·面向智能驾驶的目标检测与道路通行区域感知技术专题·

面向自动驾驶场景的多尺度车辆检测算法*

周亚罗[1,3]　王海明[1,3]　刘文广[2]　张瑞成[1,3]

（1.华北理工大学电气工程学院，唐山 063210；2.首钢京唐钢铁联合有限责任公司，唐山 063205；3.河北省风光氢储安全监测与智能运行技术创新中心，唐山 063210）

【摘要】针对自动驾驶交通场景中车辆检测目标尺度分布不均、背景干扰显著、小目标结构失真以及计算复杂度高等问题，提出一种基于YOLOv11n的改进算法ELD-YOLO算法。通过引入C3k2-EMSC模块，增强多尺度特征提取能力，在复杂交通场景下获得更丰富的上下文信息；LSDECD模块利用共享卷积与群归一化强化特征表达，减少冗余响应；通过动态采样（DySample）提升小目标与边缘特征表征能力。试验结果表明：在KITTI与BDD100K数据集上，ELD-YOLO参数量较YOLOv11n降低16%，mAP@0.5分别达到89.7%和52.7%，降低模型复杂度的同时，显著提升检测精度与鲁棒性。

主题词：自动驾驶　车辆检测　YOLOv11n　多尺度检测

中图分类号：TP391.41　　**文献标志码：**A　　**DOI:** 10.19620/j.cnki.1000-3703.20250865

A Multi-Scale Vehicle Detection Algorithm for Autonomous Driving Scenarios

Zhou Yaluo[1,3], Wang Haiming[1,3], Liu Wenguang[2], Zhang Ruicheng[1,3]

（1. North China University of Science and Technology, College of Electrical Engineering, Tangshan 063210; 2. Shougang Jingtang Iron and Steel United Co., Ltd., Tangshan 063205; 3. Hebei wind-Solor-Hydrogen-storage system Security Monitoring and Intelligent Operation Center of technology Innovation, Tangshan 063210）

【Abstract】 To address challenges in autonomous driving scenarios, such as uneven distribution of target scales, significant background interference, structural distortion of small objects, and high computational complexity, this paper proposes an improved algorithm, ELD-YOLO, based on YOLOv11n. The C3k2-EMSC module is introduced to enhance multi-scale feature extraction, capturing richer contextual information in complex traffic scenarios. The LSDECD module is designed to strengthen feature representation through shared convolutions and group normalization, reducing redundant responses. The DySample module is introduced to enhance small object and edge feature representation through dynamic sampling. Results demonstrate that on the KITTI and BDD100K datasets, the parameter quantity of ELD-YOLO is reduced by 16% compared with YOLOv11n, while achieving mAP@0.5s of 89.7% and 52.7%, respectively. This approach significantly improves detection accuracy and robustness while reducing model complexity.

Key words: Autonomous driving, Vehicle inspection, YOLOv11n, Multi-scale detection

【引用格式】 周亚罗, 王海明, 刘文广, 等. 面向自动驾驶场景的多尺度车辆检测算法[J]. 汽车技术, 2026(7): 1-10.

ZHOU Y L, WANG H M, LIU W G, et al. A Multi-Scale Vehicle Detection Algorithm for Autonomous Driving Scenarios[J]. Automobile Technology, 2026(7): 1-10.

1　前言

车辆检测作为环境感知的核心任务之一，其性能直接影响自动驾驶系统的安全性与可靠性。真实交通场景中车辆尺度分布不均，不同车型在成像尺寸与空间分布上差异显著，导致检测模型难以兼顾多尺度目标的有效表征。同时，复杂背景干扰削弱了目标与背景的判别性，小目标在特征提取中易出现结构失真与细节丢失，使得高计算复杂度的检测模型难以在实际场景中部署。因此，在计算资源受限的条件下，实现尺度不均目标的准确检测并保持小目标结构完整性，成为亟待解决的关键问题[1]。

*基金项目：河北省自然科学基金项目（F2018209201）；唐山市科技局科技计划项目（22130213G）。

基于深度学习的目标检测主要使用两阶段检测器和单阶段检测器[2]。其中，单阶段检测器依据端到端结构，实现了较高的推理速度。面对车辆与行人检测任务，基于YOLO系列的单阶段检测框架近年来不断演进[3]，郭雨等[4]在YOLOv5中引入多尺度注意力，增强了小尺度车辆的语义表征，但该方法依赖复杂的跨尺度特征交互，模型结构复杂且参数量较大。Liu等[5]通过多尺度特征融合和轻量化模块降低了模型参数量，但其在高密度车流遮挡场景下的检测性能仍有待验证。崔家礼等[6]引入多膨胀率注意力机制，增强了上下文建模能力，但在低光照或恶劣天气条件下，小目标结构表征稳定性不足。孙华等[7]通过引入复杂注意力机制，强化小目标特征选择能力，但其结构依赖全局或稀疏建模，多尺度结构信息的计算开销较高。许德刚等[8]采用多分支特征提取增强局部结构表达，但特征冗余增加易引入干扰响应，难以在轻量化条件下实现稳定的小目标检测。

现有车辆目标检测方法在尺度分布不均、复杂背景及遮挡场景中，小尺度车辆的结构特征易在多层下采样与特征压缩过程中退化。同时，在低光照、雨雪等复杂成像环境下，模型对结构信息与语义信息的协同建模能力有限，计算开销较高[9]。因此，本文提出一种基于改进YOLOv11n的增强轻量级目标检测（Enhanced Lightweight Detection You Only Look Once，ELD-YOLO）网络：通过引入C3k2-多尺度特征增强卷积模块（C3k2 Enhanced Multi-Scale Convolution，C3k2-EMSC）增强模型多尺度特征提取能力，以在复杂场景下获得更丰富的上下文信息；设计轻量级共享细节增强卷积检测头（Lightweight Shared Detail-Enhanced Convolutional Detection Head，LSDECD），利用共享卷积与群归一化强化特征表达，降低背景干扰；引入动态采样（Dynamic Sampling，DySample）模块，提升小目标与边缘特征表征能力。

2 ELD-YOLO模型

在自动驾驶环境中，检测模型需同时识别多类目标，并适应光照、遮挡等干扰。为兼顾准确性与效率，本文构建了轻量化的ELD-YOLO模型，其总体架构如图1所示。

C3k2-EMSC模块通过多尺度卷积分支提取不同感受野特征并在通道维度融合，增强车辆结构信息与细节表达，同时抑制背景干扰，提高小目标及多尺度目标的鲁棒性。

LSDECD检测头采用共享卷积实现跨尺度特征一致性，减少冗余响应，并通过4类差分卷积显式建模车辆边缘与局部细节，提升对复杂背景下多尺度车辆的判别能力。

DySample通过动态预测采样偏移自适应调整上采样位置，使车辆边缘、车灯及车牌等高频结构能够精细

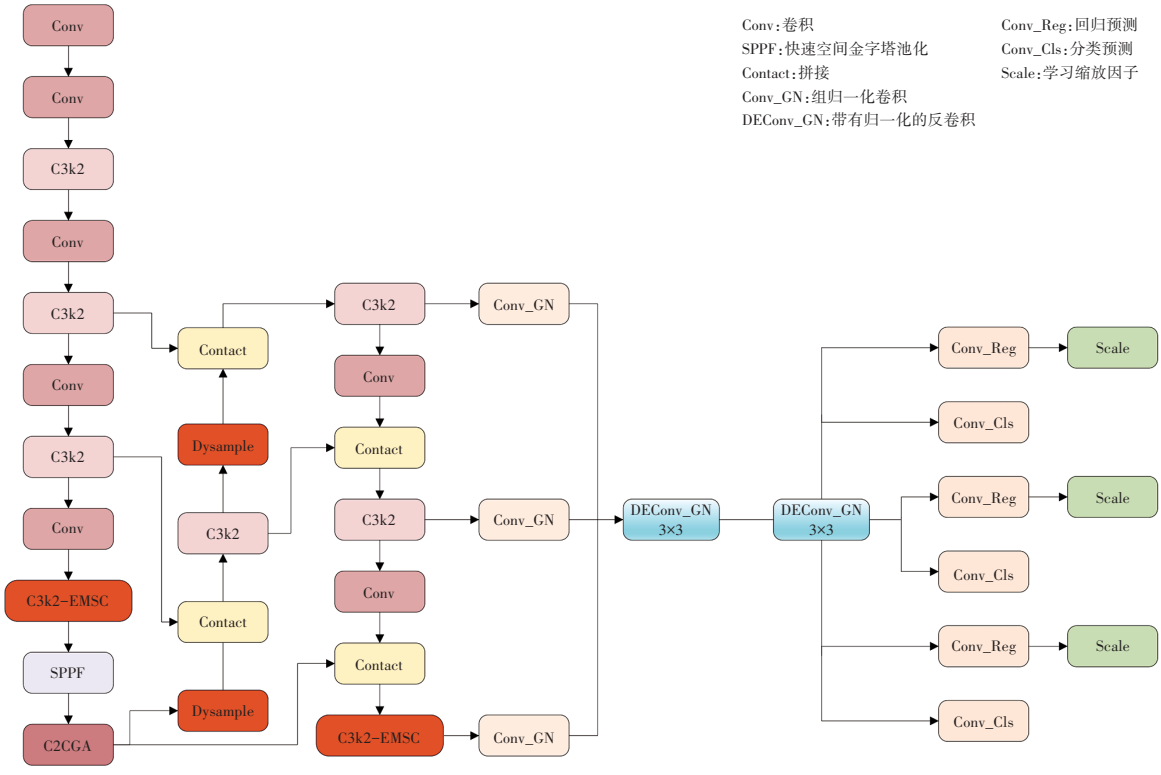


图1 ELD-YOLO结构

重建，保持空间结构完整性，提升整体检测性能。

2.1 多尺度卷积模块

虽然YOLOv11n的C3k2模块整合残差块（C3k）结构具备较强的特征提取能力，但面对车辆尺度差异显著、小目标占比高与遮挡干扰并存情况时，对多尺度目标的表征能力和特征鲁棒性不足，根本原因在于单一尺度卷积在同一层内难以兼顾建模局部细节与大尺度上下文信息，导致模型对尺度分布不均目标的响应能力受限。

为此，本文引入一种高效的多尺度卷积（Efficient Multi-Scale Convolution，EMSC），如图2所示。该模块在同一特征层内引入多种尺度的卷积核，构建不同感受野的特征映射，单层内可同时感知局部细节与大尺度上下文信息，使特征表示对目标尺度变化保持更高的稳定性。

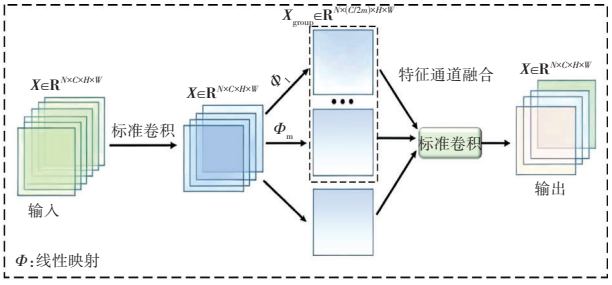


图2 EMSC模块架构

首先，使用标准卷积进行基础特征映射，统一通道维度并提取初始语义信息；随后，特征在通道维度上被划分为快速路径分支和多尺度特征分支。其中，快速分支作为快速路径直接保留，用于维持原始语义信息，在特征融合过程中相当于引入恒等映射项，有助于避免多尺度卷积导致的特征平滑与语义漂移，在增强上下文建模能力的同时维持目标定位精度。多尺度特征分支的通道特征按照预定义的组数m进行划分，对各通道组施加线性映射Φ，结合不同尺度的卷积核捕获多感受野下的空间信息。最后，将快速路径特征与各通道组输出的多尺度特征在通道维度上融合，利用逐点卷积实现通道间信息的整合与重组，生成最终的综合特征表示。

从整体映射角度，EMSC模块可表示为一组多尺度卷积映射函数与快速路径特征的融合函数，其整体输出可表示为：

$$\boldsymbol{X}=(\boldsymbol{X}_{\text{cheap}},\boldsymbol{X}_{\text{group}}),\boldsymbol{X}_{\text{cheap}}\in\mathbf{R}^{N\times\frac{C}{2}\times H\times W} \tag{1}$$

$$\boldsymbol{X}_{\text{group}}=\{\boldsymbol{X}_{\text{group}}^{(i)}\}_{i=1}^{m},\boldsymbol{X}_{\text{group}}^{(i)}\in\mathbf{R}^{N\times\frac{C}{2}\times H\times W} \tag{2}$$

式中：$\boldsymbol{X}$为完整4维特征张量，$\boldsymbol{X}_{\text{cheap}}$、$\boldsymbol{X}_{\text{group}}$分别为快速分支和多尺度分支特征张量，$N$为样本批次数量，$C$、$H$、$W$分别为特征总通道数、高度和宽度，$m$、$i$分别为分组数量和索引，$\mathbf{R}$为实数空间。

$\boldsymbol{X}_{\text{cheap}}$保留输入特征的低频基础信息，通过1×1卷积快速投影完成基础特征提取，$\boldsymbol{X}_{\text{group}}$专注于精细多尺度特征的建模，在通道维度上按照预定义的组数划分为m个子特征图。其中，每个子特征图$\boldsymbol{X}_{\text{group}}^{(i)}$对应一个特定尺度的特征子空间，在不同感受野范围内捕获目标的结构与上下文信息。其对应尺度的卷积映射为：

$$\boldsymbol{Y}_i=\boldsymbol{\Phi}_i(\boldsymbol{X}_{\text{group}}^{(i)})=conv_{k_i}(\boldsymbol{X}_{\text{group}}^{(i)}) \tag{3}$$

式中：$\boldsymbol{\Phi}_i$为第i个尺度下的线性卷积映射，k_i为对应卷积核尺寸。

各尺度分支输出$\boldsymbol{Y}_i$与快速路径$\boldsymbol{X}_{\text{cheap}}$在通道维度上进行拼接，得到：

$$\boldsymbol{Y}=concat(\boldsymbol{X}_{\text{cheap}},\boldsymbol{Y}_1,\boldsymbol{Y}_2,\cdots,\boldsymbol{Y}_m) \tag{4}$$

该融合操作在保持低频基础语义与空间结构信息的同时，引入来自不同感受野的多尺度上下文响应，使特征在单层内兼顾局部细节与全局结构。随后引入一个1×1卷积进行通道融合：

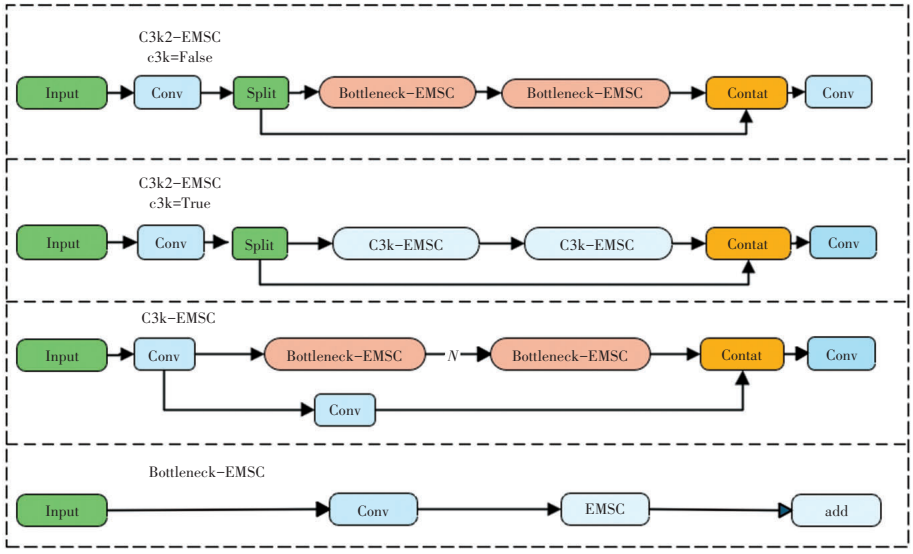
$$\boldsymbol{Z}=conv_{1\times1}(\boldsymbol{Y}) \tag{5}$$

通过跨通道线性组合实现不同尺度特征的显式交互与信息重加权，在不改变空间分辨率的前提下生成最终的多尺度增强特征表示。EMSC在单层内可实现尺度域扩展，采用通道分组方式进行多尺度并行建模，保持计算复杂度近似线性的同时，显著提升了特征表达多样性。

鉴于EMSC模块在低层特征中计算开销较大，且其多尺度分组与通道建模效果依赖于较高的通道维度，本文仅在主干网络和颈部网络中通道数大于512的高语义层引入C3k2-EMSC模块，以增强上下文信息建模能力。该模块结构如图3所示，在空间维度上利用多尺度卷积捕获不同感受野下的目标结构与细节信息，在通道维度上通过分组特征融合增强高层语义表达与上下文建模能力，有效提升模型对多尺度车辆、小目标密集分布以及复杂光照与遮挡场景下的检测鲁棒性。

2.2 共享卷积检测头

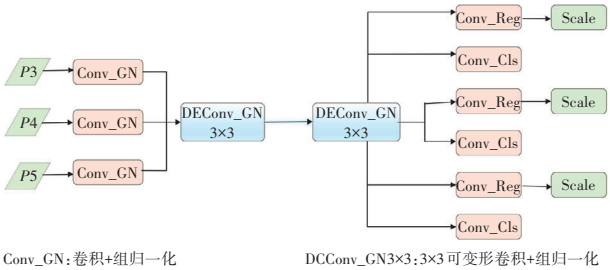
在复杂交通场景中，YOLOv11n原始检测头对不同尺度特征进行独立预测，各卷积参数相互独立，难以有效抑制复杂背景的冗余响应，导致多尺度车辆目标检测性能受限、特征冗余增加以及模型自适应能力不足等问题。为此，本文提出一种共享卷积检测头LSDECD



Conv：标准卷积　Split：通道划分操作　Concat：通道拼接操作　C3k-EMSC：基于EMSC的C3k模块　add：逐元素加法
Bottleneck-EMSC：在Bottleneck结构中以EMSC替代标准卷积的残差模块

图3　C3k2-EMSC结构

（见图4）利用共享卷积机制[10]，对不同尺度分支施加一致的参数约束，以抑制冗余特征表达并增强跨尺度语义一致性。由于检测头的批量归一化层在小批量训练情况下对批量大小（Batch size）较为敏感，本文采用对其不敏感的群归一化，保证检测头训练过程的稳定性与预测一致性。



Conv_GN：卷积+组归一化　DCConv_GN3×3：3×3可变形卷积+组归一化
ConvReg/Conv_Cls：解耦回归头/分类头　Scale：用于边界框回归的可学习缩放因子
P3、P4、P5：来自FPN的多尺度特征

图4　LSDECD结构

为了进一步增强细节信息的提取能力，LSDECD引入细节增强卷积（Detail Enhanced Convolution，DEConv），如图5所示。DEConv由5个并行的卷积分支组成：常规卷积（Vanilla Convolution，VC）、中心差分卷积（Central Difference Convolution，CDC）、角度差分卷积（Angle Difference Convolution，ADC）、水平差分卷积（Horizontal Difference Convolution，HDC）和垂直差分卷积（Vertical Difference Convolution，VDC）。其中，VC负责提取基础语义特征，4个差分卷积从不同方向捕获图像的细粒度变化与边缘信息，显著提升细节表达能力与特征的判别性。各分支特征在训练阶段进行融合，通过重新参数化技术在推理阶段折叠为等效的3×3卷积核，在不增加参数量与计算复杂度的前提下保留多分支结构的细节增强优势。

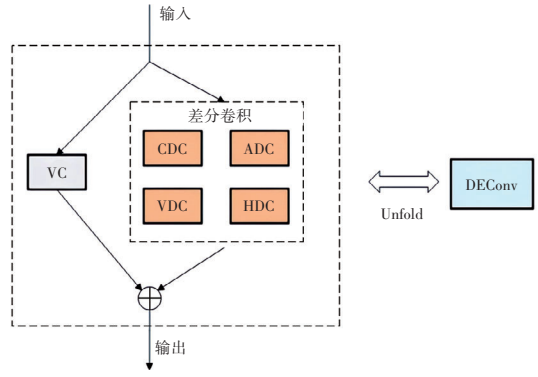


图5　DEConv结构

故融合后的等效核为：

$$\begin{cases} K_{\text{cvt}} = \sum_{i=1}^{5} K_i \\ \boldsymbol{Y} = K_{\text{cvt}} * \boldsymbol{X} \end{cases} \tag{6}$$

式中：K_1~K_5分别为VC、CDC、ADC、HDC、VDC的3×3卷积核，*为常规卷积运算。

引入DEConv后，LSDECD显式强化了车辆边缘与结构区域的局部细节表达，有效抑制了复杂交通场景中背景区域的冗余响应，增强了检测头对多尺度车辆目标的判别能力与对复杂背景变化的适应性。

2.3　动态上采样

原始YOLOv11的特征金字塔结构中，高层语义特征向高分辨率空间映射时通常采用固定插值上采样。该方式缺乏对局部结构变化的自适应能力，易在密集车流及远距离小目标场景中出现边缘失真和语义模糊问题。尤其在目标尺度较小或相邻车辆紧密分布的区域，固定插值上采样难以准确恢复目标的空间结构，削弱了后续检测头对车辆边界与关键部位的判别能力。

为提升上采样特征对车辆结构的表达能力，同时兼顾推理效率需求，本文在YOLOv11颈部网络中引入轻量化动态上采样模块DySample（见图6）替换原有固定插值上采样层。不同于依赖动态卷积核进行特征重建的方法，DySample直接对采样位置进行建模，通过内容感知的采样点重定位实现特征上采样，避免额外的动态卷积核生成增加计算开销。

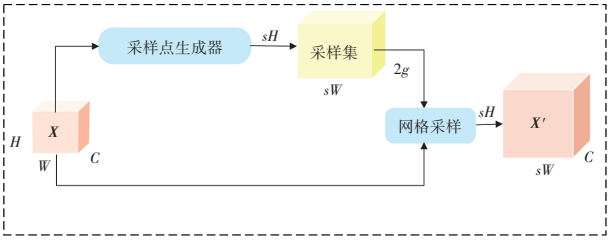
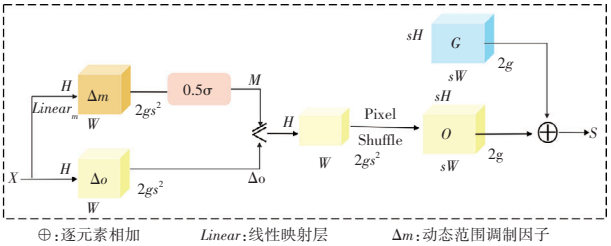


图6　DySample模块结构

DySample由采样点生成与特征重采样构成，输入特征图的尺寸为$C \times H \times W$，采样点生成器根据特征信息预测像素级采样点集合S，尺寸为$2g \times sH \times sW$，其中$2g$为x和y方向的坐标信息，其生成过程如图7所示。随后，网格采样（Grid Sampling）函数根据采样点集中各像素点的空间位置信息，结合双线性插值机制，对输入特征进行重采样与重构，得到新的上采样特征图：

$$\boldsymbol{X}' = f_{\text{grid sample}}(\boldsymbol{X}, \boldsymbol{S}) \tag{7}$$



⊕：逐元素相加　*Linear*：线性映射层　Δm：动态范围调制因子
Δo：候选采样偏移　S：最终采样点集合　Pixel Shuffle：像素重排操作
G：原始采样网络　$o(\cdot)$：Sigmoid激活函数　M：调制后的动态范围因子
O：最终偏移映射

图7　采样点集生成过程

首先，对输入特征$\boldsymbol{X}$分别进行2组独立的线性映射，生成动态调制因子Δm和候选采样偏移Δo。其中，Δm主要调节采样偏移幅度，Δo表征待学习的采样位置偏移信息，其计算过程为：

$$\begin{cases} \Delta m = Linear_m(\boldsymbol{X}), & \Delta m \in \mathbf{R}^{2S^2 \times H \times W} \\ \Delta o = Linear_o(\boldsymbol{X}), & \Delta o \in \mathbf{R}^{2S^2 \times H \times W} \end{cases} \tag{8}$$

为将调制分支输出限制在稳定范围内，首先对其进行Sigmoid非线性映射，将特征归一化至(0,1)区间；随后设置缩放因子为0.5，对动态响应范围进行约束，避免采样偏移过大导致的不稳定问题。故逐点动态范围因子为：

$$M = 0.5 \cdot \sigma(\Delta m) \tag{9}$$

式中：$\sigma()$为Sigmoid函数。

将候选偏移Δo与动态范围因子M在像素层面进行逐元素相乘，并通过线性映射后采用像素重排（Pixel Shuffle）操作进行从通道到空间的重组，生成高分辨率偏移映射O。将原始采样网络G和偏移量O相加，形成最终采样点集合：

$$S = G + O \tag{10}$$

通过对采样坐标进行内容感知建模，DySample将固定插值的规则重建方式转换为受特征引导的自适应的点采样，使采样位置能够随局部结构动态调整，在保持计算效率的同时，实现对车辆边缘、车灯及车牌等高频结构区域的精细对齐，缓解密集车流与远距离小目标场景中的结构重建模糊问题。

3　试验环境与评价指标

3.1　数据集

为评估ELD-YOLO在实际驾驶环境中的车辆与行人检测性能，本文使用KITTI数据集[11]进行模型的训练、验证与测试。该数据集涵盖城市道路、郊区街道及高速公路场景，共7 481张标注图像、80 256个目标实例，单幅图像最多约含15台车辆和30个行人。目标类别划分为行人（Pedestrians）、车辆（Car）和骑行者（Cyclists），按照8∶1∶1随机分为训练集、验证集和测试集，以保证评估公平性。

本文使用BDD100K数据集[12]验证模型的泛化能力，该数据集不参与模型训练，仅用于泛化性能评估。试验中随机选取8 000张图像进行测试，整合为行人（class0）、骑行者（class1）和机动车辆（class2）3类。

3.2　试验环境

本文试验在Ubuntu 20.04操作系统下完成，计算平台配备NVIDIA RTX3080Ti，显存为12 GB，中央处理器使用Intel Xeon Silver 4214R。试验基于PyTorch 2.0.0深度学习框架，并结合CUDA 11.8实现图形处理器加速训练。模型训练过程中，采用随机梯度下降（Stochastic Gradient Descent，SGD）优化器，初始学习率设为0.01，权重衰减参数为5×10^{-4}，动量值为0.973。

3.3　评价指标

模型性能评估指标主要包括召回率（Recall）R、精确度（Precision）P、均值平均精度（mean Average Precision，mAP）、参数量（Parameter）、每秒十亿次浮点运算（Giga Floating Point Operations per second，GFLOPs）和帧率（Frames Per Second，FPS）。其中：mAP@0.5反映模型总体检测精度；mAP@0.5∶0.95在严格的阈值范围内评估模型的边界回归能力；Recall衡量模型正样本的

检出率，Precision体现其误检情况；GFLOPs用于单次推理的计算开销。各指标公式为：

$$\begin{cases} P = \dfrac{TP}{TP + FP} \\ R = \dfrac{TP}{TP + FN} \\ A_{\mathrm{mAP}} = \dfrac{1}{N}\sum_{i=1}^{N} A_{\mathrm{AP},i} \end{cases} \tag{11}$$

式中：TP为正确识别为目标类别的样本数，FP为误判为目标类别的负样本数，FN为未被检出的真实目标样本数，N为数据集的类别数量，$A_{\mathrm{AP},i}$为第i类目标的平均精度。

4 试验结果分析

4.1 消融试验

为验证本文模型在目标检测任务中的有效性，量化各改进模块的性能，本文以YOLOv11n作为基线网络，在KITTI数据集上开展消融试验，结果如表1所示。其中，A、B、C分别为C3k2-MESC模块、LSDECD检测头和Dysample模块。

表1 KITTI数据集上各模块性能分析

A	B	C	mAP@0.5/%	mAP@0.5:0.95/%	R/%	计算量/×10⁹次	FPS/帧·s⁻¹	参数量/×10⁶个
			86.8	60.0	78.1	6.4	107.3	2.58
√			87.9	61.7	79.1	6.2	120.1	2.48
	√		88.0	61.3	79.5	6.0	135.6	2.30
		√	88.4	61.3	78.6	6.3	100.3	2.59
√	√		88.4	62.7	79.4	5.9	125.4	2.16
√		√	88.2	62.5	80.5	6.3	104.1	2.49
	√	√	88.4	62.7	79.5	6.0	110.2	2.27
√	√	√	89.7	63.0	82.7	6.0	115.5	2.17

注：√表示添加该模块。

由表1可知，相较于基线模型，单独引入C3k2-EMSC，模型的mAP@0.5和mAP@0.5:0.95分别上涨了1.1百分点和1.7百分点，FPS由107.3帧/s提升至120.1帧/s，计算量和参数量分别下降3.2%和3.87%，表明该模块能增强多尺度场景下小目标结构特征建模能力，且具备良好的推理效率。仅使用LSDECD时，mAP@0.5和mAP@0.5:0.95分别提高了1.2百分点和1.3百分点，推理速度进一步提升至135.6帧/s，计算量与参数量分别下降6.25%和7.75%，说明共享卷积与细节增强机制有效缓解了小目标特征提取过程中的细节损失。单独替换Dysample，mAP@0.5和mAP@0.5:0.95分别提高了1.6百分点和1.3百分点。

相较于基线模型，本文方法的参数量仅为2.17×10⁶个，下降16%，mAP@0.5和mAP@0.5:0.95分别提高了2.9百分点和3百分点，R提高4.6百分点，且推理速度仍保持在115.5帧/s。这表明ELD-YOLO在保证模型轻量化的同时，有效提升了复杂自动驾驶交通场景中对多尺度车辆目标的检测性能。

4.2 对比试验

为了验证ELD-YOLO模型在驾驶场景下的车辆检测性能，使用KITTI数据集，与YOLOv8n、YOLOv9-tiny、YOLOv10n、YOLOv11n、YOLOv12n[13]、RetinaNet[14]、SSD[15]、Faster-R-CNN[16]模型进行对比试验，结果如表2所示。

表2 KITTI数据集上不同模型对比结果

模型	P/%	R/%	mAP@0.5/%	参数量/×10⁶个
SSD	88.4	83.1	77.6	76.0
RetinaNet	64.5	55.6	60.1	37.8
Faster-R-CNN	82.3	79.9	84.8	41.4
YOLOv8n	89.5	78.3	86.6	2.68
YOLOv9-tiny	88.9	77.5	86.0	2.48
YOLOv10	87.1	78.2	86.7	2.26
YOLOv11n	90.3	78.1	86.8	2.58
YOLOv12	87.2	78.5	86.2	2.51
ELD-YOLO	88.5	82.7	89.7	2.17

由表2可知，ELD-YOLO以仅2.17×10⁶个参数量成为规模最小的模型，较YOLOv11n减少16%。同时mAP@0.5达到89.7%，为所有方法中最高，较YOLOv11n提升2.9百分点，Recall提升至82.7%，提高4.6百分点。上述结果表明，ELD-YOLO在低参数规模下实现了精度与召回率的同步提升，适用于资源受限的车辆检测场景。

为系统评估所提出的C3k2-EMSC模块在多尺度特征提取与细节表达方面的有效性，本文构建了多种基于C3k2结构的变体模块，并开展对比试验。在基础C3k2模块中分别引入扩张残差块（Dilated Residual Block，DRB）[17]、动态权重重参数化（Dynamic Weight Re-parameterization，DWR）[18]、动态倒置多分支模块（Dynamic Inverted Multi-Branch Block，DIMB）[19]、倒置多残差模块（Inverted Multi-Residual Block，IMRB）[20]、多尺度高效信息提取模块（Multi-Scale Efficient Information Extraction，MSEIE）[21]及动态幽灵卷积（GhostDynamicConv，GDC）[22]进行对比试验，结果如表3所示。

表3　不同模块改进C3k2的对比结果

模块	*P*/%	*R*/%	mAP@0.5/%	mAP@0.5:0.95	参数量/×10⁶个
C3k2	90.3	78.1	86.8	60.0	2.58
C3k2-DRB	87.7	78.2	86.1	60.2	2.44
C3k2-DWR	89.5	79.3	87.6	62.3	2.56
C3k2-DIMB	88.9	77.2	87.2	60.3	2.30
C3k2-IMRB	88.6	78.9	87.1	60.3	2.43
C3k2-MSEIE	87.5	79.0	86.8	61.0	2.50
C3k2-GDC	89.2	75.6	86.0	60.2	2.27
C3k2-EMSC	90.1	79.1	87.9	61.7	2.48

由表3可知，C3k2-EMSC在各项评价指标上均取得最优结果，其中，mAP@0.5达到87.9%，mAP@0.5:0.95同样为所有对比结构中最高，表明其在多尺度目标定位与泛化能力方面具有优势。同时，Precision和Recall分别达到90.1%、79.1%，有效降低了误检与漏检。

对比其他改进变体可知：部分模块（DWR、MSEIE）虽然引入了较复杂的结构设计，但参数量与计算开销的增加未能使模型性能提升；轻量化变体（DWB、IMRB）在模型规模压缩的同时，多尺度上下文建模能力受限，对小目标与远距离目标的表征不足。C3k2-EMSC在仅2.48×10⁶个参数量的条件下实现了最佳检测精度，能够实现检测性能与模型复杂度间的最优平衡。

4.3　泛化试验

在BDD100K数据集上开展模型泛化试验，该数据集涵盖多样化的驾驶场景，覆盖不同天气条件（晴天、阴天、雨天）、不同时段（白天、夜晚）及多种城市道路环境，满足评估模型在复杂真实交通场景下的稳健性与适应性，结果如表4所示。

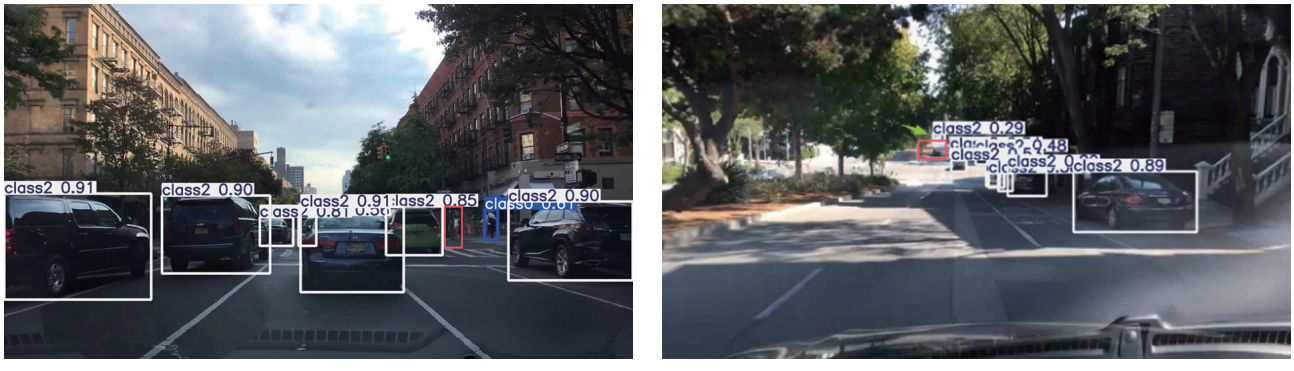
表4　BDD100K模型泛化试验

模型	*P*/%	*R*/%	mAP@0.5/%	mAP@0.5:0.95	计算量/×10⁹次	参数量/×10⁶个
YOLOv11n	59.5	46.3	49.7	25.5	6.4	2.58
ELD-YOLO	63.8	47.6	52.7	27.3	6.0	2.17

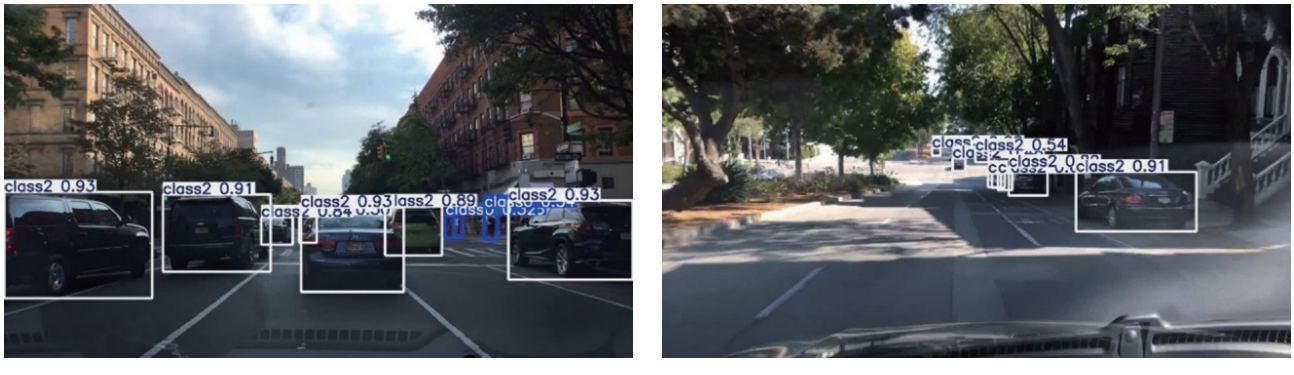
在轻量化模型约束下，ELD-YOLO的检测性能明显优于YOLOv11n，mAP@0.5提升3百分点，表明模型在多尺度目标的整体检测与特征表达方面能够更检出易被忽略的车辆与行人目标；Precision提升4.3百分点，可有效降低了误检率，提升检测结果的可靠性。同时，在雨雪及复杂光照场景中，模型表现稳定，具备较强的场景适应性与泛化能力。

4.4　可视化试验

为了直观验证改进模型在实际场景中性能，在不同光照条件、雪天、雨天以及夜晚等环境下针对行人（class 0）、骑行者（class 1）和机动车辆（class 2）的可视化检测结果与YOLOv11进行对比分析，如图8所示。



YOLOv11n



ELD-YOLO

（a）光照条件

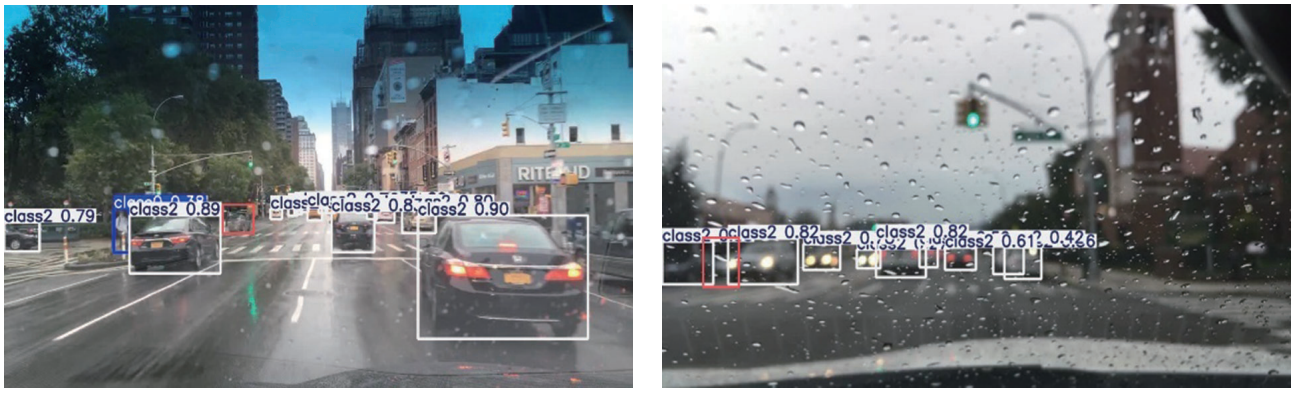


YOLOv11n



ELD-YOLO

(b)大雪影响

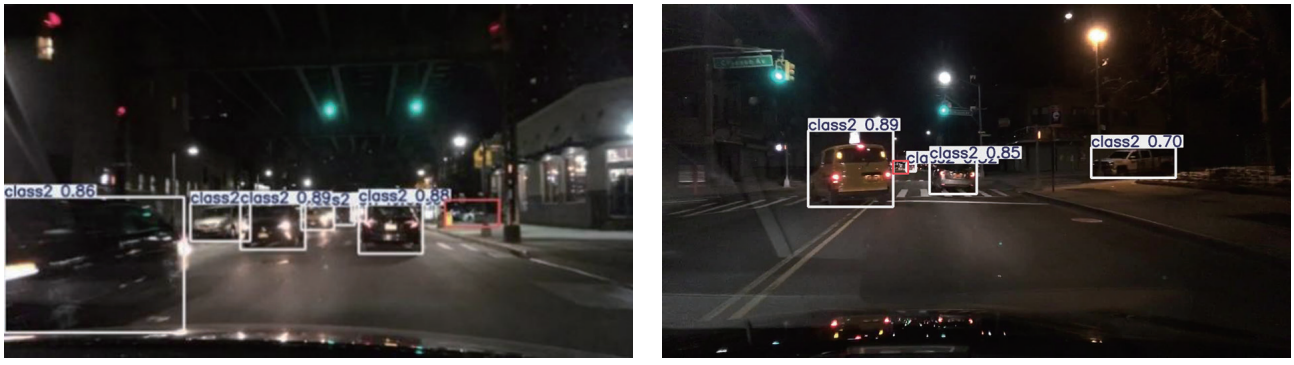


YOLOv11n

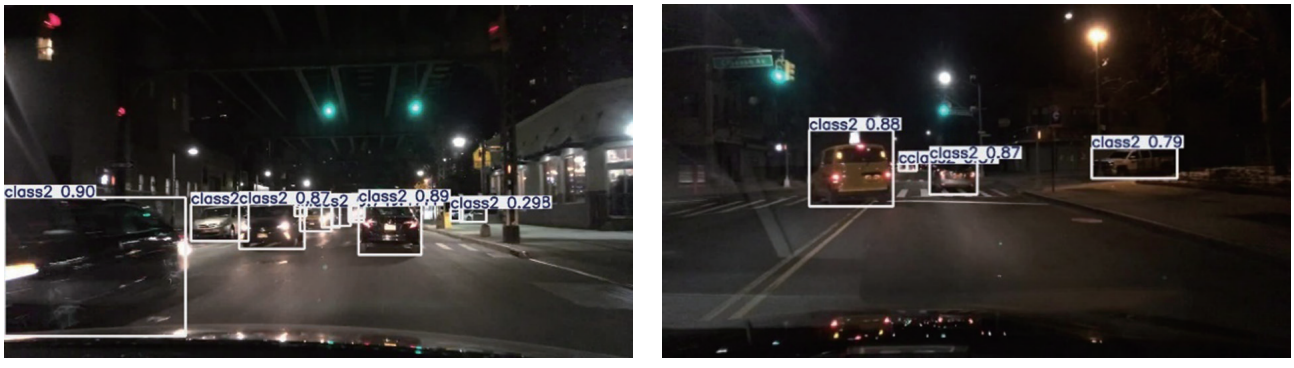


ELD-YOLO

(c)大雨影响



YOLOv11n



ELD-YOLO

(d)低光照条件

图8　不同场景效果对比

在远距离小目标检测任务中，原始YOLOv11存在明显漏检，而ELD-YOLO通过多尺度特征增强成功识别远距车辆。在背景纹理复杂或光照干扰区域，原始YOLOv11易将非目标误识为车辆，而ELD-YOLO显著抑制了这些误检，说明共享卷积检测头在降低背景干扰方面具有优势。对于弱特征目标、雪天、雨天及夜间等特殊环境条件，YOLOv11的漏检情况较严重，而ELD-YOLO通过优化采样有效提升了检测精度，确保了弱特征和远距目标被正确识别。因此，ELD-YOLO具有小目标检测、误检抑制及复杂天气光照适应性特性。

5　结束语

本文提出ELD-YOLO检测算法，在背景干扰显著、小目标结构失真、多尺度等极端条件下，提高了多目标检测的检测精度，在KITTI和BDD100K数据集上，能够有效平衡了检测精度与模型规模。未来，将针对极端环境中易出现的漏检与误检问题进行优化，进一步强化模型在真实交通场景中的实时性与可靠性。

参　考　文　献

[1] 张哲宁, 刘祯, 王化强. 基于改进Transformer-无迹卡尔曼滤波器的智能车辆多模态3D目标检测方法[J]. 汽车技术, 2026(2): 39-47.
ZHANG Z N, LIU Z, WANG H Q. Multimodal 3D Object Detection Method for Intelligent Vehicles Based on Improved Transformer-Unscented Kalman Filter[J]. Automobile Technology, 2026(2): 39-47.

[2] 董文轩, 梁宏涛, 刘国柱, 等. 深度卷积应用于目标检测算法综述[J]. 计算机科学与探索, 2022, 16(5): 1025-1042.
DONG W X, LIANG H T, LIU G Z, et al. Review of Deep Convolution Applied to Object Detection Algorithms[J]. Journal of Frontiers of Computer Science and Technology, 2022, 16(5): 1025-1042.

[3] 何泽江, 蒋淑霞, 柳霞. 基于改进YOLOv7-tiny的轻量化道路目标检测算法[J]. 汽车技术, 2025(2): 9-16.
HE Z J, JIANG S X, LIU X. Lightweight Road Object Detection Algorithm Based on Improved Yolov7-tiny[J]. Automobile Technology, 2025(2): 9-16.

[4] 郭雨. 基于轻量化改进型YOLOv5的车辆检测方法[J]. 建模与仿真, 2023, 12(3): 1840-1849.
GUO Y. Vehicle Detection Method Based on Lightweight Improved YOLOv5[J]. Modeling and Simulation, 2023, 12(3): 1840-1849.

[5] LIU Y H, HUANG Z H, SONG Q, et al. PV-YOLO: A Lightweight Pedestrian and Vehicle Detection Model Based on Improved YOLOv8[J]. Digital Signal Processing, 2025, 156.

[6] 崔家礼, 张林, 郑瀚. 基于MDA-YOLOv10的融合图像车

辆行人检测方法[J/OL]. 红外技术, (2025-04-01) [2026-03-09]. https://link. cnki. net/urlid/53.1053. TN. 20250331.1749.004.
CUI J L, ZHANG L, ZHENG H. Vehicle and Pedestrian Detection Method Based on MDA-YOLOv10 Fused Image[J/OL]. Infrared Technology, (2025-04-01) [2026-03-09]. https://link.cnki.net/urlid/53.1053.TN.20250331.1749.004.

[7] 孙华, 王蒙, 曲海成. 改进YOLOv8的遥感图像车辆目标检测[J]. 遥感信息, 2025, 40(5): 26-34.
SUN H, WANG M, QU H C. Vehicle Target Detection in Remote Sensing Images Based on Improved YOLOv8[J]. Remote Sensing Information, 2025, 40(5): 26-34.

[8] 许德刚, 李宝森, 李滨, 等. 基于高频边缘特征增强的道路车辆检测算法[J/OL]. 计算机工程与应用, (2025-07-10) [2026-03-09]. https://link.cnki.net/urlid/11.2127.TP.202510.1342.003.
XU D G, LI B S, LI B, et al. Road Vehicle Detection Algorithm Based on High-Frequency Edge Feature Enhancement[J/OL]. Computer Engineering and Applications, (2025-07-10) [2026-03-09]. https://link.cnki.net/urlid/11.2127.TP.20250710.1342.003.

[9] 王子钰, 张建成, 刘元盛. 改进YOLOv8n的尘雾环境下目标检测算法[J]. 汽车技术, 2025(6): 1-7.
WANG Z Y, ZHANG J C, LIU Y S. Object Detection Algorithm in Dust and Fog Environment Based on Improved YOLOv8n[J]. Automobile Technology, 2025(6): 1-7.

[10] PANG Y W, SUN M L, JIANG X H, et al. Convolution in Convolution for Network in Network[J]. IEEE Transactions on Neural Networks and Learning Systems, 2017, 29(5): 1587-1597.

[11] GEIGER A, LENZ P, STILLER C, et al. Vision Meets Robotics: The KITTI Dataset[J]. International Journal of Robotics Research, 2013, 32(11): 1231-1237.

[12] YU F, CHEN H F, WANG X, et al. BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, Washington, USA: IEEE, 2020: 2636-2645.

[13] TIAN Y J, YE Q X, DOERMANN D S. YOLOv12: Attention-Centric Real-Time Object Detectors[EB/OL]. (2025-02-18) [2026-03-09]. https://arxiv. org/abs/2502.12524.

[14] WANG Y Y, WANG C, ZHANG H, et al. Automatic Ship Detection Based on Retinanet Using Multi-Resolution Gaofen-3 Imagery[J]. Remote Sensing, 2019, 11(5).

[15] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single Shot Multibox Detector[C]// European Conference on Computer Vision. Amsterdam, The Netherlands: Springer Nature, 2016: 21-37.

[16] REN S 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.

[17] ZHAGN T L, PANG H L, JIANG C H. GDM-YOLO: A Model for Steel Surface Defect Detection Based on YOLOv8s[J]. IEEE Access, 2024, 12: 148817-148825.

[18] LUO J D, LUO H B, SHI X. Intelligent Detection Method for Lightning Protection Points Based on YOLOv8[C]// 2024 7th International Conference on Computer Information Science and Application Technology (CISAT). Hangzhou, China: IEEE, 2024: 1106-1112.

[19] ZHANG Y J, GUO H C, ZHAO Y, et al. GSTD-DETR: A Detection Algorithm for Small Space Targets Based on RT-DETR[J]. Electronics, 2025, 14(12).

[20] ZHANG J N, LI X T, LI J, et al. Rethinking Mobile Block for Efficient Attention-based Models[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Paris, France: IEEE, 2023: 1389-1400.

[21] ZHANG Y, LI G X, GAO H, et al. Multi-Scale Interaction Network for Multimodal Entity and Relation Extraction[J]. Information Sciences, 2025, 699.

[22] FAN Z P, WU Y Y, LIU W, et al. LG-YOLOv8: A Lightweight Safety Helmet Detection Algorithm Combined with Feature Enhancement[J]. Applied Sciences, 2024, 14(22).

（责任编辑　瑞　秋）

修改稿收到日期为2026年3月9日。