

DOI:10.19651/j.cnki.emt.2519284

融合通道剪枝的轻量化 YOLOv8 交通标志检测算法*

艾强^{1,2} 冯永安³ 王灵超⁴ 董丽欣^{1,2}(1. 青海师范大学计算机学院 西宁 810008; 2. 藏语智能全国重点实验室 西宁 810008;
3. 辽宁工程技术大学软件学院 葫芦岛 125105; 4. 辽宁理工学院信息工程学院 锦州 121000)

摘要: 针对现有交通标志检测算法在边缘计算设备上部署时面临的计算开销大、对小目标和形变目标检测能力不足的问题,本文提出一种融合通道剪枝的轻量化 YOLOv8 交通标志检测算法 LTS-YOLO。首先,在 YOLOv8n 的基础上,设计了一种结合噪声抑制与深层语义增强的特征融合模块(NSSE),以优化多尺度特征表达并抑制背景干扰。其次,在主干网络中引入多尺度通道注意力(MSCA),并在检测头前集成局部形变注意力(LDA),以分别增强模型对不同尺度目标的感知能力和对几何形变的鲁棒性,从而构建高精度模型 TS-YOLO。最后,为实现模型轻量化,基于 BatchNorm 缩放因子设计通道剪枝策略对 TS-YOLO 进行压缩,得到最终的 LTS-YOLO 模型。在 TT100K 和 CCTSDB 两个公开数据集上的实验结果表明,相比基线模型 YOLOv8n,TS-YOLO 在 mAP@50-95 指标上分别提升了 2.5% 和 1.8%,经过剪枝得到的 LTS-YOLO 模型,在保持精度优势的同时,参数量与计算量显著降低,验证了本方法的有效性和实用性。

关键词: 交通标志检测;YOLOv8;轻量化模型;注意力机制;通道剪枝

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:** 520.6040

Lightweight YOLOv8 traffic sign detection algorithm
incorporating channel pruningAi Qiang^{1,2} Feng Yongan³ Wang Lingchao⁴ Dong Lixin^{1,2}

(1. College of Computer, Qinghai Normal University, Xining 810008, China; 2. The State Key Laboratory of Tibetan Intelligence, Xining 810008, China; 3. School of Software, Liaoning Technical University, Huludao 125105, China; 4. College of Information Engineering, Liaoning Institute of Science and Engineering, Jinzhou 121000, China)

Abstract: To address the challenges of high computational overhead and insufficient detection capability for small and deformed targets when deploying existing traffic sign detection algorithms on edge computing devices, this paper proposes a lightweight YOLOv8-based traffic sign detection algorithm LTS-YOLO incorporating channel pruning. First, building upon YOLOv8n, we design a feature fusion module combining noise suppression and deep semantic enhancement (NSSE) to optimize multi-scale feature representation and suppress background interference. Second, we introduce a multi-scale channel attention (MSCA) mechanism into the backbone network and integrate local deformation attention (LDA) before the detection head, enhancing the model's perception of multi-scale targets and robustness to geometric deformations, thereby resulting in a high-precision model, TS-YOLO. Finally, to achieve model compression, we apply a BatchNorm scaling factor-based channel pruning strategy to compress TS-YOLO, obtaining the final LTS-YOLO model. Experimental results on the TT100K and CCTSDB datasets demonstrate that, compared to the baseline YOLOv8n, TS-YOLO improves mAP@50 by 2.5% and 1.8% on TT100K and CCTSDB, respectively. After pruning, the resulting LTS-YOLO model maintains its accuracy advantage while significantly reducing both parameter count and computational complexity, demonstrating the effectiveness and practicality of the proposed method.

Keywords: traffic sign detection; YOLOv8; lightweight model; attention mechanism; channel pruning

0 引言

交通标志的准确、实时检测是智能驾驶辅助系统和自

动驾驶汽车环境感知的核心任务之一^[1-2]。交通标志为车辆提供关键的道路规则和警示信息,其快速识别直接关系到行车安全与效率。然而,在真实交通场景中,交通标志检

收稿日期:2025-07-04

* 基金项目:国家自然科学基金面上项目(51874166)、国家自然科学基金青年基金(51904144)、青海师范大学中青年科研基金(2023QZR002)、青海师范大学 2025 年大学生创新训练计划项目(qhnucxy2025022, qhnucxy2025026)资助

测面临诸多挑战:1)多尺度问题^[3-4]:交通标志因距离远近呈现出巨大的尺寸差异,尤其是远景中的小目标,其像素少、特征弱,极易漏检;2)复杂背景干扰^[5-7]:标志牌常与广告牌、建筑物、树木等复杂背景融为一体,对模型的抗干扰能力要求高;3)几何形变^[8]:由于拍摄角度、镜头畸变或标志牌本身破损,交通标志常出现透视、旋转等非刚性形变,增加了检测难度;4)部署限制^[9-11]:车载平台等边缘设备的计算资源和功耗有限,要求检测模型必须兼具高效率 and 低复杂度的特点。

针对上述交通标志检测任务中普遍存在的问题,国内外学者开展了大量研究。王海瑞等^[12]提出了 MOMP-YOLOv8 算法,在网络中引入混合注意力与动态卷积,并增加小目标检测头以提升小目标识别能力。王泽玄等^[13]基于 RT-DETR 提出了一种轻量化交通标志检测方法,以 ShuffleNetV2 替代 ResNet 主干并结合通道剪枝、量化和知识蒸馏实现模型压缩。该方法在保证检测精度的同时显著降低模型规模并提升了检测速度。Wang 等^[14]在 YOLOv8 基础上,引入多尺度特征增强与遮挡感知注意力机制,并融合边缘空间信息以减弱背景干扰。该方法在复杂场景下有效提升了小目标与遮挡目标的检测性能。李牧等^[15]针对小目标交通标志易受光照和天气影响的问题,提出改进 YOLOv5 算法,引入 SPD-Conv 与 BiFPN 融合多尺度特征并增加小目标检测层。该方法显著提升了小目标检测精度并保持实时性。韩东旭等^[16]提出 YOLOv8-Faster-Ghost-GAM 模型,在主干中引入全局注意力机制并结合 Fasternet 与 Ghost 卷积以减轻模型复杂度。该方法利用 WiOU 损失提升小目标检测性能,在精度与召回率上均取得改进。

现有研究大多聚焦于模型轻量化或提升复杂场景下的检测精度,但在二者兼顾方面仍存在提升空间。为进一步实现检测模型在精度与效率之间的平衡,本文以 YOLOv8n 为基线,提出一种针对交通标志检测优化的轻量化算法。本文的主要贡献如下:

提出一种结合噪声抑制与深层语义增强的特征融合模块(noise suppression and semantic enhancement, NSSE),通过抑制背景噪声和强化前景语义,提升了网络对多尺度交通标志的特征提取能力。随后将多尺度通道注意力(multi-scale channel attention, MSCA)和局部形变注意力(local deformable attention, LDA)引入网络,前者用于增强对不同感受野下目标的响应,后者用于提升模型对几何变化的适应性,共同构建了高精度检测模型 TS-YOLO。最后对改进后的 TS-YOLO 模型应用了通道剪枝技术,在裁剪冗余参数、降低计算复杂度的同时,最大限度地保留了模型的检测精度,得到了最终的轻量化模型 LTS-YOLO。

1 YOLO 系列算法简介

YOLO(you only look once)系列算法以其卓越的速度-

精度平衡性,在目标检测领域得到了广泛应用。YOLOv8^[17]作为该系列的代表之一,通过引入新的网络结构和无锚框设计,在性能上取得了显著提升。然而,其标准版本对于资源受限的平台而言依然过于庞大,而其轻量化版本 YOLOv8n 虽然速度快,但在处理上述复杂场景,尤其是小目标和形变目标的检测精度上仍有提升空间^[18]。

2016 年,Redmon 等^[19]提出了 YOLO 目标检测框架,该方法通过将输入图像划分为固定数量的网格单元,并在每个单元中同时进行类别预测与边界框定位,从而实现了较高的检测效率。然而,该方法在处理小目标时存在一定的局限性,检测精度仍有提升空间。同年,Liu 等^[20]发布了 SSD(single shot multi-box detector)模型,采用多尺度特征图以适应不同尺寸的目标,从而显著缓解了小目标难以检测的问题。随后,YOLOv2^[21]与 YOLOv3^[22]相继问世,通过引入锚框机制、特征融合策略及批量归一化等技术手段,进一步提升了模型的准确率与稳定性。此外,RefineDet 通过对锚框的精细化回归校正,也在一定程度上增强了检测效果。

2020 年,YOLOv4^[23]结合多项优化策略,在保持高帧率的同时进一步提高了检测精度。为了突破基于锚框方法的结构限制,研究人员陆续提出了无锚框检测架构,如 CornerNet^[24]、CenterNet^[25]以及 FCOS^[26],这类方法通常通过回归目标的中心点或边界信息进行目标定位,在特定应用场景下取得了优异表现。近年来,YOLO 系列持续迭代更新,YOLOv7^[27]与 YOLOv8 在继承前代核心优势的基础上,通过引入新的网络结构与模块设计,在检测性能和运行速度上实现进一步提升,同时还支持多任务场景,如目标分割与姿态估计。后续 YOLOv9^[28]、YOLOv10^[29]、YOLOv11 与 YOLOv12^[30]则主要是在 YOLOv8 框架下做出局部优化,其改进幅度相对有限。

2 改进 YOLOv8n 网络结构

为提升 YOLOv8 网络在交通标志检测任务中的检测精度并有效降低模型复杂度,本文在轻量化版本 YOLOv8n 的基础上提出了一种改进模型 TS-YOLO,其整体结构如图 1 所示。该模型主要从特征融合、注意力机制及模型压缩 3 个方面进行了优化。首先,引入结合噪声抑制与深层语义增强的特征融合策略,以提升多尺度特征融合能力,并增强交通标志的语义表达。其次,在主干网络末端加入多尺度通道注意力机制,增强特征图对不同感受野下交通标志特征的响应能力。最后,增加一个小目标检测头并在所有检测头前端集成局部形变注意力模块,以适应由于拍摄角度变化和镜头畸变导致的目标形变问题。在上述改进基础上,为进一步提升模型推理效率,本文采用通道剪枝方法对 TS-YOLO 模型进行参数优化,得到 LTS-YOLO 从而有效减少网络冗余,降低计算开销。

2.1 结合噪声抑制与深层语义增强的特征融合

为提升 YOLOv8n 网络在交通标志图像中小目标的检

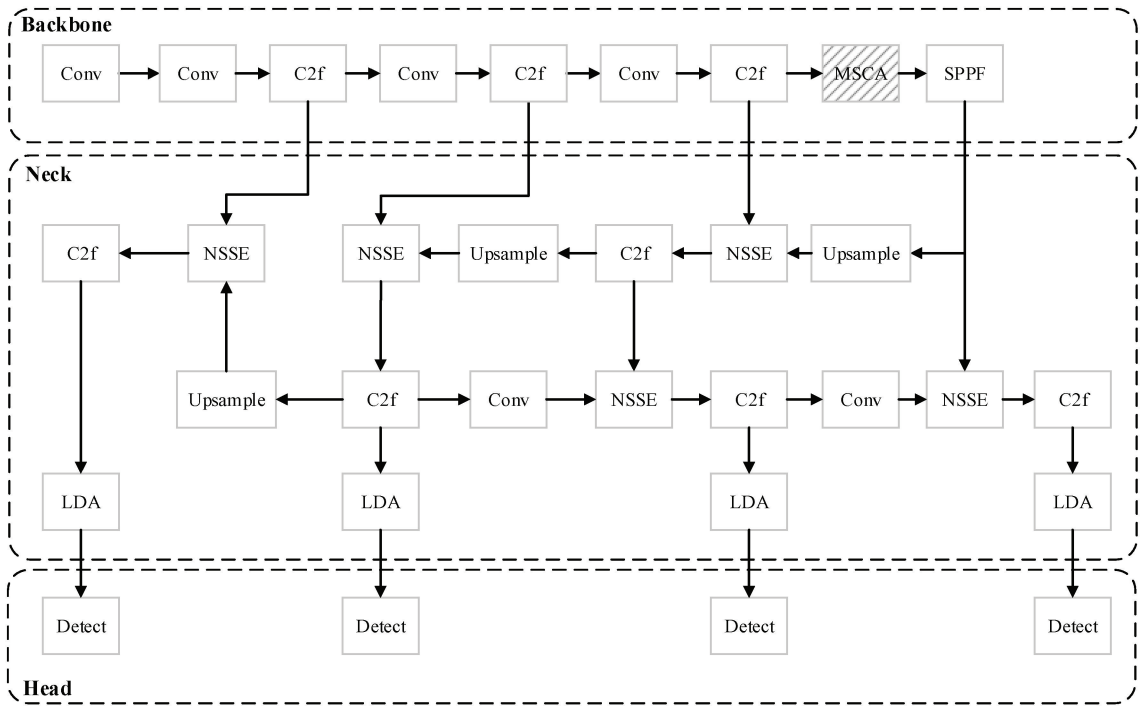


图1 TS-YOLO结构

Fig. 1 TS-YOLO structure

测性能,本文提出一种结合噪声抑制与深层语义增强的轻量级特征融合模块 NSSE,旨在优化 YOLOv8 中的 PAN-FPN 结构,增强跨尺度特征的代表能力与目标响应能力。NSSE 模块的整体结构如图 2 所示,NSSE 模块由两个关键子模块组成,分别为噪声抑制注意力单元 (noise suppression attention unit, NSAU) 与深层语义增强模块 (deep semantic enhancement, DSE)。前者通过显式抑制背景噪声提升目标区域响应,后者对融合特征进行语义建模与补全,以增强特征的判别能力。

1) 噪声抑制注意力单元 (NSAU)

NSAU 模块的目标是在多尺度特征融合的早期阶段,主动识别并抑制由复杂背景(如广告牌、建筑物纹理)引入的噪声信息,同时增强交通标志等关键目标的特征响应。设来自 YOLOv8n 颈部不同层级的浅层特征图为 $P_{i-1} \in \mathbb{R}^{C \times H \times W}$, 深层特征图为 $P_i \in \mathbb{R}^{C \times \frac{H}{2} \times \frac{W}{2}}$, 首先对深层特征进行上采样与 1×1 卷积,使其在空间分辨率和通道维度上与浅层特征 P_{i-1} 对齐:

$$\hat{P}_i = \text{Conv}_{1 \times 1}(\text{Upsample}(P_i)) \quad (1)$$

随后将对齐后的深层特征与浅层特征进行初步融合,得到粗融合特征图:

$$P_c = P_{i-1} + \hat{P}_i \quad (2)$$

为有效压制 P_c 中的背景区域干扰并提升目标区域的显著性,NSAU 引入了串联的通道注意力和空间注意力机制,这借鉴了 CBAM^[31] 的设计思想但更侧重于噪声的抑

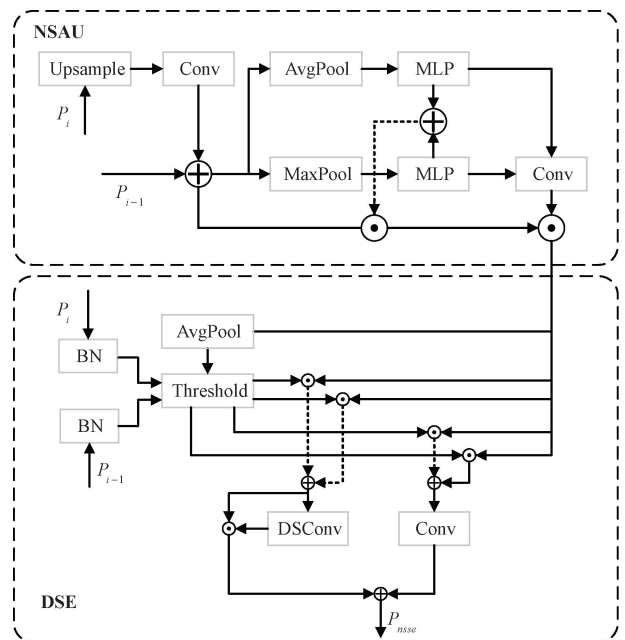


图2 NSSE结构

Fig. 2 NSSE structure

制。首先,通过通道注意力模块学习各通道的重要性权重。该模块并行使用全局平均池化 (avgpool) 和全局最大池化 (maxpool) 来聚合通道的空间信息,然后将结果送入一个共享的多层感知机 (multi layer perceptron, MLP),最后通过 Sigmoid 函数得到通道注意力权重:

$$A_{\text{avg}} = \text{AvgPool}(P_c) \quad (3)$$

$$A_{\text{max}} = \text{MaxPool}(P_c) \quad (4)$$

$$M_c = \sigma(\text{MLP}(A_{\text{avg}}) + \text{MLP}(A_{\text{max}})) \quad (5)$$

接着,在通道增强特征 M_c 的基础上应用空间注意力模块。该模块首先沿通道维度对特征进行平均池化和最大池化,并将结果拼接起来,然后通过一个卷积层(如 7×7 卷积)和 Sigmoid 函数生成空间注意力权重图。

$$M_s = \sigma(\text{Conv}_{7 \times 7}([A_{\text{avg}}; A_{\text{max}}])) \quad (6)$$

最终噪声抑制增强特征图表示为:

$$P_s = M_s \cdot (M_c \cdot P_c) \quad (7)$$

其中, \cdot 表示逐元素乘法, σ 为 Sigmoid 激活函数。该模块可有效压制低响应区域,提升特征图中目标区域的显著性。

2) 深层语义增强模块 (DSE)

NSAU 模块在多尺度融合的早期阶段对背景噪声进行抑制并增强目标响应,使输入特征图具备较高的显著性。然而,复杂场景下的小目标或模糊交通标志仍可能存在响应较弱或语义信息不足的问题。为此,本文设计 DSE 模块通过区分强响应与弱响应区域,分别采用卷积细化和轻量级语义变换(feature transformation unit, FTU)进行处理,从而进一步丰富特征的语义表达能力,提升目标判别性,为后续检测头提供更具语义信息的特征表示。

DSE 模块接收来自 NSAU 的增强特征 P_s , 其目标是进一步提升特征的语义表达能力,特别是针对在复杂场景下可能仍然响应较弱的小目标或模糊交通标志。该模块通过区分并分别处理特征中的强响应和弱响应分量来实现这一目标。

首先从 P_s 中提取通道重要性阈值,用于区分强响应与弱响应区域:

$$T_c = \sigma(\text{AvgPool}(P_s)) \in \mathbb{R}^{C \times 1 \times 1} \quad (8)$$

接着计算浅层与深层原始特征的通道激活图 ω_i , ω_{i-1} , 并根据通道阈值划分出强响应区域 ω^{up} 与弱响应区域 ω^{low} :

$$\omega_i = \sigma(\text{BN}(P_i)), \omega_{i-1} = \sigma(\text{BN}(P_{i-1})) \quad (9)$$

$$\omega_i^{up}, \omega_i^{low} = \text{Threshold}(\omega_i, T_c) \quad (10)$$

$$\omega_{i-1}^{up}, \omega_{i-1}^{low} = \text{Threshold}(\omega_{i-1}, T_c) \quad (11)$$

随后利用这两个通道图对融合特征进行掩码加权,得到强响应特征与弱响应特征:

$$P^{up} = P_s \cdot \omega_{i-1}^{up} + P_s \cdot \omega_i^{up} \quad (12)$$

$$P^{low} = P_s \cdot \omega_{i-1}^{low} + P_s \cdot \omega_i^{low} \quad (13)$$

其中,强响应区域直接使用 1×1 卷积细化边界信息:

$$\hat{P}^{up} = \text{Conv}_{1 \times 1}(P^{up}) \quad (14)$$

而弱响应区域送入 FTU 进行压缩建模,该模块由深度可分离卷积与通道注意力构成:

$$\hat{P}^{low} = A(P^{low} \cdot \text{DSConv}(P^{low})) \quad (15)$$

其中, $\text{DSConv}(\cdot)$ 表示深度可分离卷积操作,用于有

效降低卷积计算复杂度,同时保留空间结构信息;符号 \cdot 表示逐元素乘法。函数 $A(\cdot)$ 表示通道注意力调制函数,其形式为:

$$A(X) = \text{Softmax}(W_2 \cdot \delta(W_1 \cdot \text{GAP}(X))) \cdot X \quad (16)$$

其中, $\text{GAP}(\cdot)$ 表示全局平均池化, W_1 和 W_2 为两层线性变换参数, δ 表示 ReLU 激活函数, Softmax 用于生成归一化的通道注意力权重。最终输出 \hat{P}^{low} 即为弱响应区域经 FTU 模块压缩建模后的特征表示。

最终,融合强弱语义,得到 NSSE 模块的输出特征:

$$P_{\text{nsse}} = \hat{P}^{up} + \hat{P}^{low} \quad (17)$$

2.2 多尺度通道注意力

为提升网络对不同感受野下交通标志等目标特征的响应能力,本文在主干网络末端引入了多尺度通道注意力机制(multi scale channel attention, MSCA),以加强对目标语义与位置的建模能力。该模块结合通道与空间维度的信息增强策略,主要包括 3 步:全局上下文建模、通道响应融合与多尺度增强。

首先对输入特征图 $P_c \in \mathbb{R}^{C \times H \times W}$ 进行平均池化与最大池化,获取全局语义统计信息:

$$A_{\text{avg}} = \text{AvgPool}(P_c) \quad (18)$$

$$A_{\text{max}} = \text{MaxPool}(P_c) \quad (19)$$

其中, $\text{AvgPool}(\cdot)$ 与 $\text{MaxPool}(\cdot)$ 分别为全局平均池化与最大池化操作,输出维度均为 $\mathbb{R}^{C \times 1 \times 1}$ 。

随后将上述两个池化特征分别输入共享参数的 MLP, 并进行相加融合,得到通道注意力权重:

$$M_c = \sigma(\text{MLP}(A_{\text{avg}}) + \text{MLP}(A_{\text{max}})) \quad (20)$$

在通道注意力之外,为进一步增强特征图的空间响应,采用空间注意力机制。首先对 P_c 进行通道维度的平均池化与最大池化:

$$B_{\text{avg}} = \text{Mean}_{\text{chan}}(P_c) \quad (21)$$

$$B_{\text{max}} = \text{Max}_{\text{chan}}(P_c) \quad (22)$$

将上述结果按通道拼接后,输入一个 7×7 卷积操作,生成空间注意力图:

$$M_s = \sigma(\text{Conv}_{7 \times 7}([B_{\text{avg}}; B_{\text{max}}])) \quad (23)$$

最终,将通道注意力图 M_c 与空间注意力图 M_s 分别作用于原始输入特征图 P_c , 得到增强后的多尺度特征:

$$P'_c = M_c \odot M_s \odot P_c \quad (24)$$

其中, \odot 表示逐元素乘法操作。

MSCA 模块通过联合建模通道重要性与空间响应区域,从多个尺度增强了特征表达能力,尤其在小目标识别与复杂背景下对目标响应更具鲁棒性。

2.3 局部形变注意力

针对交通场景中目标因拍摄视角变化、镜头畸变等因素导致的空间形变问题,本文在检测头前引入局部形变注意力(local deformable attention, LDA)模块,以增强模型对目标局部几何变化的感知能力,提升对小目标和变形目

标的检测精度。该模块主要由3部分构成:形变偏移估计、特征采样对齐与加权融合增强。

首先给定输入特征图 $F \in \mathbb{R}^{C \times H \times W}$, 使用一个 3×3 卷积层预测每个位置的采样偏移量:

$$\Delta p = \text{Conv}_{3 \times 3}(F) \quad (25)$$

其中, $\Delta p \in \mathbb{R}^{2N \times H \times W}$ 表示 N 个偏移点的二维坐标偏移。

对于每个标准采样点位置 p_i , 其实际采样位置为:

$$\hat{p}_i = p_i + \Delta p_i \quad (26)$$

考虑 \hat{p}_i 为非整数位置, 为获取连续特征值, 采用双线性插值从原始特征图 F 中提取形变对齐特征:

$$F^D(p_i) = \sum_{q \in \mathcal{N}(\hat{p}_i)} w(q, \hat{p}_i) \cdot F(q) \quad (27)$$

其中, $\mathcal{N}(\hat{p}_i)$ 表示 \hat{p}_i 附近的4个整数位置, $w(\cdot)$ 为双线性插值权重。

为提升对局部变形特征的代表能力, 引入通道注意力机制对形变对齐后的特征图 F^D 进行加权:

$$A = \text{Softmax}(\text{MLP}(\text{AvgPool}(F^D))) \quad (28)$$

最终输出增强后的特征图为:

$$F' = A \cdot F^D \quad (29)$$

其中, $A \in \mathbb{R}^{C \times 1 \times 1}$ 为通道注意力权重, \cdot 表示逐通道加权乘法。

LDA模块通过引入可学习的采样偏移与局部注意机制, 有效缓解了固定感受野带来的对目标几何变化的不敏感问题。在复杂交通场景中, 该机制能够自适应地增强检测头对小目标、畸变目标的响应能力, 从而提升整体检测性能。

2.4 通道剪枝

为进一步提升模型在边缘设备上的推理效率, 本文在完成结构优化后, 采用通道剪枝方法对所提出的TS-YOLO模型进行压缩, 从而减少冗余计算, 提高运行速度。该方法通过评估通道的重要性, 有选择地去除对模型性能影响较小的通道, 实现参数数量和计算量的同步压缩。

首先采用基于BatchNorm中的缩放因子 γ 作为通道重要性指标, 设特征图 $F \in \mathbb{R}^{C \times H \times W}$ 通过BN层后, 其输出为:

$$F_{\text{BN}} = \gamma \cdot \frac{F - \mu}{\sqrt{\sigma^2 + \epsilon}} + \beta \quad (30)$$

其中, γ 为每个通道的缩放因子, μ 与 σ^2 分别为均值与方差, ϵ 为数值稳定项, β 为偏移项。

将 $|\gamma|$ 值从小到大排序, 并设定剪枝比例 α , 选取前 αC 个通道作为待剪枝对象。

随后, 对低重要性的通道进行裁剪, 得到压缩后的特征图:

$$F' = \text{Prune}(F, \gamma, \alpha) \quad (31)$$

其中, $\text{Prune}(\cdot)$ 表示保留 $1 - \alpha$ 比例的重要通道。裁

剪完成后, 为恢复模型性能, 需对精简后的模型进行微调训练。

为减小一次性剪枝带来的性能突降, 采用分阶段剪枝策略。每轮剪枝后执行短周期微调, 逐步收敛至稳定状态。设总迭代轮数为 T , 每阶段剪枝比例为 α_i , 则最终模型的参数量压缩率为:

$$\text{CR} = 1 - \prod_{i=1}^n (1 - \alpha_i) \quad (32)$$

其中, n 为剪枝阶段数, CR为累计压缩率。

通道剪枝作为模型压缩技术的重要手段, 有效减少了TS-YOLO的参数数量和计算开销。配合前文提出的结构优化设计, 压缩后的模型在保持精度的同时, 显著提升了在边缘设备上的部署效率与实时性。

3 实验结果分析

3.1 实验数据集

本文采用TT100K数据集^[32]作为对比试验数据集, 并在CCTSD2021数据集^[33]上进行泛化性验证。上述数据集均涵盖中国交通标志的典型场景, 包括城市道路、高速公路及乡村道路, 具有复杂背景、小目标及遮挡情况, 能够全面评估交通标志检测算法性能。

TT100K数据集通过6台高分辨率广角单反相机, 在中国多个城市的多样化天气条件下进行实地拍摄。原始数据集图片分辨率为 2048×2048 , 涵盖了指示类、禁令类和警告类三大交通标志类别, 共计包含221种细分类别, 如图3所示。针对数据分布不均衡的问题, 本研究通过Python脚本对数据进行了筛选, 最终选取了样本量均超过100张的42个交通标志类别作为研究对象。经筛选后, 数据集划分如下: 训练集包含7773张图片, 测试集1965张。

CCTSD2021数据集包含17856张图片, 分辨率由 $1000 \times 350 \sim 1024 \times 768$ 不等。其中交通标志分为3类: 强制性、禁止性和警示性。本文遵循原始数据集划分比例, 训练集包含16356张图片, 测试集包含1500张图片。数据中各交通标志类别所占比例如图4所示。

3.2 实验环境和参数

本文实验使用PyTorch深度学习框架开发, 实验环境如表1所示。

在模型训练的超参数设置上, 本研究选用Adam优化器进行参数更新, 其基础学习率配置为0.0001。训练阶段batchsize设置为8, 经过300个epoch的迭代优化, 所有输入图像均预处理为 $640 \text{ pixel} \times 640 \text{ pixel}$ 的统一尺寸。为确保实验对比的可靠性, 不同组别的对照实验均采用完全相同的超参数组合, 从而消除训练条件差异对结果分析的影响。

3.3 评价指标

本研究采用多维度评估体系对模型检测性能进行量化分析, 主要考察以下核心指标: 精确率(precision)、召回

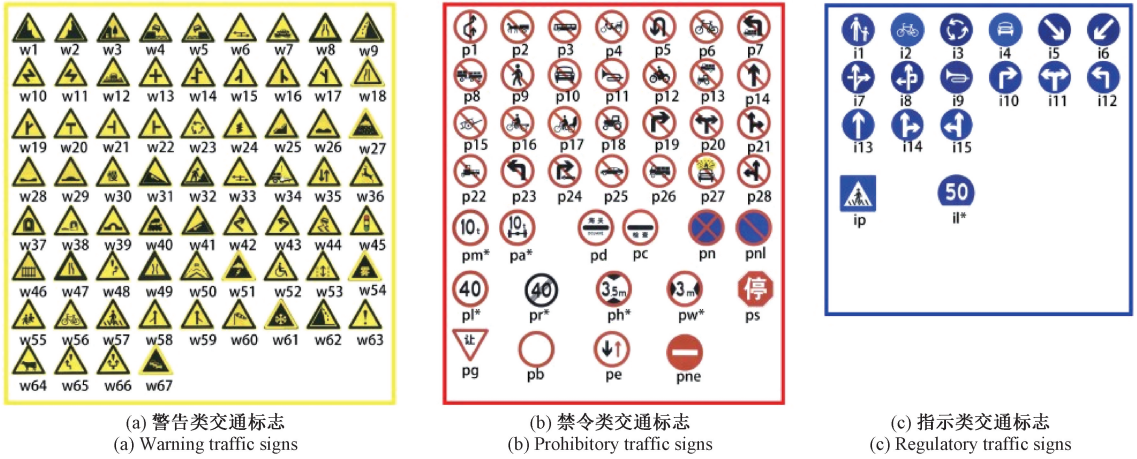


图 3 TT100K 数据集中包含的类别
Fig. 3 Categories included in the TT100K dataset

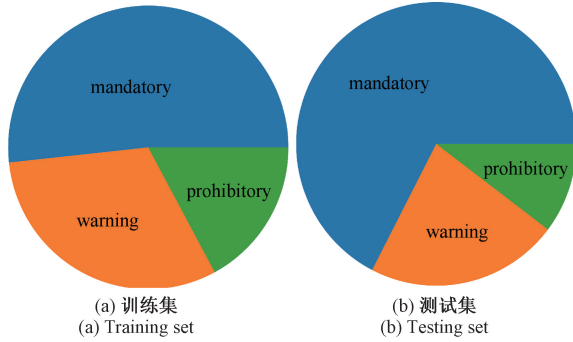


图 4 CCTSDB 数据集中类别分布
Fig. 4 Category distribution in the CCTSDB dataset

表 1 实验环境信息表

Table 1 Experimental environment setting

实验环境	配置
CPU	Xeon E5-2667 V3
GPU	NVIDIA RTX 3090 24 G
内存	32 G
开发语言	Python 3.9
深度学习框架	Pyorch2.6.0+cu118

率 (Recall)、平均精确率均值 (mean average precision, mAP)、模型参数量 (Params) 以及浮点运算量 (floating point operations, FLOPs)。各指标的具体定义和计算方法说明如下:

精确率是评估模型预测准确性的重要指标,其定义为模型预测为正类的样本中真实为正类的比例。数学表达式如下:

$$Precision = \frac{TP}{TP + FP} \quad (33)$$

其中, TP 为被正确检测的正样本数量, FP 为负样本被误检为正样本的数量。

召回率是评估模型正例识别能力的关键指标,其数学表达式为:

$$Recall = \frac{TP}{TP + FN} \quad (34)$$

其中, FN 为实际为正例但被模型漏检的样本数量。

平均精确率均值是多类别目标检测中的核心评估指标,其数学表达式为:

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (35)$$

其中, N 为待检测的类别总数, AP_i 为第 i 个类别的平均精确率。

模型参数量是评估深度学习模型复杂度的重要指标,其定义为:

$$Params = \sum_{l=1}^L (\mathbf{W}_l + b_l) \quad (36)$$

其中, L 为模型总层数, \mathbf{W}_l 为第 l 层的权重参数矩阵, b_l 为第 l 层的偏置参数向量。

浮点运算量是衡量模型计算复杂度的关键指标,表示模型完成单次前向推理所需的浮点运算总次数。该指标综合反映了模型对计算资源的需求,包括卷积、全连接等各类层中的乘法、加法等基本运算。

3.4 对比实验结果

为了全面评估所提出算法的性能,在 TT100K 数据集上当前主流轻量化目标检测算法进行了的横向对比,实验结果如表 2 所示,对比发现,本文提出的 LTS-YOLO 模型在性能和效率上均展现出显著的优势。相较于基线模型 YOLOv8n,其参数量和计算量,已大幅降低至与 YOLOv9t 相当的水平,但在 mAP@50 上明显领先。与其他主流轻量化模型相比,如参数量相近的 YOLOv9t、YOLOv10n 和 YOLOv11n,LTS-YOLO 在 mAP@50 上均取得了最优异的性能。此外,LTS-YOLO 在检测精度接近 YOLOv8n 的改良版本 TS-YOLO 的同时,通过有效的通

道剪枝策略,显著降低了计算开销,进一步提升了其在资源受限场景下的部署可行性。此外,在召回率和 mAP@75 等指标上,LTS-YOLO 也普遍优于或持平于对比模型,综

合证明了本文所提出的算法在实现高精度交通标志检测的同时,能够有效地兼顾模型轻量化和高效率的要求,具有显著的有效性和实用性。

表2 TT100K数据集上的对比实验结果

Table 2 Comparative experimental results on the TT100K dataset

模型	参数量/M	计算量/G	精度/%	召回率/%	mAP@50-95/%	mAP@75/%	mAP@50/%
YOLOv5n	2.51	7.1	74.5	66.5	55.7	65.0	73.7
YOLOv6n	4.30	11.8	63.4	59.1	47.3	55.5	62.7
YOLOv8n	3.01	8.2	73.0	68.4	56.4	65.8	74.3
YOLOv9t	1.98	7.6	79.2	65.6	57.4	67.2	75.8
YOLOv10n	2.28	6.6	74.0	64.9	55.8	65.5	73.2
YOLOv11n	2.58	6.3	71.1	66.6	54.9	64.6	71.6
YOLOv12n	2.57	6.3	71.5	63.0	53.8	62.7	70.4
TS-YOLO	2.95	13.1	79.3	69.2	58.7	70.2	76.8
LTS-YOLO	2.01	7.28	79.2	69.8	58.9	70.5	77.2

图5~8展示了本文方法与基线模型YOLOv8n的典型检测对比结果。每幅图像右侧均给出了关键区域的局部放大图,以突出检测性能差异。从图5可以观察到,本文方法在存在目标遮挡的情况下仍能够实现准确检测,基线模型遗漏的两个部分遮挡标志在改进模型中得以正确识别。图6显示了本文方法在减少误检方面的优势,基线模型将 p160、p150 和 p26 分别错误识别为 pm20、pm55 和

p11,而改进模型能够有效避免此类错误。图7表明,在远距离小目标场景中,本文方法较基线模型显著降低了错误检测的发生率。图8则进一步表明,本文方法能够成功检测出基线模型遗漏的远距离小目标,从而有效缓解了漏检问题。综合上述结果可以看出,本文方法在遮挡环境、复杂背景及远距离小目标检测等典型挑战场景下均表现出更优的鲁棒性和检测精度。

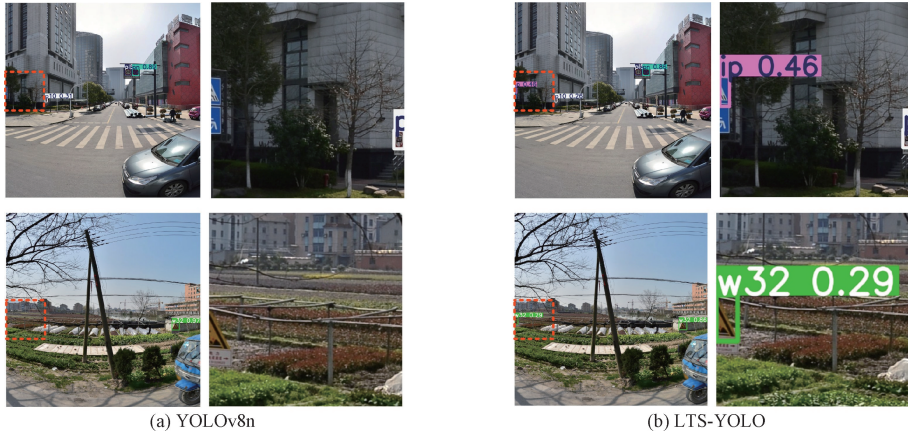


图5 改进受遮挡目标漏检问题

Fig. 5 Improvement of missed detection of obscured targets

3.5 剪枝率对比实验结果

为评估剪枝策略对 TS-YOLO 模型性能的影响,本研究设计了不同加速比下的对比实验。加速比定义为剪枝前后模型计算量的比值,其中加速比为 1.0 表示原始模型,加速比为 2.0 表示剪枝后计算量减半。每次剪枝后均进行微调训练以恢复模型性能。

实验结果如表3所示,当加速比控制在 1.8 时,模型在性能和轻量化之间达到最佳平衡,mAP@50 达到 77.2%,mAP@50-95 为 58.9%,同时参数量压缩至 2.01 M,计算

量降至 7.28 GFLOPs。值得注意的是,适度剪枝反而使部分指标如召回率提升 0.6%,表明剪枝具有正则化效果。但当加速比超过 2.0 后,性能出现明显下降,如 mAP@50-95 降低 3.8%,说明过度剪枝会损害模型特征提取能力。

3.6 消融实验结果

为了量化分析所提出各改进模块的有效性,采用 TT100K 数据进行如表4所示的消融实验,具体分析如下。实验采用 YOLOv8n 作为基线,mAP@50 达到了 74.3%,参数量和计算量分别为 3.01 M 和 8.1 GFLOPs。MSCA

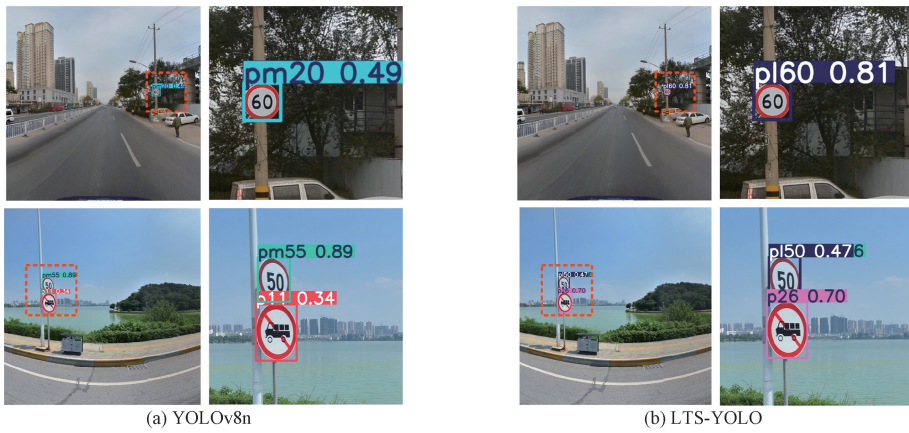


图 6 改进错检问题

Fig. 6 Improvement of misdiagnosis



图 7 改进远距离小目标错检问题

Fig. 7 Improved misdetection of small targets at long range



图 8 改进远距离小目标漏检问题

Fig. 8 Improvement of the lack of detection of small targets at long range

模块的引入使 $mAP@50$ 提升 0.8%，证明了该模块增强不同尺度特征响应的能力。LDA 模块的加入将模型的 $mAP@50$ 提升 1.2%，表明其在处理形变目标方面具有一定优势。NSSE 模块加入后的 $mAP@50$ 提升达到了 1.9%，

且参数量降低至 2.72 M，证明其在特征融合和语义增强方面的效率。当所有模块集成时， $mAP@50$ 提升了 2.5%，在不增加参数的情况下，增加了计算开销，但精度的大幅提升。这充分证明了本研究模块间协同作用的有效性。

表 3 TT100K 数据集上的模型剪枝对比实验结果

Table 3 Results of model pruning comparison experiments on the TT100K dataset

加速比	参数量/M	计算量/G	精度/%	召回率/%	mAP@50-95/%	mAP@75/%	mAP@50/%
1.0	2.95	13.1	79.3	69.2	58.7	70.2	76.8
1.6	2.16	8.19	79.5	69.5	59.1	70.3	76.9
1.8	2.01	7.28	79.2	69.8	58.9	70.5	77.2
2.0	1.92	6.55	78.6	68.4	57.4	69.5	75.9
2.2	1.83	5.95	77.3	67.2	55.8	65.5	74.6
2.4	1.78	5.46	76.1	66.6	54.9	64.6	71.3

表 4 采用不同改进策略后的检测结果

Table 4 Detection results after using different improvement strategies

实验	MSCA	LDA	NSSE	mAP@50/%	参数量/M	计算量/G
YOLOv8n				74.3	3.01	8.1
A	✓			75.1	3.23	9.3
B		✓		75.5	3.17	9.8
C			✓	76.2	2.72	8.2
D	✓		✓	76.4	2.85	10.4
E		✓	✓	76.3	2.76	11.2
F	✓	✓	✓	76.8	2.95	13.1

注:✓表示采用了改进后的策略

3.7 模型决策可视化对比

为了深入分析模型的检测机制,本研究采用 LayerCam^[34]可视化技术对模型决策过程进行可解释性分析。如图 9 所示,该方法通过生成热力图直观呈现模型关注的关键区域。



图 9 LayerCam 可视化

Fig. 9 LayerCam visualisation

观察图 9 中两种方法的可视化热力图对比可知,LTS-YOLO 相较于 YOLOv8n 展现出更优异的检测性能。特别是在远距离目标和遮挡场景下,LTS-YOLO 仍能保持清晰集中的热力响应,而 YOLOv8n 则出现响应弥散或缺失现象。这些结果验证了 LTS-YOLO 通过多尺度特征融合和注意力机制等改进,显著增强了模型对困难样本的感知能力,使其在复杂场景下仍能保持精确的检测性能。该可视化分析在定量分析基础上进一步证实了 LTS-YOLO 架构改进的有效性。

3.8 泛化性验证

为了评估 LTS-YOLO 模型在不同数据集上的适应性,本研究在 CCTSDB 数据集上进行了泛化能力验证。如表 5 所示,在 CCTSDB 数据集上,基线模型 YOLOv8n 的 mAP@50 为 79.8%,而改进后的 TS-YOLO 模型则达到了 81.6%,有效提升了检测精度,但参数量和计算量有所增加。最终的轻量化模型 LTS-YOLO 在保持了高达 81.4%的 mAP@50 的同时,成功将参数量降低至 2.14 M,计算量降至 7.5 GFLOPs,相较于 YOLOv8n 实现了显著的轻量化,同时在精度上有所超越。与其他轻量化模型如 YOLOv5n、YOLOv9t 等的对比表明,LTS-YOLO 在保持明显高精度的同时,也展现出优越的轻量化性能。该结果有效验证了所提出的方法在不同数据集上的鲁棒性和泛化能力。

表 5 CCTSDB2021 数据集上的泛化能力验证结果
Table 5 Generalization experiment results on the CCTSDB2021 dataset

模型	参数量/M	计算量/G	精度/%	召回率/%	mAP@50-95/%	mAP@75/%	mAP@50/%
YOLOv5n	2.51	7.1	87.7	69.6	49.2	57.4	77.7
YOLOv6n	4.30	11.8	85.7	70.8	49.1	55.8	77.9
YOLOv8n	3.01	8.2	89.2	69.3	50.9	59.2	79.8
YOLOv9t	1.98	7.6	86.6	71.8	48.9	55.2	77.5
YOLOv10n	2.28	6.6	85.4	70.3	51.1	59.9	78.4
YOLOv11n	2.58	6.3	90.1	69.6	50.9	58.9	79.1
YOLOv12n	2.57	6.3	88.3	70.4	51.6	61.0	78.8
TS-YOLO	2.95	13.1	90.3	73.0	52.4	62.1	81.6
LTS-YOLO	2.14	7.5	90.5	73.1	52.6	61.8	81.4

4 结 论

本文提出了一种融合通道剪枝的轻量化 YOLOv8 交通标志检测算法 LTS-YOLO,旨在解决现有检测方法在资源受限设备上部署时的计算负担及对小目标与形变目标检测能力不足的问题。基于轻量级基线模型 YOLOv8n,本研究首先设计了融合噪声抑制与深层语义增强的特征融合模块 NSSE,以强化复杂背景下的语义表达并抑制背景噪声;随后在主干网络与检测头前分别引入多尺度通道注意力 MSCA 与局部形变注意力 LDA,以提升模型对尺度变异与几何形变的感知与鲁棒性,从而构建了高精度的改进模型 TS-YOLO。最后,采用基于 BatchNorm 缩放因子的通道剪枝策略对 TS-YOLO 进行压缩,得到轻量化模型 LTS-YOLO,在显著降低模型复杂度的同时最大限度保留了检测性能。

在两个具有代表性的交通标志检测公开数据集 TT100K 与 CCTSDB 上的实验证明了所提方法的有效性。在 TT100K 数据集上,与基线模型 YOLOv8n 相比,经过通道剪枝的 LTS-YOLO 模型在保持高精度的同时,参数量减少了约 33%,计算量降低了约 10%,实现了在速度和精度上的双重优化。在 CCTSDB 数据集上,LTS-YOLO 在精度上超越了基线模型,并显著优于其他主流轻量化模型。

综上,本文提出的 LTS-YOLO 算法成功地实现了交通标志检测的高精度与轻量化兼顾,为将先进的交通标志识别技术部署到边缘计算设备提供了可行且高效的解决方案,在提升自动驾驶安全性和效率方面具有重要的应用价值。

参考文献

[1] 秦伦明,张云起,崔昊杨,等.基于改进 RT-DETR 的极端天气下交通标志检测方法[J].电子测量技术,2025,48(9):56-64.
QIN L M, ZHANG Y Q, CUI H Y, et al. Improved RT-DETR based method for traffic sign recognition in

extreme weather [J]. Electronic Measurement Technology, 2025, 48(9): 56-64.

- [2] 刘美辰,李杰,陈廷伟. DTI-YOLO:改进 YOLOv10s 的交通标志检测模型[J]. 计算机工程与应用, 2025, 61(17):112.
LIU M CH, LI J, CHEN T W. DTI-YOLO: Improved YOLOv10s traffic sign detection model[J]. Computer Engineering and Applications, 2025, 61(17): 112.
- [3] 赵磊,李栋. PMM-YOLO:多尺度特征融合的交通标志检测算法[J]. 计算机工程与应用, 2025, 61(4): 262-271.
ZHAO L, LI D. PMM-YOLO: Traffic sign detection algorithm with multi-scale feature fusion [J]. Computer Engineering and Applications, 2025, 61(4): 262-271.
- [4] 全智.基于 YOLOv7 的数据增广和多尺度注意力交通标志检测方法研究[D].南昌:南昌大学,2024.
QUAN ZH. Research on data augmentation and multi-scale attention traffic sign detection method based on YOLOv7 [D]. Nanchang: Nanchang University, 2024.
- [5] 陆维,吴锡.基于深度学习的复杂天气场景交通标志检测[J].软件导刊,2025,24(6):175-184.
LU W, WU X. Traffic sign detection in complex weather conditions based on deep learning [J]. Software Guide, 2025, 24(6): 175-184.
- [6] 赖华清.复杂场景下的交通标志小目标检测技术研究[D].武汉:武汉轻工大学,2024.
LAI H Q. The research on small object detection technology of traffic signs in complex scenes [D]. Wuhan: Wuhan Polytechnic University, 2024.
- [7] 许明,屈泰澎,姜彦吉.改进 YOLOv7 在复杂场景下的交通标志检测算法[J].计算机工程,2025,51(2): 335-343.
XU M, QU T P, JIANG Y J. Improved YOLOv7 traffic

- sign detection algorithm in complex scenarios [J]. *Computer Engineering*, 2025, 51(2): 335-343.
- [8] 杨伟林. 交通标志检测与分类算法研究[D]. 沈阳: 东北大学, 2019.
- YANG W L. Research on traffic sign detection and classification algorithm [D]. Shenyang: Northeastern University, 2019.
- [9] 解浩龙, 张孝龙, 魏培旭, 等. 基于 RT-DETR 的轻量化交通标志检测 [J]. *应用光学*, 2025, 46(2): 300-308.
- XIE H L, ZHANG X L, WEI P X, et al. Lightweight traffic sign detection based on RT-DETR [J]. *Journal of Applied Optics*, 2025, 46(2): 300-308.
- [10] 罗向龙, 吕温馨, 石镇岳, 等. 改进 YOLOv8n 的轻量化交通标志检测算法 [J]. *激光与光电子学进展*, 2025, 62(12): 422-433.
- LUO X L, LYU W X, SHI ZH Y, et al. Improved lightweight traffic sign detection algorithm for YOLOv8n [J]. *Laser & Optoelectronics Progress*, 2025, 62(12): 422-433.
- [11] 何鑫. 基于 YOLOv5 的轻量化交通标志检测算法研究[D]. 淮南: 安徽理工大学, 2024.
- HE X. Research on lightweight traffic sign detection algorithm based on YOLOv5 [D]. Huainan: Anhui University of Science and Technology, 2024.
- [12] 王海瑞, 贺存龙, 朱贵富. 基于 MOMP-YOLOv8 的交通标志检测算法研究 [J]. *计量学报*, 2025, 46(9): 1385-1394.
- WANG H R, HE C L, ZHUG F. Research on traffic sign detection algorithm based on MOMP-YOLOv8 [J]. *Acta Metrologica Sinica*, 2025, 46(9): 1385-1394.
- [13] 王泽玄, 雷雪梅. 基于 RT-DETR 的轻量化交通标志检测算法 [J]. *现代电子技术*, 2025, 48(18): 57-64.
- WANG Z X, LEI X M. Lightweight traffic sign detection algorithm based on RT-DETR [J]. *Modern Electronic Technique*, 2025, 48(18): 57-64.
- [14] WANG L CH, AI Q, SHEN X L. Multi-scale lightweight algorithm for UAV aerial target detection [J]. *Engineering Letters*, 2024, 32(12): 2324-2335.
- [15] 李牧, 陶启婷, 柯熙政. 基于改进 YOLOv5 的小目标交通标志检测算法 [J]. *计算机应用*, 2025, 45(S1): 239-244.
- LI M, TAO Q T, KE X ZH. Detection algorithm for small-target traffic signs based on improved YOLOv5 [J]. *Journal of Computer Applications*, 2025, 45(S1): 239-244.
- [16] 韩东旭, 谢雨飞. 基于改进的小目标交通标志检测算法研究 [J]. *电子测量技术*, 2025, 48(6): 28-37.
- HAN D X, XIE Y F. Research on improved small target traffic sign detection algorithm [J]. *Electronic Measurement Technology*, 2025, 48(6): 28-37.
- [17] SOHAN M, SAI R T, RAMI R C V. A review on YOLOv8 and its advancements [C]. *International Conference on Data Intelligence and Cognitive Informatics*, 2024: 529-545.
- [18] 沈学利, 王灵超. 基于 YOLOv8n 的无人机航拍目标检测 [J]. *计算机系统应用*, 2024, 33(7): 139-148.
- SHEN X L, WANG L CH. UAV aerial photography target detection based on YOLOv8n [J]. *Computer Systems & Applications*, 2024, 33(7): 139-148.
- [19] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection [C]. *IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 779-788.
- [20] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector [C]. *European Conference on Computer Vision*. Cham: Springer International Publishing, 2016: 21-37.
- [21] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger [C]. *IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 7263-7271.
- [22] FARHADI A, REDMON J. YOLOv3: An incremental improvement [C]. *Computer Vision and Pattern Recognition*, 2018: 1-6.
- [23] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection [J]. *ArXiv preprint arXiv: 2004.10934*, 2020.
- [24] LAW H, DENG J. Cornernet: Detecting objects as paired keypoints [C]. *European Conference on Computer Vision (ECCV)*, 2018: 734-750.
- [25] DUAN K W, BAI S, XIE L X, et al. Centernet: Keypoint triplets for object detection [C]. *IEEE/CVF International Conference on Computer Vision*, 2019: 6569-6578.
- [26] TIAN ZH, SHEN CH H, CHEN H, et al. Fcos: Fully convolutional one-stage object detection [C]. *IEEE/CVF International Conference on Computer Vision*, 2019: 9627-9636.
- [27] WANG C 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.
- [28] WANG C Y, YE H I H, LIAO H Y M. YOLOv9: Learning what you want to learn using programmable

- gradient information [C]. European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2024: 1-21.
- [29] WANG AO, CHEN H, LIU L H, et al. YOLOv10: Real-time end-to-end object detection [J]. ArXiv preprint arXiv:2405.14458, 2024.
- [30] TIAN Y, YE Q, DOERMANN D. YOLOv12: Attention-centric real-time object detectors[J]. ArXiv preprint arXiv:2502.12524, 2025.
- [31] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module [C]. European Conference on Computer Vision(ECCV), 2018: 3-19.
- [32] ZHU ZH, LIANG D, ZHANG S H, et al. Traffic-sign detection and classification in the wild[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016: 2110-2118.
- [33] ZHANG J, ZOU X, KUANG L D, et al. CCTSDB 2021: A more comprehensive traffic sign detection benchmark [J]. Human-centric Computing and Information Sciences, 2022, 12. DOI: 10.22967/HICIS.2022.12.023.
- [34] JIANG P T, ZHANG C B, HOU Q, et al. Layercam: Exploring hierarchical class activation maps for localization[J]. IEEE Transactions on Image Processing, 2021, 30: 5875-5888.

作者简介

艾强, 硕士研究生, 主要研究方向为时空数据挖掘、图像处理等。

E-mail: aiqiang@stu.qhnu.edu.cn

冯永安, 副教授, 主要研究方向为图像处理和大数据技术等。

E-mail: 5179554@qq.com

王灵超, 硕士, 主要研究方向为图像与视觉信息计算、目标检测等。

E-mail: 20241035@lise.edu.com

董丽欣(通信作者), 副教授, 主要研究方向为复杂网络及其应用。

E-mail: 2013040@qhnu.edu.cn