

DOI:10.19651/j.cnki.emt.2416950

自适应复杂环境噪声的多重关注联合优化检测算法*

张绪康¹ 朱 硕²

(1.南京信息工程大学电子与信息工程学院 南京 210044; 2.无锡学院电子信息工程学院 无锡 214105)

摘要: 针对自动驾驶车辆视觉感知系统在雾天、雨天等复杂天气下因环境噪声导致目标检测效果不佳的问题,提出基于自适应图像去噪与多重关注的联合优化目标检测算法(DMC-YOLO)。构建一个图像去噪网络,融合暗通道先验算法和ACE图像增强技术模块,提升复杂天气下的图像质量;进一步地,将该网络与YOLOv8主干网络相连,并在YOLOv8网络中运用SCDown卷积代替标准卷积,集成点卷积与深度卷积,降低网络计算成本,同时获得更丰富的下采样信息;采用SEAM注意力模块,整合网络局部信息和全局信息;引入SA检测头,广泛关注上下文特征以保留更多细节信息;在损失函数中引入线性区间映射重构IoU,以提升网络对于不同复杂环境的适应性。实验结果表明,相较于基线模型,改进算法在参数量降低15%的情况下,平均精度提升2.9%,有效增强了自动驾驶车辆在复杂环境下对目标的识别能力,在EC-R3588SPC和Nvidia Jetson NX边缘设备上部署效果良好,可以满足复杂天气下的实时检测需求。

关键词: 目标检测;联合优化;YOLOv8;复杂天气;图像去噪

中图分类号: TP391.4;TN919.8 **文献标识码:** A **国家标准学科分类代码:** 510.1040

Multi-attention joint optimization detection algorithm for adapting to complex environmental noise

Zhang Xukang¹ Zhu Shuo²

(1. Electronic and Information Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China;

2. College of Electronic Information Engineering, Wuxi University, Wuxi 214105, China)

Abstract: To address the issue of inadequate target detection performance in the visual perception systems of autonomous vehicles, particularly under complex weather conditions such as fog and rain that introduce environmental noise, we propose a joint optimization target detection algorithm based on adaptive image denoising and multiple attention mechanisms (DMC-YOLO). An image denoising network has been constructed that combines the dark channel prior algorithm with ACE image enhancement technology to improve image quality in challenging weather conditions. Additionally, this network is integrated with the YOLOv8 backbone, utilizing SCDown convolution to replace standard convolution. By incorporating point convolution and depth convolution, the aim is to reduce computational costs while obtaining richer down-sampling information. The SEAM attention module is employed to merge local and global information within the network. Furthermore, the SA detection head is introduced to emphasize contextual features, allowing for the retention of more detailed information. To enhance the network's adaptability to various complex environments, linear interval mapping is incorporated into the loss function for reconstructing IoU. Experimental results indicate that, compared to the baseline model, the average accuracy of the improved algorithm increases by 2.9% while reducing the number of parameters by 15%. This effectively enhances the ability of autonomous vehicles to recognize targets in complex environments. The deployment outcomes on EC-R3588SPC and Nvidia Jetson NX edge devices are promising, fulfilling real-time detection requirements even under challenging weather conditions.

Keywords: target detection; joint optimization; YOLOv8; complicated weather; image denoising

0 引 言

随着科技的不断进步,无人驾驶技术受到了工业界和

学术界的广泛关注。目标检测是无人驾驶技术中的重要一环,它通过识别和定位图像或视频中的目标物体,确保自动驾驶的安全与高效。然而,在恶劣天气条件下,能见度降低

收稿日期:2024-09-21

* 基金项目:江苏双创博士基金(JJSSCBS20210871)项目资助

等环境因素导致拍摄到的图像退化模糊,物体遮挡以及小目标像素比例较低也使得实时检测的准确性降低,削弱了视觉算法的整体性能,从而给车辆行驶带来巨大的风险。因此,如何有效预防和减少复杂天气环境影响所带来的问题具有重要意义^[1-2]。

目标检测算法分为传统算法和深度学习算法,传统算法由于依赖人工设计的复杂特征,无法完全准确的表示图像中的目标,以至检测精度低,实时精度不佳^[3]。随着图像处理与目标检测技术的快速发展,深度学习算法逐渐成熟,常见的目标检测算法按照算法的执行流程,主要可以分为单阶段检测和双阶段检测两个类别,两阶段检测法如区域卷积神经网络(region-based convolutional neural network, R-CNN)^[4],快速区域卷积神经网络(faster region-based convolutional neural network, Faster R-CNN)^[5],它基于候选框策略,将特征提取、区域提取和分类器都整合在一个网络中,使得综合性能有了很大的提高,但检测速度较慢;另一种是单阶段检测法,如实时 Transformer 检测器(real-time detection transformer, RT-DETR)^[6], YOLO 系列(you only look once, YOLO)^[7],单步骤多框检测器(single shot multibox detector, SSD)^[8]等算法,单阶段检测法基于回归方法,使用卷积神经网络结构直接从输出图像中预测回归边框和类别概率,极大地提高了检测的速度,可以实时对物体进行分类检测。

为提高复杂天气场景下检测性能,Chen 等^[9-10]提出一种基于领域适应(domain adaptation, DA)的目标检测算法模型,模型以 Faster R-CNN 作为目标检测框架,引入自注意力层(self adaptation, SA),有效地捕捉源领域和目标领域之间的相似性和差异性,提升了目标检测的适应性和准确性,但由于二阶段算法参数量较大,检测效率低。乔石丽等^[11-13]设计了全局相似性注意力模块(global similarity attention module, GSIMAM)模块来增强 YOLOv5 的特征提取能力,通过双向特征金字塔网络实现多尺度特征融合,并前置 Swin Transformer 模型优化去雨图像,在 BDD100K 数据集下识别精度相较于基线模型提升 4.2%,但在实际检测场景复杂多变,模型仍会出现适应性弱,执行速度较慢等问题。Hnewa 等^[14]以 YOLOv4 为基础,构造领域自适应网络,使用多个域自适应路径和相应的域分类器来生成域不变特征,使检测器能够在不同层产生领域不变的特征。Zhou 等^[15]利用 Mean Teacher 模型调整知识提炼框架,来帮助学生模型从未标记的目标领域中提取实例级特征,并利用场景风格转换在噪声图像领域交叉生成伪图像,弥补图像层次的差异。关家志^[16]结合不同天气图像和无监督图像转换框架训练额外编码模型,通过编码模型提取图像鲁棒特征,将提取到的鲁棒特征与主干网络的特征进行融合,增强了其在复杂环境下的适应能力,但显著增加了检测模型的算法复杂度,不利于模型应用。Li 等^[17]设计了具有注意力融合模块和自监督损失的双分支结构,注

意力模块关注重要信息,自监督损失提升模块在不同雾气浓度下的鲁棒性,联合训练去雾和检测。

在图像检测中,密集的小目标信息难以捕捉,多尺度目标信息分布不均以及目标间相互遮挡等是导致模型误检和漏检的重要原因,上述研究结合多种方法改善恶劣天气环境下的目标检测性能,这些方法主要针对检测网络模块进行优化,在复杂环境下降低误检和漏检问题,但在复杂环境下算法适应能力依旧会受到影响,尤其在多样化的环境噪声中,检测难度将大大提升,给检测模型带来巨大挑战。因此,本文选取 YOLOv8 作为基础网络,提出基于自适应图像去噪与多重关注的联合优化目标检测算法(denoising and multiple concern-based YOLOv8, DMC-YOLO),具体贡献如下:

1)结合暗通道先验算法和自动色彩均衡图图像增强技术,采用图像还原算法中的基于大气散射模型的暗通道先验算法对图像进行去噪处理,然后校正目标及其周围像素点的像素值,提升图像质量,降低模型检测难度。

2)在 YOLOv8 模型中引入空间-通道解耦下采样卷积,先利用点卷积调整通道维数,再送入深度卷积进行空间下采样,在减少模型计算量的前提下,极大丰富了下采样信息。

3)针对图像中目标间相互遮挡而产生的误检漏检问题,提出混合注意力机制,通过深度可分离卷积深度逐一通道进行卷积,以学习不同通道重要特征并减少参数量,然后使用两层全连接网络融合每个通道的信息,增强所有通道之间的联系。

4)引入多头注意力检测头,其自注意力模块使网络能够捕捉更为广泛的上下文信息,卷积模块对自注意力关注的重要信息进行整合,并减少自注意力机制产生的计算复杂度。

5)在损失函数中引入线性区间映射机制,重构交并比(intersection over union, IoU)以关注各类回归样本,动态对回归样本进行质量评估,使得模型能够在每一个时刻做出最符合当前情况的样本回归策略。

1 YOLOv8 目标检测算法原理

YOLOv8 作为一种一阶段目标检测算法,检测准确率较高而且模型较小。其结构由输入端(input),主干网络(Backbone),颈部网络(Neck)和预测头(Head)4 个部分构成, YOLOv8 网络结构如图 1 所示。

YOLOv8 的输入端沿用 YOLOv5 中的 mosaic 数据增强。经过预处理进入 Backbone 主干网络, Backbone 主要由跨阶段部分模块(cross stage partial network, C2f),标准卷积(convolutional layer, Conv)和快速空间金字塔池化(spatial pyramid pooling fast, SPPF)模块构成。C2f 模块依照 YOLOv7 中高效层级增强模块的聚合网络结构,通过多分支跨层链接,丰富了模型的梯度流,以实现更好的参数

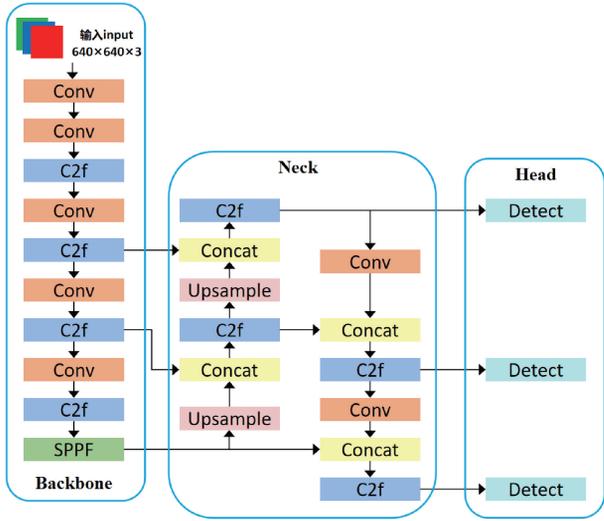


图 1 YOLOv8 网络结构图

Fig. 1 YOLOv8 network structure diagram

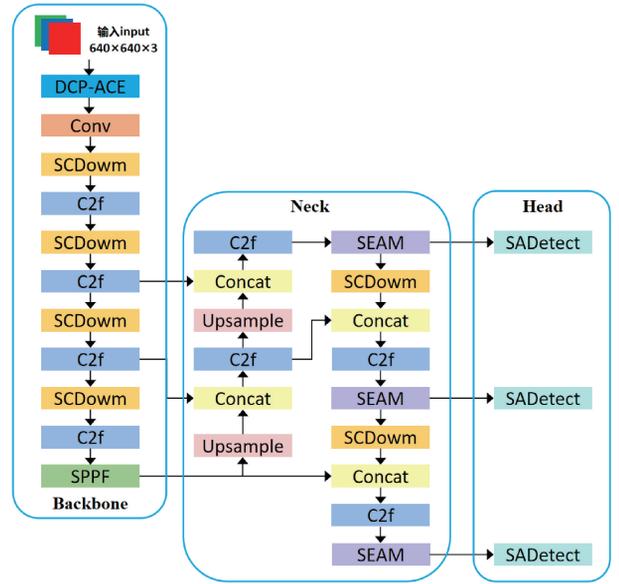


图 2 DMC-YOLO 算法结构图

Fig. 2 Structure diagram of DMC-YOLO algorithm

利用率,增加网络深度。Conv 模块内置 2d 卷积块,对传入网络的特征信息进行下采样,并其归一化以及激活。Backbone 下采样最后经过 SPPF, SPPF 将池化核大小为 13×13 、 9×9 、 5×5 和 1×1 的最大池化层进行串行结合,同时引入并行结构,分别处理不同像素大小的特征图完成特征融合。

YOLOv8 的 Neck 颈部网络采用路径聚合特征金字塔网络结构,由特征图金字塔网络 (feature pyramid networks, FPN) 和路径聚合网络 (path aggregation network, PAN) 两部分构成,特征金字塔结构采用上采样的方式将深层特征图中的强语义信息传递到浅层,以实现多尺度特征融合;路径聚合网络则以下采样的方式将浅层的位置信息传递到深层,以进一步优化多尺度特征图的表示能力和检测精度。

在 Head 网络中, YOLOv8 引入解耦头结构 (decoupled-head), 将分类和回归任务分支处理,同时采用 DFL Loss 和 CIoU Loss 作为损失函数,在模型训练和推理中取得高性能收益,更好地提取和学习图像特征信息。最后将获得的图像信息通过 3 个尺度为大、中、小的检测头输出最终目标检测数据。

2 算法改进

2.1 网络整体结构

恶劣天气下,环境噪声明显导致图像质量降低,且图像中还会出现大量小目标信息及目标间信息遮挡情况,检测难度增大。本文以 YOLOv8 算法为基础,提出自适应去噪和多重关注的联合优化目标检测算法 DMC-YOLO,整体结构如图 2 所示。

DMC-YOLO 在预处理阶段前置暗通道图像增强去噪网络,融合暗通道先验和色彩均衡模图像增强算法,优化图

像噪声,增强图像质量,减少复杂天气所带来的干扰。在卷积网络中嵌入点卷积与深度卷积,轻量化下采样图像信息。在 Neck 网络中引入混合注意力机制,从图像中提取目标纹理细节和语义含义特征,增强网络对遮挡目标的敏感度。为检测头添加多头自注意力模块,丰富特征关联性,有利于网络捕捉上下文信息。最后重构回归损失函数,提升网络的适应能力。DMC-YOLO 采用联合优化策略,经由图像去噪后进行检测任务,增强网络在恶劣环境下的鲁棒性。

2.2 DCP-ACE 图像去噪

因受恶劣天气环境影响,图像信息会产生噪声,增加目标检测算法检测难度。因此本文提出 DCP-ACE (dark channel prior-automatic color enhancement) 图像去噪算法,用以消除图像的天气噪声。DCP-ACE 融合暗通道先验算法与自动色彩均衡模块 (automatic color enhancement, ACE), 本文选取暗通道先验模块 (dark channel prior module, DCP) 作为图像去噪模块^[18], 噪声图像经由该模块先获得图像暗通道,暗通道定义如式(1)所示。

$$J^{dark}(x) = \min_{c \in \{R, G, B\}} (\min_{y \in \Omega(x)} (y)) \quad (1)$$

其中, x 是图像的像素点, $\Omega(x)$ 代表以 x 为中心的局部区域, y 就是这局部区域内的任意像素, R, G, B 表示 3 种色彩通道, J^c 是代表去噪图像的 R, G, B 三通道中的某一色彩通道值, J^{dark} 为去噪图像的暗原色通道值。

由暗通道先验方法对图像的透射率 $t(x, y)$ 和大气光值 A 进行估计,首先选取噪声图像 $I(x, y)$ 的暗通道图像中前 0.1% 最亮的那部分像素点,并将该像素点在有雾图像中对应的区域作为目标区域,定义目标区域内最大的像素值为大气光值 A 。随后引入噪声成像模型如式(2)所示,设透射率为 $t(x, y)$, 大气光值 A 为常数,对式(1)两边同

除 A , 再进行两次最小化操作, 得到式(3):

$$I(x, y) = J(x, y)t(x, y) + A(1 - t(x, y)) \quad (2)$$

$$\min \left\{ \min_{c \in \{R, G, B\}} \left(\frac{I^c(x, y)}{A^c} \right) \right\} =$$

$$t(x, y) \min \left\{ \min_{c \in \{R, G, B\}} \left(\frac{J^c(x, y)}{A^c} \right) \right\} + (1 - t(x, y)) \quad (3)$$

式中: (x, y) 表示图像像素坐标值, $I(x, y)$ 是噪声图像, $J(x, y)$ 为去噪后图像, $I^c(x, y)$ 为噪声图像某一色彩通道值。

由于正常天气大气中也有颗粒物, 引入缓冲因子 ω , 保留少量噪声, 令 J^{dark} 趋向于 0, 得到透射率公式:

$$t(x, y) = 1 - \omega \min \left\{ \min_{c \in \{R, G, B\}} \left(\frac{I^c(x, y)}{A^c} \right) \right\} \quad (4)$$

最终, 设置透射率下限 t_0 , 将透射率和大气光值带入环境噪声成像模型获得图像去噪公式:

$$J(x, y) = \frac{I(x, y) - A}{\max[t(x, y), t_0]} + A \quad (5)$$

为避免去噪模块在高亮度区域易出现光晕、色偏等问题, 采取 ACE 模块^[19]。运用区域自适应滤波对图像进行色彩和空域调整, 完成图像的色差校正, 得到空域重构图像, 区域自适应滤波计算如式(6)所示。

$$Y_c(x) = \sum_{j \in Subset, j \neq p} \frac{G[I_c(x) - I_c(z)]}{d(x, z)} \quad (6)$$

其中, $Subset$ 是图像的像素集合, z 为像素集合中的像素点, Y_c 为重构中心图像, $I_c(x) - I_c(z)$ 为 x, z 两像素点灰度差, $d(x, z)$ 为欧式距离, G 是亮度表现函数, 函数公式如式(7)所示。

$$G_\alpha(x) = \begin{cases} 1, & x < -\alpha \\ x/\alpha, & -\alpha \leq x \leq \alpha \\ -1, & x > \alpha \end{cases} \quad (7)$$

式中: α 为颜色饱和度的调节参数。

再对图像的单通道色调调整拉伸, 延伸应用到 RGB 彩色空间的 3 通道中。最后将式(1)中得到的中间量拉伸映射到 $[0, 255]$ 中, 改善了原图像的明暗程度, 增强图像质量的同时保持了目标的真实性, 设 $L(x)$ 定义域为 $[\min(Y), \max(Y)]$, 全局重整拉伸计算公式如下:

$$L(x) = \frac{Y(x) - \min(Y)}{\max(Y) - \min(Y)} \quad (8)$$

2.3 空间-通道解耦下采样模块

在 YOLOv8 网络中, 下采样模块通过一个 3×3 的卷积, 同时实现空间下采样和通道变换, 使得模型能够提取更高层次的特征, 获得更为丰富的感受野, 减少特征图尺寸, 同时模型的参数量也会相对减少, 有助于防止过拟合。但在复杂天气背景下, 图像中细节信息易受环境噪声干扰, 导致模型的检测性能下降。本文提出空间-通道解耦下采样 (spatial-channel decoupled downsampling, SCDown)^[20] 代替原下采样模块, SCDown 模块结构如图 3 所示。

空间-通道解耦下采样模块将空间缩小和通道增加操

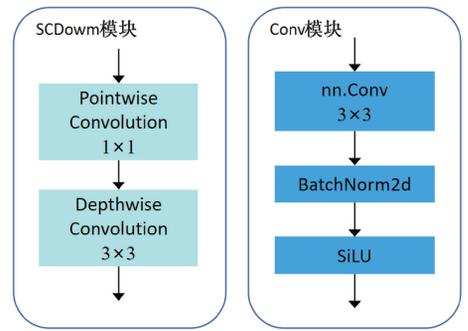


图 3 SCDown 模块结构

Fig. 3 SCDown module structure

作解耦, 首先利用逐点卷积通道维度由 C 提升至 $2C$, 然后利用深度卷积来执行高度和宽度 (从 $H \times W$ 调整至 $H \setminus 2 \times W \setminus 2$) 的空间下采样。相较于标准卷积, SCDown 将计算成本从 $9/2HWC^2$ 降低到 $2HWC^2 + 9/2HWC$, 参数数量从 $18C^2$ 降低到 $2C^2 + 18C$, 且在下采样过程中最大程度地保留了信息, 保证模型能更为准确的捕捉图像细节特征。

2.4 混合注意力机制模块

在复杂天气检测任务中, 经常出现因目标与背景之间的对比度较低或目标间相互遮挡, 导致目标漏检和误检现象, 造成检测模型的性能下降。为了解决这一问题, 在网络中引入混合注意力机制 (separated and enhancement attention module, SEAM)^[21], 具体结构如图 4 所示。

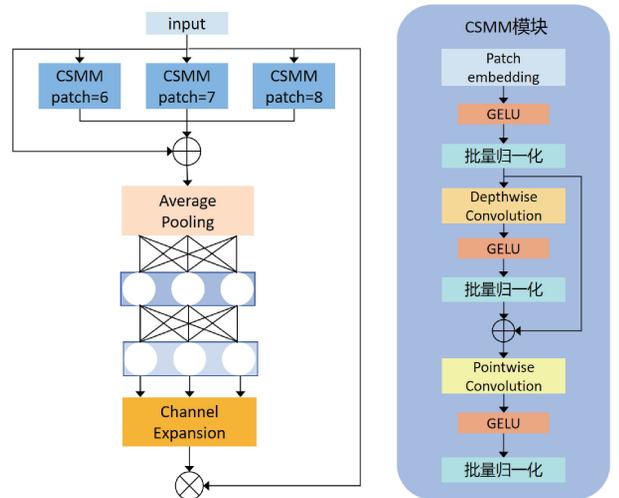


图 4 SEAM 注意力机制结构

Fig. 4 Structure of SEAM attention mechanism

输入图像先由 SEAM 进入 3 个不同尺度的 CSMM 模块, 通道注意力多尺度模块 (channel and spatial mixing module, CSMM) 模块采用深度可分离卷积 (depthwise convolution, DW) 与逐点卷积 (pointwise convolution, PW) 串联的形式, 对每个输入通道分别卷积, 学习各通道重要信息并减少参数量, 再用一个 1×1 的卷积将多个通道的信息空间混合, 学习不同通道间的关系, 并以残差连接结构融合

原始信息,深度可分离卷积信息和逐点卷积信息。3个CSMM模块输出经平均池化后进入两层全连接网络融合各通道信息,进一步加强所有通道之间的信息连接,最后通过指数归一化输出注意力信息,与原始信息相乘,从通道与空间的两个维度对特征信息进行细节处理,在保持计算效率的同时,增强网络对于物体遮挡以及目标背景干扰问题。

2.5 多头注意力检测头

在目标检测任务中,主干网络通过卷积采样可以提取丰富的图像特征信息,但卷积操作会导致特征图尺寸不断减小,小目标特征很可能会被忽略甚至丢失,处理多尺度目标时也会因目标的尺寸变化而导致检测精度下降,因此YOLOv8使用解耦头将分类子任务和回归子任务分开并独立学习,但仍缺乏针对多种复杂环境下的检测学习^[22]。针对上述问题本文引入多头注意力检测头(self-attention detect, SADetect),多头注意力检测头内置自注意力模块和卷积模块,能够捕捉更为广泛的上下文信息^[23]。

如图5(a)所示,SADetect采用自注意力机制,先由 1×1 卷积降低维度,以减少自注意力计算负荷,然后通过多头自注意力层,获得特征间的信息相关性,经 1×1 卷积恢复通道数,批量归一化后将自注意力模块与初始输入相加形成残差连接,输出带有自注意力的图像特征。卷积模块部分具体结构如图5(b)所示,同样运用残差链接思想,图像信息先经过降采样模块降低空间分辨率,用 3×3 卷积进行特征转换,然后上采样恢复空间特征,与初始输入相连,使得网络能够整合自注意力机制下的局部信息并减少计算复杂度。

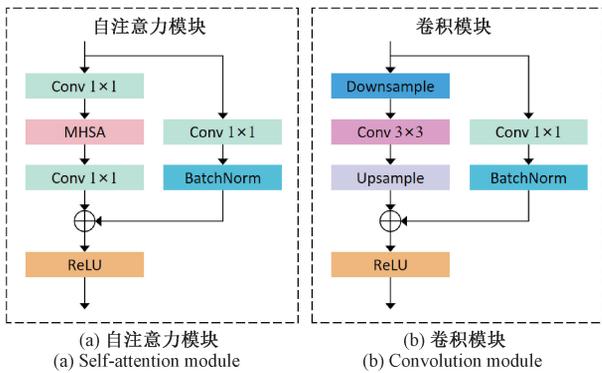


图5 SADetect检测头结构

Fig. 5 SADetect detection header structure

多头自注意力模块(multi-headed self attention, MHSA)^[24]如图6所示,输入为 $H \times W \times C$ 的图像先通过PatchEmbed模块调整空间维度,为使模块能够利用序列的顺序信息,特征序列中加入了关于序列的位置信息,并生成了3个输入,分别是查询矩阵 Q 、键矩阵 K 和值矩阵 V ,通过线性投影将特征矩阵投影 L 次到 C_q 、 C_k 和 C_v 维度来并行计算点积注意力得分,计算公式如下:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (9)$$

式中: d_k 为键矩阵维度, K^T 为键矩阵转置矩阵,图6中 P 为图像块分辨率, $N = HW/P^2$ 是图像块数量。

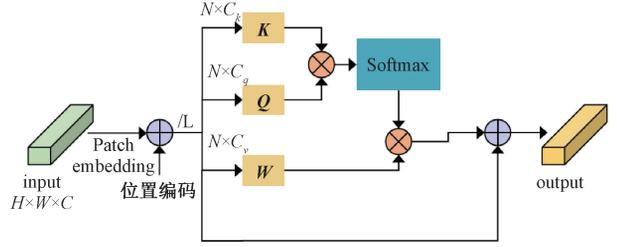


图6 多头自注意力模块

Fig. 6 Multi-head self-attention module

将注意力得分对值向量加权求和,得到最后的多头注意力输出,提高语义特征识别能力有助于学习对象间的关系,使模型能够从大邻域中收集和关联特征信息。

2.6 损失函数

YOLOv8以 $CIoU$ 作为损失函数,来衡量目标检测任务的准确性,在训练过程中能够加快模型收敛,提供更丰富的梯度信息,帮助模型更好地调整预测框^[25]。但 $CIoU$ 对背景噪声和干扰物体敏感,且在处理多尺度目标时,无法有效地适应目标的尺寸变化,导致损失计算不准确。

为使网络适应多种环境下的检测任务, $Focaler-IoU$ ^[26]采用线性区间映射方法来重构 IoU 损失,定义 IoU 下限阈值为 d ,上限阈值为 u ,约束 IoU 的值处于 d 和 u 之间,有助于网络在关注简单和困难样本的同时,更好的从中等难度样本中提取特征,保证回归样本多样性, $Focaler-IoU$ 计算公式如下:

$$IoU^{focaler} = \begin{cases} 0, & IoU < d \\ IoU - d, & d \leq IoU \leq u \\ u - d, & \\ 1, & IoU > u \end{cases} \quad (10)$$

$$L_{Focaler-IoU} = 1 - IoU^{focaler} \quad (11)$$

其中, $IoU^{focaler}$ 为重构 IoU 损失。

本文将重构 IoU 损失引入YOLOv8回归损失函数中,搭建能够关注不同样本回归的 $Focaler-CIoU$,损失定义如下:

$$L_{Focaler-CIoU} = L_{CIoU} + IoU - IoU^{focaler} \quad (12)$$

3 实验与结果分析

3.1 实验环境与数据集

本实验环境如下:操作系统为Windows10,处理器为Intel Core i7 10750H, GPU为NVIDIA GeForce RTX 2070, 8G运存。实验框架为Python3.7, PyTorch版本为1.2.1, Cuda版本为11.3, Python解释器版本3.9.1,进行100个迭代次数(epoch)的迭代训练,初始学习率为0.01,

权重衰减系数为 0.000 5, 批量大小(batchsize)设置为 4。

本文实验在 RTTS 雾天数据集, RainyDataset 雨天数据集, snow traffic 雪天数据集以及 final 黑夜数据集中进行。RTTS 数据集是一个雾天环境的道路交通图像数据集, 包含 4 322 幅自然雾天图像, 带有 5 个标注对象类即人、自行车、汽车、公共汽车和摩托车, 数据集示例如图 7 所示。其中行人所占像素较小, 车辆数量众多以及行人车辆间的堆叠问题, 都成为检测的难点, 将 RTTS 数据集按照 7:2:1 的比例随机分为训练集, 测试集和验证集部署消融实验, 并在该数据集下以相同的实验环境分别与其他网络进行对比实验。



图 7 RTTS 数据集图像示例

Fig. 7 Examples of RTTS dataset images

Snow traffic 数据集是雪天环境数据集, 包含 1 005 张训练集图像和 203 张验证集图像。RainyDataset 数据集是雨天交通数据集, 包含 1 655 张训练集图像和 350 张验证集图像。final 数据集是黑夜下的航拍交通数据集, 包含 1 018 张训练集图像和 304 张验证集图像。选择以上数据集作为泛化能力对比实验数据集, 旨在表现本文模型对于不同恶劣环境的适应能力, 数据集图像示例如图 8 所示。



图 8 泛化性实验数据集图像示例

Fig. 8 Examples of images from the generalization experiment

3.2 评价指标

为衡量改进后模型在复杂环境下检测的性能和效果,

本文选取准确率(Precision), 召回率(Recall)和平均精度均值(mean average precision, mAP)来评估模型的检测能力, 选择模型参数量(Params)和浮点计算量(GFLOPs)来评估模型的复杂程度。选择帧数(frames per second, FPS)来评估模型的计算速度。

准确率表示在识别出的物体中, 正确的正向预测到目标信息占全部正向预测信息的比例。召回率表示所有检测图像中, 模型能够预测出来的比例。指标计算公式如下:

$$Precision = \frac{TP}{TP + FP} \quad (13)$$

$$Recall = \frac{TP}{TP + FN} \quad (14)$$

其中, TP 表示实际待预测目标模型判断为正例的数量, FP 表示为错误预测为正例的数量, 即实际不是待预测目标但被网络错误预测为正例的实例数; FN 表示检测到待预测目标被分类器错误判定为负例的数量。

平均精度(average precision, AP)作为目标检测算法中常用度量指标, 它是以 Precision 和 Recall 分别作为横纵坐标绘制而成的函数图像, 用于评价模型对于目标信息的检测精度, 对各类目标信息的 AP 值进行平均, 可得平均精度均值(mAP), 用于网络模型的整体评估, 其公式如下, n 为类别个数。

$$AP = \int_0^1 P(R) dR \quad (15)$$

$$mAP = \frac{1}{n} \sum_{i=1}^n AP \quad (16)$$

其中, P 为 Precision 数值, R 为 Recall 数值。

参数量(Params)反应网络结构的复杂程度, 结构越复杂, 内存占用的参数量就越多。浮点数(GFLOPs)反应模型训练或检测时的计算性能, 浮点计算量越大, 说明模型具有更高的计算能力和处理速度。FPS 表示网络每秒处理的图像的帧数, 帧率越高说明网络的计算效率越高。

3.3 消融实验

为验证本文改进思路和方法的有效性, 以 YOLOv8 作为基础模型, 在 RTTS 数据集下以相同环境和参数训练, 获得 8 组消融实验, 以分析和验证不同改进方法对整体算法的影响, 消融实验结果如表 1 所示。

由表 1 中可以得出, 在雾天环境下, YOLOv8 的准确率为 79.9%, 召回率为 75.3%, mAP@0.5 为 82.4%, 浮点数和参数量分别是 79.1 G 和 52 M, 帧数为 71 fps。模型在卷积部分引入 SCDown 结构, 对输入图像特征进行解耦, 与基线模型相比, mAP@0.5 提升了 1.8%, 准确率提升了 9.6%, 参数量减少至 44.1 M, 帧数上升至 73.9 fps, 说明引入 SCDown 结构, 使得模型获得丰富采样信息, 增强主干网络的特征提取能力, 相较于标准卷积, SCDown 结构利用逐点卷积与深度卷积, 降低模型计算成本, 提高检测效率, 但是由于目标特征受到遮挡或部分遮挡, 导致模型出现漏检, 召回率下降了 3.5%。所以提出 SEAM 注意力机制, 运用

表1 消融实验结果

Table 1 Results of ablation experiments

实验	模型	mAP@0.5/%	P/%	R/%	GFLOPs	Params	FPS
A	YOLOv8	82.4	79.9	75.3	79.1	25.9	71.0
B	YOLOv8+SCDown	84.2	89.5	71.8	73.7	21.9	73.9
C	YOLOv8+SEAM	83.3	84.9	75.7	80.4	26.4	70.5
D	YOLOv8+SADetect	83.5	89.1	74.7	73.1	25.2	72.7
E	YOLOv8+SCDown+SEAM	84.3	88.3	72.9	75.1	22.5	72.6
F	YOLOv8+SEAM+SADetect	83.7	88.7	73.2	74.8	25.7	71.6
G	YOLOv8+SEAM+SCDown+SADetect	84.6	90.0	73.0	69.4	21.8	74.7
H	YOLOv8+SEAM+SCDown+SADetect+Focaler-CIoU	84.9	90.5	73.3	69.4	21.8	75.3
I	DMC-YOLO	85.3	91.0	73.7	70.3	22.4	74.4

通道注意力学习逐一通道细节信息,再采用空间注意力融合各通道关系,增强目标信息未遮挡部分特征,降低模型因目标受背景遮挡或目标间相互遮挡所出现的漏检和误检问题,本文在颈部网络添加 SEAM 模块,如实验 C 所示,在增加 1.3 M 参数量和 1.3 G 浮点数的基础上 mAP@0.5 达到 84.9%,准确率提升了 5%,召回率提升了 0.4%,但由于模型复杂度增加的原因,帧数下降了 0.5 fps。SADetect 内置多头自注意力模块,有助于检测头更好的关注上下文信息,并学习信息之间的关联性,尽可能降低图像噪声和图像压缩对目标信息的干扰,从而增强模型对不同尺度目标的检测能力,相较于基础模型,SADetect 改进模型在召回率基本不变的基础上,准确率上涨 9.2%,mAP@0.5 提升 1.1%。在实验 E 中,模型在标准卷积中引入 SCDown 结构,在颈部网络添加 SEAM 注意力机制,这样设计使得模型在特征提取部分能够更全面的获得特征信息,在特征融合部分添加混合注意力机制,则可以使得模型关注重要信息而忽略无关信息,更为高效和细致的处理特征提取部分传递进来的大量信息,降低漏检率,改进模型 mAP@0.5 提升至 84.3%,准确率和召回率分别达到 88.3%和 72.9%,浮点数和参数量相较于基础模型都有所下降,帧数提升到 72.7 fps,推理速度加快。实验 F 在颈部网络引入 SEAM,在头部网络引入 SADetect,mAP@0.5 达到 83.7%,准确率和召回率为 88.7%和 73.2%,说明 SEAM 模块与 SADetect 模块分别为模型的特征融合部分以及检测头部分提供注意力机制,进一步加强信息之间的关联性,多重关注信息重要特征和细节特征,扩大网络的感受野,但是由于 SEAM 模块会产生少量计算量,导致模型参数量和帧数与 YOLOv8 基本一致。实验 G 将实验 B、C、D 模块进行融合,SADetect 模块的引入,使得模型能够更有效地学习和利用不同尺度和细节的特征信息,促进多尺度信息的交互和整合,帮助模型理解在特征提取和特征融合部分被压缩的小目标信息,消除冗余信息,实验所得的 mAP@0.5 继续上涨至 84.6%,浮点数和参数量则分别下降 9.7 G 和 8.1 M,帧数为 74.7 fps。实验 H 在实验 G

的基础上,重构回归损失函数 Focaler-CIoU,帮助网络更好的识别和处理不同环境噪声下的特征样本,提升模型的适应能力,mAP@0.5 相较于先前的改进提升 0.3%,准确率与召回率提升至 90.5%和 73.3%,帧数上涨到 75.3 fps。最终在预处理阶段引入 DCP-ACE 去噪算法,降低环境噪声对于小目标信息和遮挡信息等问题的干扰,提升图像质量,大大提高了模型的鲁棒性,mAP@0.5 提升至 85.3%,相较于基线模型提升了 2.9%,准确率达到 91%,浮点数和参数量分别降低至 69.4 G 和 43.9 M,帧数为 74.4 fps,改进模型性能显著提升,但由于在预处理引入计算复杂度,帧率略有下降。

3.4 对比实验

为验证本文算法在复杂环境下目标检测的优越性能,在保证相同实验环境及数据集的前提下,本文将 YOLOv7-CDD 与现阶段主流单阶段算法进行对比实验,选择算法包括 SSD, R-FCN, CenterNet, Faster R-CNN, RetinaNet, YOLOv4, YOLOv5m, YOLOX, YOLOv7 和 YOLOv8,实验结果如表 2 所示。

表2 RTTS 对比实验

Table 2 Comparison experiments of RTTS

算法	mAP@0.5/ %	P/ %	R/ %	Params	FPS
SSD	62.18	83.0	75.9	26.3	54.9
R-FCN	63.02	87.0	76.9	50.8	36.8
CenterNet	63.69	85.0	74.5	32.7	41.2
Faster R-CNN	64.84	84.0	83.2	137.0	27.4
RetinaNet	65.36	86.0	88.0	36.2	24.1
YOLOv4	72.17	89.0	73.4	72.8	63.0
YOLOv5m	76.89	90.0	62.7	29.3	124.0
YOLOX	70.93	86.5	52.2	12.7	37.9
YOLOv7	81.10	82.1	74.0	37.2	36.0
YOLOv8	82.40	79.9	75.3	25.9	71.0
本文	85.30	91.0	73.7	22.4	74.4

通过表 2 可以看出,R-FCN 和 Faster-RCNN 等二阶段算法均展现出了较高的准确率和召回率,但其计算负担较为庞大。在单阶段算法中,SSD,CenterNet 和 RetinaNet 这 3 种算法在复杂环境下的检测速率要优于 R-FCN 和 Faster-RCNN,同时其识别精度与二阶段算法接近,但仍受环境噪声影响过大。YOLOv4 在原有模型基础上引入 MobileNet 框架,简化模型结构,使其参数量为 72.8 M,但 mAP 损失较多。开发程度较高的 YOLOv5 算法在 FPS 和平均精度均值上都超过了拥有解耦头算法的 YOLOX, mAP 达到 76.89%,但在复杂天气下效果还是不尽人意, YOLOv7 和 YOLOv8 作为近期活跃度较高的算法,在对比检测实验中表现最佳,而相较于 YOLOv7, YOLOv8 模型以较低的参数量获得更好的检测效果,检测效率较高,因此本文选择 YOLOv8 作为基线模型,提出 DMC-YOLO, mAP@0.5 达到 85.3%,参数量为 25.9 M,由此可以得出,本文模型在识别精度和计算量上取得较好的平衡,检测性能良好。对于实时性的目标检测任务,本文算法参数量只有 22.4 M,将算法嵌入到系统或移动设备中的难度较小,且帧数达到 74.4 fps,已超过大部分算法,能够快速响应实时场景下的目标信息,可以满足现实恶劣天气环境下的检测。

3.5 泛化性实验

为体现改进后模型的泛化能力,展现模型在不同噪声天气下仍具有良好的适应能力,本文在 snow traffic 数据集,RainyDataset 数据集和 final 数据集下进行实验,以测试 DMC-YOLO 在不同复杂环境下的模型鲁棒性,通用性对比实验结果如表 3~5 所示。

Snow traffic,RainyDataset 和 final 数据集图像分别具有雪天、雨天和黑夜的环境噪声,由表 3~5 可以得出,提出的 DMC-YOLO 的 mAP@0.5 在雪天环境提升至 72.9%,在雨天环境提升至 82.3%,在黑夜环境提升至

表 3 Snow traffic 泛化性实验

Table 3 Experiments on generalization of snow traffic

算法	mAP@0.5/%	P/%	R/%	Params
YOLOv8	69.7	72.6	60.8	25.9
本文	72.9	71.7	68.1	22.4

表 4 RainyDataset 泛化性实验

Table 4 Experiments on generalization of RainyDataset

算法	mAP@0.5/%	P/%	R/%	Params
YOLOv8	81.1	80.0	77.5	25.9
本文	82.3	83.7	76.8	22.4

表 5 Final 泛化性实验

Table 5 Final generalization experiments

算法	mAP@0.5/%	P/%	R/%	Params
YOLOv8	79.2	77.1	86.0	25.9
本文	83.5	79.7	87.1	22.4

84.8%,相较于基础模型均有提升,且 DMC-YOLO 改进轻量化下采样卷积结构进行更深程度的特征融合,使得模型参数量和浮点数同时下降,在实时检测部署中具有较大优势。综上,DMC-YOLO 不仅适用于雾天噪声环境的目标检测任务,在其他环境噪声干扰下仍具有良好效果。

3.6 可视化实验

为验证 DMC-YOLO 算法在实际应用中的表现,先由 YOLOv8 模型和 DMC-YOLO 在相同环境下训练获得训练权重,随机选取复杂天气下的几张图片进行检测,对检测结果进行可视化处理。图 9~12 分别展示了雾下的原始图像信息, YOLOv8 检测结果和 DMC-YOLO 检测结果目标检测结果。

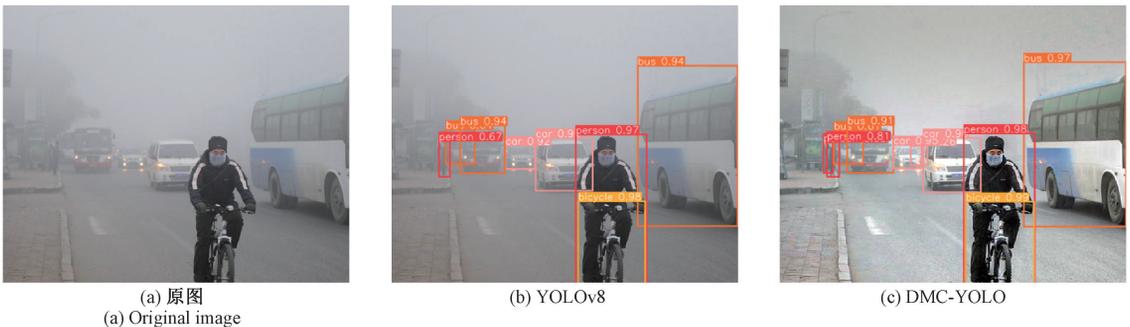


图 9 雾天环境下检测效果对比

Fig. 9 Comparison of detection effects in foggy environment

图 9 目标间堆叠现象明显,但在雾天环境干扰下, YOLOv8 网络出现漏检问题,而 DMC-YOLO 先除噪,入检测网络后运用 SCDwn 提取图像中信息的细节特征,减少漏检问题发生。在图 10 中原始网络将雪天噪声下的无

关信息误检为 person 以及 motorcycle,且远处车辆信息也没有识别出来。图 11 中车辆信息密集,同时水汽较多,导致图像中大量信息被遮盖,增加了目标检测难度,但经 DMC-YOLO 预处理水汽影响减少,通过多重注意力关注

表 6 硬件设备部署实验

Table 6 Hardware device deployment experiments

移动设备	算法	FPS	Size
PC 端	YOLOv8	71.0	49.6
	本文	74.4	41.9
EC-R3588SPC	YOLOv8	23.3	26.0
	本文	38.5	23.1
NVIDIA Jetson Orin NX	YOLOv8	43.6	26.0
	本文	50.5	23.1

EC-R3588SPC,所以在模型大小相同的基础上,NVIDIA Jetson Orin NX 中部署 DMC-YOLO 所得帧数更高,达到 50.5 fps,实时效果更好。综上,本文算法低资源环境中部署效果良好。

4 结 论

本文针对因复杂环境噪声导致的识别精度低,目标漏检误检增多等问题,提出一种目标检测算法 DMC-YOLO。首先构建基于暗通道先验算法和 ACE 图像增强的去噪网络,减少图像噪声并提升图像质量,然后进行图像检测任务,运用去噪-检测联合优化的方法提升算法性能。检测部分引入空间-通道解耦采样卷积,以更大感受野获得更深层次的图像信息,减少模型参数量,平衡模块间产生的计算复杂度,在颈部网络中添加 SEAM 注意力模块,通过其混合注意力机制整合网络的细节信息与全局信息,替换多头注意力检测头,捕捉上下文图像特征,学习信息间的关联性,降低图像噪声影响,最后重构 IoU 损失,关注不同难度回归样本,强化网络对图像噪声的适应能力。实验结果表明,相较于现阶段算法,DMC-YOLO 的平均精度均值达到 85.3%,参数量和浮点数下降,证明改进算法在恶劣天气条件下的有效性,并将 DMC-YOLO 部署于移动设备进行实时目标检测,体现改进算法实际部署的可行性。

本文采用图像去噪和图像检测联合优化的方法完成恶劣天气条件下的目标检测任务,图像噪声去除明显,误检漏检现象减少,检测精度得以提高。在未来的研究中,将在不损伤网络性能的前提下进一步轻量化网络结构,加快网络检测和部署能力。

参考文献

- [1] HUSSAIN R, ZEADALLY S. Autonomous cars: Research results, issues, and future challenges[J]. IEEE Communications Surveys & Tutorials, 2018, 21(2): 1275-1313.
- [2] TERVEN J, CORDOVA-ESPARZA D. A comprehensive review of YOLO: From YOLOv1 to YOLOv8 and beyond [J]. ArXiv preprint arXiv: 2304.00501, 2023.
- [3] 汪显东,郭继昌,王天保.一种改进的雾天图像行人和

车辆检测算法[J].西安电子科技大学学报,2020,47(4):70-77.

WANG Y D, GUO J CH, WANG T B. An improved pedestrian and vehicle detection algorithm for foggy images[J]. Journal of Xidian University, 2020,47(4): 70-77.

- [4] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.
- [5] REN SH Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 39(6): 1137-1149.
- [6] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers [C]. European Conference on Computer Vision, Cham: Springer International Publishing, 2020: 213-229.
- [7] WANG CH Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 7464-7475.
- [8] LIU W, ANGELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]. Computer Vision-ECCV 2016: 14th European Conference, 2016: 21-37.
- [9] CHEN Y H, WANG H R, LI W, et al. Scale-aware domain adaptive Faster R-CNN [J]. International Journal of Computer Vision, 2021, 129 (7): 2223-2243.
- [10] XIE X X, CHENG G, WANG J B, et al. Oriented R-CNN for object detection [C]. IEEE/CVF International Conference on Computer Vision, 2021: 3520-3529.
- [11] 乔石丽.基于改进 YOLOv5 的复杂场景下车辆目标检测算法研究[D].长春:长春工业大学,2024.
- QIAO SH L. Research on vehicle target detection algorithm in complex scenes based on improved YOLOv5[D]. Changchun: Changchun University of Technology, 2024.
- [12] HUANG K E, LERTNIPHONPHAN K, CHEN F, et al. Multi-object tracking by self-supervised learning appearance model [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 3163-3169.
- [13] TAN F X, QIAN Y R, KONG Y T, et al. DBSwin:

- Transformer based dual branch network for single image deraining [J]. *Journal of Intelligent & Fuzzy Systems*, 2022,43(4): 5109-5123.
- [14] HNEWA M, RADHA H. Multiscale domain adaptive yolo for cross-domain object detection[C]. 2021 IEEE International Conference on Image Processing(ICIP). IEEE, 2021: 3323-3327.
- [15] ZHOU H Y, JIANG F, LU H T. SSDA-YOLO: Semi-supervised domain adaptive YOLO for cross-domain object detection [J]. *Computer Vision and Image Understanding: CVIU*, 2023.
- [16] 关家志. 基于特征融合的复杂天气下图像识别鲁棒模型[J]. *信息与电脑(理论版)*, 2022,34(7):143-145.
GUAN J ZH. Robust model of image recognition in complex weather based on feature fusion [J]. *Information & Computer(Theoretical Edition)*, 2022, 34(7):143-145.
- [17] LI CH Y, ZHOU H, LIU Y, et al. Detection-friendly dehazing: Object detection in real-world hazy scenes[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023(7):45.
- [18] TANG Q F, YANG J, HE X J, et al. Nighttime image dehazing based on retinex and dark channel prior using Taylor series expansion[J]. *Computer Vision and Image Understanding*, 2021, 202: 103086.
- [19] GETREUER P. Automatic color enhancement(ACE) and its fast implementation[J]. *Image Processing on Line*, 2012, 2: 266-277.
- [20] CHEN W J, HUANG H B, PENG SH, et al. YOLO-face: A real-time face detector [J]. *The Visual Computer*, 2021,37:805-813.
- [21] WANG AO, CHEN H, LIU L H, et al. YOLOv10: Real-time end-to-end object detection [J]. *ArXiv preprint arXiv:2405.14458*, 2024.
- [22] HE Q Y, XU AO, YE ZH W, et al. Object detection based on lightweight YOLOX for autonomous driving[J]. *Sensors*, 2023, 23(17): 7596.
- [23] YU H Y, WAN CH, LIU M CH, et al. Real-time image segmentation via hybrid convolutional-transformer architecture search [J]. *ArXiv preprint arXiv:2403.10413*, 2024.
- [24] ZHANG Z X, LU X Q, CAO G J, et al. ViT-YOLO: Transformer-based YOLO for object detection [C]. 2021 IEEE/CVF International Conference on Computer Vision, Montreal: IEEE, 2021: 2799-2808.
- [25] ZHENG ZH H, WANG P, REN D W, et al. Enhancing geometric factors in model learning and inference for object detection and instance segmentation[J]. *IEEE Transactions on Cybernetics*, 2021, 52(8): 8574-8586.
- [26] ZHANG H, ZHANG SH J. Focaler-IoU: More focused intersection over union loss [J]. *ArXiv preprint arXiv:2401.10525*, 2024.

作者简介

张绪康, 硕士, 主要研究方向为图像处理。

E-mail: aa3235539536@163.com

朱硕(通信作者), 副教授, 博士, 主要研究方向为数字媒体信息处理、机器视觉、光电信息处理。

E-mail: zshuo2011@163.com