PARK J C, LEE J Y, et al. CBAM: Convolutional block attention module[C]. European Conference on Computer Vision(ECCV), Springer, 2018; 3-19.
- [31] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]. 2018 IEEE/CVF International Conference on Computer Vision and Pattern Recognition, IEEE, 2018; 7132-7141.
- [32] HOU Q B, ZHOU D Q, FENG J S. Coordinate attention for efficient mobile network design[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), IEEE, 2021; 13708-13717.
- [33] AZAD R, NIGGEMEIER L, HÜTTEMANN M, et al. Beyond self-attention: deformable large kernel attention for medical image segmentation[C]. 2024 IEEE/CVF Winter Conference on Applications of Computer Vision(WACV), IEEE, 2024; 1276-1286.
- [34] CAI X H, LAI Q X, WANG Y W, et al. Poly kernel inception network for remote sensing detection[C]. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2024; 27706-27716.
- [35] OUYANG D L, HE S, ZHANG G ZH, et al. Efficient multi-scale attention module with cross-spatial learning [C]. ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing(ICASSP), IEEE, 2023; 1-5.
- [36] YANG X, YAN J CH, MING Q, et al. Rethinking rotated object detection with gaussian wasserstein distance loss [C]. 38th International Conference on Machine Learning (PMLR), ACM, 2021; 11830-11841.
- [37] YANG X, YANG X J, YANG J R, et al. Learning high-precision bounding box for rotated object detection via kullback-leibler divergence [J]. ArXiv preprint arXiv:2106.01883, 2021.

作者简介

梁列全, 博士, 教授, 硕士生导师, 主要研究方向为机器学习、数据挖掘。

E-mail: lianglq@gdufe.edu.cn

李想, 硕士研究生, 主要研究方向为目标检测、自动泊车。

E-mail: lixiang@student.gdufe.edu.cn

何永华, 硕士研究生, 主要研究方向为目标检测、自动泊车。

E-mail: heyh2023@163.com

周璇(通信作者), 博士, 教授, 博士生导师, 主要研究方向为机器学习、模式挖掘、建筑节能。

E-mail: zhouxuan@scut.edu.cn