

中图法分类号: 文献标识码: 文章编号: 1006-8961(XXXX)XX-0001-19

论文引用格式: Cheng Yong, Zhang Zhuocheng, Meng Yifei, Han Changlin, Wu Changchun, Li Chaofan. Feature-enhanced YOLO11 for object detection in foggy scenes[J/OL]. Journal of Image and Graphics, XXXX: 1-19. DOI: 10.11834/jig.260053. (程勇, 张卓成, 孟逸飞, 韩长霖, 吴长春, 李超凡. 面向雾霾场景的特征增强 YOLO11 目标检测[J/OL]. 中国图象图形学报, XXXX: 1-19. DOI: 10.11834/jig.260053.) [DOI: 10.11834/jig.260053]

面向雾霾场景的特征增强 YOLO11 目标检测

程勇¹, 张卓成², 孟逸飞³, 韩长霖^{4*}, 吴长春⁵, 李超凡⁶

1. 南京信息工程大学, 江苏省南京市 210044; 2. 南京信息工程大学, 江苏省南京市 210044; 3. 中国水利水电科学研究院, 北京市 100048; 4. 中国水利水电科学研究院, 北京市 100048; 5. 海康智联科技有限公司, 浙江省杭州市 311100; 6. 南京信息工程大学, 江苏省南京市 210044

摘要: 针对雾霾场景下图像低对比度、目标边缘模糊以及复杂背景干扰导致目标检测性能下降的问题, 提出一种面向雾霾场景的特征增强目标检测网络 SD-YOLO11。该方法以 YOLO11 为基础, 从主干网络与颈部网络两个层面对模型进行协同优化。在主干网络中, 引入 SIBlock 模块, 通过递归空间交互增强模糊目标与小目标的特征表达能力; 结合 GhostConv 与 ADown 模块, 在降低计算冗余的同时提高关键特征保留能力。在颈部网络中, 设计改进的 BifpnCat 特征融合结构, 以增强多尺度语义信息与细节信息的交互能力; 同时引入 DASI 注意力机制, 对目标相关特征进行自适应增强, 从而提升复杂雾霾场景下的目标感知能力。此外, 通过优化特征拼接路径并增加 P2 检测头, 进一步增强网络对小目标与低分辨率目标的检测能力。为了验证所提方法的有效性, 在 HazyDet、Foggy Cityscapes 和 RTTS 数据集上开展对比实验。实验结果表明, SD-YOLO11 在 HazyDet 数据集上的 mAP₅₀ 和 mAP₅₀₋₉₅ 分别达到 78.8% 和 61.4%, 较 YOLO11 分别提升 1.4 个百分点和 2.0 个百分点; 在 Foggy Cityscapes 和 RTTS 数据集上同样取得优于主流检测方法的检测性能, 并在小目标与低对比度目标检测任务中表现出更好的鲁棒性。研究结果表明, 所提方法能够有效提升雾霾场景下目标检测的精度与稳定性, 同时保持较好的实时检测性能, 可为低能见度环境下的视觉感知任务提供有效技术支持。

关键词: 雾霾场景; 目标检测; 特征提取; 注意力机制; 特征融合

Feature-enhanced YOLO11 for object detection in foggy scenes

Cheng Yong¹, Zhang Zhuocheng², Meng Yifei³, Han Changlin^{4*}, Wu Changchun⁵, Li Chaofan⁶

1. Nanjing University of Information Science and Technology, Nanjing 210044, China; 2. Nanjing University of Information Science and Technology, Nanjing 210044, China; 3. China Institute of Water Resources and Hydropower Research, Beijing 100048, China; 4. China Institute of Water Resources and Hydropower Research, Beijing 100048, China; 5. Hikvision Intelligent Connected Technology Co., Ltd., Hangzhou 311100, China; 6. Nanjing University of Information Science and Technology, Nanjing 210044, China

Abstract: Object detection in foggy scenes is an important research topic in computer vision, aiming to improve the perception capability of intelligent systems under low-visibility conditions. However, fog-induced degradation, including low con-

收稿日期: 2026-01-25; 修回日期: 2026-07-01

* 通信作者: 韩长霖, 通信作者, 男, 正高级工程师, 主要研究方向为水电站及新能源电站监控系统研发, 水电站系统自动化、信息化、智能化。E-mail: jkhanel@iwhr.com

基金项目: 国家重点研发计划项目(项目编号: 2023YFE0208100); 中国水利水电科学研究院基本科研业务费项目(项目编号: AU0145B022021)

Supported by: National Key Research and Development Program of China (Grant No. 2023YFE0208100); Fundamental Research Funds of China Institute of Water Resources and Hydropower Research (Grant No. AU0145B022021)

trast, blurred boundaries, weakened texture details, and severe background interference, significantly reduces feature discriminability and often leads to missed detections and false alarms. In practical applications such as autonomous driving, intelligent transportation systems, unmanned aerial vehicle surveillance, and security monitoring, reliable object detection under adverse weather conditions is essential for ensuring environmental awareness and decision-making safety. Nevertheless, atmospheric scattering not only obscures object appearance but also weakens the distinction between foreground targets and background regions, making feature extraction and target localization considerably more challenging. Existing approaches mainly improve detection performance through image dehazing or network architecture optimization. However, dehazing-based methods usually introduce additional computational burden and may generate artifacts that negatively affect subsequent detection, while many detection-oriented methods still suffer from insufficient feature representation, inadequate multi-scale information fusion, and weak robustness to small or low-contrast objects in complex foggy environments. To address these issues, this paper proposes a feature-enhanced object detection network named SD-YOLO11 based on YOLO11. The proposed method improves the network from both the backbone and neck components to enhance the representation and fusion of degraded features under foggy conditions. Specifically, a Spatial Interaction Block (SIBlock) is introduced into the backbone network to strengthen the modeling capability for blurred targets and weak texture regions. By recursively aggregating spatial contextual information, SIBlock enhances high-order spatial interactions and improves the network's ability to capture structural continuity and semantic correlations in degraded scenes. Compared with conventional convolution operations that mainly focus on local receptive fields, the recursive interaction mechanism enables the network to establish richer contextual dependencies and improves the representation of degraded object features. To further improve computational efficiency while preserving feature quality, Ghost Convolution (GhostConv) is adopted to reduce redundant computation through inexpensive feature generation operations, and an Adaptive Downsampling (ADown) module is employed to enhance feature preservation during the downsampling process. These two modules jointly improve the balance between detection accuracy and computational complexity by reducing parameter redundancy while retaining essential semantic information. In the neck network, an improved feature fusion structure named BifpnCat is designed to enhance information exchange among feature maps of different resolutions through bidirectional feature aggregation and channel-wise fusion. Compared with conventional feature concatenation strategies, the proposed structure facilitates more effective integration of semantic information and spatial details, thereby improving multi-scale feature representation. In addition, a simplified DASI attention mechanism is incorporated into the feature fusion stage to adaptively emphasize target-related responses while suppressing background interference, which enhances the discriminative capability of the network in complex foggy scenes. Furthermore, the feature concatenation pathway is redesigned to retain richer channel information during feature aggregation, and a high-resolution P2 detection head is introduced to improve the detection performance for small and distant objects. The additional detection head allows the network to utilize high-resolution feature maps more effectively, thereby alleviating information loss caused by repeated downsampling operations. Unlike methods that rely on explicit image restoration, the proposed framework focuses on enhancing feature learning directly within the detector, reducing computational overhead while avoiding error propagation from image enhancement to object detection. Through the collaborative optimization of feature extraction, feature preservation, attention enhancement, and multi-scale feature fusion, SD-YOLO11 establishes a more robust detection framework for low-visibility environments. To verify the effectiveness of the proposed method, extensive experiments are conducted on three representative fog-related datasets, namely HazyDet, Foggy Cityscapes, and RTTS. Experimental results demonstrate that SD-YOLO11 achieves 78.8% mAP_{50} and 61.4% mAP_{50-95} on the HazyDet dataset, improving the original YOLO11 by 1.4 and 2.0 percentage points, respectively. The ablation experiments further confirm the effectiveness of each proposed component, showing that SIBlock, GhostConv, ADown, BifpnCat, and the DASI attention mechanism all contribute positively to the final detection performance. In addition, the proposed method achieves superior results on the Foggy Cityscapes dataset compared with mainstream one-stage detectors, fog-oriented detection models, and transformer-based detectors. It also demonstrates competitive performance in categories closely related to small and low-contrast objects, indicating strong robustness under controlled fog degradation. Detailed comparative analyses show that the proposed improvements consistently enhance feature representation and multi-scale perception capability, leading to more accurate localization and classification of degraded targets. To further evaluate

cross-dataset generalization capability, experiments are performed on the RTTS dataset, which contains real-world foggy traffic scenes. The results show that SD-YOLO11 maintains stable detection accuracy across different visibility conditions and achieves better overall performance than other competing methods. The consistent improvements obtained on both synthetic and real-world datasets demonstrate that the proposed method possesses strong adaptability and robustness under diverse foggy conditions. Moreover, despite the introduction of multiple enhancement modules, the proposed network still preserves real-time inference capability, making it suitable for practical deployment in intelligent transportation systems, autonomous driving platforms, and video surveillance applications operating in adverse weather environments. Overall, the proposed SD-YOLO11 effectively alleviates feature degradation caused by atmospheric scattering through enhanced spatial interaction modeling, lightweight feature extraction, adaptive feature preservation, attention-guided feature enhancement, and improved multi-scale feature fusion. Experimental results on multiple benchmark datasets demonstrate that the proposed method consistently improves detection accuracy, robustness, and generalization ability while maintaining favorable computational efficiency. These findings indicate that SD-YOLO11 provides an effective solution for object detection in foggy scenes and offers a practical reference for visual perception tasks in complex low-visibility environments. The proposed framework also demonstrates the feasibility of improving adverse-weather object detection through detector-oriented feature enhancement rather than relying solely on image restoration, which may provide valuable insights for future research on robust visual perception and intelligent scene understanding in challenging environments.

Key words: foggy scenes; object detection; feature extraction; attention mechanism; feature fusion

公式章 1 节 1. 图法分类号: TP391.41; TP183
文献标识码: A 文章编号: 1006-8961(年) -

论文引用格式: Cheng Y, Zhang Z C, Meng Y F, Han C L, Wu C C and Li C F. 2026. Feature-Enhanced YOLO11 for Object Detection in Foggy Scenes. Journal of Image and Graphics, 31(xx): xxxx-xxxx (程勇, 张卓成, 孟逸飞, 韩长霖, 吴长春, 李超凡. 2026. 面向雾霾场景的特征增强 YOLO11 目标检测. 中国图象图形学报, 31(xx): xxxx-xxxx) [DOI:10.11834/jig.260053]

0 引言

随着计算机视觉的快速发展,恶劣天气条件下的图像处理已成为该领域的重要研究方向。在军事领域,雾霾场景目标检测可用于低能见度环境下的入侵监测,提高环境感知能力(Binbusayis 等, 2024; Yi 等, 2023); 在民用领域,自动驾驶系统需要准确识别障碍物、行人和交通标志,小目标与弱目标检测技术的发展进一步提升了低能见度环境下的视觉感知能力(冯琪涵等, 2025; 夏波等, 2025)。此外,该技术在海洋监测、道路安防等领域也具有广泛应用价值,因此研究雾霾场景鲁棒目标检测具有重要意义。

从技术路线角度看,现有雾霾场景目标检测方法中,一类研究侧重于将图像去雾与目标检测相结

合,通过改善输入图像质量以提升下游检测性能。在雾、霾等恶劣天气条件下拍摄的图像由于气溶胶粒子的吸收和散射作用会产生可视度低、失真等退化现象。常见的解决方法是将去雾算法考虑到雾霾场景目标检测任务中,如刘春晓等人(2025)针对非均质雾霾图像提出了解耦合三阶段去雾增强网络,通过分离结构、细节与颜色信息的恢复过程,有效提升了复杂雾霾场景下图像的清晰度和可辨性,为下游目标检测提供了更可靠的输入。Zhang 等人(2023)构建了 DH-YOLO(dehazing YOLO)框架,联合优化图像恢复与目标检测任务,并引入特征过滤模块(feature filtering module, FFM)去除冗余信息。Li 等人(2023)设计的 BAD-Net(end-to-end union detection network)通过双分支结构融合去雾与原始特征,并引入自监督去雾损失,提高模型对不同雾霾浓度的适应性。Gupta 等人(2024)则通过合成天气数据增强模型鲁棒性,并引入图像去噪处理,有效提升恶劣天气下的检测性能。岳永恒等(2025)提出基于 CycleGAN(Cycle-Consistent Generative Adversarial Network)与 YOLOv8 的联合雾霾环境感知方法,通过引入自注意力和自正则化机制提升去雾质量,并利用注意力模块强化检测特征,在 RESIDE 和 BDD100k 数据集上显著提高了雾霾场景中的识别性能。尽管上述方法能一定程度上改善雾霾图像质量,但仍存在一些不足:一是检测性能高度依赖图像

恢复模块,若去雾过程中产生伪影或细节丢失,可能干扰目标识别;二是去雾模块通常计算复杂,影响检测速度,难以满足实时性要求;三是这些方法普遍在特定浓度或类型雾霾下优化,导致泛化能力不足,无法适应复杂多变的气象条件。

另一类研究不再显式引入图像去雾模块,而是通过改进目标检测模型的网络结构,使其在雾霾条件下具备更强的鲁棒特征建模与多尺度感知能力。现有主流目标检测框架主要分为双阶段与单阶段两类。双阶段检测器如 R-CNN(region-based convolutional neural network)(Girshick 等, 2014)、Fast R-CNN(Girshick 等, 2015)和 Faster R-CNN(Ren 等, 2015),先提取候选区域后分类,精度较高,但推理慢、计算复杂。单阶段检测器如 YOLO 系列(Redmon 等, 2016; Ge 等, 2021; Redmon 和 Farhadi, 2017; Redmon 等, 2018; Bochkovskiy 等, 2020; Jocher 等, 2022; Li 等, 2022; Wang 等, 2023; Hermens 等, 2024)和 SSD(single shot multibox detector)(Liu 等, 2016)则以端到端方式直接预测类别与位置,具备更高的实时性优势。部分研究者也对 YOLOX 结构进行了改进,以适应遥感图像或水域监测等特殊任务(He 等, 2024; Cheng 等, 2023)。近年来,特征增强、多尺度融合及 Transformer 结构逐渐应用于复杂视觉场景。Liu 等(2025)提出 FETNet,通过特征增强模块与轻量 Transformer 提升复杂背景下的小目标分割性能; Fu 等(2026)提出 DTL-IceNet,利用多尺度融合与注意力机制抑制背景干扰,提高复杂环境下的覆冰检测性能。然而,这些方法主要面向遥感分割或覆冰检测任务,对雾霾场景目标检测中的退化特征建模缺乏针对性。此外,一些研究尝试通过模型结构改进提升检测性能。Yue 等人(2025)提出 AW-YOLO(all weather YOLO),在 YOLOv8 基础上引入 Dilation-wise Residual 模块和动态上采样器,增强了复杂天气条件下的感知能力。Kiran 等人(2024)改进 YOLOv4,融合多尺度 Retinex、双树复小波变换(dual-tree complex wavelet transform, DTCWT)与脉冲耦合神经网络(pulse coupled neural network, PCNN),显著提升动态光照和噪声场景下的车辆检测精度。Meng 等人(2023)提出 YOLOv5s-Fog,利用 SwinFocus 检测层与解耦头配合软非极大值抑制(soft non-maximum suppression, Soft-NMS)提升了雾霾场景下模糊小目标的检测能力。Luo 等人(2024)

提出 SDNIA-YOLO(stylization data driven neural image adaptive YOLO),借助神经风格迁移增强图像质量,并通过自适应建模提升 YOLOv3 在极端天气下的表现。Muhammad 等人(2023)通过融合 ACDC 与 DAWN 数据集训练 YOLOv8,显著提升了模型的泛化能力。Wang 等人(2025)提出 PC-YOLO(Physics-Aware and Dynamic Network Structures for YOLO),结合物理感知模块与条件宽度分支,强化雨雾下特征并自适应调节网络宽度,从而提升恶劣天气中的稳健性。尽管已有方法在一定程度上提升了雾霾场景中的检测性能,但仍存在三方面问题:其一,多数方法缺乏对低质量图像细节建模的能力,尤其在小目标与模糊边界识别上仍有短板;其二,结构复杂度较高,不利于部署在算力受限的边缘设备中;其三,部分特征融合策略未能充分捕捉多尺度语义信息,影响检测结果的稳定性。

基于之前 YOLO 模型的优势, YOLO11(Khanam 等, 2024)在网络结构设计、特征提取能力与工程部署方面进行了进一步优化。该版本引入了多项架构增强,包括改进的注意力机制、更深层次的特征提取层以及无锚定检测范式,从而在面对多变且复杂的环境时,能够更好地平衡检测效率与准确性。虽然 YOLOv9 在若干公开基准上的结果更强,但其核心创新属于训练范式与骨干的研究型升级,如可编程梯度信息(programmable gradient information, PGI)和广义高效层聚合网络(generalized efficient layer aggregation network, GELAN),与随之而来的数据增强、损失与默认超参构成一整套配方;直接采用将引入较多实现与配方差异。基于上述两类研究路线可以看出,现有方法要么依赖图像恢复以换取检测性能,要么通过局部结构改进提升鲁棒性,但在兼顾雾霾场景特征建模、多尺度融合与计算效率方面仍存在改进空间。相较之下, YOLO11 更强调工程化落地与体系化优化,如 C3k2/C2PSA(C2 position-sensitive attention)、解耦且无锚的检测头、训练一导出一部署的一体化工具链,与本文提出的设计方案在张量维与计算图层面高度兼容,并在统一实现与配方内评估结构贡献、降低混杂因素,提升可复现与可维护性。针对雾霾场景中普遍存在的对比度降低、边缘模糊以及多尺度目标特征退化等问题,本文从特征建模与特征融合两个层面对 YOLO11 进行系统性改进。整体设计思路并非简单叠加现有模

块,而是围绕雾霾环境下目标特征易受干扰、细节信息易丢失这一核心问题,对网络主干与颈部结构进行针对性重构。在主干网络中,重点增强空间上下文建模能力并抑制冗余特征计算;在颈部网络中,则通过改进多尺度特征融合与注意力机制,强化目标区域的判别性表达。与现有工作直接堆叠注意力或多尺度模块不同,本文更关注雾霾场景下特征退化机理与网络结构之间的匹配关系。针对低对比度条件下空间结构信息易缺失、小目标特征在下采样过程中易衰减以及多尺度语义融合不充分等问题,本文从空间建模、轻量化特征提取与跨尺度特征交互三个层面对 YOLO11 进行协同优化。各模块并非孤立引入,而是在统一检测框架下围绕“雾霾退化目标的鲁棒表征”这一核心目标进行结构重构与功能互补。基于上述思路,本文构建了改进网络 SD-YOLO11 (Spatial Interaction and Dual-path Feature Fusion YOLO11)。

本文的贡献可以总结如下:

针对雾霾场景下目标边缘模糊、低对比度及复杂背景干扰导致的特征退化问题,本文从空间交互建模、轻量化特征提取与多尺度特征融合三个层面对 YOLO11 进行协同优化,构建了面向低能见度场景的目标检测网络 SD-YOLO11,在保证实时性的同时提升了模型对模糊目标与小目标的检测能力。

在主干网络部分,本文引入空间交互模块(Spatial Interaction Block, SIBlock),通过递归空间交互增强网络对模糊边界与弱纹理区域的特征建模能力;结合 GhostConv 与自适应下采样模块(Adaptive Downsampling, ADown),在降低计算冗余的同时提升特征保留能力,实现检测精度与推理效率之间的平衡。

在颈部网络部分,本文提出改进的双向特征金字塔特征融合结构(Bidirectional Feature Pyramid Concatenation, BifpnCat),通过双向特征融合与通道拼接增强多尺度信息交互能力,以提升复杂雾霾场景下的小目标检测性能。同时,引入动态自适应空间交互注意力机制(Dynamic Adaptive Spatial Interaction, DASI)以增强目标相关特征表达,降低背景噪声干扰。

此外,本文重新设计了颈部特征拼接路径,并新增 P2 检测头,以增强网络对高分辨率细粒度特征的利用能力,从而进一步提升小目标与低分辨率目标

的检测精度。

1 方法

1.1 整体网络结构设计

YOLO11 在速度、准确性和特征提取方面具有较好的性能。它在继承之前版本优势的基础上,引入了一些架构上的增强。YOLO11 通常由三个关键组成部分构成:主干网络、颈部模块和头部模块。首先,输入图像通过主干网络提取不同尺度的关键特征。在主干部分,YOLO11 采用了 C3K2 块,替代了 YOLOv8 中使用的 C2f 块(Hermens 等, 2024)。C3K2 块通过实现跨阶段部分(cross stage partial network, CSP),在计算效率上做出了优化,并有效地增强了特征提取能力。在主干网络的最后两个阶段,YOLO11 引入了空间金字塔池化(spatial pyramid pooling-fast, SPPF)和具有空间注意力机制的交叉阶段部分(C2PSA),以进一步提高特征提取的能力。SPPF 模块通过将输入图像分割成多个网格,并独立地从每个网格中提取特征,再通过最大池化操作处理多尺度信息。这一过程有效地整合了来自不同尺度的信息。C2PSA 模块则利用两个位置敏感注意力(position-sensitive attention, PSA)模块来处理输入张量及其位置敏感的注意力。该模块有助于模型有选择性地关注图像中的细节,进一步增强了特征图的表达能力。两个 PSA 模块分别在特征图的不同部分上工作,且其输出被连接起来,形成更为丰富的特征表示。在颈部模块中,YOLO11 聚合了来自不同尺度的特征,并将这些特征传递到解耦头部,该头部模块用于确定目标类别、计算目标置信度,并精确预测已识别目标的边界框(Rasheed 等, 2024)。在 YOLO11 的衍生模型中,YOLO11-S 通过减少参数数量并提高部署的便捷性,展现了较好的性能。因此,本研究选择对 YOLO11-S 进行进一步优化和改进。该网络的改进结构如图 1 所示。

1.2 面向退化特征的主干网络增强设计

1.2.1 基于递归空间交互的退化特征增强

在雾霾条件下,目标边缘与纹理信息易被雾气遮蔽,单次局部卷积难以形成稳定的空间响应。为增强网络对模糊轮廓和上下文关系的建模能力,本文在主干网络中设计 SIBlock 模块。现有注意力机制如 SE(squeeze and excitation)、ECA(efficient chan-

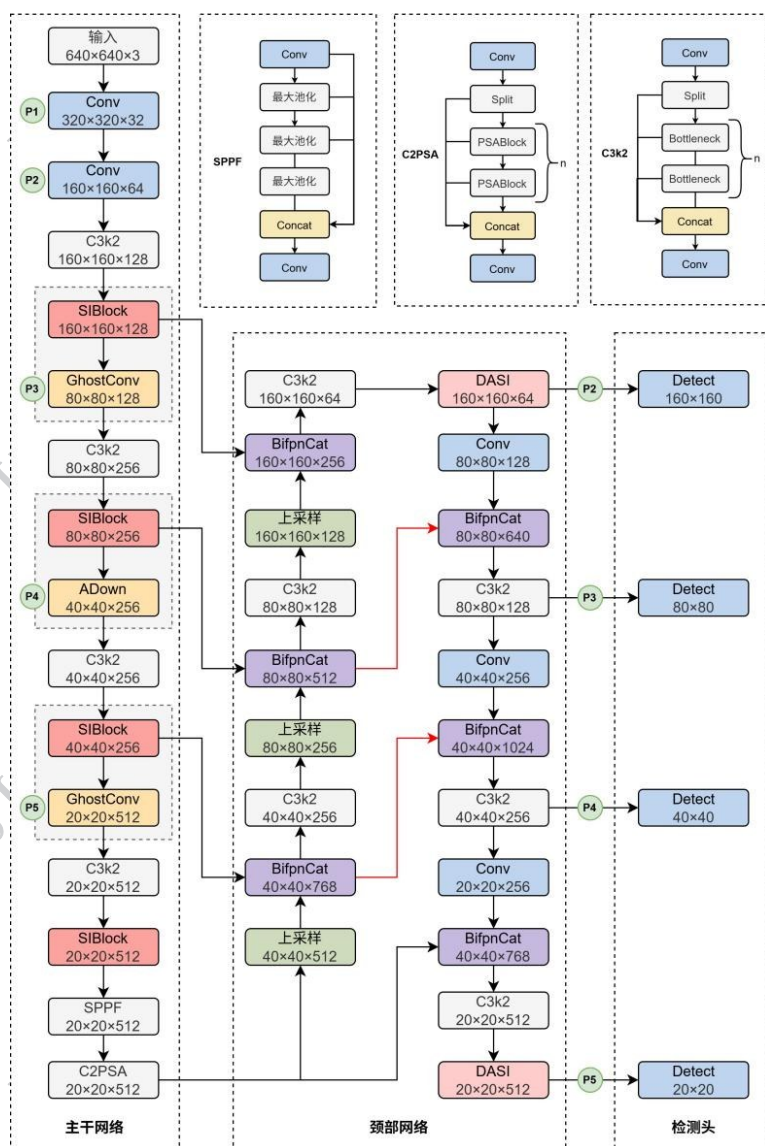


图1 SD-YOLO11模型图

Fig. 1 SD-YOLO11 Model Diagram

nel attention) 和 CBAM (convolutional block attention module) 虽能增强特征表达, 但分别存在空间信息利用不足、全局相关性建模欠缺以及计算冗余较高等问题。

针对上述不足, 本文构建 SIBlock 模块。该模块通过引入递归卷积 $g^n\text{Conv}$ (generalized n-order convolution) (Rao 等, 2022),

增强空间特征交互能力, 捕捉长期空间依赖关系, 提高复杂背景下模糊目标与低对比度目标的特征表达能力。模块首先采用层归一化统一特征分布, 再利用 $g^n\text{Conv}$ 实现高阶空间交互。考虑到递归交互易在雾霾等噪声场景中产生分支共适应与梯度放大, 本文在 $g^n\text{Conv}$ 与多层感知机 (multilayer per-

ceptron, MLP) 之间引入 DropPath 作为结构化正则, 以减弱分支相关性并提高模块稳健性。在 $g^n\text{Conv}$ 中, 首先使用 Conv_{in} 对输入特征进行投影, 得到初步特征, 如公式 (1) 所示:

$$[\lambda_0^{H \times C_0}, \mu_0^{H \times C_0}, \dots, \mu_{n-1}^{H \times C_{n-1}}] = \text{Split}(\text{Conv}_{in}(x)) \quad (1)$$

式中, x 表示输入特征; $\mu_i^{H \times C_i}$ 表示第 i 阶递归交互对应的特征映射, 其中 H 、 W 分别表示特征图的高度和宽度, C_i 表示第 i 阶特征的通道数; Conv_{in} 表示输入投影卷积; $\text{Split}(\cdot)$ 表示通道划分操作; n 表示递归交互次数。

然后, 通过递归方式更新特征, 第 k 次递归过程如公式 (2) 所示:

$$\lambda_{k+1} = \text{DWConv}_k(\mu_k) \odot \lambda_k / \alpha \quad (2)$$

式中, λ_k 表示第 k 次递归得到的特征; λ_{k+1} 表示更新后的特征; μ_k 表示第 k 阶输入特征; $DWConv_k(\cdot)$ 表示第 k 阶深度卷积操作; $\odot$ 表示逐元素乘法; α 表示融合权重常数, 用于平衡递归分支与残差分支, 本文固定取 0.5。

通过递归卷积的多次迭代, 逐步引入更高阶的空间上下文信息, 从而实现特征在空间维度上的深层交互。如图 2 所示。通过这种局部交互, $g^n Conv$ 在空间上对特征进行了增强, 使得每个目标的特征不仅仅依赖于它自己的局部信息, 还能利用邻近目标的空间信息。之后再通过 MLP 处理非线性关系。它能够将输入的特征进行高维映射, 从而捕捉更加复杂的模式。

与标准卷积仅在固定感受野内建模局部响应不同, SIBlock 通过递归卷积在多次迭代过程中逐步聚合更高阶的空间上下文信息, 使特征响应不再局限于单一尺度或局部邻域。在雾霾场景中, 目标边缘和纹理往往因散射效应而呈现连续模糊状态, 单次卷积易将目标与背景混合建模, 而递归空间交互能够在更大范围内捕捉轮廓走向与结构一致性, 从而增强对模糊边界的整体感知能力。这种高阶空间建模机制有助于在低对比度条件下稳定目标结构响应, 为后续检测头提供更加判别性的特征基础。

1.2.2 面向轻量化特征保留的卷积与下采样设计

在雾霾场景中, 特征表达本身已受到退化影响, 冗余计算不仅难以带来有效增益, 还可能放大噪声

干扰。基于此, 本文通过引入 GhostConv 与 ADown 模块, 在保持特征表达能力同时抑制冗余计算, 并增强多尺度特征在下采样过程中的有效保留。

标准卷积通常包含卷积、批归一化和激活函数, 容易生成大量相似特征图, 造成较高的计算开销。为此, GhostConv 采用两步式特征生成策略 (Han 等, 2020)。首先, 利用较小卷积核的标准卷积从输入特征图 $X \in R^{h \times w \times c}$ 中生成一组通道数较少的内在特征 $Y' \in R^{h' \times w' \times c'}$, 其中, h 和 w 分别表示输入特征图的高度和宽度, c 表示输入通道数, h' 、 w' 和 c' 分别表示输出特征图的高度、宽度和通道数, 通常取 $c' = c/2$ 。随后, 对每个内在特征通道采用深度卷积生成附加特征, 记为 φ , 最后将两者在通道维进行拼接得到最终输出。图中“Identity”表示恒等映射直连分支, 用于保留原始信息并稳定梯度传播, 不引入额外参数, 如图 3 所示。基于此, 本文采用 GhostConv 优化主干网络, 以降低计算量并提升推理速度。

此外, 本文采用 ADown 下采样卷积块替代原有 P4 位置的标准卷积模块 (Wang 等, 2024)。首先, 对输入特征进行平均池化 (average pooling 2D, Avg-Pool2d) 实现初步下采样; 随后, 将特征划分为两路, 一路经最大池化 (max pooling 2D, MaxPool2d) 和卷积提取局部细节信息, 另一路直接通过卷积提取全局上下文信息; 最后, 将两路特征进行通道拼接, 如图 4 所示。该结构融合了不同池化方式的特征信息, 增强了模型对不同尺度目标的感知能力, 尤其适用于背景复杂、目标尺度差异较大的雾霾场景。

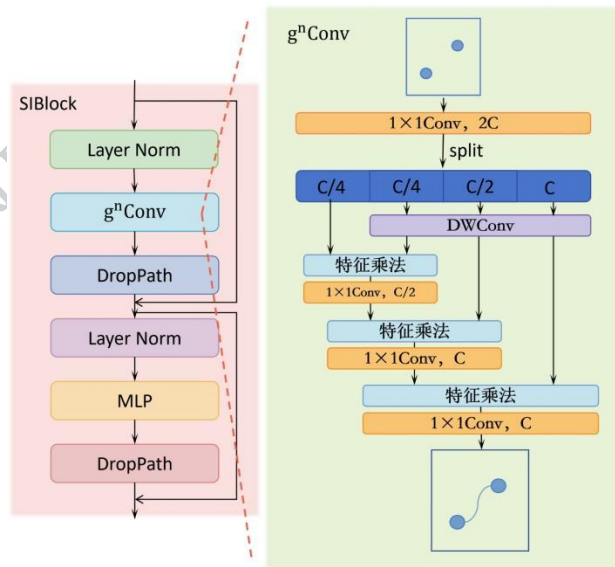


图2 SIBlock 结构图

Fig. 2 SIBlock Structure Diagram

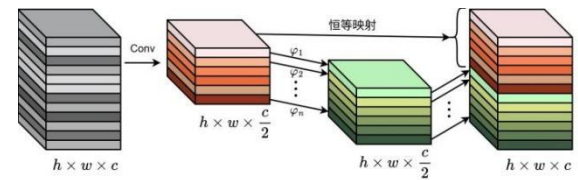


图3 GhostConv 模型

Fig. 3 GhostConv Module

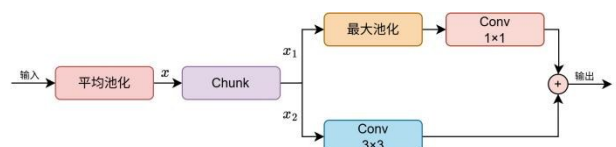


图4 ADown 模块

Fig. 4 ADown Module

1.3 面向多尺度特征的融合增强设计

1.3.1 基于双向融合的跨尺度特征融合

在雾霾环境中,目标尺度分布复杂且语义信息易随层级加深而衰减,简单的特征拼接难以兼顾高层语义与低层细节。为此,本文在颈部网络中重点优化多尺度特征的交互方式,通过改进特征融合拓扑并引入注意力机制,使不同尺度特征在融合过程中能够更有效地保留目标相关信息,从而提升网络在复杂雾霾场景下的检测稳定性。在 YOLO 系列网络中,颈部常将两路输入特征直接拼接。然而当尺度差异较大时,简单拼接容易造成信息冲突或细粒度特征的弱化。为此,BiFPN(bidirectional feature pyramid concatenation)引入自上而下/自下而上的双向路径进行特征融合,使低分辨率特征能够向高分辨率层传递语义信息,从而提升多尺度学习能力与融合效果(Tan 等, 2020);同时采用加权融合策略,由网络自学习分配各输入的权重,避免直接求和/拼接引发的特征丢失与信息冲突,如图 5a 所示。

在此基础上,本文提出 BifpnCat。与传统 BiFPN 的单步加权融合不同,BifpnCat 首先对相邻层特征分别完成同层聚合,形成中间节点 p_1^{mid} 和 p_2^{mid} 。其中,高层特征 p_2 经上采样后与低层特征 p_1 进行加权融合得到 p_2^{mid} ,而 p_1^{mid} 则由低层特征 p_1 与来自高层路径的中间特征 p_2^{mid} 共同生成,其计算过程如公式(3)与(4)所示:

$$p_2^{mid} = Conv\left(\frac{w_1 \cdot Up(p_2) + w_2 \cdot p_1}{w_1 + w_2 + \varepsilon}\right) \quad (3)$$

$$p_1^{mid} = Conv\left(\frac{w_3 \cdot p_1 + w_4 \cdot Down(p_2^{mid})}{w_3 + w_4 + \varepsilon}\right) \quad (4)$$

式中, p_1 和 p_2 分别表示相邻尺度的输入特征; p_1^{mid} 和 p_2^{mid} 分别表示第一阶段融合得到的中间特征; w_1 、 w_2 、 w_3 和 w_4 表示网络自学习得到的融合权重; $Up(\cdot)$ 表示上采样操作; $Down(\cdot)$ 表示下采样操作; $Conv(\cdot)$ 表示卷积操作; ε 表示防止分母为零的极小常数。

随后,在中间特征的基础上进一步生成输出节点 p_1^{out} 与 p_2^{out} ,其计算过程如公式(5)与(6)所示:

$$p_1^{out} = Concat(p_1^{mid}, p_2^{mid}, p_1) \quad (5)$$

$$p_2^{out} = Concat(p_1^{out}, p_2^{mid}, p_2) \quad (6)$$

式中, p_1^{out} 和 p_2^{out} 分别表示最终融合后的输出特征; $Concat(\cdot)$ 表示通道拼接操作。

通过两条跨级捷径,来缩短监督/梯度回路路

径、减少上/下采样错配带来的信息衰减,如图 5 红色流程线所示;在输出侧采用通道拼接保留两路统计而非一步相加。该拓扑在重复堆叠($\times n$)时能持续对小目标与弱边缘进行补充。因此,BifpnCat 在保持计算复杂度可控的同时,使不同尺度间的信息传递更为充分,增强了多尺度特征的整合与表达能力;每个自上而下/自下而上的双向单元被视为一层并重复利用,从而促进更高层语义与低层细节的协同融合,如图 5b 所示。在复杂背景与多尺度目标场景下表现更稳健。

然而,我们也注意到,传统结构中,颈部网络中后两个拼接的输入分别来自于 C3K2 和 C2PSA,这导致输入的通道数大幅减少,在一定程度上限制了网络对细节信息的捕捉能力。为了解决这一问题,我们将后两个拼接的输入替换为前两个拼接的输出结果,从而保留更多的通道数,如图 1 红色流程线所示。这一改进的优点在于,它能够有效提升网络对细节特征的感知能力,避免了由于通道数过少导致的特征表达能力不足。通过这种调整,网络可以更好地捕捉到图像中的细节信息,提升了整体检测精度。

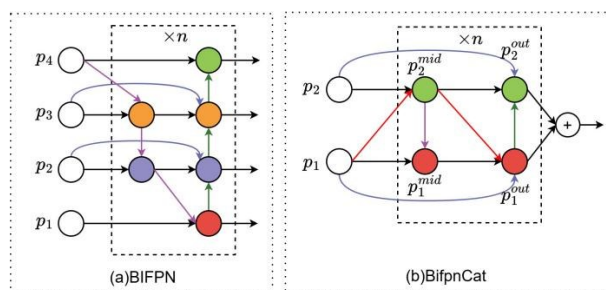


图 5 BiFPN 与 BifpnCat 的对比

Fig. 5 Comparison between BiFPN and BifpnCat

1.3.2 基于特征选择的注意力增强机制

在雾霾图像中,背景区域与目标区域在低频信息上具有相似统计特性,容易在特征空间中产生响应混叠。DASI 模块通过通道选择与残差结构,对特征响应进行选择性地增强与抑制,使目标相关特征在整体特征分布中占据更高权重,从而提升模型在复杂雾霾背景下的判别稳定性。

本文在颈部网络中引入一种基于 DASI 思想的简化注意力机制(Xu 等, 2024),通过卷积操作增强特征表示能力。与原始 DASI 侧重于多分支层级特征交互不同,本文采用单分支输入,并将注意力计算

限制在特征融合阶段,以更好地适配 YOLO 系列检测网络的颈部结构。在保留特征选择思想的同时,有效降低了计算开销。

模块流程如下:首先,输入特征经过卷积块(convolution block, ConvBlock)提取基础特征,卷积块由卷积层(Conv)、批归一化层(batch normalization, BN)和 ReLU 激活函数组成;随后,再经过第二个卷积块进一步增强特征表达,并将两级特征进行残差相加融合;最后,经批归一化与 ReLU 激活后输出最终特征,如图 6 所示。

本实现保留了 DASI 中“残差 + 特征选择”的核心思想,因此仍称为 DASI。该模块通过卷积块之间的顺序特征提取与融合操作,有效增强了特征的多层次表达能力。相加操作能够融合不同卷积层提取的特征信息,从而提升模型对复杂场景的特征学习能力。

在雾霾场景中,大量背景区域与目标区域在低频信息上具有相似统计特性,容易在特征空间中产生响应混叠。DASI 模块通过残差结构与通道选择机制,对目标相关特征进行增强,同时削弱背景主导的冗余响应。该机制并非依赖显式的像素级注意力定位,而是在特征层面对重要信息进行自适应筛选,从而在不引入额外监督或可视化约束的前提下,提高模型在复杂雾霾背景中的判别稳定性。

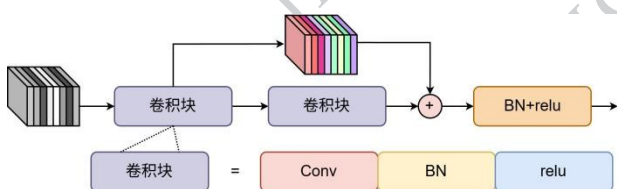


图 6 DASI 注意力机制

Fig. 6 DASI Attention Mechanism

此外,本文通过引入额外的检测层进一步增强了网络对小目标的感知能力。该检测层融合了来自 P2 的高分辨率特征以及原有的 P3 至 P5 多尺度特征映射,使网络能够同时利用细粒度空间信息与高层语义信息。如图 1 所示,分辨率为 160×160 的 P2 特征有助于缓解小目标在下采样过程中信息丢失的问题。通过增加相应的检测头,网络结构能够更加有效地聚焦于与小目标相关的特征,从而提升整体检测性能。

2 实验

2.1 数据集

HazyDet 数据集是针对雾霾环境下无人机目标检测任务构建的大规模数据集(Feng 等, 2024),包含合成与真实雾霾图像,可为相关目标检测算法提供统一评测基准。数据集涵盖城市、乡村和沿海等多种场景,并包含不同飞行高度与拍摄角度下的雾霾图像。其中,合成数据通过大气散射模型(atmospheric scattering model, ASM)生成,真实数据则通过实地采集与网络收集获得。整个数据集包含 11,000 张合成图像和 600 张真实图像,共标注约 365,000 个目标实例,类别包括汽车、卡车和公交。数据集按照 8:1:1 划分为训练集、验证集和测试集,用于模型训练、参数调优与性能评估。

Foggy Cityscapes. 为评估方法在合成雾环境下的表现,我们引入 Foggy Cityscapes 数据集。该数据集以 Cityscapes 的精标注图像为基础,依据大气散射/消光模型在保持原有标注与图像几何不变的前提下合成不同雾密度的图像(常见雾参数为 $\beta \in \{0.005, 0.01, 0.02\}$)。我们采用其与 Cityscapes 一致的 train/val/test 划分与类别体系,默认使用中等雾密度版本进行训练与模型选择,用以检验在受控雾化强度下的检测性能及对小目标/低对比细节的鲁棒性。

RTTS(Real-world Task-driven Testing Set). 为验证跨数据集泛化能力,我们进一步在 RTTS 上进行测试。RTTS 由真实雾霾交通场景构成,包含城市道路、十字路口等多种拍摄条件与能见度,提供检测任务所需的边界框标注。我们沿用其公开类别,并与本文任务所用类别体系做一致性映射,用于跨数据集零迁移评估。本研究实际使用的 RTTS 统计为 4,322 张图像、41,201 个实例(以本文标注映射后的类别集合为准),用于衡量模型在真实雾霾场景中的可迁移性与稳健性。

2.2 实验环境和评价指标

实验操作系统是 Windows 11,处理器为 Intel i5-13600KF,显卡为 NVIDIA GEFORCE RTX 4070 SUPER,显存为 12 G,内存为 32G。训练环境配置为 Python3.9.20,深度学习框架为 Pytorch 2.1.0 和 CUDA 12.1。模型初始输入图像大小设置为 $640 \times$

640, 模型训练周期(epoch)设置为 200, batch 大小设置为 4。初始学习率设置为 0.01, 学习率动量设置为 0.937。训练过程中使用随机梯度下降(stochastic gradient descent, SGD)作为优化器, 并采用余弦退火学习率调度策略。

为了准确评估所提方法的检测性能, 本文选取平均精度(mean average precision, mAP)、精确率(precision, P)、召回率(recall, R)、F1 分数以及每秒帧数(frames per second, FPS)作为性能评价指标; 同时采用参数量(Parameters)和每秒十亿次浮点运算数(giga floating-point operations, GFLOPs)评价模型复杂度。各评价指标计算如式(7)~式(10)所示:

$$P = \frac{TP}{TP + FP} \tag{7}$$

$$R = \frac{TP}{TP + FN} \tag{8}$$

$$F1 = \frac{2 \times P \times R}{P + R} \tag{9}$$

$$mAP = \frac{\sum_{i=1}^k \int_0^1 P_i(R) dR}{k} \tag{10}$$

式中, P 表示精确率; T 表示召回率; F1 表示精确率与召回率的调和平均值; TP(true positive)表示正确

检测的目标数量; FP(false positive)表示误检目标数量; FN(false negative)表示漏检目标数量; $P_i(R)$ 表示第 i 类目标 Precision-Recall 曲线对应的精确率函数; k 表示目标类别总数。

由于精确率和召回率分别反映模型不同方面的检测性能, 单独使用某一指标难以全面评价检测效果, 因此采用 F1 分数综合衡量模型性能。当精确率和召回率均较高时, F1 分数也相应较高, 说明模型具有较好的综合检测能力。mAP₅₀ 表示在交并比 IoU(intersection over union)阈值为 0.5 时各类别平均精度的均值; mAP₅₀₋₉₅ 表示 IoU 阈值从 0.50 到 0.95(步长为 0.05)范围内平均精度的均值。

IoU 用于衡量预测框与真实框之间的重叠程度, 其计算公式如式(11)所示:

$$IoU = \frac{A \cap B}{A \cup B} \tag{11}$$

式中, A 表示预测框区域, B 表示真实框区域, $A \cap B$ 表示预测框与真实框的交集面积, $A \cup B$ 表示预测框与真实框的并集面积。

2.3 消融实验

为评估各改进模块在整体网络结构中的作用,

表 1 改进模块的消融实验
Table 1 Ablation Study of the Improved Modules (√ indicates use of the module)

Model						Para/ (M)	GFLOPs	Preci- sion/(%)	Recall/ (%)	F1/ (%)	mAP ₅₀ / (%)	mAP ₅₀₋₉₅ / (%)	FPS/ (f/s)
4D	SIBlock	GhostConv	ADown	BifpnCat	DASI								
						9.4	21.3	83.3	65.6	73.4	77.4	59.4	159.9
√						9.8	30.7	83.3	66.3	73.8	77.4	59.8	144.2
√	√					14.7	56.2	82.6	<u>68.1</u>	74.6	78.0	60.7	119.8
√	√	√				13.8	53.9	83.7	67.2	74.5	77.8	60.4	123.0
√	√	√	√			13.7	53.4	83.9	68.2	75.2	78.4	61.0	118.6
√	√	√	√	√		13.7	53.4	84.5	67.8	75.2	78.3	60.7	118.2
√	√	√	√	√	√	16.0	54.8	<u>84.2</u>	68.2	75.4	78.8	61.4	<u>118.7</u>
√	√		√	√	√	16.6	56.1	83.9	68.2	<u>75.2</u>	<u>78.7</u>	<u>61.3</u>	115.6
√	√	√	√		√	16.2	55.8	84.2	68.2	75.4	78.3	60.7	117.2

注: √表示采用对应模块, 加粗字体为每列最优值, 下划线表示每列次优值。

本文采用逐步引入模块的累加式消融实验设计, 即每一步仅在前一网络配置基础上增加一个模块, 其余训练策略与超参数保持一致。该设计旨在分析各模块在当前结构体系中的边际影响及其协同

效果, 而非孤立评估单一模块的独立性能。本文以 YOLO11 为基线模型, 在 HazyDet 数据集上依次引入第四个检测头、SIBlock、GhostConv、ADown、BifpnCat 以及 DASI 注意力机制, 实验结果如表 1 所示。

表 2 颈部网络拼接层输入改进的消融实验
Table 2 Ablation Study on Modified Concatenation Input in the Neck

Model	Para/ (M)	GFLOPs	Preci- sion/(%)	Recall/ (%)	F1/(%)	mAP ₅₀ / (%)	mAP ₅₀₋₉₅ / (%)
SD-YOLO11 (Baseline)	15.6	53.7	83.4	68.0	74.9	78.3	60.9
SD-YOLO11 (Modified Concatenation Input)	16.0	54.8	84.2	68.2	75.4	78.8	61.4

注:加粗字体为每列最优值。

在消融实验中,原始 YOLO11 模型的基准性能为 mAP₅₀77.4% 和 mAP₅₀₋₉₅59.4%。在引入第四个检测头后,模型的 mAP₅₀ 保持不变,而 mAP₅₀₋₉₅ 提升了 0.4%,表明额外检测头在一定程度上增强了模型在不同 IoU 阈值下的检测稳定性,尤其对小目标相关场景具有积极影响。尽管该改动带来了一定的计算开销增加,但从表 1 所示的 FPS 结果来看,整体推理速度仍能够满足实时检测需求。

进一步引入 GhostConv 后,模型参数量和计算量明显下降,其中参数量减少约 0.9M,GFLOPs 降低约 2.3G,而检测性能基本保持稳定。结合完整模型的移除式对照实验可以看出,去除 GhostConv 后参数量与计算量均有所增加,并伴随检测精度下降,说明该模块能够有效降低冗余计算并维持特征表达能力。

随后在引入 ADown 模块后,模型的 mAP₅₀ 提升至 78.4%,mAP₅₀₋₉₅ 提升至 61.0%,反映出该模块在多尺度特征下采样过程中对信息保留具有一定的积极作用。进一步加入 BifpnCat 模块后,通过对完整模型进行移除式对照可以发现,去除该模块会导致 mAP₅₀ 与 mAP₅₀₋₉₅ 均出现下降,说明改进后的特征融合拓扑在整体结构中对多尺度特征整合具有重要影响。

在完整模型中引入 DASI 注意力机制后,mAP₅₀ 和 mAP₅₀₋₉₅ 分别进一步提升 0.5% 和 0.7%,表明该注意力机制在复杂背景和特征退化条件下能够增强目标相关特征的表达效果。总体来看,随着各模块的逐步引入,模型的 mAP₅₀ 和 mAP₅₀₋₉₅ 相较基线模型分别提升了 1.4 和 2.0 个百分点,体现了所提出结构改进在雾霾场景目标检测任务中的整体有效性。

此外,为进一步评估颈部特征拼接方式改进的影响,本文对比了基线模型(SD-YOLO11, Baseline)与改进拼接输入后的模型(SD-YOLO11, Modified

Concatenation Input),结果如表 2 所示。可以看出,改进后的模型在 mAP₅₀ 和 mAP₅₀₋₉₅ 上均提升约 0.5%,说明通过调整拼接输入来源,在特征融合过程中能够保留更多通道与细节信息,从而改善目标特征的表达能力。尽管该改动带来了参数量与 GFLOPs 的轻微增加,但相对于检测性能的提升,其额外计算开销仍处于可接受范围内。

2.3 对比实验

为了进一步验证改进的 YOLO11 在雾霾图像中的对象检测有效性与合理性,本研究在相同的测试环境下,采用 HazyDet 数据集与当前主流的目标检测算法进行了对比实验,实验结果详见表 3。首先,将本文改进的模型与经典检测模型 YOLOv8(Hermens 等, 2024)、YOLOv9(Wang 等, 2024)和 YOLOv10(Wang 等,2024)进行对比,结果显示,经典 YOLO 模型在雾霾场景下的检测性能(以 mAP₅₀ 和 mAP₅₀₋₉₅ 为主要指标)均低于本文改进后的算法。接下来,将本文改进的模型与目前流行的先进改进算法 YOLO11(Khanam 等, 2024)、YOLO9tr(Youwai 等, 2024)、LeYOLO(lightweight and efficient YOLO)(Hollard 等, 2024)、YOLOv10DS(Ahmed 等, 2023)、CAF-YOLO(Chen 等, 2025)、SCoralDet(Lu 等, 2025)和 MHAF-YOLO(multi branch heterogeneous auxiliary fusion YOLO)(Yang 等, 2025)进行对比。相比之下,本文改进模型在 HazyDet 数据集上的检测性能表现更为出色。具体而言,与 YOLO11、YOLO9tr、LeYOLO、YOLOv10DS、CAF-YOLO、SCoralDet 和 MHAF-YOLO 相比,本文模型的 mAP₅₀ 分别提升了 1.4%、0.9%、1.8%、4.2%、6.3%、1.3% 和 3.6%,而 mAP₅₀₋₉₅ 则分别提高了 2.0%、2.2%、2.5%、6.1%、4.0%、1.0% 和 6.1%。这些数据充分表明,本文提出的改进策略在雾霾场景下能够

有效提高目标检测精度,同时在保证较低计算复杂度的前提下,实现了较高的实时性。通过对各

表 3 HazyDet数据集中的实验结果

Table 3 Experimental Results on the HazyDet Dataset

Model	Para/(M)	GFLOPs	F1/(%)	AP on Test-set					FPS/(f/s)
				Car/(%)	Truck/(%)	Bus/(%)	mAP ₅₀ /(%)	mAP ₅₀₋₉₅ /(%)	
YOLOv8	9.8	23.3	73.2	87.5	57.9	84.6	76.7	59.3	95.6
YOLOv9	6.2	22.1	73.7	87.7	59.1	<u>85.3</u>	77.4	60.0	92.3
YOLOv10	8.0	24.5	72.5	87.4	57.8	83.6	76.3	58.9	135.2
YOLO11	9.4	21.3	73.4	87.7	59.6	84.8	77.4	59.4	110.4
YOLO9tr	10.0	40.2	74.2	88.1	<u>60.1</u>	85.4	<u>77.9</u>	59.2	122.3
LeYOLO	33.6	73.9	73.4	<u>87.8</u>	58.8	84.4	77.0	58.9	97.4
YOLOv10DS	16.5	63.4	75.4	87.0	54.0	82.9	74.6	55.3	<u>128.6</u>
CAF-YOLO	26.6	73.2	68.1	76.6	57.5	83.3	72.5	57.4	107.9
SCoralDet	29.2	73.3	73.5	87.8	60.0	84.7	77.5	<u>60.4</u>	101.5
MHAF-YOLO	15.8	67.6	75.9	86.7	55.8	83.2	75.2	55.3	104.1
SD-YOLO11	16.0	54.8	<u>75.4</u>	90.7	60.3	85.5	78.8	61.4	118.7

注:加粗字体为每列最优值,下划线表示每列次优值。

表 4 Foggy-Cityscapes数据集中的实验结果

Table 4 Experimental Results on the Foggy Cityscapes Dataset

Model	Para/(M)	GFLOPs	F1/(%)	mAP ₅₀ /(%)	mAP ₅₀₋₉₅ /(%)	FPS/(f/s)
YOLOv8	9.8	23.4	49.2	55.4	35.7	95.6
YOLOv9	6.2	22.1	50.1	56.5	<u>37.6</u>	92.3
YOLOv10	8.0	24.5	47.5	53.8	35.4	135.2
YOLO11	9.4	21.3	49.4	54.9	36.6	110.2
PC-YOLO	22.3	98.7	53.9	49.9	30.0	67.3
DNE-YOLO	10.5	25.5	50.4	56.6	37.2	112.6
SCoralDet	29.3	73.3	50.0	<u>56.8</u>	37.4	101.5
MHAF-YOLO	15.8	67.6	50.7	47.0	28.2	104.1
RT-DETRv2	20.1	30.7	54.6	54.5	33.5	94.0
Relation-DETR	18.3	41.1	44.9	42.1	23.5	79.5
DR-YOLO	15.5	31.5	47.3	46.7	28.0	85.8
DA-Detect	17.2	38.8	49.8	45.5	26.8	76.8
SD-YOLO11	16.0	54.8	<u>54.1</u>	58.5	37.7	<u>118.7</u>

注:加粗字体为每列最优值,下划线表示每列次优值。

模型在 mAP₅₀ 和 mAP₅₀₋₉₅ 指标上的综合比较,可以得出结论,本文所提出的改进算法在雾霾环境下的目标检测任务中具有更优的检测性能和更高的实际应用价值。

为避免“仅在单一数据集得出结论”的偏差,并更充分地验证方法在合成域与真实域的适用性,本

文在 HazyDet 之外,新增在 Foggy Cityscapes 与 RTTS 上的对比评测:前者用于在受控雾密度下验证改进带来的精度提升,后者用于评估跨数据集泛化与真实场景下的鲁棒性。除评估对象外,新增对比还涵盖雾霾专用检测器 PC-YOLO (Wang 等, 2025) 与 DNE-YOLO (Wu 等, 2024), 以及基于 Transformer 的

表 5 Foggy-Cityscapes 数据集中各个类别的实验结果
Table 5 Experimental Results for Each Category in the Foggy Cityscapes Dataset

Model	person	rider	car	truck	bus	train	motorcycle	bicycle
YOLOv8	57.6	61.5	75.1	40.6	61.1	52.2	<u>42.3</u>	52.8
YOLOv9	57.2	58.7	75.3	45.0	60.8	<u>61.9</u>	41.2	<u>51.7</u>
YOLOv10	57.8	58.8	74.1	37.8	62.8	47.9	41.6	49.3
YOLO11	58.5	59.2	75.6	39.9	61.1	53.7	38.4	52.7
PC-YOLO	53.2	55.0	72.5	33.9	61.6	44.6	33.0	45.0
DNE-YOLO	58.2	<u>61.5</u>	75.4	40.9	63.8	59.3	42.6	51.1
SCoraldet	<u>59.7</u>	62.4	76.9	38.0	<u>64.3</u>	59.0	41.0	52.8
MHAF-YOLO	50.0	51.9	72.1	31.6	55.5	39.4	33.2	41.9
RT-DETRv2	56.8	60.3	<u>77.3</u>	<u>41.5</u>	63.8	42.7	42.1	51.6
Relation-DETR	49.3	49.4	71.1	24.8	40.5	29.0	33.4	39.4
DR-YOLO	52.1	52.3	73.4	33.2	50.2	44.4	31.6	36.4
DA-Detect	46.5	51.1	65.2	32.4	49.7	34.2	39.4	45.6
SD-YOLO11	60.8	61.4	77.6	40.9	65.3	71.1	40.2	50.7

注:加粗字体为每列最优值,下划线表示每列次优值。

表 6 RTTS 数据集中各个类别的实验结果
Table 6 Experimental Results for Each Category in the RTTS Dataset

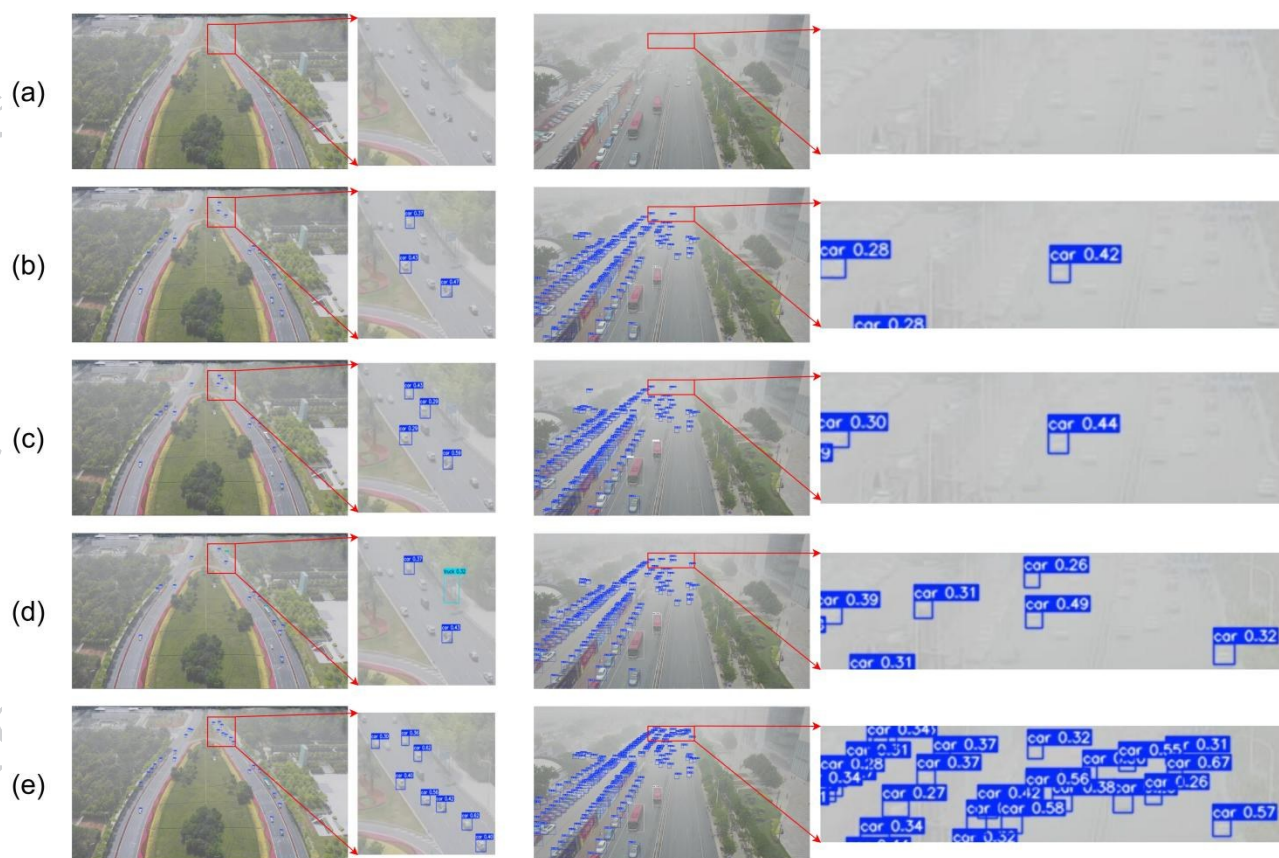
Model	person	bicycle	car	bus	motorcycle	mAP ₅₀ /(%)	mAP ₅₀₋₉₅ /(%)
YOLOv8	<u>84.8</u>	70.3	90.7	68.4	72.3	77.3	57.2
YOLOv9	84.5	70.5	90.1	<u>70.2</u>	73.7	77.8	57.3
YOLOv10	83.7	68.8	<u>89.6</u>	69.9	72.5	76.9	56.7
YOLO11	84.7	67.0	90.5	69.0	<u>74.7</u>	77.2	56.3
PC-YOLO	84.2	64.4	91.7	69.7	73.0	76.6	52.9
DNE-YOLO	84.2	69.2	90.3	68.8	74.4	77.4	<u>57.4</u>
SCoraldet	83.9	66.0	89.8	69.0	72.3	76.2	56.6
MHAF-YOLO	83.8	65.0	90.5	70.1	74.4	76.7	53.1
RT-DETRv2	84.5	<u>70.4</u>	<u>90.8</u>	70.4	74.5	<u>78.1</u>	55.4
Relation-DETR	79.2	51.4	83.3	54.8	56.7	65.1	40.8
DR-YOLO	80.7	60.8	84.9	55.4	<u>57.2</u>	67.8	47.5
DA-Detect	80.6	61.7	85.4	56.1	58.2	68.4	48.0
SD-YOLO11	85.1	70.8	92.3	70.4	74.8	78.7	57.9

注:加粗字体为每列最优值,下划线表示每列次优值。

检测器 RT-DETRv2 (real-time detection transformer) (Lv 等, 2024)、Relation-DETR (Hou 等, 2025)、DR-YOLO (dehazing and reasoning YOLO) (Zhong 等, 2024) 和 DA-Detect (domain adaptive object detection)

(Li 等, 2023)。

Foggy Cityscapes 以 Cityscapes 的精标注为基础,在受控雾密度下合成图像,能够在标注一致、几何不变的条件考察雾化对检测的影响,弥补仅在



LeYOLO. (d) SCoralDet. (e) SD-YOLO11

图7 HazyDet数据集的检测结果。(a) 输入图像;(b) YOLO11;(c) LeYOLO;(d) SCoralDet;(e) SD-YOLO11

Fig. 7 Detection results on the HazyDet dataset. (a) Input image. (b) YOLO11. (c)

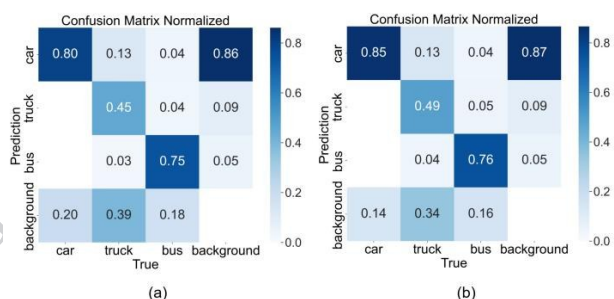
单一数据集下评估带来的局限。表4报告整体 mAP_{50} 与 mAP_{50-95} , 表5进一步给出了各类别检测结果。可以看出, SD-YOLO11 在整体检测指标上均优于现有主流方法, 并在 person、rider、car、bus、train 及 motorcycle 等多数类别上取得了较好的检测性能。不同类别的性能增益存在一定差异, 这主要与目标尺寸、纹理特征及雾霾退化程度等因素有关。其中, bicycle 类目标通常尺寸较小、形态细长, 在复杂雾霾背景下更容易受到遮挡和背景干扰, 因此类别间性能变化具有一定的合理性。综合来看, 所提方法在整体指标及多数类别上均表现出较好的检测性能与鲁棒性。

RTTS 由真实雾霾交通场景构成, 覆盖多地点、多能见度与多拍摄条件, 用于检验模型在真实场景中的跨数据集泛化能力。表6给出各类别 AP_{50} 与整体 mAP 指标。SD-YOLO11 在两个 mAP 指标上均取得优于其他对比方法的结果, 同时在车辆与行人相关类别上表现稳健, 说明本文方法在真实雾霾域具有良好的迁移效果。

为了更清晰地展示低分辨率与模糊目标的检测效果, 图7中红色框标注的区域为对应可视化的局部放大结果。在 HazyDet 数据集的实验中, 当场景复杂且目标分布密集或目标在图像中分辨率较低时, YOLO11 的检测效果存在明显不足, 容易出现漏检和误检等问题。例如, 图7a展示了低分辨率目标、因浓雾遮挡而模糊的目标以及边缘模糊的目标, 而图7b、7c和7d中均可观察到漏检现象。这表明其他算法在处理雾霾复杂场景时, 无法充分捕捉微小和模糊目标的信息。相比之下, 本文所提出的改进模型在图7e中有效解决了这些问题。通过对模型的改进, 使其在特征提取和多尺度融合方面得到了显著增强, 从而提高了低分辨率目标和被浓雾遮挡目标的检测能力, 显著减少了漏检和误检情况。这些可视化结果充分验证了所提出改进方法在雾霾环境下目标检测任务中的优越性和实用性。

图8展示了SD-YOLO11在HazyDet数据集上的检测性能。与YOLO11相比, 混淆矩阵在最后一行显示了更浅的阴影, 这表明减少了作为背景的对象

的错误分类。然而,在准确识别汽车上面仍然存在挑战,这些汽车往往被错误地标记为背景。尽管存在这些问题,SD-YOLO11 在主对角线上显示出较暗的阴影,表明正确检测到的物体总体上有所增加。

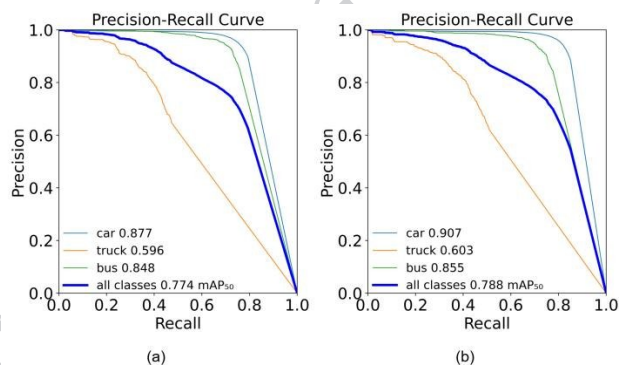


(b) Confusion matrix of SD-YOLO11

图8 混淆矩阵可视化对比(a)YOLO11的混淆矩阵;(b)SD-YOLO11的混淆矩阵

Fig. 8 Confusion matrix visualization comparison. (a) Confusion matrix of YOLO11.

图9展示了基础模型与改进模型在 HazyDet 数据集上每个类以及总的 mAP_{50} 值和 P-R 曲线,每个类别对应的曲线按数值排列,标注在图中。从图中曲线的面积可知,每个类实现不同程度的提升,彰显改进模型良好的性能。



curves of SD-YOLO11

图9 P-R 曲线图可视化对比(a)YOLO11的P-R 曲线图;(b)SD-YOLO11的P-R 曲线图

Fig. 9 P-R curve visualization comparison. (a) P-R curves of YOLO11. (b) P-R

3 讨论

本文通过构建改进型检测网络 SD-YOLO11,提升了模型在雾霾等低能见度环境下对模糊目标和小

目标的检测能力。实验结果表明,各项改进有效缓解了低对比度、信息缺失及多尺度干扰带来的检测困难,在多个评价指标上均优于现有主流方法,验证了所提方法在复杂雾霾场景中的有效性与泛化能力。

在主干网络方面,SIBlock 通过递归卷积不断聚合高阶空间上下文信息,增强了模型对低对比度与弱纹理目标的特征建模能力。与普通卷积主要依赖固定局部感受野不同,递归卷积能够在多次空间交互过程中逐步扩大有效感受范围,从而更充分地捕获模糊边界与弱纹理区域中的连续结构信息。在雾霾场景下,目标边缘往往因雾气散射而呈现低对比度与轮廓模糊的特点,普通卷积容易将目标区域与背景区域混合建模,导致特征响应减弱。而 SIBlock 通过高阶空间交互能够增强邻域间的结构关联性,使网络更加关注目标轮廓的连续变化,从而提升模型对模糊目标与小目标的判别能力。Rao 等人(2022)指出,递归卷积能够以有限复杂度实现高阶空间交互,相比普通卷积具有更强的复杂空间建模能力。

为降低模型计算冗余,本文引入 GhostConv 与 ADown 模块。GhostConv 通过生成冗余特征映射替代部分传统卷积操作,在减少参数量与计算量的同时保留关键特征信息。Cao 等人(2023)通过 GhostConv 实现了轻量化与检测精度的平衡。ADown 则在下采样阶段自适应保留更具判别性的通道特征,有助于提升雾霾场景下小目标的信息保留能力(Huo 等,2024)。

在颈部网络方面,BifpnCat 通过双向特征融合与通道拼接策略保留更多语义与细节信息,增强了不同尺度特征间的信息交互能力,从而提升模型对低分辨率和模糊目标的感知能力。John 等人(2025)指出,BiFPN 结构能够有效增强多尺度特征融合能力。此外,DASI 注意力机制进一步强化了模型对关键目标区域的关注能力,可有效抑制复杂背景干扰,提升低可视度场景中的目标提取能力(Xu 等,2024)。

同时,重新设计的拼接路径增强了深层语义与浅层细节之间的信息流动,减少了特征融合过程中的信息损失,进一步提升了模型整体表达能力。实验结果表明,“结构改进+注意力机制”的协同设计能够有效增强模型在复杂天气场景中的检测性能,也

验证了从网络结构优化角度提升低质量图像检测能力的可行性。

尽管本文方法在检测精度方面取得了较好效果,但在图7中的部分边缘区域,由于目标特征过于模糊,仍存在漏检现象。此外,SIBlock与BifpnCat的引入增加了网络复杂度,导致模型推理速度受到一定影响。针对上述问题,后续工作可考虑结合Laplacian边缘算子增强模糊边界信息(Bui等,2024),并借鉴LitE-SNN(lightweight and efficient spiking neural network)中的结构压缩与联合优化策略(Liu等,2024),在保持特征表达能力的同时进一步降低计算冗余,实现更加轻量化和高效的目标检测。

4 结 论

针对雾霾场景下目标检测面临的低对比度、复杂背景和小目标漏检等问题,本文提出了一种改进目标检测网络——SD-YOLO11。首先在主干网络部分加入了SIBlock模块,使模型性能提升了0.6%,验证了该模块有效提升了模型对低对比度及模糊目标的识别能力;其次,为了在保证检测精度的同时平衡检测速度与模型复杂度,本文引入了GhostConv和ADown模块,模型GFLOPs下降了2.8,同时检测准确率提升了0.4%,表明这两种模块在降低计算量和参数量的同时,确保了有效的特征提取。随后,在颈部网络部分,本文改进了BifpnCat模块并添加了DASI注意力机制,从而进一步增强了多尺度特征融合和关键区域的注意能力,准确率进一步提升了0.4%。这些改进使得SD-YOLO11在雾霾场景下能够稳定提升检测精度同时保证实时性,验证了各结构改进在复杂低能见度环境中的有效性。实验结果表明,在HazyDet、Foggy Cityscapes与RTTS三个具有代表性的雾霾数据集上,SD-YOLO11均取得了优良的检测性能,其中在HazyDet数据集上 mAP_{50} 和 mAP_{50-95} 分别达到了78.8%和61.4%。未来的研究将进一步关注边缘化图像的处理,以提升目标检测算法在极端环境下的适应性和泛化能力。

参考文献(References)

Ahmed A and Manaf A. 2024. Pediatric wrist fracture detection in X-

rays via YOLOv10 algorithm and dual label assignment system[EB/OL]. (2024-07-22)[2025-09-10].

<https://arxiv.org/abs/2407.15689>

Binbusayyis A. 2024. Hybrid VGG19 and 2D-CNN for intrusion detection in the fog-cloud environment. *Expert Systems with Applications*, 238 (A): 121758 [DOI: 10.1016/j.eswa.2023.121758]

Bochkovskiy A, Wang C Y and Liao H Y M. 2020. YOLOv4: Optimal speed and accuracy of object detection [EB/OL]. (2020-04-23) [2025-09-10].

<https://arxiv.org/abs/2004.10934>

Bui N T, Hoang D H, Nguyen Q T, Tran M T and Le N. 2024. Meganet: Multi-scale edge-guided attention network for weak boundary polyp segmentation//*Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. Waikoloa, HI, USA: IEEE: 7985-7994

Cao J S, Bao W, Shang H, Yuan M and Cheng Q. 2023. GCL-YOLO: A GhostConv-based lightweight YOLO network for UAV small object detection. *Remote Sensing*, 15 (20): 4932 [DOI: 10.3390/rs15204932]

Chen D, Dai L, Yue X, Gu Q, Huang S, Pan J, et al. 2025. AW-YOLO: A multi-object detection network for autonomous driving under all weather conditions. *IET Image Processing*, 19: e70111 [DOI: 10.1049/ipr2.70111]

Chen Z and Lu S. 2025. CAF-YOLO: A robust framework for multi-scale lesion detection in biomedical imagery//2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Hyderabad, India: IEEE: 1-5 [DOI: 10.1109/ICASSP49660.2025.10888358]

Cheng Y, Wang W, Zhang W, Yang L, Wang J, Ni H, et al. 2023. A multi-feature fusion and attention network for multi-scale object detection in remote sensing images. *Remote Sensing*, 15 (8): 2096 [DOI: 10.3390/rs15082096]

Ding Y and Luo X. 2024. SDNIA-YOLO: A robust object detection model for extreme weather conditions [EB/OL]. (2024-06-19) [2025-09-10].

<https://arxiv.org/abs/2406.12395>

Doherty J, Gardiner B, Kerr E and Siddique N. 2025. BiFPN-YOLO: One-stage object detection integrating bi-directional feature pyramid networks. *Pattern Recognition*, 160: 111209 [DOI: 10.1016/j.patcog.2024.111209]

Feng C, et al. 2026. HazyDet: Open-source benchmark for drone-view object detection with depth cues in hazy scenes. *IEEE Transactions on Geoscience and Remote Sensing*, 64: 5006414 [DOI: 10.1109/TGRS.2026.3701819]

Feng Q H, Wang Z X, Sun C C and Shao Z W. 2025. Small object detection in drone images via foreground refinement and multidimensional inductive bias self-attention. *Journal of Image and Graphics*, 30 (11): 3547-3563 (冯琪涵,王志晓,孙成成,邵志文. 2025. 融

- 合前景细化和多维归纳偏置自注意力的无人机图像小目标检测. 中国图象图形学报, 30(11): 3547-3563 [DOI:10.11834/jig.250017]
- Fu Y, Cheng Y, Cao S Y, Tan L, He J, Wang M and Zhang W. 2026. DTL-IceNet: A dual-task learning architecture with multi-scale fusion mechanisms for enhanced ice detection on transmission lines. *Atmospheric Measurement Techniques*, 19: 3407-3426 [DOI:10.5194/amt-19-3407-2026]
- Ge Z, Liu S T, Wang F, Li Z M and Sun J. 2021. YOLOX: Exceeding YOLO series in 2021 [EB/OL]. (2021-07-17) [2025-09-10]. <https://arxiv.org/abs/2107.08430>
- Girshick R, Donahue J, Darrell T and Malik J. 2014. Rich feature hierarchies for accurate object detection and semantic segmentation// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Columbus, OH, USA: IEEE: 580-587
- Girshick R. 2015. Fast R-CNN//*Proceedings of the IEEE International Conference on Computer Vision*. Santiago, Chile: IEEE: 1440-1448
- Gupta H, Kotlyar O, Andreasson H and Lilienthal A J. 2024. Robust object detection in challenging weather conditions//*Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. Waikoloa, HI, USA: IEEE: 7523-7532
- Han K, Wang Y H, Tian Q, Guo J Y, Xu C J and Xu C. 2020. GhostNet: More features from cheap operations//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, WA, USA: IEEE: 1580-1589
- He J X, et al. 2024. EC-YOLOX: A deep-learning algorithm for floating objects detection in ground images of complex water environments. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17: 7359-7370 [DOI: 10.1109/JSTARS.2024.3367713]
- Hermens F. 2024. Automatic object detection for behavioural research using YOLOv8. *Behavior Research Methods*, 56 (7) : 7307-7330 [DOI:10.3758/s13428-024-02420-5]
- Hollard L, Mohimont L, Gaveau N, et al. 2024. LeYOLO: New scalable and efficient CNN architecture for object detection [EB/OL]. (2024-06-25) [2025-09-10]. <https://arxiv.org/abs/2406.14239>
- Hou X, Liu M, Zhang S, Wei P, Chen B and Lan X. 2025. Relation DETR: Exploring explicit position relation prior for object detection//*Computer Vision - ECCV 2024. Lecture Notes in Computer Science*, Vol. 15108. Cham: Springer: 95-113 [DOI:10.1007/978-3-031-72973-7_6]
- Huo Y, et al. 2025. FA-YOLO: Research on efficient feature selection YOLO improved algorithm based on FMDS and AGMF modules// *2025 International Joint Conference on Neural Networks (IJCNN)*. Rome, Italy: IEEE: 1-8 [DOI: 10.1109/IJCNN64981.2025.11227997]
- Jocher G, Chaurasia A, Stoken A, et al. 2022. Ultralytics/YOLOv5: v7.0—YOLOv5 SOTA realtime instance segmentation [EB/OL]. Zenodo. (2022-11-06) [2025-09-10]. [DOI:10.5281/zenodo.3908559]
- Keerthi Kiran V, Dash S and Parida P. 2024. Vehicle detection in varied weather conditions using enhanced deep YOLO with complex wavelet. *Engineering Research Express*, 6(2): 025224 [DOI: 10.1088/2631-8695/ad507d]
- Khanam R and Hussain M. 2024. YOLO11: An overview of the key architectural enhancements [EB/OL]. (2024-10-25) [2025-09-10]. <https://arxiv.org/abs/2410.17725>
- Kumar D and Muhammad N. 2023. Object detection in adverse weather for autonomous driving through data merging and YOLOv8. *Sensors*, 23(20): 8471 [DOI:10.3390/s23208471]
- Li C, Li L, Jiang H, et al. 2022. YOLOv6: A single-stage object detection framework for industrial applications [EB/OL]. (2022-09-06) [2025-09-10]. <https://arxiv.org/abs/2209.02976>
- Li C Y, Li B, Xu X, Zhang Z, Xu Z and Feng J, et al. 2023. Detection-friendly dehazing: Object detection in real-world hazy scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(7): 8284-8295 [DOI:10.1109/TPAMI.2023.3234976]
- Li J L, Xu R S, Ma J, Zou Q, Ma J Q and Yu H K. 2023. Domain adaptive object detection for autonomous driving under foggy weather// *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. Waikoloa, HI, USA: IEEE: 612-622
- Liu C X, Hu P J, Li S C, Wang C H and Ling Y. 2025. Decoupled triple-stage enhancement network for non-homogeneous image dehazing. *Journal of Image and Graphics*, 30(1): 83-94 (刘春晓, 胡鹏靖, 厉世昌, 王成骅, 凌云. 2025. 面向非均质图像去雾的解耦合三阶段增强网络. 中国图象图形学报, 30(1): 83-94) [DOI:10.11834/jig.240069]
- Liu H, Zhang W, Cheng Y, He J, Shao H, Bai S, et al. 2025. Shrub extraction in arid regions based on feature enhancement and transformer network from high-resolution remote sensing images. *Forests*, 16: 1288 [DOI:10.3390/f16081288]
- Liu Q H, Yan J, Zhang M, et al. 2024. LITE-SNN: Designing lightweight and efficient spiking neural network through spatial-temporal compressive network search and joint optimization [EB/OL]. (2024-01-29) [2025-09-10]. <https://arxiv.org/abs/2401.14652>
- Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C Y, et al. 2016. SSD: Single shot MultiBox detector//*Computer Vision - ECCV 2016. Lecture Notes in Computer Science*, Vol. 9905. Cham: Springer: 21-37 [DOI:10.1007/978-3-319-46448-0_2]
- Lu Z X, Liao L C, Xie X G and Yuan H. 2025. SCoralDet: Efficient real-time underwater soft coral detection with YOLO. *Ecological Informatics*, 85: 102937 [DOI:10.1016/j.ecoinf.2024.102937]
- Lv W Y, et al. 2024. RT-DETRv2: Improved baseline with bag-of-

- freebies for real-time detection transformer[EB/OL]. (2024-07-25) [2025-09-10].
<https://arxiv.org/abs/2407.17140>
- Meng X, Liu Y, Fan L and Fan J. 2023. YOLOv5s-Fog: An improved model based on YOLOv5s for object detection in foggy weather scenarios. *Sensors*, 23(11): 5321 [DOI:10.3390/s23115321]
- Rao Y M, Zhao W L, Tang Y S, Zhou J, Lim S N and Lu J W. 2022. HorNet: Efficient high-order spatial interactions with recursive gated convolutions. *Advances in Neural Information Processing Systems*, 35: 10353-10366
- Rasheed A F and Zarkoosh M. 2025. YOLO11 optimization for efficient resource utilization. *The Journal of Supercomputing*, 81: 1085 [DOI:10.1007/s11227-025-07520-3]
- Redmon J, Divvala S, Girshick R and Farhadi A. 2016. You only look once: Unified, real-time object detection//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, NV, USA: IEEE: 779-788
- Redmon J and Farhadi A. 2017. YOLO9000: Better, faster, stronger//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, HI, USA: IEEE: 7263-7271
- Redmon J and Farhadi A. 2018. YOLOv3: An incremental improvement [EB/OL]. (2018-04-08) [2025-09-10].
<https://arxiv.org/abs/1804.02767>
- Ren S Q, He K M, Girshick R and Sun J. 2015. Faster R-CNN: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems*, 28
- Tan M X, Pang R M and Le Q V. 2020. EfficientDet: Scalable and efficient object detection//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, WA, USA: IEEE: 10781-10790
- Wang A, Chen H, Liu L H, Chen K, Lin Z J and Han J G, et al. 2024. YOLOv10: Real-time end-to-end object detection. *Advances in Neural Information Processing Systems*, 37: 107984-108011 [DOI:10.52202/079017-3429]
- Wang C Y, Bochkovskiy A and Liao H Y M. 2023. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Vancouver, BC, Canada: IEEE: 7464-7475
- Wang C Y, Yeh I H and Liao H Y M. 2025. YOLOv9: Learning what you want to learn using programmable gradient information//*Computer Vision - ECCV 2024. Lecture Notes in Computer Science*, Vol. 15089. Cham: Springer: 3-20 [DOI: 10.1007/978-3-031-72751-1_1]
- Wang Y, Ke H P and Cai H H. 2025. PC-YOLO: Enhancing object detection in adverse weather through physics-aware and dynamic network structures. *Journal of Electronic Imaging*, 34(2): 023049 [DOI:10.1117/1.JEI.34.2.023049]
- Wu H T, Mo X T, Wen S J, Wu K L, Ye Y and Wang Y M, et al. 2024. DNE-YOLO: A method for apple fruit detection in diverse natural environments. *Journal of King Saud University - Computer and Information Sciences*, 36(9): 102220 [DOI:10.1016/j.jksuci.2024.102220]
- Xia B, Xue W T, Zhou X Y and Huang H. 2025. Remote sensing object detection via local parameter-free attention and combined loss. *Journal of Image and Graphics*, 30(3): 683-695 (夏波, 薛卫涛, 周新尧, 黄鸿. 2025. 局部无参注意力和联合损失的遥感目标检测. 中国图象图形学报, 30(3): 683-695) [DOI: 10.11834/jig.240316]
- Xu S, et al. 2024. HCF-Net: Hierarchical context fusion network for infrared small object detection//2024 IEEE International Conference on Multimedia and Expo (ICME). Niagara Falls, ON, Canada: IEEE: 1-6 [DOI: 10.1109/ICME57554.2024.10687431]
- Yang Z, Guan Q, Yu Z, et al. 2025. MHAF-YOLO: Multi-branch heterogeneous auxiliary fusion YOLO for accurate object detection[EB/OL]. (2025-02-09) [2025-09-10].
<https://arxiv.org/abs/2502.04656>
- Yi L, Yin M and Darbandi M. 2023. A deep and systematic review of the intrusion detection systems in the fog environment. *Transactions on Emerging Telecommunications Technologies*, 34(1): e4632 [DOI: 10.1002/ett.4632]
- Youwai S, Chaiyaphat A and Chaipetch P. 2024. YOLO9tr: A lightweight model for pavement damage detection utilizing a generalized efficient layer aggregation network and attention mechanism. *Journal of Real-Time Image Processing*, 21: 163 [DOI: 10.1007/s11554-024-01545-2]
- Yue Y H and Lei W P. 2025. Joint foggy-road environment perception algorithm based on improved CycleGAN and YOLOv8. *Journal of South China University of Technology (Natural Science Edition)*, 53(2): 48-57 (岳永恒, 雷文朋. 2025. 基于改进 CycleGAN 和 YOLOv8 联合雾天道路环境感知算法. 华南理工大学学报(自然科学版), 53(2): 48-57)
- Zhang K, Yan X, Wang Y and Qi J. 2023. Adaptive dehazing YOLO for object detection//*Artificial Neural Networks and Machine Learning - ICANN 2023. Lecture Notes in Computer Science*, Vol. 14260. Cham: Springer: 19-31 [DOI:10.1007/978-3-031-44195-0_2]
- Zhong F J, Shen W X, Yu H, Wang G Y and Hu J. 2024. Dehazing & reasoning YOLO: Prior knowledge-guided network for object detection in foggy weather. *Pattern Recognition*, 156: 110756 [DOI: 10.1016/j.patcog.2024.110756]

作者简介

张卓成,男,硕士研究生在读,主要研究方向为深度学习与计算机视觉。E-mail:18332042262@163.com

孟逸飞,男,硕士研究生在读,主要研究方向为水电站自动化智能化系统,水电站设备智能异常监测。E-mail:mengy-

ifei@live.com

李超凡,男,科技产业处副处长,主要研究方向为资源环境遥感应用。E-mail:lcf@nuist.edu.cn

吴长春,男,华东区总经理,主要研究方向为多感融合在车路云一体化领域的应用。E-mail:changchunchina@163.com