

# 融合视觉与几何推理的地下车位全场景检测算法

刘平<sup>1\*</sup>, 马春蕾<sup>1</sup>, 刘明杰<sup>1</sup>, 朴昌浩<sup>1</sup>, 黄龙航<sup>2</sup>, 颜鲁鹏<sup>2</sup>

(1. 重庆邮电大学自动化学院, 重庆 400065; 2. 重庆市建设信息中心, 重庆 400010)

**摘要:** 在智能停车管理系统中, 复杂环境下高精度车位感知是实现自动化引导与资源优化的核心技术。传统检测方法成本高、维护困难且车位状态信息单一, 而基于视觉的方法虽能提供丰富信息, 但在地下停车场面临光照昏暗、强光源干扰、背景复杂、车辆遮挡严重等挑战。为此, 本文提出一种融合视觉与几何推理的轻量化地下停车场车位实时检测网络 (Real-Time Underground Parking Space Occupancy DEtection TRansformer, RT-UPSODETR)。首先, 设计了膨胀空间注意力轻量级残差 F-DBlock (Fusion DBlock) 模块, 在扩大感受野捕获全局模糊特征的同时显著减少冗余计算。其次, 构建了多机制融合的 Transformer 编码层 CSAL (Cognitive Spatial Attention Layer) 模块, 替换 RT-DETR-R18 中原有 AIFI 模块, 显著提升了模型在低光照、模糊图像中的特征判别能力。进一步, 针对车辆相互遮挡导致的后排车位漏检问题, 设计了几何推理透视拓扑补全模块 (Perspective Topology Completion Module, PTCM), 利用前排已检测车位作为几何锚点估计透视灭点, 基于径向投影推断被遮挡车位的近似位置, 有效恢复缺失车位框。为验证算法有效性, 本文构建了地下停车场车位检测数据集以及遮挡补全专项测试集, 并在公开低光照数据集 ExDark 上开展泛化实验。实验结果表明, RT-UPSODETR 在自建数据集上相较于基线 RT-DETR-R18 模型, 在精度、召回率、mAP@0.5 和 mAP@0.5-0.95 指标上分别提升 1.5%、3.3%、0.7% 和 0.9%; 同时参数量降低 22.1%, GFLOPs 降低 18.8%, 推理速度为 52.4 FPS, 满足实时检测需求。在 ExDark 数据集泛化实验中, 本文模型较 RT-DETR-R18 在精确率、召回率、mAP@0.5 和 mAP@0.5-0.95 上分别提升 2.1%、1.3%、2.1% 和 1.0%, 验证了模型在低光照条件下的鲁棒性。车位补全方面, PTCM 模块在遮挡补全实验中在漏检召回率、补全准确率和平均中心点误差指标上达到了 82.48%、74.83% 和 23.27 pixels, 证明其能有效降低漏检率。本文提出的 RT-UPSODETR 网络通过轻量化设计与多机制注意力融合, 有效解决了地下停车场低光照、复杂背景和遮挡场景下的车位检测难题, 透视拓扑补全模块进一步增强了系统对遮挡的鲁棒性, 为实际智能停车系统提供了可行的视觉解决方案。

**关键词:** RT-DETR 网络; 视觉融合; 几何推理; 车位检测; CSAL 模块; F-DBlock 模块; 遮挡补全

**基金项目:** 国家重点研发计划 (No.2022YFE0101000); 重庆市建设信息中心项目 (No.JSXXCQ-CY01)

**中图分类号:** TP391.41; U495

**文献标识码:** A

**文章编号:** 0372-2112(2026)04-1723-13

**电子学报 URL:** <http://www.ejournal.org.cn>

**DOI:** 10.12263/DZXB.20260063

## A Full-Scene Detection Algorithm for Underground Parking Spaces Fusing Vision and Geometric Reasoning

LIU Ping<sup>1\*</sup>, MA Chunlei<sup>1</sup>, LIU Mingjie<sup>1</sup>, PIAO Changhao<sup>1</sup>, HUANG Longhang<sup>2</sup>, YAN Lupeng<sup>2</sup>

(1. College of Automation, Chongqing University of Posts and Telecommunications, Chongqing 400065, China;

2. Chongqing Construction Information Center, Chongqing 400010, China)

**Abstract:** In intelligent parking management systems, achieving high-precision parking space perception in complex environments is a core technology for automated guidance and resource optimization. Traditional detection methods suffer from high costs, difficult maintenance, and limited parking status information. Although vision-based approaches can provide rich information, they face challenges in underground parking lots, such as dim illumination, strong light interference, complex backgrounds, and severe vehicle occlusion. To address these issues, this paper proposes a lightweight real-time underground parking space occupancy detection network integrating vision and geometric reasoning, termed real-time underground parking space occupancy detection transformer (RT-UPSODETR). First, a lightweight residual F-DBlock (Fusion DBlock) module with dilated spatial attention is designed, which expands the receptive field to capture global blurred features while significantly reducing redundant computation. Second, a multi-mechanism Transformer encoder layer, cognitive spatial attention layer (CSAL), is constructed to replace the original AIFI module in RT-DETR-R18, significantly enhancing the model's feature discrimination ability in low-light and blurred images. Furthermore, to address the missed detection of rear parking spaces caused by mutual vehicle occlusion, a geometric reasoning perspective topology completion module

(PTCM) is devised, which uses the detected front parking spaces as geometric anchors to estimate the vanishing point and then infers the approximate positions of occluded spaces via radial projection, effectively recovering missing bounding boxes. To validate the effectiveness of the proposed algorithm, we construct an underground parking space detection dataset and a dedicated occlusion completion test set, and conduct generalization experiments on the public low-light dataset ExDark. Experimental results on the self-built dataset show that compared with the baseline RT-DETR-R18, RT-UPSO-DETR improves precision, recall, mAP@0.5, and mAP@0.5-0.95 by 1.5%, 3.3%, 0.7%, and 0.9%, respectively. Meanwhile, the number of parameters is reduced by 22.1%, GFLOPs by 18.8%, and the inference speed reaches 52.4 FPS, satisfying real-time detection requirements. On the ExDark dataset, compared with RT-DETR-R18, the proposed model improves precision, recall, mAP@0.5, and mAP@0.5-0.95 by 2.1%, 1.3%, 2.1%, and 1.0%, respectively, verifying its robustness under low-light conditions. For parking space completion, the PTCM achieves a missed recall rate of 82.48%, a completion accuracy of 74.83%, and a mean center error of 23.27 pixels, demonstrating its effectiveness in reducing missed detections. The proposed RT-UPSO-DETR network, through lightweight design and multi-mechanism attention fusion, effectively addresses the challenges of parking space detection in underground parking lots with low illumination, complex backgrounds, and occlusions. The perspective topology completion module further enhances the system's robustness to occlusions, providing a feasible vision-based solution for practical intelligent parking systems.

**Keywords:** RT-DETR network; fusing vision; geometric reasoning; parking space detection; CSAL module; F-DBlock module; occlusion completion

**Foundation Item(s):** National Key Research and Development program (No.2022YFE0101000); Chongqing Construction Information Center Project (No.JSXXCQ-CY01)

## 0 引言

近年来,随着城市化进程的加速与机动车保有量的激增,停车难已成为制约智慧城市发展的瓶颈问题。智能停车管理系统(Intelligent Parking Management System, IPMS)作为解决该问题的关键技术,取得了长足进步。一个高效的IPMS不仅需要实时、准确地判断车位的占用状态,更依赖于对车位空间位置的精确感知。通过将目标检测结果与摄像机标定参数结合,系统可以将图像坐标映射为物理空间的经纬度坐标,从而实现车位引导与反向寻车功能。因此,如何在复杂环境下实现高精度的车位感知,成为计算机视觉与智能交通领域的研究热点。

传统的实时车位状态检测设备包括激光雷达<sup>[1-2]</sup>、超声波<sup>[3-4]</sup>、磁感应传感器<sup>[5-6]</sup>、红外传感器<sup>[7-8]</sup>等,但是这些传统设备仍存在着一定的不足,比如基于激光雷达和超声波传感器容易造成行人被误判,红外传感器则容易被环境温度影响,基于磁感应传感器的方法则需要将传感器埋入地下,导致设备维护困难。因此,亟需一种实现简单,安装成本和改装成本低的方法。近年来,随着深度学习的快速发展,基于卷积神经网络(Convolutional Neural Network, CNN)的目标检测算法在车位检测得到了重视。当前主流的快速车位目标检测算法可以分为三个类别:以Faster RCNN<sup>[9]</sup>为代表的双阶段检测算法,以YOLO(You Only Look Once)<sup>[10]</sup>系列模型为代表的单阶段检测算法和以DETR(DEtection TRansformer)系列模型<sup>[11]</sup>为代表的端到端检测算法。端到端检测算法将目标检

测的多个子任务整合为单一网络流程,省略人工设计的中间步骤,在精度、效率和工程落地性上均有突出表现,因此成为研究的主流。但是由于transformer计算量较大,因此导致了其训练资源要求较高,训练时间长。为了解决这个问题,Zhao等人<sup>[12]</sup>提出了实时目标检测算法(Real-Time DEtection TRansformer, RT-DETR),在DETR模型的基础上设计了高效混合编码器(efficient hybrid encoder)和IoU感知查询选择(IoU-aware query selection),有效减少了DETR模型的冗余计算,在保持高精度基础同时提高了模型推理速度。

尽管RT-DETR模型在通用数据集上取得了显著成效,但在将其迁移至地下停车场等特定场景时,仍然存在诸多问题。首先,地下环境普遍存在光照昏暗、强光源干扰及背景复杂等问题,导致提取的特征信噪比低;其次,监控视角下车位目标尺度变化剧烈,RT-DETR模型在处理小目标或多尺度目标时往往力不从心。为解决此类问题,文献[13]在RT-DETR的基础上使用重参数化网络RepViT作为模型的主干网络并且设计了基于通道注意力的双向特征融合模块(Channel Attention Fusion Module, CAFM)块使得模型能关注到长距离的上下文信息,提升了模型的识别精度;文献[14]通过引入多阶门控聚合模块、卷积加性标记混合器和CoreProximity-IoU损失函数,有效增强了复杂场景下小目标的特征表达与抗背景干扰能力,同时降低了模型参数量和计算复杂度。然而,地下停车场受限于层高和建筑结构,监控摄像头通常安装位置较低且俯仰角较小,这种特殊的采集视

角带来了严重的单目透视效应(perspective effect),远离摄像头的后停车位极易被前排车辆遮挡从而导致目标漏检,即使检测模型具有极强的特征提取能力,也无法从像素层面恢复被完全遮挡的目标,导致系统无法生成完整的车位分布地图,严重影响智慧停车场的感知与管理。

文献[13-14]在一定程度上增强了 RT-DETR 在复杂场景下的性能,但在应用于地下停车场环境时仍存在一定局限性。具体而言,文献[13]通过 CAFM 模块建模长距离上下文信息提升了识别精度,但并未针对低光照和强光源干扰造成的特征退化问题进行处理;文献[14]通过多阶门控聚合和 CoreProximity-IoU 损失降低了模型复杂度并增强了小目标特征表达,但其设计并未明确针对昏暗或光照不均区域的特征清晰度提升。为克服上述局限性,本文方法在两方面进行了改进:首先,在 Feijoo 等人<sup>[15]</sup>提出的 DBlock 模块基础上进行 F-DBlock 模块设计,以此减少模型计算复杂度与参数量;同时,融合多认知适配器(Multi-cognitive visual adapters, Mona)<sup>[16]</sup>、极性感知线性注意力机制(Polarity-aware Linear Attention, POLA)<sup>[17]</sup>和空间增强前馈网络(Spatially-Enhanced Feedforward Network, SEFN)<sup>[18]</sup>开发一种多机制 Transformer EncoderLayer 模块 CSAL,用以增强模型对模糊特征与低光照区域的识别能力;此外,针对透视遮挡问题,设计一种透视拓扑补全模块(Perspective Topology Completion Module, PTCM),从几何层面出发建立透视场和深度比例一致性原理反向推断并恢复被遮挡的后停车位,通过将深度学习的感知能力与几何学的推理能力相结合,实现对地下停车场车位的全量感知;最后,本文构建了包含多种类别和场景的地下停车场车位检测数据集,并制作了遮挡补全专用测试集,在多种评价指标下对所提出模型进行测试以验证其性能。

## 1 改进 RT-UPSO-DETR 网络

为了应对地下停车场车位检测复杂环境,本文以 RT-DETR-R18 为基准模型,提出一种名为 RT-UPSO-DETR 的网络模型,其结构如图 1 所示。该模型骨干网络采用 ResNet-18<sup>[19]</sup>提取多尺度特征,通过简化卷积层,移除归一化和激活函数,减少计算冗余;同时,改进设计 F-DBlock 模块,通过引入多尺度扩张卷积来同时捕获局部和全局信息,增强车位检测的识别能力,为后续模块提供清晰和丰富的特征信息并且减少模型参数量;进一步,融入 CSAL 模块,通过多认知适配器、极性感知线性注意力机制和空间增强前馈网络,捕获多尺度细节信息并增强空间感知能力;最后,Transformer 解码器通过 IoU 感知查询选择机制筛

选 300 个高质量初始查询,并利用交叉注意力迭代优化,输出预测结果。

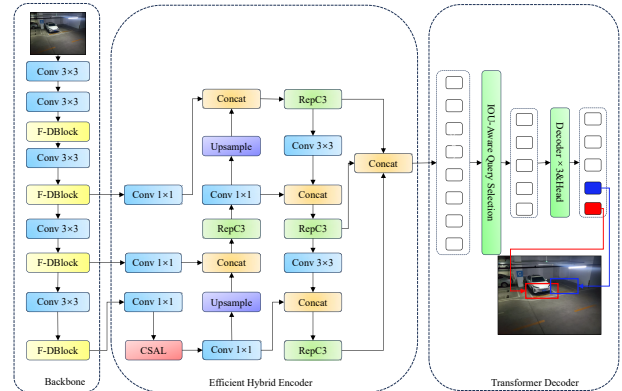


图 1 RT-UPSO-DETR 网络结构示意图

Figure 1 Schematic diagram of the RT-UPSO-DETR network structure

### 1.1 改进 F-DBlock 模块

本文研究的原始结构图为  $\mathbb{R}^{BS \times c \times h \times w}$ ,其中 BS 表示纬度(batch size),由于后续计算中 BS 不存在纬度变换,为简洁起见,正文公式省略 BS,而在伪代码中保留 BS 以表示实际输入张量形状。

#### 1.1.1 F-DBlock 模块结构

F-DBlock 模块用于解决复杂地下停车场目标检测边界模糊的问题,其结构图如图 2 所示。该结构主要基于 CSPNet 模块<sup>[20]</sup>和 DBlock 模块<sup>[15]</sup>改进,核心思想是通过特征图的拆分和部分跨阶段融合来减少计算量。

首先 F-DBlock 原始特征图通过一个  $1 \times 1$  的卷积进行特征重组(如图 2(a)所示),然后特征图拆分(split)将空间信息进行拆分,这个过程定义如下:

$$\begin{cases} \mathbf{Y} = \text{conv}1 \times 1(\mathbf{x}), \mathbf{Y} \in \mathbb{R}^{2c \times h \times w} \\ \mathbf{c} = C_2 \cdot e \\ \mathbf{Y}_0, \mathbf{Y}_1 = \text{split}(\mathbf{Y}), \mathbf{Y}_0, \mathbf{Y}_1 \in \mathbb{R}^{c \times h \times w} \end{cases} \quad (1)$$

其中,  $C_1$  为输入通道数;  $C_2$  为目标输出通道数;  $e$  为扩展因子(本文取  $e=0.5$ );  $c$  为隐藏通道数。扩展因子  $e$  决定了通过 DBlock 处理的通道比例,当  $e=0.5$  时,拆分后的两部分通道数相等,  $\mathbf{Y}_0$  保留原始特征信息不进行处理;  $\mathbf{Y}_1$  输入到由  $k$  个 DBlock 级联而成的通路中进行处理,产生  $k$  个特征图  $\{\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_k\}$ 。随后,将原始特征  $\mathbf{Y}_0$ 、初始特征  $\mathbf{Y}_1$  以及 DBlock 处理输出的特征  $\{\mathbf{Z}_k\}$  进行拼接,得到特征维度为  $(k+2) \cdot c$  的融合特征张量。最后通过  $1 \times 1$  卷积将拼接后的高维特征  $(k+2) \cdot c$  投影到目标维度  $C_2$ ,实现跨通道特征融合与维度适配,处理过程定义为

$$\begin{cases} \mathbf{Z}_k = \text{DBlock}^{(k)}(\mathbf{Z}_{k-1}), k = 1, 2, \dots, \mathbf{Z}_k \in \mathbb{R}^{c \times h \times w} \\ \mathbf{F}_{\text{concat}} = \{\mathbf{Y}_0, \mathbf{Y}_1, \mathbf{Z}_1, \dots, \mathbf{Z}_k\}, \mathbf{F}_{\text{concat}} \in \mathbb{R}^{(k+2)c \times h \times w} \\ \mathbf{X}_{\text{out}} = \text{conv} 1 \times 1(\mathbf{F}_{\text{concat}}), \mathbf{X}_{\text{out}} \in \mathbb{R}^{C_2 \times h \times w} \end{cases} \quad (2)$$

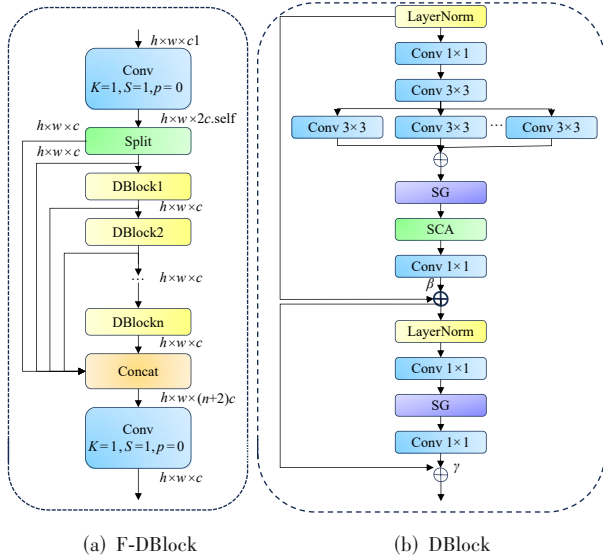


图2 F-DBlock 模块网络结构

Figure 2 Network structure of F-DBlock module

F-DBlock 模块由于只对一半的通道进行了复杂处理,因而减少了计算量,提高了模型的计算效率。DBlock 的结构图如图 2(b)所示,其过程为首先对于输入的初始特征经过层归一化处理,随后通过一个深度可分离卷积进行局部空间特征预精炼,之后通过  $1 \times 1$  卷积进行通道扩展,产生中间特征  $\mathbf{Z}$ ;进一步,特征通过多个并行的空洞卷积分支进行多尺度上下文提取,各分支输出求和融合后应用简单门控机制(Simple Gate, SG)实现特征自适应筛选;筛选后的特征经过简化通道注意力模块(Simplified Channel Attention, SCA)生成空间权重图,通过逐通道缩放实现特征精炼,再经  $1 \times 1$  卷积降维操作后与原始输入进行带可学习参数  $\beta$  的加权残差连接;随后,通过简单门控进行二次增强:归一化后的特征通过  $1 \times 1$  卷积变换,经过门控筛选后压缩回原通道维度,最终与中间特征进行带可学习参数  $\gamma$  的残差融合输出与输入

同维度的增强特征  $\mathbf{Y}$ ,多尺度扩张卷积和特征融合的过程为

$$\mathbf{M} = \sum_i^3 \text{conv}_{3 \times 3}^{(d_i=i)}(\mathbf{Z}), \mathbf{M} \in \mathbb{R}^{C \times h \times w} \quad (3)$$

其中,  $\mathbf{Z}$  为输入特征;  $d_i$  表示第  $i$  个空洞卷积分支的膨胀率;  $\mathbf{M}$  表示多尺度卷积分支融合后的特征。该结构通过并行空洞卷积分支提取不同感受野下的空间上下文信息,并在后续处理中完成特征筛选与残差增强。

### 1.1.2 F-DBlock 模块算法实现

结合上述分析,本文提出的 F-DBlock 模块算法实现过程伪代码如算法 1 所示。

算法 1 F-DBlock 模块算法

输入: SRC[BS,  $C_1$ ,  $h$ ,  $w$ ]

输出: 输出特征图[BS,  $C_2$ ,  $h$ ,  $w$ ]

参数:  $C_2$ ,  $n$ ,  $e$

1. 计算隐藏通道:  $c = C_2 \times e$
2. 初步卷积:  $Y = \text{Conv} 1 \times 1(x) \leftarrow [\text{BS}, 2c, h, w]$
3. 特征分割:  $Y_0, Y_1 = \text{Split}(Y, \text{dim} = \text{channel}) \leftarrow [\text{BS}, c, h, w]$
4. 主干特征列表初始化:  $\text{feat\_list} \leftarrow [Y_0, Y_1]$
5. 循环处理分支特征: for  $i = 1$  to  $k$  do
6. 分支处理  $y_2 = \text{DBlock}_k(\mathbf{Z}_{k-1}) \leftarrow [\text{BS}, c, h, w]$
7. 特征追加:  $\text{feat\_list.append}(\mathbf{Z}_k)$
8. 结束循环
9. 特征拼接:  $y = \text{Concat}(\text{features}, \text{dim} = \text{channel}) \leftarrow [\text{BS}, (k+2) \times C, h, w]$
10. 最终卷积:  $\text{output} = \text{Conv} 1 \times 1(y_{\text{concat}}) \leftarrow [\text{BS}, C_2, h, w]$

### 1.2 基于改进 SEFN 的 CSAL 模块

CSAL 模块用于解决复杂地下停车场环境下模型空间感知能力低的问题,该模块结构如图 3 所示。其运行过程为:首先,保存输入特征  $\mathbf{X}_0$  供后续处理;然后,引入 POLA 模块<sup>[17]</sup>,将特征图执行选择性空间压缩的线性注意力,恢复空间维度后与原始特征图进行残差连接;进一步,对归一化结果进行 Mona<sup>[16]</sup>空间增强;随后,将增强特征与初始特征输入 SEFN 模块<sup>[18]</sup>,执行结合通道注意力机制的特征变换,将结果与当前特征进行残差连接;最后,通过层归一化和 Mona 空间增强输出最终特征图。

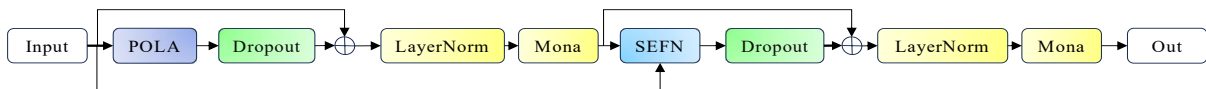


图3 CSAL 模块网络结构

Figure 3 Network structure of CSAL module

#### 1.2.1 基于 POLA 和 Mona 模块的图像预处理

考虑到对地下停车场环境进行图像识别时,立柱、管道等背景元素易被误检为车辆或行人,导致假

阳性,而 Pola 模块能够通过区分强/弱信号,减少对背景的均匀激活并通过  $G_s/G_o$  矩阵动态平衡交互权重,适应不同光照场景来提升检测效果泛化性,因此本文

采用此模块进行图像处理。整个过程分为三个部分,分别是线性拆分、极性分解和线性注意力,具体过程如下。

(1)线性拆分:将输入的特征图  $\mathbf{X} \in \mathbb{R}^{B \times N \times C}$  拆分为  $\mathbf{k} \in \mathbb{R}^{B \times n \times C}$ 、 $\mathbf{v} \in \mathbb{R}^{B \times n \times C}$ 、 $\mathbf{q} \in \mathbb{R}^{B \times n \times C}$  以便后续处理,其中  $B$ 、 $N$ 、 $n$ 、 $C$  分别为 batch size、输入的空间位置数、采样后的空间位置数、通道维数。

(2)极性分解:将特征图  $\mathbf{v}$  拆分为  $\mathbf{v}_1$  和  $\mathbf{v}_2$ ,并将  $\mathbf{k}$  和  $\mathbf{q}$  进行位置编码增强后通过位置缩放和 ReLU 激活后与超参数进行幂运算得到正极性和负极性特征  $\mathbf{k}^+$ 、 $\mathbf{k}^-$ 、 $\mathbf{q}^+$  和  $\mathbf{q}^-$ 。

(3)线性注意力:通过构建相似性流和对立性流后极性线性注意力计算,最后进行合并流,通过重还原形状。其中相似性流和对立性流分别用式(4)和式(5)进行计算。

$$\begin{cases} \mathbf{z}_{\text{sim}} = \left( [\mathbf{q}^+, \mathbf{q}^-] \cdot \frac{1}{N} \sum_{i=1}^N ([\mathbf{k}^+, \mathbf{k}^-])_i^T + \varepsilon \right)^{-1} \\ \mathbf{KV}_{\text{sim}} = \frac{1}{n} ([\mathbf{k}^+, \mathbf{k}^-]^T) \mathbf{v}_1 \\ \mathbf{X}_{\text{sim}} = \mathbf{z}_{\text{sim}} \odot ([\mathbf{q}^+, \mathbf{q}^-] \cdot \mathbf{KV}_{\text{sim}}) \end{cases} \quad (4)$$

$$\begin{cases} \mathbf{z}_{\text{opp}} = \left( [\mathbf{q}^-, \mathbf{q}^+] \cdot \frac{1}{N} \sum_{i=1}^N ([\mathbf{k}^+, \mathbf{k}^-])_i^T + \varepsilon \right)^{-1} \\ \mathbf{KV}_{\text{opp}} = \frac{1}{n} ([\mathbf{k}^+, \mathbf{k}^-]^T) \mathbf{v}_2 \\ \mathbf{X}_{\text{opp}} = \mathbf{z}_{\text{opp}} \odot ([\mathbf{q}^-, \mathbf{q}^+] \cdot \mathbf{KV}_{\text{opp}}) \end{cases} \quad (5)$$

其中,  $1/n$  为缩放因子;  $[\mathbf{q}^+, \mathbf{q}^-]$  和  $[\mathbf{k}^+, \mathbf{k}^-]$  表示将两组对应特征图  $\mathbf{q}^+$  和  $\mathbf{q}^-$ 、 $\mathbf{k}^+$  和  $\mathbf{k}^-$  进行拼接;  $\varepsilon$  为数值稳定项;相似性流  $\mathbf{X}_{\text{sim}}$  与对立性流  $\mathbf{X}_{\text{opp}}$  分别反映了输入特征在同向响应和反向响应下的线性注意力结果。通过式(6)进行分支合并和维度还原:

$$\mathbf{X}_{\text{out}} = \text{reshape}(\text{concat}(\mathbf{X}_{\text{sim}}, \mathbf{X}_{\text{opp}})), \mathbf{X}_{\text{out}} \in \mathbb{R}^{B \times N \times C} \quad (6)$$

最终,将相似性流和对立性流在特征维上进行拼接,并通过 reshape 恢复其序列结构,得到输出特征  $\mathbf{X}_{\text{out}}$ 。进一步,在 POLA 模块处理的基础上,针对地下停车场环境物体相互遮挡无法准确提取空间信息问题,采用 Mona 模块对不同尺度的空间特征进行融合。首先保留输入特征图  $\mathbf{x}$  (即  $\mathbf{X}_{\text{out}}$ ) 作为残差连接初始值,之后将输入的特征图输入到归一化层,将归一化层的输出与初始特征值进行残差连接,计算过程如下:

$$\begin{cases} \mathbf{x}' = \text{LayerNorm}(\mathbf{x}) \odot \alpha + \mathbf{x} \odot \delta \\ \mathbf{x}' \in \mathbb{R}^{C \times h \times w} \end{cases} \quad (7)$$

其中,  $\alpha$  和  $\delta$  为学习参数;  $\mathbf{x}'$  表示归一化和缩放后的特征。随后,通过一个  $1 \times 1$  的卷积将通道数压缩到 64,

将输出的特征图输入到多认知卷积组进行特征融合,输入的特征图分别进行一次卷积核为  $3 \times 3$ 、 $5 \times 5$ 、 $7 \times 7$  的深度可分离卷积,将得到的三个特征进行求平均后与初始特征图  $\mathbf{x}$  进行残差连接,计算公式如下:

$$\begin{cases} \mathbf{x}_{\text{down}} = \text{conv}1 \times 1(\mathbf{x}') \\ \text{conv}x_{1\text{down}} = \text{Dwconv}3 \times 3(\mathbf{x}_{\text{down}}) \\ \text{conv}x_{2\text{down}} = \text{Dwconv}5 \times 5(\mathbf{x}_{\text{down}}) \\ \text{conv}x_{3\text{down}} = \text{Dwconv}7 \times 7(\mathbf{x}_{\text{down}}) \\ \mathbf{x}_{\text{avg}} = \left( \sum_{i=1}^3 \text{conv}x_{i\text{down}} \right) / 3 + \mathbf{x}_{\text{down}} \\ \mathbf{x}_{\text{down}}, \mathbf{x}_{\text{avg}} \in \mathbb{R}^{h \times w \times 64} \end{cases} \quad (8)$$

进一步,将平均特征通过一个  $1 \times 1$  的卷积和一次残差连接,得到最终融合特征  $\mathbf{x}''$  如下:

$$\mathbf{x}'' = \text{conv}1 \times 1(\mathbf{x}_{\text{avg}}) + \mathbf{x}_{\text{avg}}, \mathbf{x}'' \in \mathbb{R}^{64 \times h \times w} \quad (9)$$

将融合特征  $\mathbf{x}''$  进行 GeLU 激活函数并 Dropout,最后通过一个  $1 \times 1$  的卷积进行升维,将通道数还原为  $C$ ,使模块输出的特征图为  $\mathbb{R}^{C \times h \times w}$ 。

### 1.2.2 改进 SEFN 模块

地下停车场环境常常存在低光照或车辆移动导致运动模糊,使关键细节难以识别,而模糊图像则会降低物体检测的精度,本文采用改进 SEFN 结构解决此问题。首先,设主分支输入特征图为  $\mathbf{X}$  和空间分支输入特征图为  $\mathbf{Y}$ ,二者的通道维均为  $C$ 。首先将主分支特征图  $\mathbf{X}$  输入  $1 \times 1$  卷积完成通道扩展,再进行深度卷积并沿通道维拆分为两路特征,产生  $\mathbf{X}_1$  和  $\mathbf{X}_2$ ,其中  $\mathbf{X}_1$  用于与空间分支特征融合,  $\mathbf{X}_2$  则在后续作为门控乘子参与特征重标定。计算公式为

$$\begin{cases} \mathbf{X}' = \text{conv}1 \times 1(\mathbf{X}), \mathbf{X} \in \mathbb{R}^{C \times h \times w}, \mathbf{X}' \in \mathbb{R}^{4C \times h \times w} \\ \mathbf{X}_1, \mathbf{X}_2 = \text{Dwconv}3 \times 3(\mathbf{X}'), \mathbf{X}_1, \mathbf{X}_2 \in \mathbb{R}^{2C \times h \times w} \end{cases} \quad (10)$$

分支输入特征图  $\mathbf{Y}$  通过 2 倍下采样后输入到卷积块中进行特征提取,之后再进行 2 倍上采样恢复到与主分支一致的空间分辨率,这个过程定义为

$$\begin{cases} \mathbf{Y}_{\downarrow} = \text{AvgPool}(\mathbf{Y}), \mathbf{Y}_{\downarrow} \in \mathbb{R}^{C \times \frac{h}{2} \times \frac{w}{2}} \\ \mathbf{Y}' = \text{ReLU}(\text{LayerNorm}(\text{conv}3 \times 3(\mathbf{Y}_{\downarrow}))) \\ \mathbf{Y} = \text{Upsample} \mathbf{Y}', \mathbf{Y} \in \mathbb{R}^{C \times h \times w} \end{cases} \quad (11)$$

进一步,将  $\mathbf{X}_1$  与  $\mathbf{Y}$  在通道维上拼接得到特征图  $\mathbf{Z}$ ,并通过全局平均池化获得通道描述向量  $\mathbf{M}$ :

$$\begin{cases} \mathbf{M}_b^{(k)} = \frac{1}{h \times w} \sum_{i=1}^h \sum_{j=1}^w \mathbf{Z}_b^{(k)}(i, j) \\ \mathbf{Z} \in \mathbb{R}^{B \times 2C \times h \times w}, \mathbf{M} \in \mathbb{R}^{B \times 2C} \end{cases} \quad (12)$$

其中,  $\mathbf{M}$  后续会用于通道注意力权重的计算,具体如下:

$$\begin{cases} \mathbf{W}_1 \in \mathbb{R}^{(2C/4) \times 2C} \\ \mathbf{u}_b^{(d)} = \text{ReLU} \left( \sum_{d=1}^{2C} \mathbf{W}_1^{(d,k)} \cdot \mathbf{M}_b^{(k)} + \mathbf{B}_1^{(d)} \right) \\ \mathbf{W}_2 \in \mathbb{R}^{(2C/4) \times 2C}, \quad \boldsymbol{\beta} \in \mathbb{R}^{B \times 2C \times 1 \times 1} \\ \boldsymbol{\alpha}_b^{(k)} = \sigma \left( \sum_{d=1}^{C/2} \mathbf{W}_2^{(k,d)} \cdot \mathbf{u}_b^{(d)} + \mathbf{B}_2^{(k)} \right) \\ \mathbf{Y}_b^{(k)}(i,j) = \boldsymbol{\alpha}_b^{(k)} \cdot \mathbf{Z}_b^{(k)}, \mathbf{Y} \in \mathbb{R}^{B \times 2C \times h \times w} \end{cases} \quad (13)$$

其中,  $\mathbf{W}_1$  和  $\mathbf{W}_2$  分别表示第一个和第二个全连接层的权重矩阵;  $\mathbf{B}_1$  和  $\mathbf{B}_2$  分别表示第一个和第二个全连接层的偏置;  $\boldsymbol{\beta}$  表示经过 sigmoid 激活得到的通道注意力权重张量;  $\mathbf{u}_b^{(d)}$  和  $\boldsymbol{\alpha}_b^{(k)}$  为进行 ReLU 激活和 Sigmoid 激活的特征响应;  $\boldsymbol{\alpha}_b^{(k)}$  表示映射得到的通道注意力权重, 通过该过程, 可对融合特征  $\mathbf{Z}$  进行动态通道重标定, 突出关键通道信息。最后将通道注意力处理后的特征图  $\mathbf{Y}$  进行 GeLU 激活后与特征图  $\mathbf{X}_2$  逐元素相乘, 最后通过一个  $1 \times 1$  卷积通道维度恢复为  $C$ , 从而得到模块的输出特征  $\mathbf{O}$ , 其计算式为

$$\begin{cases} \mathbf{O}' = \text{GeLU}(\mathbf{Y}) \odot \mathbf{X}_2, \mathbf{O}' \in \mathbb{R}^{B \times 2C \times h \times w} \\ \mathbf{O} = \text{conv1} \times 1(\mathbf{O}'), \mathbf{O} \in \mathbb{R}^{B \times C \times h \times w} \end{cases} \quad (14)$$

SEFN 结构通过多尺度空间特征融合与动态通道注意力门控机制, 在保留关键细节的同时增强了低光照与运动模糊场景下的特征判别性, 可以有效提升地下停车场的物体检测精度。

### 1.2.3 CSAL 模块算法实现

通过融合上述各子模块, 本文提出的 CSAL 模块算法实现过程伪代码如算法 2 所示。

算法 2 CSAL 模块算法

---

输入: SRC[BS, C, h, w]  
 输出: 编码特征[BS, C, h, w]

1. 虚保存原始输入: src\_original
2. 特征序列化: src\_flat ← 展平空间维度
3. 注意力计算: attn\_output ← pola\_attention(src\_flat)
4. 恢复空间结构: attn\_spatial ← 还原空间维度(attn\_output, h, w)
5. 残差连接 1: src ← src + dropout1(attn\_spatial)
6. 层归一化 1: src ← norm1(src)
7. 特征变换 1: src ← mona1(src)
8. 前馈计算: sefn\_output ← sefn(src, src\_original)
9. 残差连接 2: src ← src + dropout2(sefn\_output)
10. 层归一化 2: src ← norm2(src)
11. 特征变换 2: src ← mona2(src)

---

## 2 透视拓扑补全模块

### 2.1 几何推理补全理论

#### 2.1.1 场景分层与锚点对提取

首先, 依据图像纵坐标  $y$  对所有检测框进行排序

与聚类, 将车位划分为近端(前排)集合  $\{\mathbf{B}_n\}$  与远端(后排)集合  $\{\mathbf{B}_f\}$ 。随后, 在两组之间寻找可靠的几何锚点对  $\mathbf{P} = \{(\mathbf{b}_n^i, \mathbf{b}_f^j)\}$ 。一对车位被判定为锚点对的条件是满足平面投影对齐约束:

$$\frac{|x_c^n - x_c^f|}{\max(w^n, w^f)} < \varepsilon \quad (15)$$

其中,  $(x_c, y_c)$  为检测框中心坐标;  $w$  为检测框宽度;  $\varepsilon$  为对齐阈值。锚点对构成了连接图像近处与远处的视觉测量基准线, 为后续计算提供基础。

#### 2.1.2 灭点估计与透视场景建模

灭点 (Vanishing Point, VP) 是三维空间中平行线在二维图像平面的交汇点, 是描述场景透视结构的关键参数, 每个锚点对定义了一条空间线在图像上的投影线段  $\mathbf{L}_k$ 。本文通过求解以下最小二乘问题, 以此得到最优灭点估计  $\mathbf{v}^*$ :

$$\min_{\mathbf{v}} \sum_{k \in p} [\mathbf{n}_k^T \mathbf{v} - d_k]^2, \mathbf{v} = \{(\mathbf{b}_n^i, \mathbf{b}_f^j)\} \quad (16)$$

其中,  $\mathbf{n}_k$  为线段  $\mathbf{L}_k$  的单位法向量;  $d_k$  为原点到该直线距离。此解表征了车位排列方向的消失方向。

#### 2.1.3 基于深度比例一致性的遮挡补全

对于近端集合  $\mathbf{B}_n$  中未能与任何远端车位形成锚点对的孤立车位  $\mathbf{b}_n^*$ , 假设其对应的后停车位  $\mathbf{b}_f^*$  因遮挡而未被检测。利用透视几何中深度比例一致性原理, 可推断  $\mathbf{b}_f^*$  位置。基于可靠锚点对集合  $\mathbf{p}$ , 可计算平均深度收缩比  $\eta$ :

$$\eta = \frac{1}{\bar{p}} \sum_{(\mathbf{b}_n^i, \mathbf{b}_f^j \in \mathbf{p})} \frac{\|\mathbf{c}_f^* - \mathbf{v}^*\|_2}{\|\mathbf{c}_n^* - \mathbf{v}^*\|_2} \quad (17)$$

其中,  $\bar{p}$  表示集合  $\mathbf{p}$  的元素数量;  $\mathbf{c}_n^*$  为车位中心坐标;  $\mathbf{c}_f^*$  为被遮挡远端车位中心, 通过下式进行预测计算:

$$\mathbf{c}_f^* = \mathbf{v}^* + \zeta (\mathbf{c}_n^* - \mathbf{v}^*) \quad (18)$$

其中,  $\zeta$  为纵坐标比例系数, 本文取  $-0.5$ , 检测框的尺寸  $(\mathbf{w}_f^*, \mathbf{h}_f^*)$  则取远端集合  $\mathbf{B}_f$  中已检测车位的平均尺寸。由此, 生成一个推测的检测框  $\mathbf{b}_f^* = (\mathbf{c}_f^*, \mathbf{w}_f^*, \mathbf{h}_f^*)$ , 并将其初始状态标记为“占用”, 置信度为  $0.5$ 。

### 2.2 补全算法实现

结合上述补全算法理论分析, 本文提出的透视拓扑补全模块算法实现过程伪代码如算法 3 所示。

## 3 车位端到端实时检测算法实现

综合上述改进的 RT-UPSO-DETR 网络和透视拓扑补全模块, 本文提出基于 RT-UPSO-DETR 的停车场车位端到端检测算法, 其实现过程如下, 其流程图如图 4 所示。

**算法 3 透视拓扑补全模块算法**

输入:初始检测框集合  $B$

输出:输出最终检测框集合

1. 场景分层:按  $y$  坐标排序聚类,分为近端  $B_n$  与远端  $B_f$
2. 锚点提取:基于水平对齐约束,构建锚点对集合  $P$
3. 灭点估计:若  $|P| > 1$  则最小二乘法求解灭点  $v$ , 否则使用默认值
4. 补全目标:识别未匹配的孤立近端车位  $B_{isolated} \subseteq B_n$
5. 收缩比计算:基于  $P$  计算平均深度收缩比  $\eta$
6. 远端尺寸:计算远端框平均尺寸  $(w_{avg}, h_{avg})$
7. 遍历补全:for each  $b_{iso}$  in  $B_{isolated}$  do
8. 预测中心:  $c_{far} \leftarrow v + (c_{iso} - v) \times \eta$
9. 生成补全框:  $b_{new} \leftarrow (c_{far}, w_{avg}, h_{avg})$
10. 设置状态:  $b_{new.state} \leftarrow$ “占用”,  $conf \leftarrow 0.5$
11. 标记推测:  $b_{new.is\_inferred} \leftarrow True$
12. 加入集合:  $B_{completion}.add(b_{new})$
13. 结束遍历
14. 输出:补全框集合  $B_{completion}$

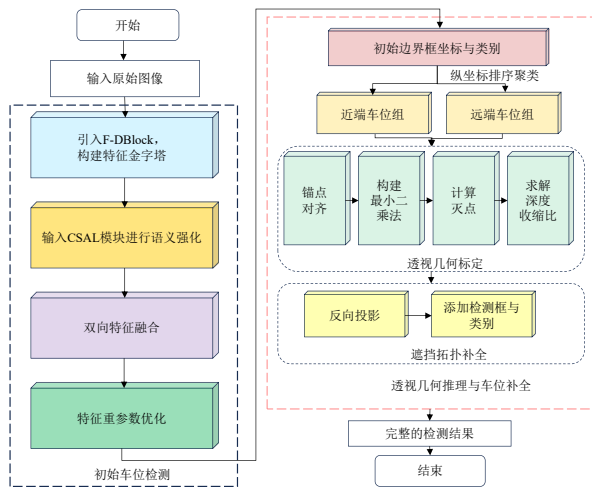


图 4 基于 RT-UPSODETR 的停车场车位端到端实时检测算法

Figure 4 End-to-end real-time parking space detection algorithm based on RT-UPSODETR

步骤 1:特征金字塔构建。输入原始图像,通过五级卷积下采样构建特征金字塔,并引入 F-DBlock 模块增强每一级特征的非线性表达能力,提取多尺度空间特征。

步骤 2:深层语义强化。将深层语义特征输入 CSAL 模块抑制背景噪声,聚焦车位区域的关键语义特征。

步骤 3:双向特征融合。执行自顶向下与自底向上的双向特征融合策略,将 CSAL 模块输出的深层语义信息与浅层细节纹理信息进行对齐与互补。

步骤 4:特征重参数化优化。利用 RepC3 模块对融合后的特征图进行跨级特征重组与优化,提升特征的鲁棒性与表达效率。

步骤 5:基线检测生成。将优化后的特征图送入解耦检测头,通过端到端解码回归出车位的初始边界框坐标及其占用状态类别。

步骤 6:场景结构解析。以基线检测结果为输入,依据检测框中心纵坐标的分布特征,自适应地将车位划分为近端集合与远端集合。

步骤 7:透视几何标定。在近、远端集合间筛选空间对齐的锚点对,构建最小二乘问题求解场景全局灭点,并解算深度收缩比系数。

步骤 8:遮挡拓扑补全。针对近端未匹配的孤立车位,基于灭点约束与深度收缩比执行径向反向投影,恢复被遮挡车位的位姿与尺度,输出最终的完整检测结果。

## 4 实验测试

### 4.1 实验数据集构建

由于公开数据集 PKLot<sup>[21]</sup> 和 CNRPark+EXT<sup>[22]</sup> 聚焦于户外停车场不同光照和天气下的车位占用情况检测,为构建适用于地下场景的高鲁棒性检测模型,针对地下停车场光照条件差、背景复杂度高、摄像头部署位置受限等特点,本文采集并制作了地下停车场车位检测数据集用于算法的测试和验证。本数据集共包含 909 张高分辨率图像,并涵盖了六类(密集停放、边缘区域、非机动车占用、异物占用、强光干扰、结构遮挡)典型停车场景,数据集采用随机抽样的方法将图片按照 7:2:1 的比例划分为训练集,验证集和测试集。同时,为全面评估本位提出模型在广泛低光照环境下的泛化能力与特征鲁棒性,选用公开的 ExDark 数据集<sup>[23]</sup> 作为泛用性检测基准,该数据集共包含 7 363 张图像,涵盖了极暗夜间、弱光室内到晨昏过渡等 10 种光照等级,并包含复杂天气(雾、雨)及动态光源(车灯、霓虹)干扰,数据集中标注了车辆、行人、动物、日常物品等 12 类常见目标,总计 40 673 个实例,分为训练集(3 000 张图像)、验证集(1 800 张图像)和测试集(2 563 张图像)。此外,针对本文提出的透视拓扑补全模块,本文还构建了包含 129 张图像的遮挡补全专项测试数据集,由于透视拓扑补全模块是独立于目标检测模型的后处理模块,无需训练,因此 129 张图像全部作为测试集使用。为了明确该专用测试集的物理属性并为后续评价指标的设定提供依据,本文对数据集中所有遮挡车位的投影尺寸进行了统计分析,结果如图 5 所示。结合统计结果,测试中车位在图像平面的平均宽度设置为 390.5 像素,平均高度设置为 111.4 像素。

### 4.2 实验参数与评价指标

本文测试在安装 Windows 11(64 位)系统,配置

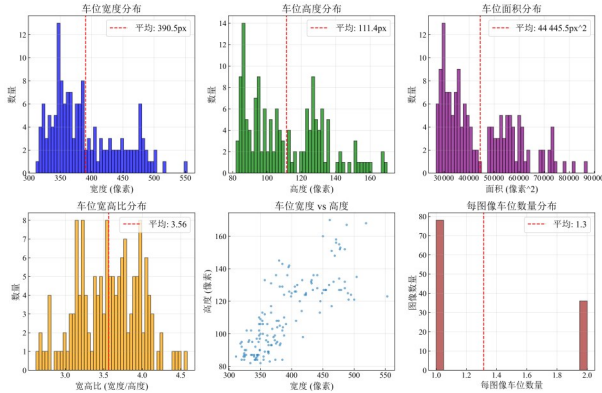


图5 遮挡补全专项测试集分析

Figure 5 Analysis of the special test set for obstruction completion

Intel Core i5-12600kf CPU 和 NVIDIA GeForce RTX 4060Ti (8 GB) GPU 的电脑上进行, 软件环境为: Python 3.9.12, CUDA 12.1 和 PyTorch 2.2.2。为保证实验一致性, 具体实验参数如表 1 所示。

本文所采用的损失函数依旧包括分类损失和回归损失两部分, 其中分类损失使用标准交叉熵损失, 回归损失由 L1 损失和 GIoU 损失构成, 总损失函数定义为

$$L = \lambda_{cls} L_{cls} + \lambda_{L1} L_{L1} + \lambda_{giou} L_{giou} \quad (19)$$

其中,  $L_{cls}$  为分类损失;  $L_{L1}$  为预测边界框与真实框坐标的 L1 距离损失;  $L_{giou}$  为广义交并比损失, 各损失项

表 2 目标检测模型评价指标

Table 2 Evaluation indicators of the target detection model

指标(Metric)	计算公式(Formula)	指标(Metric)	计算公式(Formula)
精确度(P)	$P = TP / (TP + FP)$	参数量(Parameters)	$Parameters = \sum_{l=1}^L [C_{in_l} \times C_{out_l} \times k_{w_l} \times k_{h_l}]$
召回率(R)	$R = TP / (TP + FN)$	平均精度(mAP)	$AP = \frac{1}{n} \sum_{i=1}^n AP_i$
计算量(GFLOPs)	$GFLOPs = C_{in} \times C_{out} \times K_w \times K_h \times H_{out} \times W_{out}$	每秒帧数(FPS)	$FPS = \frac{1}{T_{avg}}$

此外, 针对透视拓扑补全模块的评价, 本文采用三项专用指标: 漏检召回率 (Missed Recall Rate, MRR)、补全准确率 (Completion Accuracy, CA) 和平均中心点误差 (Mean Center Error, MCE) 进行评测, 指标计算公式如表 3 所示, 其中  $N_{TPcomp}$  为有效补全数;  $N_{MISS}$  为被标记为遮挡且基线模型未检出的车位数;  $N_{comp}$  为模块输出的总补全框数 (含误补);  $C_{gt}^{(i)}$  为真实检测框中心点坐标。

### 4.3 测试结果与分析

#### 4.3.1 消融实验

为了验证提出模块的有效性, 选取 RT-DETR-R18 为基准模型进行消融实验, 实验环境、参数设置和测试数据集均保持一致, 实验结果如表 4 所示。分析可

的权重系数设置为  $\lambda_{cls} = 2.0$ ,  $\lambda_{L1} = 5.0$ ,  $\lambda_{giou} = 2.0$ 。上述损失函数在训练过程中通过反向传播优化模型参数, 使检测框具备准确的分类置信度和定位。

表 1 实验参数设置

Table 1 Experimental parameter settings

参数名称	参数设置	参数名称	参数设置
Epochs	200	Optimizer	AdamW
Batchsize	4	lr0	0.000 1
Wokers	4	lrf	1.0
Imgpsz	640	momentum	0.9

此外, 为了全面验证本文所提出算法的有效性, 本文沿用目标检测领域的通用标准, 采用精确率 (P)、召回率 (R) 及平均精度 (mAP) 来衡量模型对可见车位的识别能力并引入每秒帧数 (FPS) 作为实时性能参考指标, 同时使用参数量 (Parameters) 和浮点运算次数 (GFLOPs) 作为模型复杂度与计算开销的衡量标准, 评价指标计算公式如表 2 所示。表 2 中, TP 为模型正确检测到的目标; FP 为模型误将非目标区域检测为了目标区域; FN 为模型未能正确检测到的目标;  $L$  表示模型的总层数;  $C_{in_l}$ 、 $C_{out_l}$ 、 $K_{w_l}$ 、 $K_{h_l}$ 、 $H_{out}$  和  $W_{out}$  分别表示第  $l$  层输入输出特征图的通道数、卷积核的宽度和高度和输出特征图的高度和宽度;  $T_{avg}$  为模型处理单张图像的平均推理时间。

表 3 透视拓扑补全模块评价指标

Table 3 Evaluation indicators of the perspective topology completion module

指标(Metric)	计算公式(Formula)
漏检召回率(MRR)	$MRR = N_{TPcomp} / N_{MISS}$
补全准确率(CA)	$CA = N_{TPcomp} / N_{comp}$
中心点误差(MCE)	$MCE = \frac{1}{N_{TPcomp}} \sum_{i=1}^{N_{TPcomp}} \left\  C_{comp}^{(i)} - C_{gt}^{(i)} \right\ _2$

知, 在加入了 F-DBlock 模块后, 提出模型检测精度提升了 1%, mAP@0.5 提升了 1%, 模型的参数量下降了 31.6%, GFLOPs 降低了 20.6%, 召回率提升并不显著, mAP@0.5-0.95 降低了 1%, FPS 降低了 0.015。而在添加了 CSAL 模块后, 精度和召回率分别提高了 0.9% 和 1.1%,

mAP@0.5、mAP@0.5-0.95 和 FPS 分别提升了 0.5%、0.2%、4.48,但参数量和 GFLOPs 有小幅增加。当同时添加 F-DBlock 模块和 CSAL 模块之后,模型的精度、召回和 mAP@0.5-0.95 都为最佳,相较于基准模型分别提高了 1.5%、3.3% 和 0.9%,虽然 mAP@0.5 和 FPS 稍逊于仅加入了 F-DBlock 的模型,但模型的参数量和 GFLOPs

分别减少了 22.1% 和 18.8%。实验结果显示所提出的 F-DBlock 模块在降低计算复杂度的同时有效维持了检测精度,而提出的 CSAL 模块可以有效增强特征表达,且通过优化注意力机制实现了更高效的推理,两个模块协同工作在增强识别能力的同时能有效控制计算开销,且满足 30 FPS 以上的实时检测需求。

表 4 消融实验结果

Table 4 Results of the ablation experiment

RT-DETR-R18	F-DBlock	CSAL	P/%	R/%	mAP@0.5/%	mAP@0.5-0.95/%	Parameters/10 <sup>6</sup>	GFLOPs	FPS
√			90.8	93.3	95.5	82.8	19.87	56.9	54.54
√	√		91.8	93.4	<b>96.5</b>	81.8	<b>13.59</b>	<b>45.2</b>	54.39
√		√	91.7	94.4	96.0	83.0	21.72	57.7	<b>59.02</b>
√	√	√	<b>92.3</b>	<b>96.6</b>	96.2	<b>83.7</b>	15.47	46.2	52.40

注:“√”表示采用对应模块;加粗数字为该列最优实验结果。

### 4.3.2 对比实验

为了验证改进 RT-UPSO-DETR 模型的有效性,本文在相同数据集与实验环境下,将模型与以下两类模型进行对比:(1)主流通用检测模型,包括 RT-DETR-R18、RT-DETR-R34 和从 YOLOv8、YOLOv10<sup>[24]</sup>、YOLO11<sup>[25]</sup> 系列模型;(2)专用车位检测模型,包含 YOLO-ResNet-18 和 YOLO-EfficientNet<sup>[26]</sup>。所有对比模型均采用官方配置文件,并按照统一的训练设置在本文构建的地下停车场车位检测数据集上重新训练,以确保对比的公平性。如表 5 所示的实验结果表明,本文提出的模型在目标检测任务中展现出精度与效率的良好平衡。在精度方面,本文模型的精确率和召回率达到了 92.3% 和 96.6%,均优于所有对比模型,尤其召回率显著领先,表明模型在漏检抑制方面具有明显优势。在综合精度指标 mAP@0.5-0.95 上,本文模型达到 83.7%,略低于 YOLO-ResNet-18 的 83.9% 和 YOLOv10b 与 YOLOv10m 的 83.8%,与 RT-DETR-R34

持平,体现了良好的定位与分类性能。从计算开销看,本文模型参数量和计算量均低于 YOLOv8m、YOLOv10m、YOLO-EfficientNet 等模型,仅高于轻量化模型 YOLO-ResNet-18 与 YOLOv10s。本文模型推理速度相对较低(52.4 FPS),与基准模型 RT-DETR-R18 模型基本持平且都逊于 YOLO 系列模型,这主要源于 CSAL 模块在深层特征图上引入的注意力计算开销。整体而言,本文模型在高召回率与中等计算开销之间取得了很好的折中,同时具备进一步轻量化设计以提升推理速度的潜力。图 6 给出了所提出模型的对比测试结果展示,可以看出本文提出的模型检测出了在强光和结构柱遮挡影响下的占用车位,YOLO EfficientNet、YOLOv10m 模型未检测出最左边的空闲车位,YOLOv11m 则不仅未能检测出最左边的空闲车位,还存在着一个误报,对比结果显示改进模型在地下停车场车位检测任务中的有效性。

表 5 不同方法测试性能指标对比

Table 5 Comparison of performance indicators tested by different methods

模型	P/%	R/%	mAP@0.5/%	mAP@0.5-0.95/%	Parameters/10 <sup>6</sup>	GFLOPs	FPS
YOLOv8m	91.5	93.3	<b>96.7</b>	81.9	25.84	78.7	94.93
YOLO-ResNet-18	92.0	93.4	95.7	<b>83.9</b>	13.32	35.2	<b>137.97</b>
YOLO EfficientNet	92.2	92.6	95.2	83.4	23.40	56.4	68.54
YOLOv10b	<b>92.3</b>	90.9	96.1	83.8	20.41	97.9	71.66
YOLOv10m	89.9	92.2	96.6	83.8	16.45	63.4	82.61
YOLOv11s	91.5	93.2	96.0	82.5	<b>9.41</b>	<b>21.3</b>	113.29
YOLOv11m	91.6	94.2	96.4	81.9	20.03	67.7	90.06
RT-DETR-R18	90.8	93.3	95.5	82.8	19.87	56.9	54.54
RT-DETR-R34	90.9	93.5	95.6	83.7	31.10	88.8	43.98
RT-UPSO-DETR	<b>92.3</b>	<b>96.6</b>	96.2	83.7	15.47	46.2	52.40

注:加粗数字为该列最优实验结果。

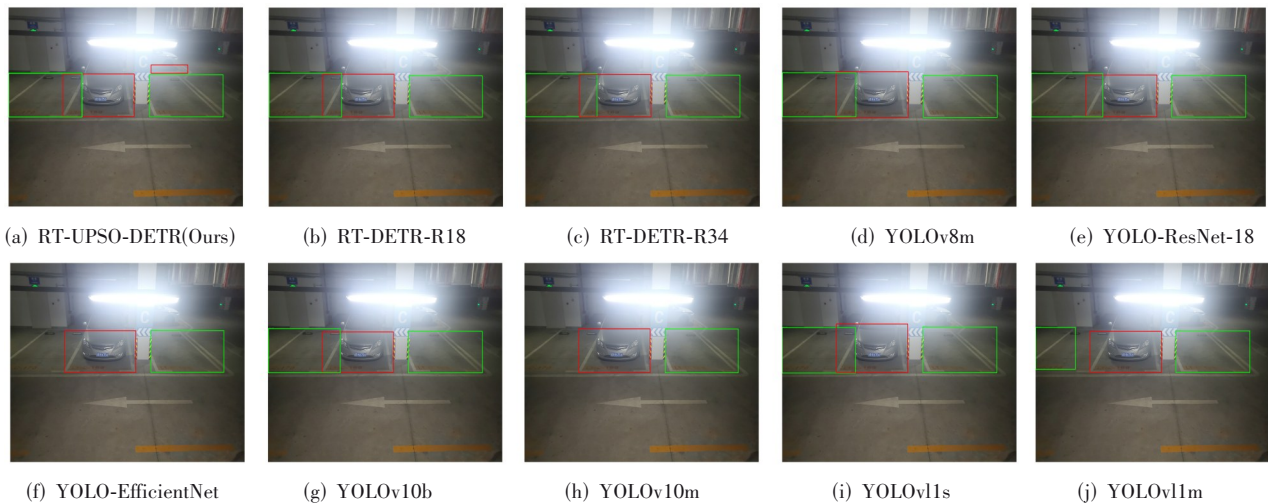


图6 车位检测对比结果

Figure 6 Comparison results of parking space detection

#### 4.3.3 对比实验模型泛化实验

为验证本文提出模型在跨域低光照场景下的泛化能力,在 ExDark 数据集<sup>[27]</sup>上进行模型泛化测试,结果如表6所示,部分典型场景的可视化检测结果如图7所示。结果表明,所改进模型在未知低光照数据上仍保持性能优势,相较于未改进 RT-DETR-R18 模型,精度、召回率、mAP@0.5 和 mAP@0.5-0.95 分别提升了 2.1%、1.3%、2.1%、1.0%,显示出改进模块有效提升了模型在低光照条件下的鲁棒性。

#### 4.3.4 车位补全实验

车位补全实验在构建的遮挡补全专项测试数据集上进行,实验结果如表7所示。分析结果可知,提出的透视拓扑补全模块成功恢复了 82.48% 被基线模型漏检的遮挡车位,并且模块生成的补全框中,

表6 EXDark 数据集测试结果 单位:%

Table 6 Test results of the EXDark dataset unit:%

模型/提升度	$P$	$R$	mAP@0.5	mAP@0.5-0.95
RT-DETR-R18	67.3	47.0	51.2	31.2
RT-UPSO-DETR	<b>69.4</b>	<b>48.3</b>	<b>53.3</b>	<b>32.2</b>
提升度	2.1	1.3	2.1	1.0

注:加粗数字代表该列指标最优结果。

74.83% 是正确且有效的。有效补全框的中心点与真实车位中心点的平均距离为 23.27 像素,这表明模块解在决透视遮挡导致漏检这一核心问题具有有效性,表明模块的几何推理的作用,验证了灭点约束与深度比例估计的准确性。为了直观展示所提出的模型的工作机制与效果,图8给出了车位检测及补全测试结果。

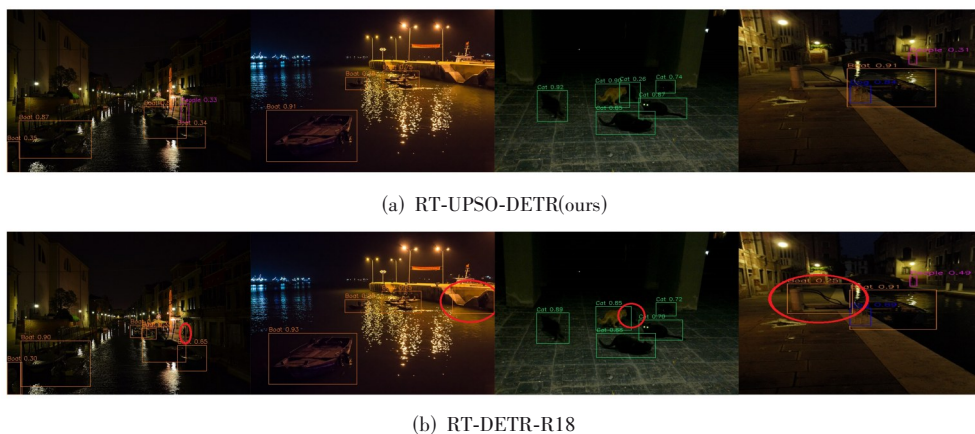


图7 EXDark 数据集检测对比结果

Figure 7 Detection comparison results of the EXDark dataset

由图8(a)可知,RT-UPSO-DETR 检测模型成功检测到了近端的3个车位,其中2个被标记为占用(红

色框),1个为空闲(绿色框)。由于透视投影和遮挡效应,远端的3个车位中只有1个被检测到,标记为

表 7 透视拓扑补全模块定量评估结果

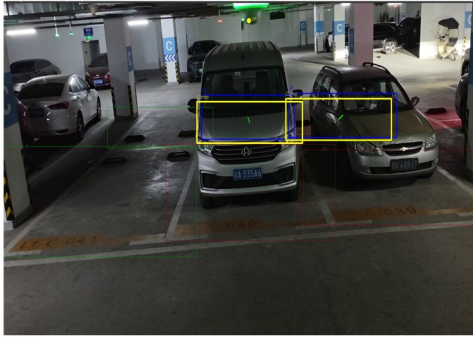
单位: %

Table 7 Quantitative evaluation results of the perspective topology completion module

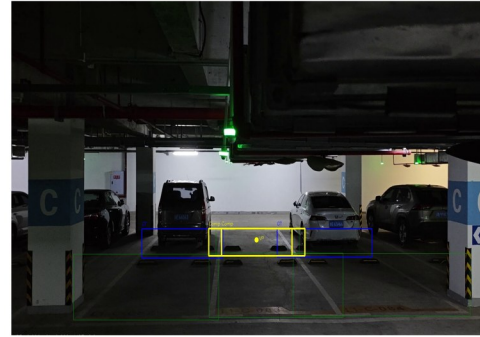
unit: %

算法	MRR	CA	MCE
纯 RT-UPSO-DETR 算法	0.0	—	—
基于 RT-UPSO-DETR 和透视补全的停车场车位检测算法	82.48	74.83	23.27

注:“—”表示该项指标无结果。



(a) 透视拓扑补全模块成功案例



(b) 透视拓扑补全模块失败案例

(b) Successful cases of the perspective topology completion module

(b) Failure cases of the perspective topology completion module

图 8 透视拓扑补全模块结果展示

Figure 8 Result display of the perspective topology completion module

空闲,漏检的车位位于图像上方,且其对应的近端车位有车辆停放,形成了典型的垂直遮挡。然而,经过本文提出的透视拓扑补全模块处理后,系统成功补全了被遮挡的后排车位,并将它们标记为“占用”(黄色框),补全框与人工标注的真值框(蓝色框)高度重合,且在车辆完全遮挡的情况下依然能够准确定位,进一步验证了模型的准确性。然而,所提透视拓扑补全模块也存在一定局限性,如图 8(b)展示了一个典型的补全失败案例,该场景中,前排车位被成功检测到三个有效框,而后排仅有一个车位被检测到,这是因为后排检测框过少,算法在构建锚点连线时,将多个前排队位与同一个后排队位进行匹配,导致所有连线均交汇于该后排队位的中心点,最小二乘法由此求解出的灭点被错误地定位在后排队位内部,而非真实的透视消失方向,最终导致补全的失败。上述结果表明当前方法对锚点数量及分布存在一定的敏感性:当某一排检测框过少时,无法形成正确的透视平行线约束,灭点估计易发生畸变,进而影响整个补全过程,这也将是后续研究需要改进的方向。

## 5 结论

本文提出了一种轻量的地下停车场车位实时检测网络 RT-UPSO-DETR 来提高检测精度并降低计算成本。首先参考了 CSPNet 模块设计了 F-DBlock 模块建设了计算量,降低了模型参数量,并通过扩大感受野来更好捕获全局特征并;其次设计了多机制融合的

Transformer 编码层 CSAL 模块,通过引入新的注意力机制和残差网络,有效地提升了模型在地下停车场环境下的处理画面模糊和低光照图片的能力;最后设计了基于灭点约束的透视拓扑补全模块,通过检测并利用近端与远端车位之间的对齐关系和场景透视结构实现了从图像模式识别到场景结构化理解的方法论提升。在地下停车场数据集上模型实现了 92.3% 的精确率、96.6% 的召回率、96.7% 的  $mAP@0.5$ 、83.7% 的  $mAP@0.5-0.95$ 、 $15.3 \times 10^6$  的参数数量和 46 G 的 GFLOPs。同时,模块实现了多种场景下的漏检车位补全,且在漏检召回率、补全准确率和平均中心点误差指标上达到了 82.48%、74.83% 和 23.27 pixels,显示出所提出的 RT-UPSO-DETR 实时检测模型在地下停车场车位检测任务中的有效性以及提出的模块在视觉透视遮挡导致漏检这一核心问题具有有效性。

尽管本文提出的透视拓扑补全模块在多数场景下能有效恢复被遮挡车位,但失败案例分析表明,该方法对锚点数量及分布较为敏感,当后排检测框过少时易导致灭点估计错误。未来透视拓扑补全模块将从以下方面进行改进:(1)引入 RANSAC 等鲁棒估计算法剔除异常匹配,提升灭点估计的稳定性;(2)融合车位线等几何先验信息,增强锚点匹配的可靠性。RT-UPSO-DETR 模型方面将进一步扩展多光照、多视角场景数据规模,提升模型泛化能力。此外,本文研究相关的数据及代码已经开源,网址为: <https://gitlink.org.cn/maxre1/RT-UPSO-DETR>。

## 参考文献

- [1] Lang A H, Vora S, Caesar H, et al. PointPillars: Fast encoders for object detection from point clouds[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019: 12689-12697.
- [2] Chiang K W, Tsai S, Chen J A, et al. AI-driven mapping system for smart parking management applications using an INS-GNSS-solid-state LiDAR-monocular camera fusion engine empowered by HD maps[J]. IEEE Open Journal of Intelligent Transportation Systems, 2025, 6: 995-1008.
- [3] Yan S, O'Connor N E, Liu M M. U-park: A user-centric smart parking recommendation system for electric shared micromobility services[J]. IEEE Transactions on Artificial Intelligence, 2024, 5(10): 5179-5193.
- [4] Bazzaza T, Tohidypour H R, Wang Y X, et al. Accurate detection and localization of individual free street parking spaces using AI and innovative global motion estimation[J]. IEEE Transactions on Intelligent Vehicles, 2025, 10(2): 1263-1272.
- [5] Zhang Z S, He X M, Huang J W, et al. Parking detection using combined magnetic sensor and pulsed coherent radar[J]. IEEE Internet of Things Journal, 2022, 9(18): 17210-17219.
- [6] Zhang Y, Wu C, Liu S Y, et al. A parking detection algorithm based on multitransitory finite-state machine using magnetic wireless sensor network[J]. IEEE Internet of Things Journal, 2024, 11(5): 8360-8372.
- [7] Dheeven T A, Kumar P M, Venkatesh V, et al. IoT based sensor enabled vehicle parking system[J]. Measurement: Sensors, 2024, 31: 100953.
- [8] 黄伟杰, 张希, 赵柏暄, 等. 基于视觉的停车场车位检测与分类算法[J]. 计算机系统应用, 2022, 31(3): 234-240. Huang Weijie, Zhang Xi, Zhao Baixuan, et al. Vision-based parking space detection and classification algorithm[J]. Computer Systems & Applications, 2022, 31(3): 234-240. (in Chinese)
- [9] Sharma A, Kumar V, Longchamps L. Comparative performance of YOLOv8, YOLOv9, YOLOv10, YOLOv11 and Faster R-CNN models for detection of multiple weed species[J]. Smart Agricultural Technology, 2024, 9: 100648.
- [10] Wang S B, Chen R H, Wu H Y, et al. YOLOH: You only look one hourglass for real-time object detection[J]. IEEE Transactions on Image Processing, 2024, 33: 2104-2115.
- [11] Jrondi Z, Moussaid A, Hadi M Y. Exploring End-to-End object detection with transformers versus YOLOv8 for enhanced citrus fruit detection within trees[J]. Systems and Soft Computing, 2024, 6: 200103.
- [12] Zhao Y A, Lv W Y, Xu S L, et al. DETRs beat YOLOs on real-time object detection[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2024: 16965-16974.
- [13] 蒋智臣, 胡俐蕊. 基于改进 RT-DETR 的浅水海洋生物识别方法[J]. 电子测量技术, 2024, 47(18): 155-163. Jiang Zhichen, Hu Lirui. Marine life identification method based on improved RT-DETR[J]. Electronic Measurement Technology, 2024, 47(18): 155-163. (in Chinese)
- [14] 程鑫淼, 张雪松, 曹冰洁, 等. 改进 RT-DETR 的小目标检测方法研究[J]. 计算机工程与应用, 2025, 61(15): 144-155. Cheng Xinmiao, Zhang Xuesong, Cao Bingjie, et al. Research on small object detection method of improved RT-DETR[J]. Computer Engineering and Applications, 2025, 61(15): 144-155. (in Chinese)
- [15] Feijoo D, Benito J C, Garcia A, et al. DarkIR: Robust low-light image restoration[C]//2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2025: 10879-10889.
- [16] Yin D S, Hu L Y, Li B, et al. 5%>100%: Breaking performance shackles of full fine-tuning on visual recognition tasks[C]//2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2025: 20071-20081.
- [17] Lv H Z, Yang W Z, Yin Y B, et al. MDF-FND: A dynamic fusion model for multimodal fake news detection[J]. Knowledge-Based Systems, 2025, 317: 113417.
- [18] Chen S, Zhang H Z, Atapour-Abarghouei A, et al. SEM-net: Efficient pixel modelling for image inpainting with spatially enhanced SSM[C]//2025 IEEE/CVF Winter Conference on Applications of Computer Vision. Piscataway: IEEE, 2025: 461-471.
- [19] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 770-778.
- [20] Wang C Y, Mark Liao H Y, Wu Y H, et al. CSPNet: A new backbone that can enhance learning capability of CNN[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Piscataway: IEEE, 2020: 1571-1580.
- [21] Nguyen H, Nawara D, Kashef R. Connecting the indispensable roles of IoT and artificial intelligence in smart

cities: A survey[J]. Journal of Information and Intelligence, 2024, 2(3): 261-285.

- [22] 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 Xudong, Qian Li. Parking slot detection based on surround-view image[J]. Journal of Central China Normal University (Natural Sciences), 2024, 58(5): 526-532. (in Chinese)
- [23] 王满利, 张航, 张长森. 基于深度学习的低光照目标检测算法[J]. 北京邮电大学学报, 2024, 47(5): 59-65.  
Wang Manli, Zhang Hang, Zhang Changsen. A low light target detection algorithm based on deep learning[J]. Journal of Beijing University of Posts and Telecommunications, 2024, 47(5): 59-65. (in Chinese)
- [24] 邓冬冬, 许建民, 孟寒, 等. 基于蚁群算法与人工势场法融合的移动机器人路径规划[J]. 仪器仪表学报, 2025, 46(2): 1-16.

Deng Dongdong, Xu Jianmin, Meng Han, et al. Mobile robot path planning based on the fusion of ant colony algorithm and artificial potential field method[J]. Chinese Journal of Scientific Instrument, 2025, 46(2): 1-16. (in Chinese)

- [25] Wang D R, Tan J S, Wang H, et al. SDS-YOLO: An improved vibratory position detection algorithm based on YOLOv11[J]. Measurement, 2025, 244: 116518.
- [26] Pokhrel A, Dao G. Optimizing YOLOv8 for parking space detection: Comparative analysis of custom YOLOv8 architecture[PP/OL]. V1. arXiv(2025-05-23) [2026-03-04]. <https://doi.org/10.48550/arXiv.2505.17364>.
- [27] 秦嘉奇, 江泽涛, 雷晓春. 基于ICFIE-YOLO的低照度图像目标检测方法[J]. 电子学报, 2025, 53(2): 514-526.  
Qin Jiaqi, Jiang Zetao, Lei Xiaochun. Low illumination image object detection method based on ICFIE-YOLO[J]. Acta Electronica Sinica, 2025, 53(2): 514-526. (in Chinese)

#### 作者简介



**刘 平** 男, 1990年6月出生于重庆市。现为重庆邮电大学自动化学院副教授、硕士生导师。主要研究方向为无人系统轨迹规划、智慧泊车。  
E-mail: liuping\_cqupt@cqupt.edu.cn



**朴昌浩** 男, 1978年11月出生于吉林省延边。现为重庆邮电大学教授、博士生导师。主要研究方向为智能网联汽车。  
E-mail: piaoch@cqupt.edu.cn



**马春蕾** 男, 2001年3月出生于重庆市。2023年毕业于重庆邮电大学电气自动化专业。现为重庆邮电大学研究生。主要研究方向为机器视觉与图像识别。  
E-mail: S240333022@stu.cqupt.edu.cn



**黄龙航** 男, 1997年9月出生于重庆市铜梁区。2024年毕业于内蒙古工业大学电子信息系, 现为重庆市建设信息中心科员。主要研究方向为智慧建筑。  
E-mail: 1511097975@qq.com



**刘明杰** 男, 1987年出生于河南省辉县市。现为重庆邮电大学副教授、博士生导师。主要研究方向为智能网联汽车环境智能感知及控制、多源信息融合、机器学习。  
E-mail: liumj@cqupt.edu.cn



**颜鲁鹏** 男, 1995年2月出生于甘肃省兰州市。2020年毕业于重庆交通大学建筑与土木工程专业。现为重庆市建设信息中心中级工程师。主要研究方向为住建领域数字化发展。  
E-mail: 954518155@qq.com