DOI: 10.19650/j.cnki.cjsi.J2513946

基于双层级显著性驱动的车辆部件检测方法*

翟永杰,吴梓沣,周迅琪,魏乐涛,王乾铭
(1.华北电力大学自动化系 保定 071003; 2.保定市电力系统智能机器人感知与控制重点实验室 保定 071003)

摘 要:高精度的车辆部件检测与分割技术对智能定损系统中辅助定位损伤部件至关重要,但常面临着复杂场景下存在的背景干扰抑制难题,以及传统检测方法局限于单层次特征表征而导致的检测效能瓶颈。为解决这一问题,提出了一种基于双层级显著性驱动的车辆部件检测方法。在图像层面,引入 DeepLabV3 结合 3 种损失函数提取显著前景以削弱背景干扰;在特征层面,基于 YOLOv11 构建检测与分割框架,在特征提取阶段融合空间注意力金字塔池化结构以提升多尺度特征聚合能力,并设计注意力引导的显著性图模块以实现全局建模与空间增强。为验证方法有效性,构建了一个面向多部件检测任务的车辆部件数据集,并在该数据集上进行了大量实验,消融实验验证了各模块的有效性。在对比实验中,检测准确率和分割准确率分别较基线模型提升 3.5% 和 3.7%,结合可视化结果进一步表明该方法更聚焦于部件显著区域,能有效减少复杂背景引起的误检与漏检。此外,该方法在公共数据集 Car Seg 上展现出良好的泛化能力,在多个评价指标上均取得最优性能。因此,双层级显著性驱动架构通过显著前景提取和注意力引导多尺度特征聚合,显著提升了对车辆部件的检测精度,为车辆保险行业的智能定损技术提供了新的实践参考。

关键词: 智能定损;车辆部件检测;显著性检测;注意力机制;多尺度特征
中图分类号: TP391.41 TH39 **文献标识码**: A **国家标准学科分类代码**: 520.20

A dual-level saliency-driven vehicle component detection method

Zhai Yongjie, Wu Zifeng, Zhou Xunqi, Wei Letao, Wang Qianming
(*1. Department of Automation, North China Electric Power University, Baoding 071003, China; 2. Baoding Key Laboratory of Intelligent Robot Perception and Control in Electric Power System, Baoding 071003, China*)

Abstract: High-precision vehicle component detection and segmentation play a vital role in intelligent damage assessment systems by assisting in the accurate localization of damaged parts. However, challenges remain due to complex backgrounds and the performance bottlenecks of traditional detection methods constrained by single-level feature representations. To address these issues, this article proposes a dual-level saliency-driven vehicle component detection method. At the image level, DeepLabV3 is employed with a combination of three loss functions to extract salient foreground regions and suppress background interference. At the feature level, a detection and segmentation framework is formulated based on YOLOv11, where a spatial attention pyramid pooling structure is integrated during feature extraction to enhance multi-scale feature aggregation. Additionally, an attention-guided saliency map module is designed to achieve global modeling and spatial enhancement. To evaluate the effectiveness of the proposed method, a customized vehicle component dataset for multi-part detection is constructed, and extensive experiments are conducted. Ablation studies confirm the contribution of each module. In comparative experiments, the method achieves improvements of 3.5% in detection accuracy and 3.7% in segmentation accuracy over the baseline model. Visualization results further show that the proposed approach focuses more accurately on salient component regions and effectively reduces false detections and missed detections caused by complex backgrounds. Moreover, the method shows strong generalization capability on the public Car Seg dataset, achieving superior performance across multiple evaluation metrics. Overall, the dual-level saliency-driven architecture significantly enhances vehicle component detection performance through salient

收稿日期:2025-04-18 Received Date: 2025-04-18
*基金项目:国家自然科学基金(62373151,U21A20486)、河北省自然科学基金(F2023502010)、中央高校基本科研业务费专项(2024MS136)项目资助

foreground extraction and attention-guided multi-scale feature aggregation, providing new practical insights for intelligent damage assessment in the vehicle insurance industry.

Keywords: intelligent loss assessment; vehicle component detection; saliency detection; attention mechanism; multi-scale features

0　引　　言

在车辆评估与保险行业中,智能定损是一个不可或缺的环节,而车辆部件的检测与分割则是实现智能定损的关键步骤。近年来,随着车辆保有量的快速增长以及道路安全意识的不断提升,保险公司亟需提高车辆部件损伤评估与理赔处理的效率和准确性[1-5]。为更好地为车主提供高效、精准的理赔服务,同时减少定损过程中定损员的主观偏差,快速且精确的车辆部件检测与分割技术对于辅助定位损伤部件显得尤为重要。图像处理与机器学习技术相结合的车辆部件检测和分割方法[6-8]在这一过程中发挥着重要作用,这些方法能够自动识别损坏部件的位置,从而辅助保险公司进行准确的损失评估。因此,综合检测多类车辆部件的算法研究具有重要的理论意义与应用价值。

目前,车辆部件检测领域已有多种方法尝试利用部件间的空间结构关系和先验知识提升检测性能。这类方法通常通过建模相对位置、显著性区域或空间约束,增强模型对结构信息的理解。文献[9]将显著部件定义为关键点并构建三级级联卷积网络进行检测,尽管提升了关键部件的定位精度,但由于依赖固定的关键点定义,难以应对视角变化或遮挡带来的不确定性;文献[10]通过构建位置先验约束优化部件检测,但其核心依赖相对位置假设,导致在车辆姿态剧烈变化或部件被遮挡时性能明显下降;文献[11]则基于部件间的相对关系进行空间建模,然而缺乏图像层级前景建模机制,使得在尺度变化大或背景干扰严重的情况下检测效果不稳定;文献[12]提出解耦式两阶段多任务结构,结合动态注意力和双分支特征金字塔以提升特征提取能力,尽管在一定程度上增强了检测与分割能力,但对结构先验的强依赖在复杂环境下易导致误检或漏检,限制了模型泛化性能;文献[13]提出轻量级两阶段识别方法,融合检测与姿态估计、再结合判别分类,但在部件遮挡或检测误差情况下识别精度下降明显。总体而言,现有方法普遍存在对复杂背景适应性不足的问题,尤其在部件遮挡或背景干扰强烈的情况下,难以准确感知前景与背景,导致检测精度下降。

为提升模型在此类复杂场景中的感知能力,显著目标检测作为一种有效的视觉感知手段逐渐受到关注。它旨在从自然图像中检测并分割出最具辨识度的目标,已被广泛应用于复杂环境下的目标定位与感知任务中[14]。早期的显著性模型由 Itti 等[15]提出,基于颜色、强度和方向 3 种生物学特征构建了初始的显著性框架。随后,考虑到视觉注意力可能由特定目标引导,一些特定目标检测器(如人脸、车辆和人体检测器)也被引入显著性特征提取中。然而,这些传统方法主要依赖手工特征,严重依赖显著区域的间接定义,如背景先验或人类视觉规律,导致在前景与背景具有相似浅层特征时易产生混淆,泛化能力较差。

随着卷积神经网络(convolutional neural networks, CNN)[16]的发展,基于 CNN 的显著性检测方法[17-20]在融合局部与全局信息方面取得显著进展。然而,多数方法仅从特征层或图像层单一角度出发,忽视了两者间的协同作用,限制了性能进一步提升。例如,谢东升[21]结合空间与通道注意力强化车辆部件特征,但未在图像层面优化显著性信息;谢永生[22]通过简化 EFDet-SPP (efficientdet-spatial pyramid pooling)结构提取局部特征,并以多尺度卷积获取全局信息,但融合仍局限于单层面,缺乏跨尺度交互;翟永杰等[23]将改进的 Transformer 与 Swin Transformer 结合增强特征提取,但同样忽略了图像层显著性信息的融合,导致全局语义与局部细节结合不足,影响复杂背景下的检测效果。

面对多部件检测需求,针对现有车辆部件检测方法在复杂背景下适应性不足以及传统检测方法过于依赖单层次特征表征的问题,提出了一种基于双层级显著性驱动的车辆部件检测方法,通过图像层面的前景提取突出显著性目标,在特征层面使用基于空间注意力的金字塔池化结构在多个尺度上聚合特征信息,并使用注意力引导的特征显著性图生成模块自适应调整不同空间区域的特征表达,双层级协同作用形成的显著性驱动框架进一步提升检测和分割精度,从而提高对车辆部件的检测能力。

1　研究方法

双层级显著性驱动框架如图 1 所示,第 1 阶段先输入车辆部件图像,并结合数据集中的掩码信息,通过 DeepLabV3 网络联合交叉熵、Dice 与边界损失函数进行训练,实现背景去除,得到显著性图像,从而完成第 1 层级显著性建模。在第 2 阶段,原始图像和显著性图像共同输入 YOLOv11(you only look once version 11)骨干网络,构建第 2 层级显著性增强机制。

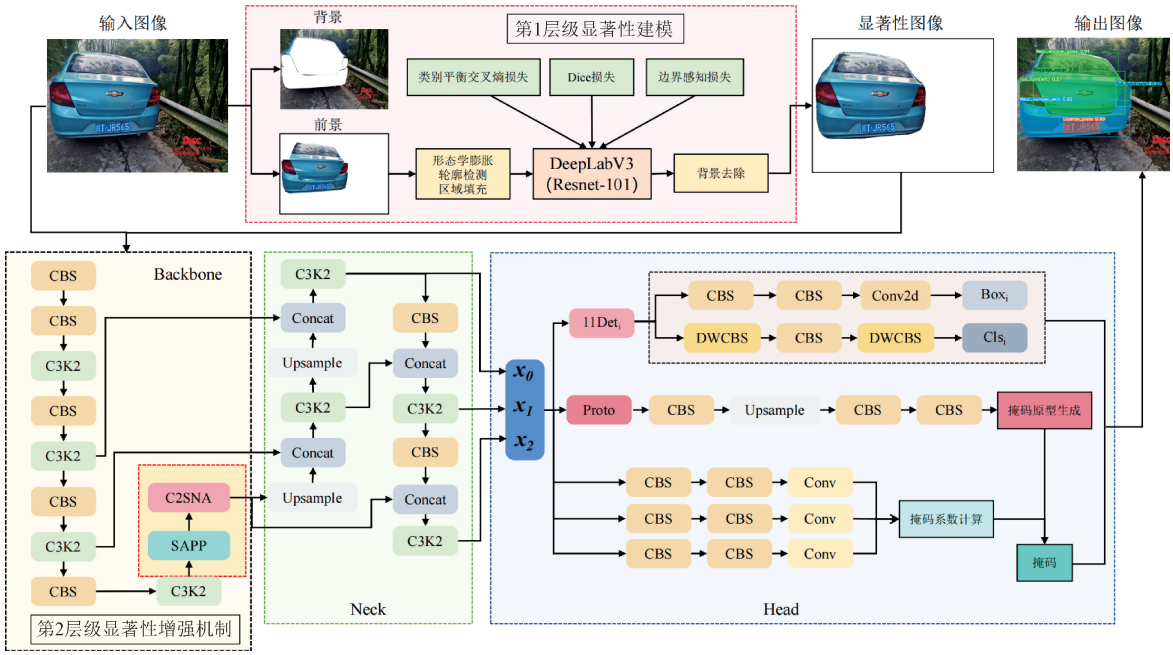


图 1 双层级显著性驱动框架

Fig. 1 Dual-level saliency-driven framework

其中,引入基于空间注意力的金字塔池化结构(spatial attention-based pyramid pooling,SAPP)提升多尺度目标感知能力,并利用注意力引导的特征显著性图生成模块(convolutional block-based saliency network with attention,C2SNA)强化特征的全局建模能力,使不同空间区域的特征表达更加充分。在 Neck 部分,通过多级卷积、上采样及特征拼接进行多尺度特征融合,以提高检测的鲁棒性。最终,在 Head 部分,采用深度可分卷积与注意力增强卷积模块进一步细化目标特征,并输出目标的检测框与分类信息,再通过掩码原型生成分支和掩码系数计算分支输出最终的目标掩码。

1.1 基于 DeepLabV3 的显著性目标获取

为解决原始数据集中车辆部件掩码存在的结构缝隙和标注不完整从而无法直接作为前景的问题,结合形态学处理与 DeepLabV3 训练实现显著性目标获取。首先应用形态学膨胀操作填补掩码中的结构缝隙,提升其连通性。接着,利用轮廓检测与区域填充方法进一步完善前景形状,从而生成初步前景图。针对掩码标注的不完整性,进一步构建了一个新的训练集,利用形态学处理后的图像作为伪标签,训练 DeepLabV3 模型,以获得更完整、准确的前景预测结果。

1)基于形态学处理的前景伪标签生成

(1)形态学膨胀操作:采用形态学膨胀操作填补掩码中的结构缝隙。使用 5×5 的矩形结构核对掩码进行膨胀处理,通过结构核 $\boldsymbol{K}$ 对图像 $\boldsymbol{I}$ 进行最大值操作,扩展目标区域,从而增强边缘信息的表达能力。其定义如式(1)所示。

$$(\boldsymbol{I} \oplus \boldsymbol{K})(x,y) = \max_{(i,j)\in \boldsymbol{K}} \boldsymbol{I}(x-i, y-j) \tag{1}$$

其中,(x,y) 是图像中的像素坐标,(i,j) 是结构核 $\boldsymbol{K}$ 中的偏移量,max 表示取最大值操作,即如果结构核覆盖的区域内存在至少一个前景像素(值为 1),则输出像素值为 1,否则为 0。

(2)轮廓检测与区域填充:为提升掩码连通性并进一步增强伪标签的完整性,在形态学膨胀操作后,引入基于轮廓检测的空洞填补策略。使用 OpenCV[24] 中的 cv2.findContours 函数,基于 Suzuki85 算法精确提取目标边界轮廓。通过设置 cv2.RETR_EXTERNAL 模式,仅保留最外层轮廓,并结合 cv2.CHAIN_APPROX_SIMPLE 方式压缩冗余点,从而提高轮廓提取效率与精度。随后,采用 cv2.fillPoly 函数对检测轮廓进行填充,不仅补全因膨胀引起的边界不连续问题,还进一步提升目标区域的整体性。该流程在确保语义一致性的同时,显著提升了伪标签的质量,为后续 DeepLabV3 模型训练提供了更加准确的监督信号。

2)多任务损失函数设计

(1)类别平衡交叉熵损失 L_{CE}:引入基于像素频率倒数的动态类别权重,缓解模型因背景占比过大而对前景关注不足的问题,从而提升前景检测能力。类别平衡交叉熵损失函数还能加快训练收敛,减小因类别不平衡带来的训练不稳定性[25]。L_{CE} 如式(2)所示。

$$L_{CE} = -\frac{1}{N}\sum_{i=1}^{N}[\beta \cdot y_i \cdot \log(p_i) + (1-\beta) \cdot (1-y_i) \cdot \log(1-p_i)] \tag{2}$$

其中,N 是像素总数,y_i 是第 i 个像素的真实标签,

p_i 是模型预测的第 i 个像素属于前景的概率,N_{fg} 和 N_{bg} 分别是前景和背景像素的数量,β 是类别权重,β 的求取公式如式(3)所示。

$$\beta = \frac{N_{bg}}{N_{fg} + N_{bg}} \tag{3}$$

(2)Dice 损失 L_{Dice}:通过增强目标区域与预测区域的重叠性,有效提升分割效果,尤其在处理小目标或边界不规则目标时表现突出。Dice 损失有助于模型更精准地学习区域特征,从而改善目标边界划分,提升分割精度[26]。L_{Dice} 如式(4)所示。

$$L_{Dice} = 1 - \frac{2\sum_{i=1}^{N} y_i p_i + \varepsilon}{\sum_{i=1}^{N} y_i + \sum_{i=1}^{N} p_i + \varepsilon} \tag{4}$$

其中,y_i 是第 i 个像素的真实标签,p_i 是模型预测的第 i 个像素属于前景的概率,ε 是一个平滑系数(取值为 1×10^{-5}),用于避免分母为 0 的情况。

(3)边界感知损失函数 $L_{Boundary}$:采用 3×3 拉普拉斯卷积核提取预测与真值边界,提升边界检测精度,并结合二元交叉熵损失强化边缘对齐。该卷积核中心值为-4、周围值为 1。边界感知损失使模型更加关注目标边界细节,有助于复杂背景下的目标定位与边界恢复。通过增强对边界信息的敏感性,边界感知损失能在分割过程中保留更清晰的边界,减少因模糊边界导致的误分割[27]。$L_{Boundary}$ 如式(5)所示。

$$L_{Boundary} = -\frac{1}{N}\sum_{i}^{N}\left[y_i^{\,b}\log(p_i^{\,b}) + (1-y_i^{\,b})\log(1-p_i^{\,b})\right] \tag{5}$$

其中,$y_i^{\,b}$ 是第 i 个像素的真实边界标签,$p_i^{\,b}$ 是模型预测的第 i 个像素属于边界的概率。

(4)通过线性组合类别平衡交叉熵损失、Dice 损失和边界感知损失,综合考虑类别平衡、区域重叠和边界对齐,从而优化模型的收敛性并提高检测精度。同时,对各损失项进行最大值标准化,确保优化过程中各部分贡献均衡,使得加权总损失函数 L_{total} 有效提升模型性能。L_{total} 如式(6) 所示。

$$L_{total} = L_{CE} + L_{Dice} + L_{Boundary} \tag{6}$$

3)基于 DeepLabV3 的显著性目标提取

显著性目标提取在 PyTorch 框架下实现,选用 torchvision 中预训练的 DeepLabV3 模型[28]作为基础架构。该模型引入空洞卷积与空间金字塔池化模块,能够融合多尺度上下文信息,在复杂背景下表现出更强的边缘保持与区域感知能力,显著优于传统的全卷积神经网络(fully convolutional networks,FCN)方法,尤其适用于前景显著性区域提取任务[29]。

为了更好地适应前景/背景二分类任务,对 DeepLabV3 的分类头进行了调整,将输出通道数设置为 2,确保其能够精准区分前景和背景。首先将输入图像调整为 512×512 的分辨率,并通过 DeepLabV3 模型获取初步分割掩码。然后,采用双线性插值将分割结果恢复至原始图像尺寸,并通过形态学闭运算去除小噪点。接着,将预测掩码二值化,并与原始图像的 RGBA 数据融合。在融合过程中,非目标区域的 Alpha 通道设为 0,确保仅保留前景区域的颜色信息,从而实现清晰的前景提取与背景去除。

如图 2 所示,显著性目标的获取过程得以直观展示。第 1 列图 2(a)为原始图像,第 2 列图 2(b)为基于掩码生成的前景图,第 3 列图 2(c)展示了经过形态学处理及轮廓提取后的前景图,第 4 列图 2(d)则为 DeepLabV3 模型预测得到的显著性前景图。

(a) 原图　(b) 基于掩码生成图　(c) 形态学处理及轮廓提取图　(d) DeepLabV3预测图
(a) Original　(b) Mask-based foreground　(c) Morphology-contour result　(d) DeepLabV3 prediction

图 2　显著性目标获取过程可视化

Fig. 2　Significant object extraction process visualization

从图 2 可以看出,形态学与轮廓处理有效增强了目标区域的连通性,而 DeepLabV3 的训练预测进一步修复了掩码标注不全的问题,得到更完整的前景区域。这种结合形态学处理与深度学习模型的显著性目标获取方法,能够有效区分显著目标(车辆部件)与背景,从而有效应对复杂背景干扰问题。该方法为后续模型训练与评估提供了可靠的数据支持,并为车辆部件的精确检测与分割任务奠定了坚实的基础。

1.2 基于空间注意力的金字塔池化结构

本研究提出了一种基于空间注意力的金字塔池化(SAPP)结构,旨在通过注意力机制增强多尺度特征的表征能力。该模块在标准空间金字塔快速池化结构的基础上引入了级联空间注意力机制,其核心结构如图 3 所示。

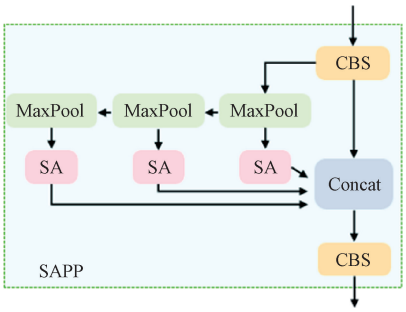


图 3 基于空间注意力的金字塔池化结构

Fig. 3 Spatial attention-based pyramid pooling

具体实现包括 3 个阶段:

1)通过 1×1 卷积将输入通道数压缩为原始值的 1/2,减少计算复杂度;

2)采用三级串行最大池化(max pooling,MaxPool)(核尺寸 5×5,步长 1,填充 2),逐步扩展感受野,高效捕获多尺度信息;

3)对每级池化输出施加空间注意力(spatial attention,SA)加权,动态调整各空间位置的贡献度。最终将多尺度特征沿通道维度拼接,并通过 1×1 卷积恢复通道维度,实现特征重构。

空间注意力结构如图 4 所示,采用双路特征聚合机制,通过通道维度的均值与最大值运算捕获空间显著性信息。

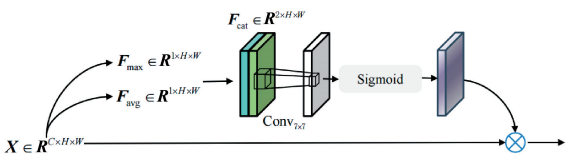


图 4 空间注意力结构

Fig. 4 Diagram of spatial attention architecture

具体而言,给定输入特征图 $\boldsymbol{X} \in \boldsymbol{R}^{C\times H\times W}$,首先沿通道维度分别计算均值特征 $\boldsymbol{F}_{avg} \in \boldsymbol{R}^{1\times H\times W}$ 和最大值特,征 $\boldsymbol{F}_{max} \in \boldsymbol{R}^{1\times H\times W}$,实现跨通道信息压缩。随后,将双路特征在通道维度拼接为 $\boldsymbol{F}_{cat} \in \boldsymbol{R}^{2\times H\times W}$,并通过 7 × 7 卷积核进行空间上下文建模,生成空间注意力图 $\boldsymbol{A}_s \in \boldsymbol{R}^{1\times H\times W}$。该过程可形式化为:

$$\boldsymbol{A}_s = \sigma(\mathrm{Conv}_{7\times7}([\boldsymbol{F}_{avg};\boldsymbol{F}_{max}])) \tag{7}$$

其中,σ 表示激活函数 Sigmoid,$\mathrm{Conv}_{7\times7}$ 卷积核尺寸 7×7、填充 3 的单通道卷积操作。最终,通过逐元素乘法将注意力权重与原始特征融合,强化关键区域特征响应。

1.3 注意力引导特征显著性图生成模块

为了增强模型对特征图关键区域的关注能力,并实现显著性信息与注意力特征的自适应融合,提出了注意力引导特征显著性图生成模块(C2SNA),如图 5 所示。

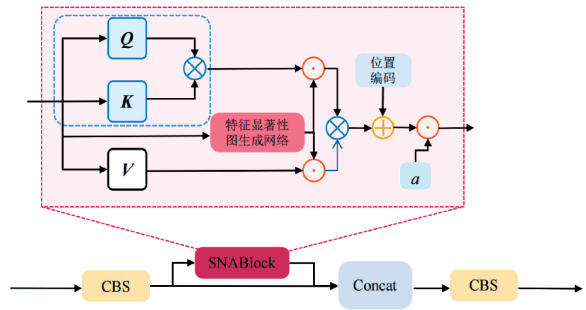


图 5 注意力引导特征显著性图生成模块

Fig. 5 Convolutional block-based saliency network with attention

注意力显著性网络块(saliency network with attention block,SNABlock)核心在于通过注意力机制引导特征显著性图的生成,使模型在特征提取过程中更加关注图像中的关键区域,并通过显著性调节因子 α 实现双层级注意力特征与原始特征的动态融合,从而增强不同空间区域的特征表达。

1)特征显著性图生成网络

如图 6 所示,设计的特征显著性图生成网络通过多层卷积操作提取图像的局部特征,并结合双层级注意力机制优化特征提取过程,从而生成具有空间自适应的显著性图。其中每层卷积后接批量归一化(batch normalization,BN)和激活函数 ReLU,构成 CBR 块。前 3 层卷积用于逐步提取图像的多层次特征,第 4 层卷积则生成单通道的显著性图。

为了保留全局信息,网络在输出端引入残差连接结构,将输入图像经 1×1 卷积变换后与显著性图相加,最终通过 Sigmoid 函数将显著性图归一化到[0,1]范围内。

在第 3 层卷积后,引入了双层级注意力机制,包括:

(1)多头可变形注意力机制(multi-head deformable attention,MHDA)融合了多头注意力结构与可变形注意

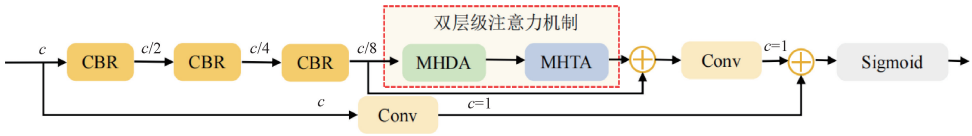


图 6　特征显著性图生成网络

Fig. 6　Feature saliency map generation network

力模块,能够通过动态采样点更灵活地建模图像中关键区域之间的关系。在该机制中,首先将输入特征映射为多组查询(Query)、键(Key)和值(Value),随后,通过偏置生成网络对每个注意力头提取的值 $\boldsymbol{V}$ 特征进行卷积操作,得到对应的二维空间偏移量。该偏移量与基础采样网格相加,以生成动态采样点的坐标,并利用双线性插值方法,从值特征中提取稀疏且关键的特征信息。提取结果作为更新后的值 $\boldsymbol{V}$ 特征,参与局部注意力的点积计算。最终,各注意力头的输出经拼接与线性变换融合,得到局部注意力模块的输出特征 MHDA(X),如图 7 所示。

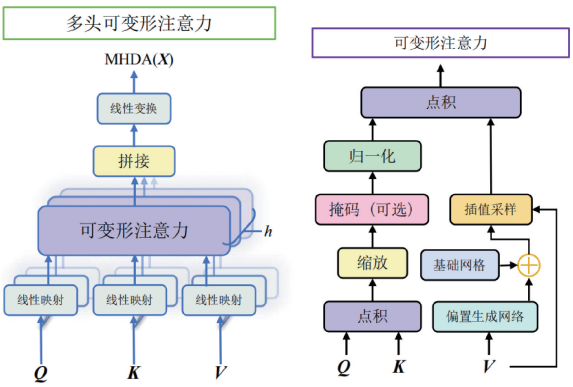


图 7　多头可变形注意力

Fig. 7　Multi-head deformable attention

(2)多头可调注意力机制(multi-head tunable attention,MHTA)结合了多头自注意力和可学习的显著性调节因子 f_t,通过动态捕捉全局依赖关系并自适应地增强关键特征,进一步在全局范围内优化局部特征的分布,从而提升模型的特征表达能力,其整体结构如图 8 所示。

在 MHTA 模块里,首先将来自局部注意力模块输出的注意力特征 MHDA(X)进行展平操作,将其空间维度重塑为一维序列,以适配区域级注意力的计算需求。随后,针对展平后的特征分别施加线性变换,生成区域级的查询(Query)和键(Key),用于刻画图像不同区域间的全局依赖关系。接着,依据区域级查询与键之间的注意力权重,对原始特征序列进行加权汇聚,得到区域级注意力表示。最终,将该表示与初始输入进行残差连接,并引入可学习的显著性调节因子 β 对融合结果进行动态缩放,输出最终的注意力增强特征。

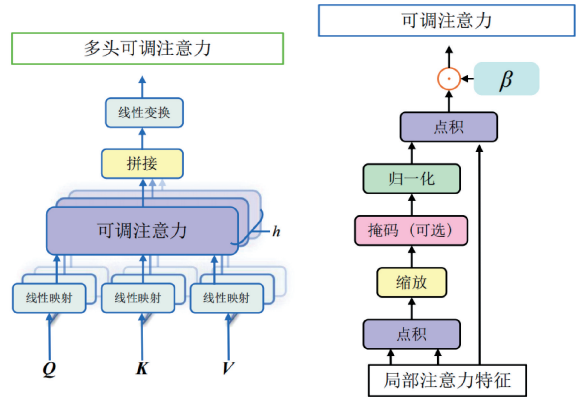


图 8　多头可调注意力

Fig. 8　Multi-head tunable attention

双层级注意力机制先由 MHDA 来实现局部自注意力,通过动态调整采样点的位置,捕捉局部区域的细节信息;再由 MHTA 来实现区域级注意力,在全局范围内重新分配局部注意力资源,平衡不同区域的关注程度。该双层级注意力机制实现了从局部细节捕捉到全局上下文建模的渐进融合。

2)带显著性调节因子 f_t 的特征融合机制

为实现双层级注意力特征与原始特征之间更具适应性的动态融合机制,通过引入显著性调节因子 f_t 以实现对模块输出的动态调整,其实现过程如图 9 所示。

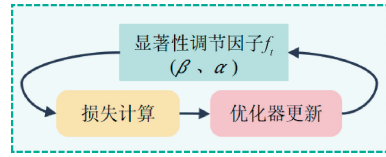


图 9　显著性调节因子 f_t 流程

Fig. 9　Flowchart of the saliency modulation factor f_t

其中,调节因子 β 负责调控 MHTA 模块的输出特征权重,在区域级注意力融合过程中对关键特征实施动态加权优化;调节因子 α 则作用于 SNABlock 的输出特征,通过特征缩放机制动态调节最终生成的特征显著性图在不同空间位置的响应强度分布。这种双因子协同调控的机制有效提升了特征融合过程的自适应能力,使得关键区域的特征表达在不同层级均能获得最优的显著性增强。

首先,在参数初始化阶段,利用 torch. tensor 函数创建一个具有预设初始值且支持梯度计算的张量 f_t,并通过 nn. Parameter 将其封装为可学习参数,使其能够在训练过程中通过优化器进行更新。

其次,在前向传播过程中,显著性调节因子 f_t 与模块输出特征进行逐元素相乘,从而实现对显著性特征的动态缩放。

最后,在训练过程中,显著性调节因子 f_t 作为模型参数的一部分,在每次损失计算时自动计算其梯度。基于计算得到的梯度 $\eta \nabla f_t L$ 以及预设的学习率 η,采用随机梯度下降(stochastic gradient descent,SGD)优化算法对 f_t 进行迭代更新,其实现如式(8)所示。更新后的 f_t 重新输入模块中,最终实现带显著性调节因子的特征融合机制,从而进一步提升模块的性能。

$$f_t = f_t - \eta \nabla f_t L \tag{8}$$

2 实验设置

2.1 数据集来源

实验采用了自建车辆部件数据集 VehicleParts59(VP59)和公开基准数据集 Car Seg 两类数据集,图 10 以可视化方式展示了这两个数据集中不同车辆部件类别的掩码实例数量分布。自建数据集 VP59 涵盖的车辆部件类别更加全面,且实例数量较多;两个数据集均呈现长尾分布的特点。

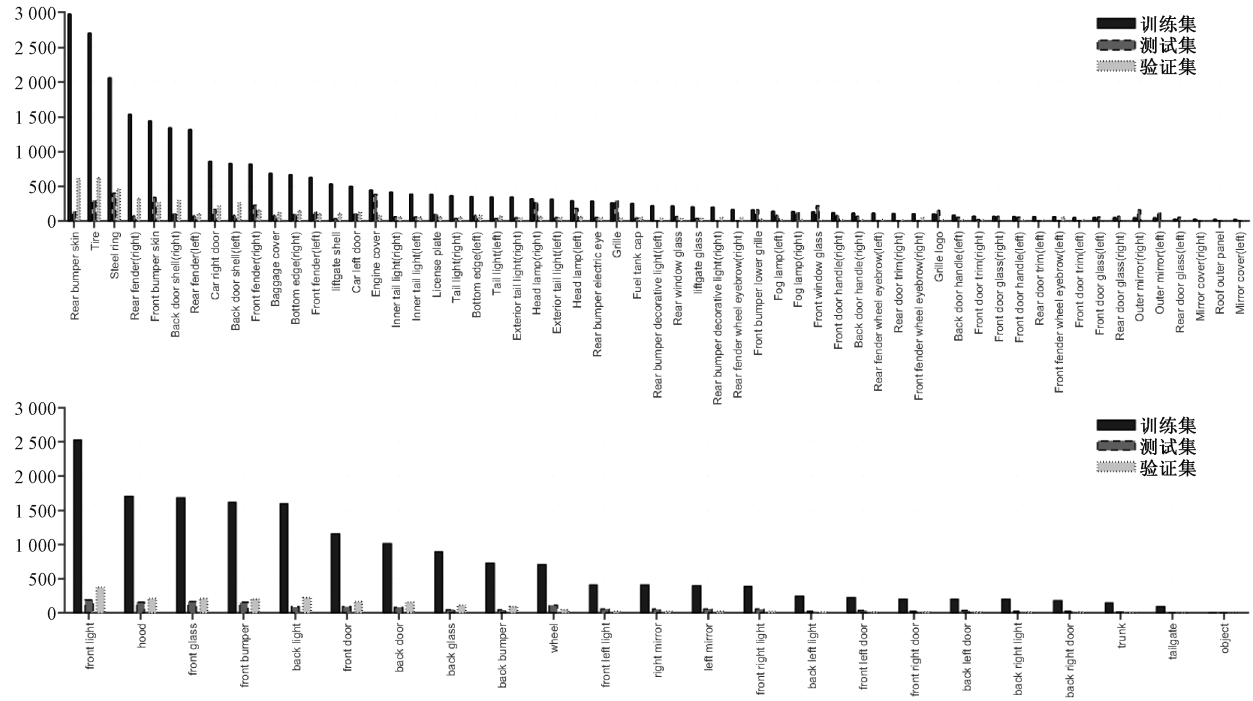


图 10 VP59 数据集(上)和 Car Seg 数据集(下)的不同类别掩码实例数量分布

Fig. 10 Distribution chart of the number of different category mask instances in the VP59 dataset (top) and the Car Seg (bottom)

VP59 数据集来源于某保险公司车辆智能定损系统的真实案例照片,涵盖 59 类车辆部件类别,全面覆盖车辆外部部件系统,具体类别如表 1 所示。按照 COCO 数据集的构建规范,VP59 数据集中的 8 965 张图像被划分为训练集 6 694 张、验证集 1 674 张和测试集 597 张。目前,该数据集尚未公开。

为进一步验证方法的泛化性能,本研究还采用了 Roboflow Universe 平台上的公开数据集 Car Seg 进行对比实验。该数据集由 Gianmarco Russo 创建,包含 3 833 张标注图像,涵盖前/后保险杠、车门、车灯等 23 类常见车辆部件,并依据官方划分方案分为训练集 3 156 张、验证集 401 张和测试集 276 张。作为车辆实例分割领域的标准测试平台,Car Seg 数据集为本研究提供了可靠的性能对比基准。(数据集访问链接为:https://universe. roboflow. com/gianmarco-russo-vt9xr/car-seg-un1pm。)

2.2 评价指标

为了验证双层级显著性驱动的车辆部件检测方法的有效性,使用目前目标检测和实例分割模型中常用的评价指标平均精度均值(mean average precision, mAP)对模型进行总体的评估,平均精度均值计算为:

表 1　车辆部件数据集类别情况

Table 1　Category of vehicle components dataset

序号	名称	序号	名称	序号	名称
1	钢圈	21	外尾灯(左)	41	前保险杠皮
2	轮胎	22	外尾灯(右)	42	后保险杠皮
3	车牌	23	内尾灯(左)	43	前风挡玻璃
4	中网	24	内尾灯(右)	44	后风挡玻璃
5	油箱盖	25	前雾灯(左)	45	倒车镜护盖(左)
6	尾灯(左)	26	前雾灯(右)	46	倒车镜护盖(右)
7	尾灯(右)	27	前大灯(左)	47	后保险杠装饰灯(左)
8	车顶外板	28	前大灯(右)	48	后保险杠装饰灯(右)
9	行李箱盖	29	前门饰条(左)	49	后门外拉手(左)
10	中网徽标	30	前门饰条(右)	50	后门外拉手(右)
11	发动机罩	31	后门饰条(左)	51	后叶子板轮眉(左)
12	举升门壳	32	后门饰条(右)	52	后叶子板轮眉(右)
13	倒车镜(左)	33	前门玻璃(左)	53	前门外拉手(左)
14	倒车镜(右)	34	前门玻璃(右)	54	前门外拉手(右)
15	底大边(左)	35	后门玻璃(左)	55	前叶子板轮眉(左)
16	底大边(右)	36	后门玻璃(右)	56	前叶子板轮眉(右)
17	后门壳(左)	37	前叶子板(左)	57	举升门玻璃
18	后门壳(右)	38	前叶子板(右)	58	后保险杠电眼
19	前门壳(左)	39	后叶子板(左)	59	前保险杠下格栅
20	前门壳(右)	40	后叶子板(右)	—	

$$R_n = \frac{TP}{TP + FN} \tag{9}$$

$$P_n = \frac{TP}{TP + FP} \tag{10}$$

$$mAP = \frac{1}{c}\sum_{n=1}^{C}\int_0^1 P_n(R_n)\,\mathrm{d}R_n \tag{11}$$

其中,C 是车辆部件的类别数量,TP 表示真阳性的数量,FN 表示假阴性的数量,FP 表示假阳性的数量,Rn 表示类别 n 的召回率,$Pn(Rn)$ 是当类别 n 的召回率是 Rn 时对应的类别 n 的精度。对召回率-精度曲线下的面积进行积分会得到单个类别的 AP,mAP 是所有类别 AP 值的平均值,用于定量评估模型的训练效果,能反映出训练模型的性能。在检测中,各类参数对应为 Bbox 的 IoU,得到检测平均精度 $\mathrm{Det\_mAP}^{50}$;在分割中,各类参数对应于像素级 Segm 的 IoU,得到分割平均精度 $\mathrm{Seg\_mAP}^{50}$。

2.3　实验环境与训练参数

所述模型使用的设备与版本号如表 2 所示。模型的训练和测试均在 NVIDIA 3090Ti 专业加速卡上进行。操作系统为 Ubuntu 24.04.1 LTS,采用 CUDA 11.3 加速训练。计算机语言为 Python 3.8.12,网络开发框架为 PyTorch 1.11.0。输入模型的图片分辨率为 640×640,选择了随机梯度下降算法,并将学习率衰减系数为 0.000 5,初始学习率设置为 0.01,学习动量为 0.937,批次大小为 16,迭代次数设为 300。

表 2　设备与版本号

Table 2　Hardware and version numbers

设备	版本号
显卡	NVIDIA 3090Ti
操作系统	Ubuntu 24.04.1 LTS
编程环境	Python3.8.12、Pytorch1.11.0

3　实验结果与分析

3.1　消融实验

为了证明模型中每个模块的有效性,设计了 2 组消

融实验:1)证明图像层面显著性驱动的有效性:在基线模型的基础上,通过加入 DeepLabV3 获取的显著性图片进行检测;2)证明特征层面显著性驱动的有效性:在基线模型的基础上,分别使用基于空间注意力的金字塔池化结构 SAPP 和注意力引导特征显著性图生成模块 C2SNA 进行检测。最终检测和分割结果如表 3 所示。

表 3 模块的消融实验结果

Table 3 Ablation test results of module

方法	前景获取	SAPP	C2SNA	Det_mAP^{50}/%	Seg_mAP^{50}/%
基线模型				52.0	51.5
模型 1	√			54.1	53.1
模型 2		√		52.7	52.2
模型 3			√	53.2	52.7
模型 4		√	√	53.6	53.1
所提模型	√	√	√	55.5	55.2

表 3 展示了不同模块组合下的消融实验结果,以验证各模块对模型性能的贡献。基线模型在 Det_mAP^{50} 和 Seg_mAP^{50} 上的性能分别为 52.0% 和 51.5%。在此基础上,模型 1 仅引入显著性前景获取模块,使检测和分割性能分别提升至 54.1% 和 53.1%,表明显著性前景信息能够有效增强前景目标区域,从而提高检测和分割精度。模型 2 单独加入 SAPP 结构后,检测与分割性能分别为 52.7% 和 52.2%,相较于基线模型略有提升。模型 3 仅添加 C2SNA 模块,检测与分割性能分别达到 53.2% 和 52.7%,相较于基线模型有所提升,表明 C2SNA 能够有效引导特征显著性,提高检测性能。模型 4 同时结合 SAPP 和 C2SNA,检测和分割性能分别达到 53.6% 和 53.1%,表明两者的结合在一定程度上能够提升特征表达能力,证明了特征层面显著性驱动的有效性。最终,整合全部模块后的完整模型在检测和分割任务上分别取得 55.5% 和 55.2% 的最佳性能,相较于基线模型在检测和分割上分别提升了 3.5% 和 3.7%,充分验证了所提方法在目标检测与分割任务中的有效性,各模块间的协同作用进一步提升了模型的整体性能。

在验证图像层面显著性驱动方法的有效性时,针对显著性目标获取方法进行了 4 组消融实验,实验结果如表 4 所示。

实验中所使用的模型来自 torchvision.models 的分割模型及其特征提取网络。实验结果表明,基于 DeepLabV3 架构的模型在检测和分割任务上的精度显著优于 FCN。此外,采用 ResNet101(residual networks 101)作为特征提取网络的性能也明显优于 ResNet50。

表 4 显著性目标获取的消融实验结果

Table 4 Ablation experiment results of saliency object acquisition (%)

方法	Det_mAP^{50}	Seg_mAP^{50}
FCN-ResNet50	53.7	53.0
FCN-ResNet101	54.1	53.5
DeepLabV3-ResNet50	54.5	53.9
DeepLabV3-ResNet101	55.5	55.2

在验证特征层面显著性驱动方法的有效性时,针对显著性调节因子 β 和 α 进行了 4 组消融实验,结果如表 5 所示。

表 5 显著性调节因子 f_t 的消融实验结果

Table 5 Ablation experiment results of the saliency modulation factor f_t

模型	β	α	Det_mAP^{50}/%	Seg_mAP^{50}/%
1			54.1	53.5
2	√		54.4	53.9
3		√	54.5	53.9
4	√	√	55.5	55.2

模型 2 通过在多头可变形注意力中加入显著性调节因子,能够更好地调整注意力权重,提升了检测和分割精度,相较于模型 1 有所改进。模型 3 通过调整 SNABlock 中主路和支路的信息交互,优化了特征融合,进一步提升了精度。模型 4 通过 β 初步调节区域级注意力输出和 α 能有效引导特征显著性图中不同区域间的特征融合与信息交互,实现了最优的检测和分割精度。

显著性调节因子在训练过程中的动态变化趋势如图 11 所示,图 11 中,x1 代表调节因子 β,x2 代表调节因子 α。

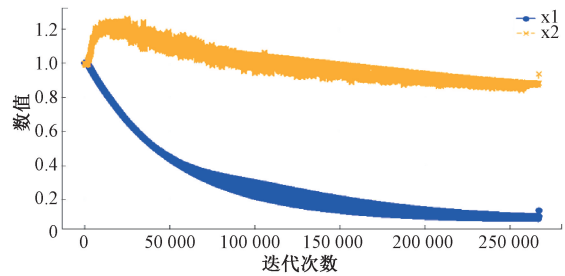


图 11 显著性调节因子训练迭代图

Fig. 11 The training iteration graph of the saliency modulation factor

β 在训练初期较大,并随着迭代次数的增加逐步下降,最终趋于稳定的较低值(接近 0.1),表明模型在初

期依赖较强的显著性调节进行特征选择,而在后期随着网络学习到稳定的特征表示,该依赖性逐渐减弱。相比之下,α 先略有上升并迅速超过 1,随后呈现缓慢下降趋势,但整体仍维持在较高水平(接近 0.9),反映出特征融合机制在整个训练过程中持续保持较高的显著性权重,以确保不同特征之间的信息交互。两者的变化趋势表明,模型在不同阶段对显著性调节的依赖程度存在差异,多头可变形注意力在训练后期降低了对特征显著性的强调,而特征融合机制仍保持较高的显著性调节能力,以强化特征交互,从而在特征选择与融合之间实现动态平衡。这一变化趋势验证了显著性驱动方法的有效性,表明所提出方法能够自适应地调整显著性权重,从而优化特征学习过程,提高模型的泛化能力。

3.2　模型检测结果可视化对比实验

为了更直观地展示不同类别的性能差异,将基线模型与所提模型的 59 类车辆部件检测的 mAP^{50} 结果通过图 12 表示。

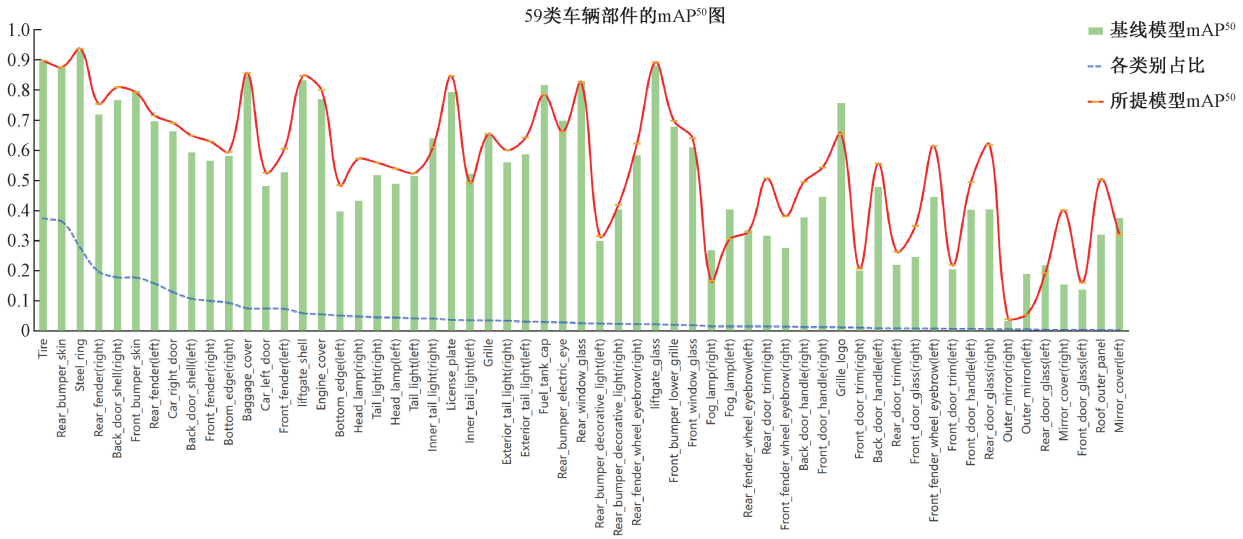


图 12　59 类车辆部件的 mAP^{50} 图

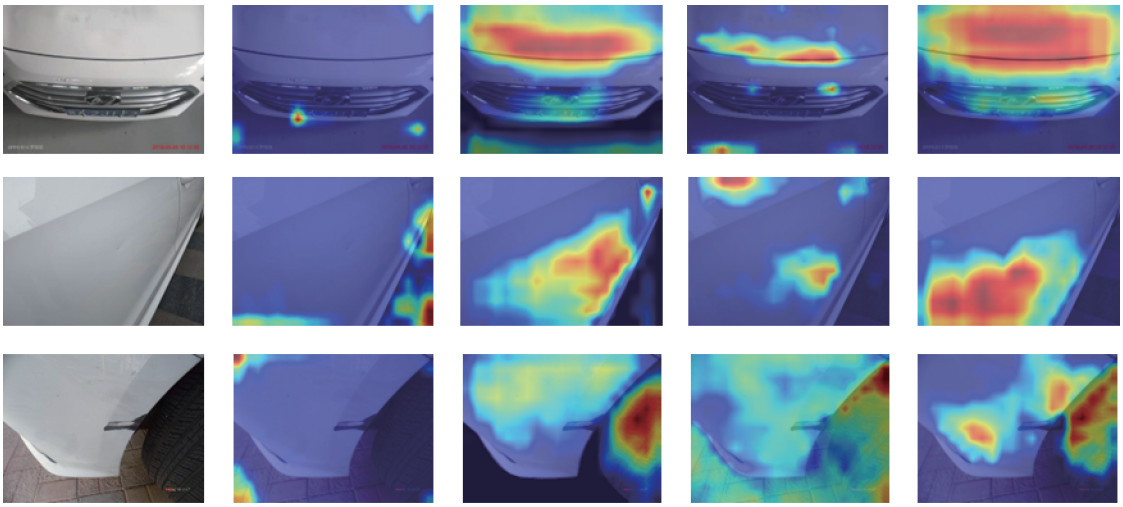
Fig. 12　mAP^{50} chart for 59 vehicle components

图 12 中采用柱状表示基线模型的 mAP^{50} 结果,实曲线表示所提模型的 mAP^{50} 结果,虚曲线则表示各类别在数据集中的占比(按降序排列)。从结果可以看出,所提模型在 59 个车辆部件类别中,有 44 个类别的 mAP^{50} 评分高于基线模型,表明所提出的方法在大多数类别上具备更强的检测能力。进一步分析发现,在数据集中占比相对较高的前 15 个类别中,所提模型在 14 个类别上取得了优于基线模型的性能,提升效果尤为显著。而在占比较小的后 15 个类别中,尽管整体提升幅度较小,所提模型仍在 10 个类别上超越了基线模型,表现出较强的竞争力。这些结果充分验证了所提模型在不同类别上的适应性和,尤其在主要部件的检测任务中展现出了更优的性能。

本研究采用热力图可视化技术对双层级显著性驱动模型进行对比分析,通过视觉注意力分布验证模型在车辆部件检测任务中的性能优势。如图 13 所示,可视化对比实验包含 5 个关键组别:图 13(a)为原始输入图像作为基准参照;图 13(b)为基线模型热力图显示,虽然存在局部激活响应,但注意力分布呈现离散化特征,缺乏对目标部件的精确定位能力;图 13(c)为背景去除后的前景热力图,通过空间约束显著提升了目标结构的关注度,其热力图基本关注在前景区域;图 13(d)为经特征显著性驱动的热力图,相比基线模型,能够更精准地聚焦于目标部件区域,但存在背景信息的特征干扰。图 13(e)为提出的双层级显著性驱动模型热力图,通过双层级显著性驱动引导,不仅剔除了背景特征干扰,同时在前景区域聚焦于关键性区域部件。

为了定性地分析所提模型的检测效果,图 14 显示了所提模型与基线模型进行分析对比,其中图 14(a)~(d)为基线模型检测结果,图 14(e)~(h)为所提模型检测结果。图 14(a)~(e)基线模型在前叶子板和前大灯的检测中出现左右区分错误,而所提模型能够准确识别并正确区分左右部件。

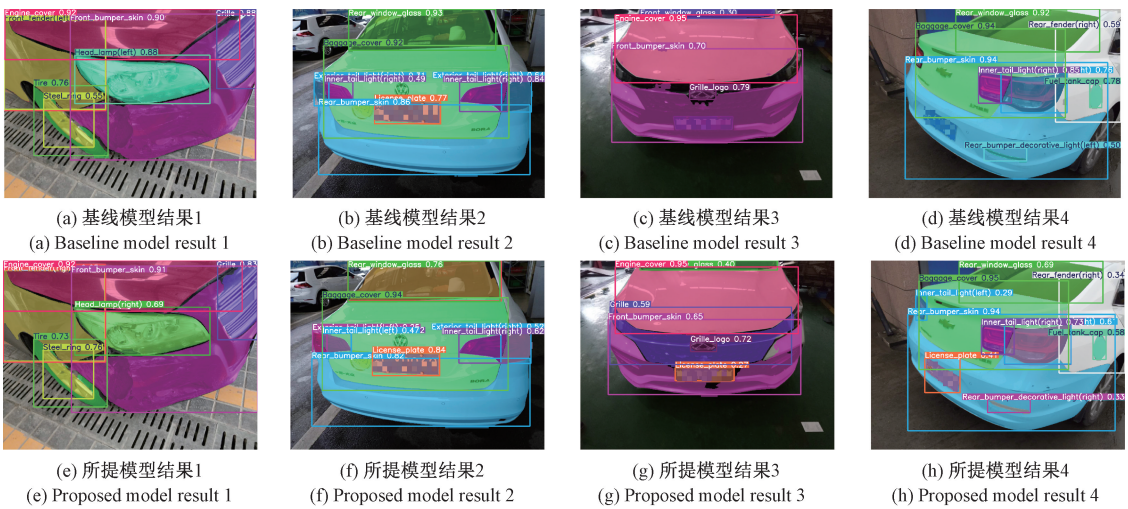
在图 14(b)、(f)中基线模型同样在内外尾灯的检测中存在左右区分错误,导致部件错检,而所提模型有效减少了此类误检情况,提高了检测精度。图 14(c)、(g)中基线模型未能成功检测出车牌和中网,而所提模型能够准确识别这两个部件,表现出更强的检测能力,尤其在关键部件的识别上更为稳定。图 14(d)、(h)中基线模型未



(a) 原图 (a) Original image　(b) 原热力图 (b) Original heatmap　(c) 前景热力图 (c) Prospect heatmap　(d) 特征热力图 (d) Feature heatmap　(e) 双层级热力图 (e) Dual level heatmap

图 13　多层级热力图对比可视化

Fig. 13　Multi-level heatmap comparison visualization



(a) 基线模型结果1 (a) Baseline model result 1　(b) 基线模型结果2 (b) Baseline model result 2　(c) 基线模型结果3 (c) Baseline model result 3　(d) 基线模型结果4 (d) Baseline model result 4

(e) 所提模型结果1 (e) Proposed model result 1　(f) 所提模型结果2 (f) Proposed model result 2　(g) 所提模型结果3 (g) Proposed model result 3　(h) 所提模型结果4 (h) Proposed model result 4

图 14　车辆部件检测定性结果对比

Fig. 14　Qualitative comparison of vehicle component detection result

能检测出车牌和内尾灯,并且错误地将右侧的后保险杠装饰灯识别为左侧的后保险杠装饰灯。而相比之下,所提模型能够更有效地克服部件的漏检和错检问题,使检测结果更加准确和可靠。

3.3　与先进模型对比实验

为全面验证所提模型的性能优势,本研究在 VP59 车辆部件数据集上进行了广泛的对比实验。对比实验设计包含 2 个层面的比较:1)经典通用检测模型,具体包括 Pointrend[30] (point-based rendering)、Mask R-CNN[31] (region-based convolutional neural network)、Cascade Mask R-CNN[32]、Mask Scoring R-CNN[33]、Condinst[34] (conditional convolutions for instance segmentation)、YOLOv8[35]、YOLOv9[36]、YOLOv11[37] 进行对比;2)车辆部件检测领域的针对性模型,具体包括文献[9]提出的 MTFDN (multi-task feature decoupling network) 和文献[28]提出的融合改进 Transformer 的车辆部件检测方法 (enhanced with transformer integration, Transformer Integration)进行对比,对比结果如表 6 所示。

根据表 6 的实验结果,可以看出所提模型在目标检测和实例分割任务上均取得了显著优势。在检测性能 (Det_mAP^{50})方面,所提模型达到了 55.5%,相比于当前最优的 YOLOv11(52.0%),提升了 3.5 个百分点,相较于

表 6　与先进分割算法性能对比

Table 6　Performance comparison with advanced segmentation algorithms　(%)

方法	Det_mAP^{50}	Seg_mAP^{50}
Pointrend	37.5	36.8
Mask R-CNN	38.0	36.5
Cascade Mask R-CNN	38.3	37.1
Mask Scoring R-CNN	38.6	38.0
Condinst	40.8	39.9
Transformer Integration	41.6	40.5
MTFDN	46.8	45.7
YOLOv8	48.4	47.9
YOLOv9	51.9	51.6
YOLOv11	52.0	51.5
所提模型	55.5	55.2

经典的 Mask R-CNN(38.0%)提高了 17.5 个百分点。此外,所提模型还超越了文献[9]的 MTFDN(46.8%)和文献[28]的 Transformer Integration(41.6%),表现出更强的检测能力。在实例分割性能(Seg_mAP^{50})方面,所提模型达到了 55.2%,同样超过了所有对比方法,其中相比于表现最优的 YOLOv11(51.5%),提升了 3.7 个百分点。此外,所提模型也在实例分割任务中超越了 MTFDN(45.7%)和 Transformer Integration(40.5%)。所提模型在目标检测与实例分割任务上不仅显著超越了 YOLO 系列、R-CNN 系列等主流通用模型,同时也优于文献[9,28]针对车辆部件检测专门设计的模型,展现出更强的泛化性能和领域适应性。这表明由于双层级显著性驱动的引入,模型能够应对多部件检测的问题,克服部件错检、漏检的问题,从而提升总体的检测效果。

3.4　公共数据集对比实验

为进一步验证所提方法的泛化性,本研究在公开数据集 Car Seg 上与现有车辆部件检测领域常用的方法进行了对比试验。所选择的对比方法涵盖了近年来在车辆部件分类、检测和分割领域表现优异的代表性模型,这些模型已在多个车辆相关实际应用场景中得到验证[38-45]。具体包括特征提取网络为 ResNet50、ResNet101、GCNet(global context network)和 ConvNext(convolutional neural network next)的 Mask R-CNN 系列[38-41],还有 Cascade Mask R-CNN[42]、HTC[40](hybrid task cascade)、YOLACT[43](you only look at coefficients)以及多种 YOLO 系列模型[44-45]等。从表 7 可以看出,在 Car Seg 数据集上的实验结果相比,所提模型在多个评价指标上均取得了最优性能,这一结果说明,所设计的双层级显著性驱动架构有更为准确和鲁棒的检测与分割性能。

表 7　公共数据集上与现有车辆部件检测常用方法对比

Table 7　Comparison with existing common vehicle component detection methods on public datasets　(%)

方法		评价指标							
		Box(P)	Box(R)	Det_mAP^{50}	Det_mAP^{50-95}	Mask(P)	Mask(R)	Seg_mAP^{50}	Seg_mAP^{50-95}
Mask R-CNN	ResNet50			69.2	45.9			68.6	47.8
	ResNet101			72.3	50.5			72.2	52.3
	GCNet			72.3	52.5			71.6	53.0
	ConvNext			75.1	51.3			74.6	54.2
Cascade Mask R-CNN				72.6	56.7			72.5	55.5
HTC				69.9	55.7			70.0	53.1
YOLACT				66.6	44.3			66.7	49.7
YOLOv5		65.6	76.0	69.2	55.4	66.8	77.5	70.7	55.9
YOLOv8		62.0	78.2	70.5	60.4	62.8	79.3	71.8	58.7
YOLOv9		61.6	77.5	71.1	60.5	62.7	78.5	72.7	59.3
YOLOv11		60.6	78.3	70.5	62.2	61.5	79.6	71.8	60.2
所提模型		66.5	79.5	76.1	67.1	67.8	80.4	77.6	64.7

3.5　实际测试图片及分析

为了验证所提出方法在车辆部件检测实际应用中的有效性,对车辆智能定损系统中车辆部件检测的测试集图片进行了可视化展示和对比分析。对比了 Car Seg 数据集(23 类部件)和 VP59 数据集(59 类部件)所训练的基线模型和最佳模型在相同测试集图片上的检测效果。

如图 15 所示,基于 VP59 数据集训练的模型在车辆部件检测上表现更为出色,覆盖了更多车辆部件类型,

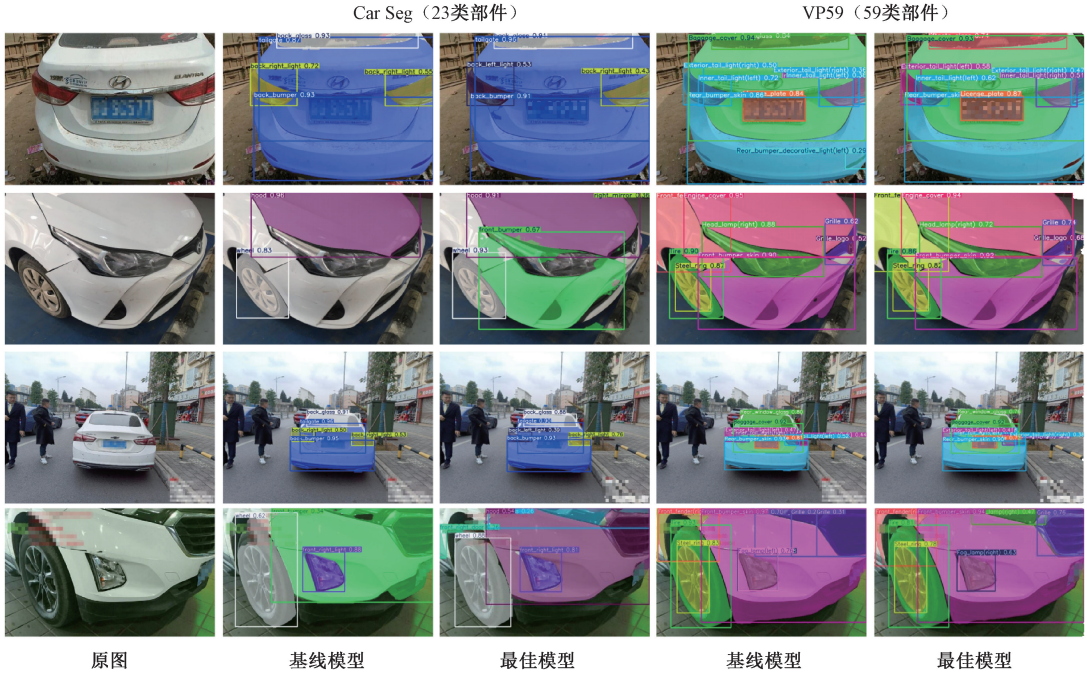


图 15 实际测试图片检测结果对比

Fig. 15 Comparison of detection results on actual test images

适应性更强。这表明 VP59 数据集在实际车辆部件检测中更具优势,能够全面覆盖不同场景中的部件类型,从而提升检测精度和可靠性。

通过对比基线模型和最佳模型的测试结果可以发现,基线模型在复杂场景中存在漏检和误检问题,尤其是在区分左右部件时效果较差。而最佳模型通过引入双层级显著性驱动架构,显著提升了目标部件的识别能力,尤其是在区分细微差异的部件时表现更为出色,有效减少了误检并增强了在各种场景下的鲁棒性。

4 结 论

针对现有车辆部件检测方法在复杂背景下适应性不足以及传统显著性检测对单层次特征表征的依赖问题,提出了一种基于双层级显著性驱动的车辆部件检测方法。首先,利用 DeepLabV3 结合多种损失函数去除背景,生成显著性图像。随后,将原始图像和显著性图像共同输入引入了基于空间注意力的金字塔池化结构的 YOLOv11 骨干网络中,并利用注意力引导的特征显著性图生成模块实现双层级协同作用的显著性驱动框架。实验结果表明,该方法有效降低了部件漏检和错检率,在多类别车辆部件数据集上的检测性能优于现有先进模型。双层级显著性驱动框架显著提升了检测与分割精度,为车辆保险行业的智能定损技术提供了新的思路。

下一步工作可聚焦于显著性图像与显著性特征之间的相互作用,探索类似于孪生网络的架构,以增强其交互性与协同效应。此外,图像层面的显著性目标获取方法仍然存在进一步研究的空间。

参考文献

[1] HOANG V D, HUYNH N T, TRAN N, et al. Powering AI-driven car damage identification based on VeHIDE dataset [J]. Journal of Information and Telecommunication, 2025, 9(1): 24-43.

[2] KUMAR M R, ADITHIYAN P, SENDUR G J, et al. Analyzing the potential of ReXNet-150: A novel architecture for automobile parts classification[C]. 2024 3rd International Conference on Artificial Intelligence for Internet of Things, 2024: 1-6.

[3] 胡聪, 李超, 周甜, 等. 基于改进深度相对距离学习框架的车辆再识别算法[J]. 仪器仪表学报, 2020, 41(12): 245-252.
HU C, LI CH, ZHOU T, et al. Vehicle re-recognition algorithm based on improved deep relative distance learning framework [J]. Chinese Journal of Scientific Instrument, 2020, 41(12): 245-252.

[4] 贺晓东, 王春艳, 孙昊, 等. 基于局部特征与视点感知的车辆重识别算法[J]. 仪器仪表学报, 2022, 43(10): 177-184.

HE X D, WANG CH Y, SUN H, et al. Local-features and viewpoint-aware for vehicle re-identification [J]. Chinese Journal of Scientific Instrument, 2022, 43(10): 177-184.

[5] 孙伟, 赵宇煌, 张小瑞, 等. 基于弱监督注意力和知识共享的车辆重识别[J]. 电子测量与仪器学报, 2023, 37(9): 179-189.
SUN W, ZHAO Y H, ZHANG X R, et al. Weakly supervised attention and knowledge sharing for vehicle re-identification[J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(9): 179-189.

[6] 梁继然, 陈壮, 董国军, 等. 结合注意力机制和密集连接网络的车辆检测方法[J]. 电子测量与仪器学报, 2022, 36(3): 210-216.
LIANG J R, CHEN ZH, DONG G J, et al. Vehicle detection method combining attention mechanism and dense connection network [J]. Journal of Electronic Measurement and Instrumentation, 2022, 36(3): 210-216.

[7] ZHENG L, REN D D. A light-weight deep neural network for vehicle detection in complex tunnel environments[J]. Instrumentation, 2023, 10(1): 32-44.

[8] 翟永杰, 刘璇, 王新颖, 等. 基于全局与局部注意力的车辆方位场景识别[J]. 电子测量技术, 2024, 47(14): 96-107.
ZHAI Y J, LIU X, WANG X Y, et al. Vehicle orientation scene recognition based on global-local attention [J]. Electronic Measurement Technology, 2024,47(14): 96-107.

[9] 舒娟. 基于深度学习的车辆部件检测[D]. 武汉: 华中科技大学, 2017.
SHU J. Vehicle component detection based on deep learning[D]. Wuhan: Huazhong University of Science and Technology, 2017.

[10] 吴建雄. 基于卷积神经网络的车辆部件检测[D]. 武汉: 华中科技大学, 2017.
WU J X. Detection of vehicle parts based on convolution neural network [D]. Wuhan: Huazhong University of Science and Technology, 2017.

[11] 吴伟鑫. 基于部件的车辆检测与运动趋势估计[D]. 长沙: 国防科技大学, 2020.
WU W X. Vehicle detection and motion trend estimation based on components[D]. Changsha: National University of Defense Technology, 2020.

[12] ZHAI Y J, ZHOU X Q, CHEN N H, et al. Multi-task feature decoupling network with clear division of labor for vehicle component detection[J]. Advanced Engineering Informatics, 2024, 62: 102601.

[13] LIU CH L, PU Z Y, LI Y SH, et al. Enabling edge computing ability in view-independent vehicle model recognition[J]. International Journal of Transportation Science and Technology, 2024, 14(2): 73-86.

[14] BORJI A, CHENG M M, HOU Q B, et al. Salient object detection: A survey [J]. Computational Visual Media, 2019, 5(2): 117-150.

[15] ITTI L, KOCH C, NIEBUR E. A model of saliency-based visual attention for rapid scene analysis[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(11): 1254-1259.

[16] DOUGLAS J S. Convolutional neural networks [J]. Demystifying Deep Learning: An Introduction to the Mathematics of Neural Networks, 2024: 111-131.

[17] 张冬明, 靳国庆, 代锋, 等. 基于深度融合的显著性目标检测算法[J]. 计算机学报, 2019, 42(9): 2076-2086.
ZHANG D M, JIN G Q, DAI F, et al. Saliency object detection based on deep fusion of hand-crafted features[J]. Chinese Journal of Computers, 2019, 42(9): 2076-2086.

[18] JI Y ZH, ZHANG H J, ZHANG ZH, et al. CNN-based encoder-decoder networks for salient object detection: A comprehensive review and recent advances[J]. Information Sciences, 2021, 546: 835-857.

[19] 何伟, 潘晨. 注意力引导网络的显著性目标检测[J]. 中国图象图形学报, 2022, 27(4): 1176-1190.
HE W, PAN CH. Saliency object detection with attention-guided networks [J]. Journal of Image and Graphics, 2022, 27(4), 1176-1190.

[20] PENG Y B, FENG M K, ZHAI ZH N. Visual saliency detection based on global guidance map and background attention[J]. IEEE Access, 2024, 12: 95434-95446.

[21] 谢东升. 基于深度学习的车辆智能定损算法研究[D]. 天津: 天津大学, 2019.
XIE D SH. Research on research on vehicle intelligent damage location algorithm based on deep learning[D]. Tianjin: Tianjin University, 2019.

[22] 谢永生. 基于改进 EfficientDet 的车辆部件检测与重识别研究[D]. 重庆: 西南大学, 2023.
XIE Y SH. Research on vehicle component detectionand re-identification based on improved EfficientDet algorithm[D]. Chongqing: Southwest University, 2023.

[23] 翟永杰, 李佳蔚, 陈年昊, 等. 融合改进 Transformer 的车辆部件检测方法[J]. 图学学报, 2024, 45(5): 930-940.
ZHAI Y J, LI J W, CHEN N H, et al. The vehicle parts detection method enhanced with Transformer integration[J]. Journal of Graphics, 2024, 45(5): 930-940.

[24] VILLÁN A F. Mastering OpenCV 4 with Python: A practical guide covering topics from image processing, augmented reality to deep learning with OpenCV 4 and Python 3.7[M]. Birmingham, 2019.

[25] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]. 2017 IEEE International Conference on Computer Vision, 2017: 2999-3007.

[26] MILLETARI F, NAVAB N, AHMADI S A. V-net: Fully convolutional neural networks for volumetric medical image segmentation[C]. 2016 Fourth International Conference on 3D Vision, 2016: 565-571.

[27] PRATT W K. Digital image processing: PIKS scientific inside[M]. Hoboken, New Jersey: Wiley-interscience, 2007.

[28] YURTKULU S C, Ş AHIN Y H, UNAL G. Semantic segmentation with extended DeepLabv3 architecture[C]. 2019 27th Signal Processing and Communications Applications Conference, 2019: 1-4.

[29] SHELHAMER E, LONG J, DARRELL T. Fully convolutional networks for semantic segmentation[C]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4):640-651.

[30] KIRILLOV A, WU Y X, HE K M, et al. PointRend: Image segmentation as rendering[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 9796-9805.

[31] HE K M, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]. 2017 IEEE International Conference on Computer Vision, 2017: 2980-2998.

[32] CAI ZH W, VASCONCELOS N. Cascade R-CNN: High quality object detection and instance segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021,43(5): 1483-1498.

[33] HUANG ZH J, HUANG L CH, GONG Y CH, et al. Mask scoring R-CNN[C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 6402-6411.

[34] TIAN ZH, SHEN CH H, CHEN H. Conditional convolutions for instance segmentation[J]. European Conference on Computer Vision, 2020: 282-298.

[35] SOHAN M, SAI RAM T, RAMI REDDY C V. A review on YOLOv8 and its advancements[C]. International Conference on Data Intelligence and Cognitive Informatics, 2024: 529-545.

[36] 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, 2025: 1-21.

[37] HE L H, ZHOU Y ZH, LIU L, et al. Application of the YOLOv11-seg algorithm for AI-based landslide detection and recognition[J]. Scientific Reports, 2025, 15(1): 12421.

[38] GONZAGA L M, ALMEIDA G. Artificial intelligence usage for identifying automotive products[C]. XLI Ibero-Latin American Congress on Computational Methods in Engineering, 2020, 2(2).

[39] ANUPAMA M A, CHHABRA K, GHOSH A, et al. Comparative analysis of deep learning models for car part image segmentation[C]. International Conference on Data Management, Analytics & Innovation, 2024: 267-279.

[40] PASUPA K, KITTIWORAPANYA P, HONGNGERN N, et al. Evaluation of deep learning algorithms for semantic segmentation of car parts[J]. Complex & Intelligent Systems, 2022, 8(5): 3613-3625.

[41] KHANAL S R, AMORIM E V, FILIPE V. Classification of car parts using deep neural network[C]. APCA International Conference on Automatic Control and Soft Computing, 2020: 582-591.

[42] CUI J H, WU ZH CH, WANG X K. Vehicle appearance parts identification based on instance segmentation algorithms[C]. International Conference on Algorithms, Software Engineering, and Network Security, 2024: 306-311.

[43] YUSUF S A, ALDAWSARI A A, SOUISSI R.

Automotive parts assessment: Applying real-time instance-segmentation models to identify vehicle parts[J]. ArXiv preprint arXiv:2202.00884, 2022.

[44] MAO M, HONG M. YOLO object detection for real-time fabric defect inspection in the textile industry: A review of YOLOv1 to YOLOv11[J]. Sensors, 2025, 25(7): 2270.

[45] DWIVEDI M, MALIK H S, OMKAR S N, et al. Deep learning-based car damage classification and detection[J]. Advances in Artificial Intelligence and Data Engineering, 2021: 207-221.

作者简介

翟永杰,1994 年于华北电力大学获得学士学位,1999 年于华北电力大学获得硕士学位,2004 年于华北电力大学获得博士学位,现为华北电力大学(教授),主要研究方向为电力视觉。

E-mail:zhaiyongjie@ncepu.edu.cn

Zhai Yongjie received his B. Sc. degree from North China Electric Power University in 1994, received his M. Sc. degree from North China Electric Power University in 1999, and received his Ph. D. degree from North China Electric Power University in 2004. He is currently a professor at North China Electric Power University. His main research interest is power computer vision.

王乾铭(通信作者),2017 年于华北电力大学获得学士学位,2020 年于中国舰船研究院获得硕士学位,2023 年于华北电力大学获得博士学位,现为华北电力大学(讲师),主要研究方向为电力视觉、输电线路巡检和视觉知识推理。

E-mail:qianmingwang@ncepu.edu.cn

Wang Qianming (Corresponding author) received his B. Sc. degree from North China Electric Power University in 2017, received his M. Sc. degree from China Ship Research and Development Academy in 2020, and received his Ph. D. degree from North China Electric Power University in 2023. He is currently a lecturer at North China Electric Power University. His main research interests include power computer vision, transmission line inspection, and visual knowledge reasoning.