DOI:10.19651/j.cnki.emt.2518055

改进 YOLO11 的多尺度上下文增强注意力车辆检测模型*

刘 炜 皮建勇 胡 倩 胡伟超
(贵州大学计算机科学与技术学院公共大数据国家重点实验室 贵阳 550025)

摘 要:为了提高高性能多尺度目标检测,特别是小目标检测的精度,以减少交通事故的发生概率。本研究提出了一种改进 YOLO11 模型的多尺度上下文增强注意力机制的汽车检测方法。首先,在主干网络中设计并引入了 RPCSPELAN5 结构替换 C3k2 模块,提升特征提取能力和信息聚合。其次,在颈部网络中创建并新增 DSM 模块,该模块通过动态上采样器和无参数注意力机制,增强小目标的特征融合。最后,进一步改进颈部网络,采用了基于 Haar 小波的下采样模块,提升语义分割表现和上下文连续性。在 VOC2012 和 COCO 数据集上的实验表明,所提出的算法在多个评估指标上均取得了显著的提升。VOC2012 数据集上的 P、R、mAP50 和 mAP50-95 分别提高了 0.2%、5.3%、3.4%和 4.2%,而 COCO 数据集上的提升幅度分别为 7.7%、6.0%、8.7%和 6.5%。本研究提出的算法在多尺度目标检测,特别是小目标检测精度上表现出优越性,有效提高了车辆检测精度,有助于降低交通事故发生的概率。

关键词:目标检测;YOLO11;多尺度;上下文增强;注意力机制
中图分类号:TP391.4;TN98 **文献标识码**:A **国家标准学科分类代码**:520.20

Multi-scale context-enhanced attention vehicle detection model based on improved YOLO11

Liu Wei Pi Jianyong Hu Qian Hu Weichao
(State Key Laboratory of Public Big Data, College of Computer Science and Technology, Guizhou University, Guiyang 550025, China)

Abstract: To improve high-performance multi-scale object detection, particularly the accuracy of small object detection, and reduce the probability of traffic accidents, this study proposes an enhanced YOLO11 model with a multi-scale context-enhanced attention mechanism for vehicle detection. Firstly, the RPCSPELAN5 structure is designed and introduced in the backbone network to replace the C3k2 module, enhancing feature extraction capability and information aggregation. Secondly, a DSM module is created and added to the neck network, which incorporates a dynamic upsampling mechanism and a simple, parameter-free attention mechanism to improve feature fusion for small objects. Finally, the neck network is further improved by adopting a Haar wavelet-based downsampling module, which enhances semantic segmentation performance and contextual continuity. Experiments on the VOC2012 and COCO datasets demonstrate significant improvements across multiple evaluation metrics. On the VOC2012 dataset, the improvements in P, R, mAP50, and mAP50-95 were 0.2%, 5.3%, 3.4% and 4.2%, respectively. On the COCO dataset, the improvements were 7.7%, 6.0%, 8.7% and 6.5%, respectively. The proposed algorithm exhibits superior performance in multi-scale object detection, particularly in small object detection accuracy, effectively enhancing vehicle detection precision and contributing to the reduction of traffic accidents.

Keywords: object detection; YOLO11; multi-scale; context enhancement; attention mechanism

0 引 言

随着交通流量的增加及道路环境日益复杂,交通事故的发生率逐步上升。因此,精确的车辆检测[1]在智能交通[2]、自动驾驶[3]、安防监控[4]等领域中具有至关重要的意义。检测精度不高可能导致严重的安全事故[5],从而造成

收稿日期:2025-02-11
* 基金项目:贵州省科技支撑计划(黔科合支撑[2023]一般 430)项目资助

不可挽回的生命和财产损失。由于车辆是交通系统中的主要参与者,因此,本研究的目标是提升算法在车辆目标检测中的表现。

车辆目标检测的发展经历了从传统的人工方法到基于深度学习的目标检测方法的转变。传统汽车检测算法依赖图像预处理(如灰度转换、高斯滤波去噪、Canny 边缘检测[6])、特征提取(SIFT[7]、HOG[8])和分类器设计(SVM[9]、AdaBoost[10])。尺度不变特征变换(SIFT)通过在球面上进行卷积实现尺度不变性,方向梯度直方图(HOG)适用于复杂背景和不同光照条件,用于提取车辆轮廓和纹理。支持向量机(SVM)通过最优超平面分类,自适应增强(AdaBoost)提升特征分类性能。但传统算法计算复杂度高,实时性差,光照或遮挡影响精度,泛化能力受限。

在基于深度与学习的车辆检测算法中,根据检测阶段的不同,目标检测算法一般可分为两类:两阶段(Two-stage)算法和单阶段(One-stage)算法[11]。两阶段目标检测算法包括 R-CNN[12](由于 R-CNN 需要对每个候选区域单独进行卷积操作,导致其计算量大、检测速度慢)、Fast R-CNN[13](Fast R-CNN 仍然依赖于选择性搜索算法来生成候选区域,这在一定程度上制约了其检测效率)、Faster-RCNN[14]等,而单阶段目标检测算法则包括 YOLO[15]系列(其中,YOLOv2[16]在处理小目标和复杂背景时,效果逊于两阶段算法;YOLOv3[17]的网络复杂,检测速度慢,难满足实时监测;YOLOv7[18]精度高于 YOLOv4[19],但模型仍较大)和 SSD[20]等。

近年来,针对小目标车辆检测[21]精度提升的问题,研究者们提出了一些基于 YOLO 系列算法的改进方案[22-26]。虽然目标检测的精度和速度都有了显著提高,但在部分复杂应用场景下存在细粒度信息丢失、小目标漏检、特征融合信息丢失和特征提取不完整等诸多问题,仍有很大的优化改进空间。为了解决这一问题,本研究基于 YOLO11[27]算法进行了若干改进,主要贡献如下:

1)设计并引入了 RPCSPELAN5 结构替代 C3k2 模块。该结构增强了主干网络的特征提取能力,减少了关键信息的丢失,从而提升了对小目标的检测精度。

2)创建并新增了 DSM 模块。该模块中的动态上采样器(dynamic upsampling,DySample)增强了颈部网络的上采样质量,使得模型能够更有效地利用特征图中的信息。同时,免参数注意力机制(simple,parameterless attention mechanism, SAM)帮助算法准确定位并识别小目标,同时减少噪声干扰,提升整体检测精度。

3)在颈部网络添加下采样模块(haar downsampling, HDWT)。该模块利用 Haar 小波进行下采样,有助于提升语义分割性能,减少特征图信息的丢失,并增强上下文的连续性。

1　基本原理

本节对 YOLO11 网络模型以及损失函数的相关基础知识进行一系列介绍。

YOLO11 是 YOLO 系列的最新成员,是 Ultralytics 团队[28]于 2024 年 9 月发布,该网络由 4 个部分组成:Input、Backbone、Neck 和 Head。如图 1 为其结构图。

Backbone 层负责提取图像特征,核心模块包括卷积、C3K2、SPPF 和 C2PSA。C3K2 通过分支处理与快捷连接融合特征,减少冗余,提升效率;SPPF[29]基于 SPP[30]结构,将并行池化改为串行池化,扩大感受野,增强多尺度目标检测能力;C2PSA 延续 CSPNet[31]分支架构,结合轻量化卷积与直接传递,通过特征融合整合浅层与深层信息。该设计高效提取关键特征,支撑后续检测任务。

Neck 层的主要作用是融合主干网络提取的不同尺度的特征信息。在 YOLO11 中,Neck 层对 YOLOv8 的 C2F 模块进行了优化,替换为 C3K2 模块。C3K2 模块的动态性源于 C2F 模块,其灵活性体现在代码参数设置上。当 c3k 参数为 FALSE 时,C3K2 模块表现为传统的 C2F 模块;而当其为 TRUE 时,Bottleneck 层将被替换为 C3K 模块。通过上采样和下采样操作,Neck 层有效融合了不同尺度的特征信息,从而捕获更丰富的语义信息,为目标检测提供支持。

Head 层主要用于目标的回归与预测,生成最终的目标检测结果。YOLO11 在 Head 部分的 cls 分支上使用深度可分离卷积(DWConv)[32]这样可以大幅度减少参数量和计算量。继续沿用 YOLOv8 的 Anchor-Free 将分类与检测头相分离,通过两条并行的分支结构,分别提取类别特征与位置特征,最后各用不同的卷积来完成分类与定位的任务,从而进一步提高网络的收敛速度与检测精度。

YOLO11 的边框回归损失由 CioULoss+DFLLoss 组成。CioULoss 的计算如下:

$$CIoU = 1 - IoU + \frac{p^2(b,b^{gt})}{c^2} + av \tag{1}$$

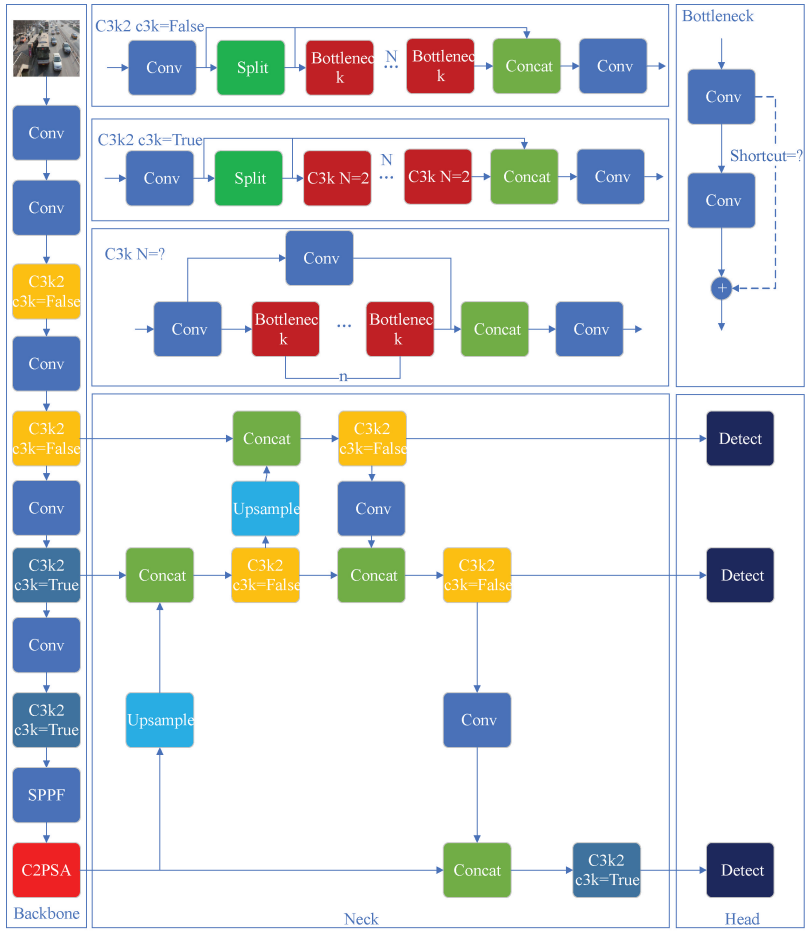
$$v = \frac{4}{\pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h}\right)^2 \tag{2}$$

$$\alpha = \frac{v}{(1-IoU)+v} \tag{3}$$

其中,IoU 为交并比,b,b^{gt} 分别为预测框的中心点坐标与真实框的中心点坐标,$p^2(b,b^{gt})$ 为 b 与 b^{gt} 两点之间的欧式距离,c 为预测框和真实框最小外接矩形的对角线距离,w 和 h 为预测框的宽和高,w^{gt} 和 h^{gt} 为真实框的宽和高。v 是惩罚项,α 是权重函数,DFL 损失通过交叉熵的形式来优化与标签最接近的左右两侧位置的概率,从而更准确地识别并解析目标位置附近区域的分布情况,其值的计算如下:

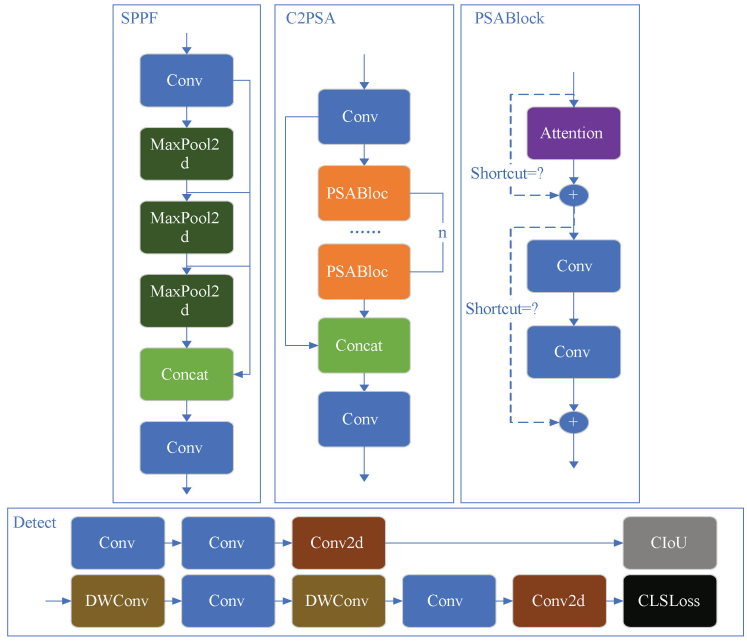
$$DFL(S_i,S_{i+1}) = -(y_{i+1} - y)\log(S_i) - (y - y_i)\log(S_{i+1}) \tag{4}$$

$$S_i = \frac{y_{i+1} - y}{y_{i+1} - y_i} \tag{5}$$



(a) YOLO11整体结构

(a) Overall structure of YOLO11



(b) YOLO11分支结构

(b) Branch structure of YOLO11

图 1　YOLO11 结构图

Fig. 1　YOLO11 architecture diagram

$$S_{i+1}=\frac{y-y_i}{y_{i+1}-y_i} \tag{6}$$

其中，y 表示标签值，y_i 和 y_{i+1} 为最接近 y 的两个数值，S_i 和 S_{i+1} 表示全局最小解。

2　改进 YOLO11 模型

为了优化 YOLO11 算法在多尺度场景下对车辆的检测精度，本研究提出一种基于改进 YOLO11 的目标检测算法，其网络结构如图 2 所示。首先，针对不同尺度的车辆对目标检测的影响，尤其是小尺度目标车辆，在 Backbone 层设计了 RPCSPELAN5 结构，该结构用于增强主干网络的特征提取能力，以此来减少重要信息的丢失，从而提高对小目标的检测精度。其次，为了使模型可以更好的利用特征图信息和增加对一些重要的目标区域的关注，同时减少复杂背景的干扰，在 Neck 层创建 DSM 模块。最后，为了提高语义分割性能减少特征图信息丢失，增强上下文连贯的作用，在颈部网络添加下采样模块 HDWT。

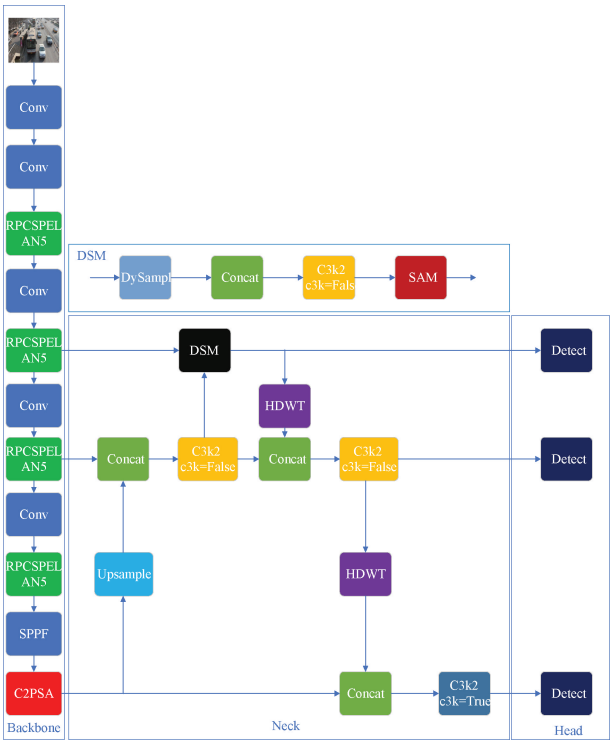


图 2　YOLO11 改进后结构图

Fig. 2　YOLO11 improved architecture diagram

2.1　RPCSPELAN5 结构

为了使模型能够更有效的解决多尺度目标图像难以检测或漏检的问题，本研究设计了 RPCSPELAN5 结构，主要包含 CBS 和 RPCSP[33] 结构，其中 CBS 表示 Conv2d＋BatchNorm2d＋SiLU（默认激活函数）而 RPCSP 结构如图 3 所示。

RPCSPELAN5 结构主要参考跨阶段局部网络（CSPNet）和高效层聚合网络（ELAN）[34] 设计而成，并参考了重参数

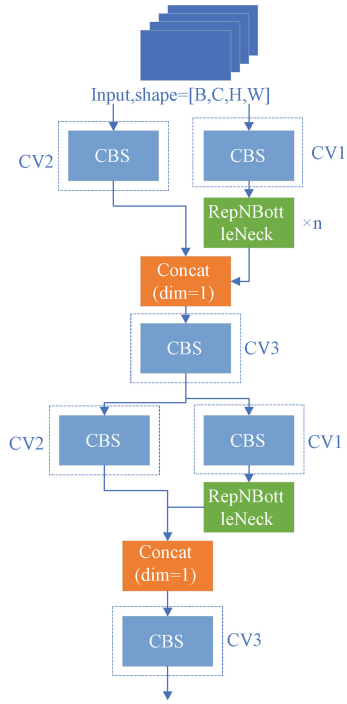


图 3　RPCSP 结构图

Fig. 3　RPCSP architecture diagram

(Re-parameter)方法，其结构如图 4 所示。设计 CSPNet 的主要目的是使该体系结构能够实现更丰富的梯度组合，同时减少计算量。通过将基础层的特征图划分为两个部分，然后通过提出的跨阶段层次结构将它们合并，可以实现此目标。具体操作是通过分割梯度流，使梯度流通过不同的网络路径传播。通过切换串联和过渡步骤，传播的梯度信息可以具有较大的相关性差异。此外，CSPNet 可以大大减少计算量，并提高推理速度和准确性。ELAN 网络的核心设计理念是通过有效地聚合不同层的特征信息，提高目标检测算法的准确性和鲁棒性。在深度学习中，不同层的特征图包含了不同尺度和语义级别的信息，而 ELAN 网络旨在充分利用这些信息，通过特定的聚合模块将不同层的特征图进行融合，从而实现多尺度特征的聚合。Re-parameter 的目的是使模型可微分(differentiable)，以便使用梯度下降等反向传播算法来训练模型，也就是将随机采样的过程转换为可导的运算，从而使得梯度下降算法可以正常工作即梯度可以在这个过程中传播。此方法可以减少特征信息的丢失和使生成模型更容易优化。

2.2　DSM 模块

1）DySample

考虑到动态卷积引入的繁重工作量，绕过基于核的范式，回归到上采样的本质，即点采样，重新考虑上采样过程。具体来说，假设输入特征通过双线性插值为连续的特征图，并生成内容感知的采样点来重新采样连续的特征图。从这个角度来看，首先提出一个简单的设计，其中点

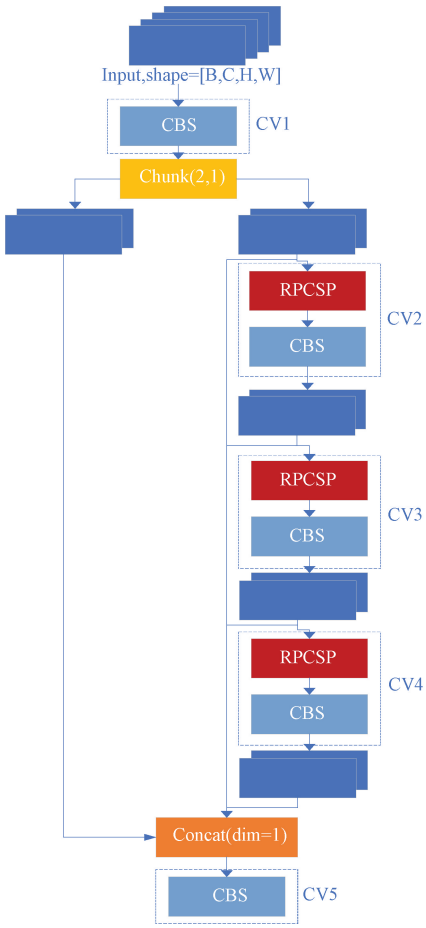


图 4　RPCSPELAN5 结构图

Fig. 4　RPCSPELAN5 architecture diagram

偏移通过线性投影生成,并用 PyTorch 中的 grid_sample 函数重新采样点值。然后,展示如何通过逐步调整来改进它,包括控制初始采样位置;调整偏移的移动范围;将上采样过程分成几个独立的组,并获得新的上采样器 DySample。与其他动态上采样器相比,DySample 不需要高分辨率引导特征作为输入,也不需要除 PyTorch 之外的任何额外 CUDA 软件包,同时不需要太大的内存占用,参数数量也比其他的上采样器少得多,其原理如图 5 所示。

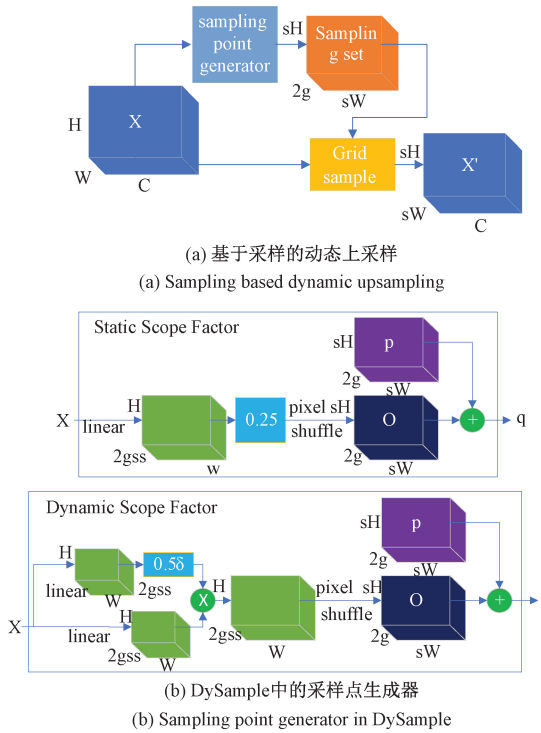
由图 5 知 DySample 中基于采样的动态上采样和模块设计。在 Static Scope Factor 模块中主要公式为:

$$O = 0.25linear(X) \tag{7}$$

在 Dynamic Scope Factor 模块中主要公式为:

$$O = 0.5\delta(linear_1(X)) \times linear_2(X) \tag{8}$$

其中,输入特征、上采样特征、生成的偏移和原始网格分别用 X、X'、O 和 p 表示。图 5(a)采样集由采样点生成器生成,输入特征通过 grid_sample 函数重新采样。图 5(b)在生成器中,采样集是生成的偏移和原始网格位置的和。上方框显示了带有静态范围因子的版本,其中偏移由线性层生成。下方描述了带有动态范围因子的版本,首先生成范围因子,然后用于调制偏移,δ 表示 sigmoid 函数。



(a) 基于采样的动态上采样

(a) Sampling based dynamic upsampling

(b) DySample中的采样点生成器

(b) Sampling point generator in DySample

图 5　DySample 原理图

Fig. 5　DySample schematic diagram

2)SAM

在颈部网络,本研究创建 DSM 模块,其结构在图 2 中可见。该模块在 DySample 的基础上新增了免参数注意力机制(SAM),SAM 与现有的基于通道和空间的注意力机制不同,此模块通过推断特征图中的三维注意力权重来工作,而无需向原始网络添加参数。具体而言,该注意力机制基于一些著名的神经科学理论,并提出了优化能量函数以找到每个神经元重要性的方法。还进一步推导了能量函数的快速闭合形式解,并展示了该解可以用不到十行的代码实现。该注意力机制的另一个优点是大多数运算符是基于定义的能量函数解的选择,避免了过多的结构调整的工作。

成功实现注意力需估算单个神经元的重要性。信息丰富的神经元通常表现出独特的放电模式,并可能抑制周围神经元活动(空间抑制)。具有明显空间抑制效应的神经元在视觉处理中应被赋予更高优先级。通过测量目标神经元与其他神经元的线性可分离性,可识别这些神经元。基于此,为每个神经元定义能量函数如下:

$$\bar{t} = w_t t + b_t \tag{9}$$

$$\bar{x}_i = w_t x_i + b_t \tag{10}$$

$$e_t(w_t, b_t, y, x_i) = (y_t - \bar{t})^2 + \frac{1}{M-1}\sum_{i=1}^{M-1}(y_0 - \bar{x}_i)^2 \tag{11}$$

其中,$\bar{t}$ 和 $\bar{x}_i$ 是 t 和 x_i 的线性变化,t 和 x_i 是输入特征 $X \times R^{C\times H\times W}$ 的单个通道中的目标神经元和其他神经元。

i 是空间维度上的索引，$M = H \times W$ 是该通道上神经元的数量，w_t 和 b_t 是权重和偏置变换，公式中所有值都是标量。经过使用二值标签、添加正则项和求偏导可得最小能量式子如下：

$$e_t^* = \frac{4(\hat{\sigma}^2 + \lambda)}{(t - \hat{u})^2 + 2\hat{\sigma}^2 + 2\lambda} \tag{12}$$

上述公式意味着：能量越低，神经元 t 与周围神经元的区别越大，也就是两个神经元越线性可分，重要性越高。因此，神经元的重要性可以通过式(13)计算得到：

$$\frac{1}{e_t^*} = \frac{(t - \hat{u})^2 + 2\hat{\sigma}^2 + 2\lambda}{4(\hat{\sigma}^2 + \lambda)} = \frac{(t - \hat{u})^2}{4(\hat{\sigma}^2 + \lambda)} + 0.5 \tag{13}$$

其中，$\hat{u}$ 和 $\hat{\sigma}^2$ 为该通道中所有神经元计算的均值和方差。

2.3 HDWT

下采样操作如最大池化或步幅卷积神经网络(CNNs)中被广泛应用，用于聚合局部特征、扩大感受野并减少计算负担。然而，对于语义分割任务，对局部领域的特征进行池化可能导致重要的空间信息丢失。为了解决这个问题，引入一种简单而有效的池化操作，称为基于 Haar 小波的下采样(HDWT)模块。该模块可以轻松集成到 CNNS 中，以提高语义分割模型的性能。HDWT 的核心思想是应用 Haar 小波变换来降低特征图的空间分辨率，同时尽可能保留更多信息。其与传统的下采样方法相比，有效地降低信息不确定性，并且显著提高不同模态图像数据集上各种 CNN 架构的分割性能。其原理如图 6 所示。

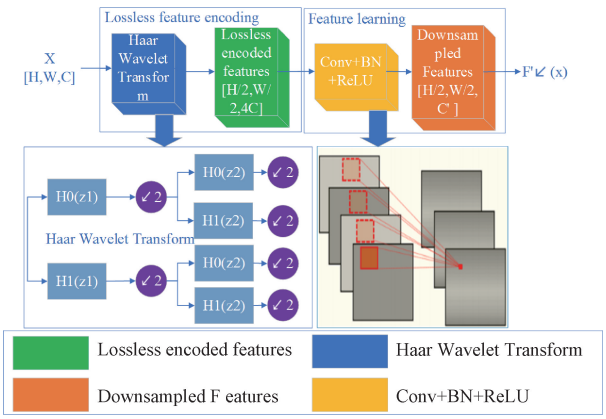


图 6　HDWT 原理图

Fig. 6　HDWT schematic diagram

3　实验设计与结果分析

3.1　实验环境

实验操作系统为 unbantu22.04，GPU 显卡使用 RTX3090、NVIDIA A10 和 NVIDIA V100，深度学习框架为 PyTorch2.3.0，运行 CUDA 版本为 CUDA=11.3，使用 YOLO11n 作为基础模型，具体实验参数如表 1 所示。

表 1　训练参数设置

Table 1　Training parameter settings

参数	值
epochs	100
batch	16
image_size	640
workers	8
optimizer	SGD
close_mosaic	0
lr0	0.01
weight_decay	0.000 5

3.2　数据集

本研究使用 VOC2012 和 COCO 公共数据集来进行训练和验证。

VOC2012 数据集全称为 PASCAL Visual Object Classes Challenge[35]，由 PASCAL (pattern analysis, statistical modelling and computational learning)组织发布，其包含人类、动物、车辆、家具等 20 个物体类别，总共有 11 530 张图像，这些图像的分辨率和场景各异，涵盖了多种不同的日常生活场景，每个图像可能包含多个物体实例，并且物体的标签和位置会被标注。由于本研究主要是做车辆目标检测，所以在本实验中只取自行车(bike)、公交车(bus)、汽车(car)和摩托车(motorbike)这几个物体类别作为最终的检测类别。数据集总共包含 5 113 张图片，其中自行车 898 张约占总数的 17.6%、公交车 671 张约占总数的 13.1%、汽车 2 673 张约占总数的 52.3%和摩托车 871 张约占总数的 17.0%。按 7∶2∶1的比例随机划分成训练集(3 579 张)验证集(1 022 张)和测试集(512 张)。

COCO(common objects in context)数据集[36]由微软研究院(Microsoft Research)于 2014 年推出，其包含多达 330 K 张图像，其中约 20 万张图像被标注为有目标的图像。数据集涵盖了 80 个类别的对象，且每个对象都有详细的标注信息。它不仅包含常见的物体(如人、车、动物等)，还包括一些更为复杂的物体(如书籍、建筑等)。数据集中的对象数量超过 250 万，并且每个图像通常含有多个标注对象。本研究主要研究的是车辆目标检测算法，所以从中提取了公交车(bus)、自行车(bicycle)、汽车(car)、摩托车(motorcycle)和卡车(truck)。该数据集总共有 20 629 张图片，其中公交车 1 898 张约占总数的 9.2%，自行车 3 289 张约占总数的 15.9%，汽车 8 476 张约占总数的 41.1%，摩托车 2 679 张约占总数的 13.0%，卡车 4 287 张约占总数的 20.8%。按 8∶1∶1的比例随机划分成训练集(16 503 张)验证集(2 063 张)和测试集(2 063 张)。

3.3　数据预处理

为了解决样本不均衡问题对性能所产生的影响以及防止过拟合和提高泛化性，本研究通过对图像数据进行几

何变换(对图片进行旋转、缩放、翻转和裁剪等)、颜色变换(调整图像的亮度、对比度、饱和度等)、噪声添加(向图像或音频数据中添加高斯噪声、椒盐噪声等)和混合图像(拼接多张图片等)等方法进行数据增强。

由于数据集中图片分辨率并不统一,在使用数据集进行训练或测试之前,对图片进行预处理,将图片通过裁剪、填充或缩放等操作调整为固定分辨率(640×640)以此来确定每张图片的分辨率。

3.4　评价指标

本实验选择 YOLO 系列算法中常用的性能评估指标:精度(precision,P)、召回率(recall,R)、所有类别的平均精度(mean average precision,mAP)、parameters(模型参数量)和 GFLOPS(每秒 10 亿次的浮点运算数)。P、R、AP 以及 mAP 计算公式如下:

$$P = \frac{TP}{TP + FP} \tag{14}$$

$$R = \frac{TP}{TP + FN} \tag{15}$$

$$AP = \int_0^1 P(R)\mathrm{d}R \tag{16}$$

$$mAP = \frac{1}{n}\sum_{i=1}^{n} AP \tag{17}$$

其中,TP 为正确的正例,FP 为错误的正例,FN 为错误的负例,精度 P 表示模型预测的正例样本中正确的正例所占比例;P 的值越大证明检测框与实际标注框吻合程度越高。召回率 R 表示在真实正例样本中模型预测正例所占的比例;R 值越大证明较多的正例被预测正确。AP 表示单个目标类别的平均精度,其值是 P-R 曲线所围图像的面积。mAP 是将数据集中所有类别的 AP 取平均值,用于全面评估算法在各个类别上的检测精度,其中 mAP50 是指当 IoU 的值为 0.5 时,所类别的平均精度。mAP50-95 表示 IoU 的值从 0.5 到 0.95,步长为 0.05,然后计算这些 IoU 下的平均 AP。如果 mAP50 值与 mAP50-95 值越大,则表明该算法的整体性能效果越好。

3.5　对比实验与结果分析

为了验证改进检测算法的性能,将其与当前几种主流算法进行对比,包括 YOLOv5n、YOLOv6n、YOLOv8n、YOLOv10n、YOLO11n 和基于 YOLO11 改进的算法等,在 VOC2012 和 COCO 数据集上进行对比实验,所有实验均在统一的实验环境下进行。实验结果如表 2 和 3 所示。

表 2　不同模型在 VOC2012 数据集上的对比实验

Table 2　Comparison experiments of different models on the VOC2012 dataset

模型	P/%	R/%	mAP50/%	mAP50-95/%	parameters/M	GFLOPS/G
YOLO11n(baseline)	87.7	75.7	86.0	65.3	2.6	6.6
YOLOv5n	86.7	75.8	85.3	62.9	2.2	5.8
YOLOv6n	87.9	76.1	85.6	65.0	4.2	11.5
YOLOv8n	87.7	77.1	86.2	65.1	3.2	8.9
YOLOv10n	87.8	71.2	83.3	63.5	2.7	8.2
YOLO11s	90.8	78.1	88.2	68.3	9.5	21.7
YOLO11n+Halo[37]	87.9	76.2	85.9	64.5	3.7	29.6
YOLO11n+ShapeIoU[38]	88.8	75.9	86.1	65.3	2.6	6.3
LeYOLO[39]	86.9	77.1	85.7	65.3	2.7	6.9
YOLO11n+文献[40]	87.6	76.5	86.0	64.5	2.2	6.1
YOLO11n-ours	**87.9(↑0.2)**	**81.0(↑5.3)**	**89.4(↑3.4)**	**69.5(↑4.2)**	**7.2**	**29.5**

表 3　不同模型在 COCO 数据集上的对比实验

Table 3　Comparison experiments of different models on the COCO dataset

模型	P/%	R/%	mAP50/%	mAP50-95/%	parameters/M	GFLOPS/G
YOLO11n(baseline)	61.1	51.6	54.3	36.7	2.6	6.6
YOLOv5n	59.1	49.0	51.2	34.2	2.2	5.8
YOLOv6n	61.8	50.4	53.5	36.4	4.2	11.5
YOLOv8n	60.6	51.8	54.2	36.4	3.2	8.9
YOLOv10n	62.1	51.1	54.6	37.0	2.7	8.2
YOLO11s	67.2	55.4	61.1	42.2	9.5	21.7
YOLO11n+Halo	60.8	49.7	53.3	36.3	3.7	29.6
YOLO11n+ShapeIoU	61.4	50.8	54.7	37.0	2.6	6.3
LeYOLO	65.5	51.4	55.7	37.4	2.7	6.9
YOLO11n+文献[40]	61.4	52.3	54.5	36.7	2.2	6.1
YOLO11n-ours	**68.8(↑7.7)**	**57.6(↑6.0)**	**63.0(↑8.7)**	**43.2(↑6.5)**	**7.2**	**29.5**

分析表 2 和表 3 数据可知，从 mAP50 方面看，在数据集 VOC2012 上，YOLO11n-ours 的 mAP50 达到 89.4%，相比 YOLOv5n(85.3%)和 YOLOv6n(85.6%)有显著优势；在数据集 COCO 上，YOLO11n-ours 的 mAP50(63.0%)远高于 YOLOv5n(51.2%)和 YOLOv6n(53.5%)，这表明，YOLO11n-ours 在检测精度上优于现有模型，特别是在复杂背景和多目标场景下的表现更加突出。

在 mAP50-95 方面，在数据集 VOC2012 上，YOLO11n-ours 达到了 69.5%，相比 YOLOv10n(63.5%)和 YOLOv8n(65.1%)有显著提升；在数据集 COCO 上，YOLO11n-ours 的 mAP50-95 为 43.2%，远超 YOLOv10n(37.0%)和 YOLOv8n(36.4%)。这一结果显示，YOLO11n-ours 在不同尺度的目标检测中，尤其是小目标检测方面具有更好的表现。

本研究主要是想解决小目标车辆检测精度低或者漏检的问题，从召回率 R 上可以看出，提出的改进算法在两个数据集上的 R 远高于其他模型，分别达到了 81.0%和 57.6%，高召回率表明，YOLO11n-ours 在检测小目标和难检测目标时具有显著优势，有效减少了漏检现象。

与基础模型 YOLO11n 相比，在数据集 VOC2012 上，其 P、R、mAP50 和 mAP50-95 分别提升了 0.2%、5.3%、3.4%和 4.2%；在数据集 COCO 上，虽然参数量有所增加但在 P、R、mAP50 和 mAP50-95 上分别提升了 7.7%、6.0%、8.7%和 6.5%，这些结果表明，提出的改进算法不仅保持了高准确率，同时显著提高了车辆检测精度，尤其在小目标检测方面效果更为突出。

综合分析可以得出本研究提出算法在精度和召回率上全面超越现有模型；尽管性能提升，但并未显著增加计算负担，适用于实际应用场景；显著提高了小目标和难检测目标的检测精度，解决了小目标漏检率高的问题；同时本研究设计的改进算法在复杂场景中展现出更强的适应性，为车辆目标检测提供了一个性能优异且实用性强的解决方案。

3.6　消融实验与结果分析

为了进一步直观地了解提出的各模块对 YOLO11 网络结构的增益效果，在 VOC2012 和 COCO 数据集上进行了多组的消融实验。消融实验的结果如表 4 和 5 所示。

表 4　不同模块在 VOC2012 数据集上的消融实验

Table 4　Ablation experiments of different modules on the VOC2012 dataset

模型	P/%	R/%	mAP50/%	mAP50-95/%	parameters/M	GFLOPS/G
YOLO11n(baseline)	61.1	51.6	54.3	36.7	2.6	6.6
YOLO11n+RPCSPELAN5	87.3	81.5	88.7(↑**2.7**)	69.1(↑**3.8**)	7.2	29.1
YOLO11n+DSM	86.0	79.8	87.2(↑**1.2**)	66.2(↑**0.9**)	2.5	6.3
YOLO11n+HDWT	86.9	77.6	86.9(↑**0.9**)	65.4(↑**0.1**)	2.4	6.3
YOLO11n+RPCSPELAN5+DSM	90.6	79.0	88.7(↑**2.7**)	69.1(↑**3.8**)	8.4	52.4
YOLO11n+RPCSPELAN5+HDWT	89.4	79.6	88.2(↑**2.2**)	68.2(↑**2.9**)	7.1	29.7
YOLO11n+DSM+HDWT	88.7	76.1	86.5(↑**0.5**)	65.4(↑**0.1**)	2.4	6.3
YOLO11n-ours	**87.9**(↑**0.2**)	**81.0**(↑**5.3**)	**89.4**(↑**3.4**)	**69.5**(↑**4.2**)	**7.2**	**29.5**

表 5　不同模块在 COCO 数据集上的消融实验

Table 5　Ablation experiments of different modules on the COCO dataset

模型	P/%	R/%	mAP50/%	mAP50-95/%	parameters/M	GFLOPS/G
YOLO11n(baseline)	61.1	51.6	54.3	36.7	2.6	6.6
YOLO11n+RPCSPELAN5	67.2	57.7	62.0(↑**7.7**)	42.9(↑**6.2**)	7.2	29.1
YOLO11n+DSM	60.5	52.3	55.1(↑**0.8**)	37.1(↑**0.4**)	2.5	6.3
YOLO11n+HDWT	61.0	50.7	54.5(↑**0.2**)	37.0(↑**0.3**)	2.4	6.3
YOLO11n+RPCSPELAN5+DSM	68.2	57.1	62.4(↑**8.1**)	43.2(↑**6.5**)	8.4	52.4
YOLO11n+RPCSPELAN5+HDWT	67.0	57.8	62.9(↑**8.6**)	43.0(↑**6.3**)	7.1	29.7
YOLO11n+DSM+HDWT	59.9	52.3	54.7(↑**0.4**)	36.8(↑**0.1**)	2.4	6.3
YOLO11n-ours	**68.8**(↑**7.7**)	**57.6**(↑**6.0**)	**63.0**(↑**8.7**)	**43.2**(↑**6.5**)	**7.2**	**29.5**

从表 4、5 可知,分别将 RPCSPELAN5、DSM 和 HDWT 嵌入原 YOLO11 网络模型中验证本研究提出的网络模型效果。首先,在 YOLO11 网络模型中引入 RPCSPELAN5 后,在数据集 VOC2012 和 COCO 上,其 R、mAP50 和 mAP50-95 分别提升了 5.8%、2.7%、3.8%和 6.1%、7.7%、6.2%,这些结果表明,RPCSPELAN5 结构能够有效减少特征信息的丢失,显著提高了网络在目标检测任务中的整体性能。将 DSM 模块加入 YOLO11 的 Neck 部分后,其 R、mAP50 和 mAP50-95 在数据集 VOC2012 和 COCO 上分别提升了 4.1%、1.2%、0.9%和 0.7%、0.8%、0.4%,这说明 DSM 模块能够有效地将小目标的特征信息融合到其他特征层中,提升了小目标检测的精度,尤其在提高召回率方面表现突出。

其次,将 RPCSPELAN5+DSM 组合网络模型嵌入 YOLO11 网络模型之后,在数据集 VOC2012 和 COCO 上,其 P、R、mAP50 和 mAP50-95 分别提升了 2.9%、3.3%、2.7%、3.8%和 7.1%、5.5%、8.1%、6.5%,说明了 RPCSPELAN5+DSM 组合模型在尽可能保留特征信息的同时,进一步加深了多尺度特征图的融合,从而提高了网络的综合性能。将 RPCSPELAN5+HDWT 组合网络模型引入 YOLO11 网络模型之后,其 P、R、mAP50 和 mAP50-95 在数据集 VOC2012 和 COCO 上分别提升了 1.7%、3.9%、2.2%、2.9%和 5.9%、6.2%、8.6%、6.3%,这一组合能够显著提高多尺度目标的检测能力,尤其在复杂环境下的目标检测表现更为出色。

最后,将 DSM+HDWT 组合模块加入 YOLO11 网络模型之后,在数据集 VOC2012 和 COCO 上,其 R、mAP50 和 mAP50-95 分别提升了 0.4%、0.5%、0.1%和 0.7%、0.4%、0.1%,尽管 DSM+HDWT 组合模块的提升幅度相对较小,但它依然有效提高了车辆检测的召回率和精度,尤其对精度和召回率的平衡做出了贡献。

综上所述,本研究的改进方法在各个评价指标上都表现出了明显的提升,表明所提改进算法能够有效增强 YOLO11n 在车辆检测任务中的鲁棒性和准确性,而且在实际应用中具备更强的适应性和实用价值。

3.7　模型泛化性分析

为了体现本模型的泛化性,在训练时除了使用 VOC2012 数据集还使用了 COCO 数据集,这两个数据集都是公共数据集而且 COCO 数据集的图片数量也远多与 VOC2012 数据集,这使得模型接触到了更广泛的数据分布同时确保了数据集的多样性,此外本研究在每个数据集都设置了独立的测试集,它与训练集分开,且代表了模型在实际中可能遇到遇到的情况。通过对实验结果的分析也确实证明了本模型有很好的泛化性。

3.8　检测结果可视化

图 7 展示了在相同实验环境下,YOLO11n 原始模型与 YOLOv6n、YOLOv10n 以及提出的改进模型在多尺度、密集、高遮挡[41]环境中车辆的检测效果。其中,每个数据集中从上到下分别展示了场景一、场景二和场景三的 4 种模型的车辆检测效果,从左至右分别是:YOLO11n 模型、YOLOv6n 模型、YOLOv10n 模型以及本研究改进的模型。从图 7 中可以观察到,在处理密集和多尺度目标的检测场景任务中,改进的模型能够更准确地识别出更多目标,特别是小尺寸目标。

在场景一中,由于远处车辆的尺寸较小,原始 YOLO11 模型未能有效检测到这些小目标,导致许多远处的小目标车辆漏检,甚至一些较大的遮挡目标也未能被检测到。然而,改进后的模型成功地检测出了更多的小目标车辆,并且在遮挡条件下,大目标车辆也被成功识别。在场景二中,场景中存在车辆密集、目标尺寸变化大以及遮挡等复杂情况。原始的 YOLO11n 模型、YOLOv6n 模型和 YOLOv10n 模型均漏检了部分遮挡目标和图像远端的

YOLO11n　YOLOv6n　YOLOv10n　改进的YOLO11n

(a) VOC2012数据集

(a) VOC2012 dataset

YOLO11n　YOLOv6n　YOLOv10n　改进的YOLO11n

(b) COCO数据集

(b) COCO dataset

图 7　检测结果可视化图

Fig. 7　Detection results visualization

小目标车辆。然而，改进后的模型成功检测出了所有目标车辆，解决了遮挡和尺度变化带来的检测难题。在场景三中，车辆之间存在较强的遮挡，原始 YOLO11n 模型漏检了许多被遮挡的车辆。相比之下，改进后的模型能够识别出更多的被遮挡目标，并且对小目标车辆的检测精度有显著提升。

综上所述，提出的改进模型显著提升了 YOLO11n 对多尺度车辆的检测能力，并有效解决了原始模型在小目标检测精度低和遮挡情况下的漏检问题。在不同场景中，改进后的模型展现出更强的鲁棒性和精度，能够适应更为复杂的环境，检测出更多目标，尤其是在小目标和遮挡情况下的表现更加优异。这表明，改进方案不仅优化了检测精度，也增强了模型在实际应用中的适用性和稳定性。

4　结　　论

针对高性能多尺度目标的检测，尤其是小目标检测精度低的问题。本研究提出了一种基于 YOLO11 改进模型，首先通过 RPCSPELAN5 结构提升主干网络的特征提取能力使其能够将更多关键信息提取出来，也使不同特征层的信息能够充分聚合，有效解决目标图像存在不同尺度问题带来的干扰。其次，将 DSM 模块加入到特征金字塔网络之中，使小目标的特征信息能够更多地融合到其他特征层中，提升小目标的检测精度。最后，采用基于 Haar 小波的下采样(HDWT)模块，用于降低特征图的空间分辨率，同时尽可能保留更多信息，同时增强上下文的连续性。

该模型在 VOC2012 数据集上的 P、R、mAP50 和 mAP50-95 分别提高了 0.2%、5.3%、3.4% 和 4.2%，而 COCO 数据集上的提升幅度分别为 7.7%、6.0%、8.7% 和 6.5%。这说明提出的算法在多尺度目标检测，特别是小目标检测精度上表现出优越性，有效提高了车辆检测精度，有助于降低交通事故发生的概率。

然而，本模型也存在一定的局限性，如该模型主要是针对正常天气情况下的车辆检测，没有对大雾、雨雪等恶劣天气条件下进行实时性的测试，未来的研究可以尝试提升模型在大雾、雨雪等低能见度环境下的鲁棒性，确保在各种复杂天气条件下也能保持高精度的检测能力；开发适合边缘设备运行的轻量级版本，优化模型的计算效率和功耗，使其能够在低计算资源的环境下高效执行，满足实时检测需求。

参考文献

[1]　郑少武，李巍华，胡坚耀. 基于激光点云与图像信息融合的交通环境车辆检测[J]. 仪器仪表学报，2019，40(12)：143-151.

ZHENG SH W，LI W H，HU J Y. Vehicle detection in the traffic environment based on the fusion of laser point cloud and image information[J]. Chinese Journal of Scientific Instrument，2019，40(12)：143-151.

[2]　梁天添，杨淞淇，钱振明. 基于改进 YOLOv8s 的恶劣天气车辆行人检测方法[J]. 电子测量技术，2024，47(9)：112-119.

LIANG T T，YANG S Q，QIAN ZH M. Improved YOLOv8s method for vehicle and pedestrian detection in adverse weather [J]. Electronic Measurement Technology，2024，47(9)：112-119.

[3]　BOUKERCHE A，SHA M ZH. Design guidelines on deep learning-based pedestrian detection methods for supporting autonomous vehicles[J]. ACM Computing Surveys(CSUR)，2021，54(6)：1-36.

[4] KIM S,KWAK S,KO B C. Fast pedestrian detection in surveillance videobased on softtarget training of shallow random forest[J]. IEEE Access, 2019, 7: 12415-12426.

[5] 陈虹,郭露露,宫洵,等.智能时代的汽车控制[J].自动化学报,2020,46(7):1313-1332.
CHEN H, GUO L L, GONG X, et al. Automotive control in intelligent era[J]. Acta Automatica Sinica, 2020,46(7):1313-1332.

[6] CANNY J. A computational approach to edgedetection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,1986(6): 679-698.

[7] CRUZ-MOTA J, BOGDANOVA I, PAQUIER B, et al. Scale invariant feature transformon the sphere: Theory and applications[J]. International Journal of Computervision, 2012,98:217-241.

[8] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection [C]. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE,2005: 886-893.

[9] OPTIMIZATION S M. A fast algorithm for training support vector machines [J]. CiteSeerX, 1998, 10(43):4376.

[10] FREUND Y, SCHAPIRE R E. A decision-theoretic generalization of on-line learning and an application to boosting [J]. Journal of Computer and System Sciences, 1997,55(1):119-139.

[11] 王永生,姬嗣愚.基于深度学习的目标检测算法综述[J].算机与数字工程,2023,51(6):1231-1237.
WANG Y SH, JI S Y. Review of target detection algorithms based on deep learning[J]. Computer & Digital Engineering, 2023,51(6): 1231-1237.

[12] 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.

[13] GIRSHICK R. Fast R-CNN [C]. International Conference on Computer Vision,2015:1440-1448.

[14] 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 & Machine Intelligence, 2017, 39(6): 1137-1149.

[15] REDMON J,DIVVALA S,GIRSHICK R,et al. You only look once: Unified, real-time objectdetection[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.

[16] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017:7263-7271.

[17] REDMON J, FARHADI A. YOLOv3: Anincremental improvement [J]. ArXiv preprint arXiv: 1804.02767,2018.

[18] 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.

[19] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. Optimal speed and accuracy of object detection[J]. ArXiv preprint arXiv: 2004.10934, 2020.

[20] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]. In Computer Vision ECCV2016: 14th European Conference, Amsterdam, The Netherlands, October11-14, 2016, Proceedings, PartI14 2016. Springer International Publishing,2014: 21-37.

[21] 张蕊,高诗博,赵霞,等.基于改进 YOLOv5s 的无人驾驶夜间车辆目标检测算法[J].电子测量技术,2023,46(17):87-93.
ZHANG R,GAO SH B,ZHAO X,et al. Algorithm on nighttime target detection for unmanned vehicles based on an improved YOLOv5s[J]. Electronic Measurement Technology, 2023, 46(17):87-93.

[22] 宋绍剑,夏海姐,李刚,等. YOLOv5 的改进算法及其在自动驾驶多目标检测的应用研究[J].计算机工程与应用,2023,59(15):68-75.
SONG SH J, XIA H J, LI G, et al. Research on improved YOLOv5 algorithm and its application in multi-object detection for automatic driving [J]. Computer Engineering and Applications, 2023, 59(15): 68-75.

[23] 刘辉,刘鑫满,刘大东.面向复杂道路目标检测的 YOLOv5 算法优化研究[J].计算机工程与应用,2023,59(18):207-217.
LIU H,LIU X M,LIU D D. Research on optimization of YOLOv5 detection algorithm for object in complex road[J]. Computer Engineering and Applications, 2023,59(18): 207-217.

[24] 魏陈浩,杨睿,刘振丙,等.具有双层路由注意力的 YOLOv8 道路场景目标检测方法[J].图学学报,2023,44(6):1104-1111.
WEI CH H, YANG R, LIU ZH B, et al. YOLOv8 with bi-level routing attention for road scene object detection[J]. Journal of Graphics, 2023, 44(6):

1104-1111.

[25] 熊恩杰,张荣芬,刘宇红,等.面向交通标志的 Ghost-YOLOv8 检测算法[J].计算机工程与应用,2023,59(20):200-207.
XIONG EN J,ZHANG R F,LIU Y H,et al. Ghost-YOLOv8 detection algorithm for traffic signs[J]. Computer Engineering and Applications, 2023, 59(20):200-207.

[26] 张利丰,田莹.改进 YOLOv8 的多尺度轻量型车辆目标检测算法[J].计算机工程与应用,2024,60(3):129-137.
ZHANG L F,TIAN Y. Improved YOLOv8 multi-scale and lightweight vehicle object detection algorithm[J]. Computer Engineering and Applications, 2024,60(3):129-137.

[27] JOCHER G, CHAURASIA A, QIU J. YOLO by ultralytics[EB/OL]. (2024-09-30)[2025-02-11]. https://github.com/ultralytics/ultralytics.

[28] TERVEN J, CORDOVA-ESPARZA D M. A comprehensive review of YOLO: From YOLOvl to YOLOv8 and beyond[J]. ArXiv preprint arXiv: 2304.00501, 2023.

[29] NELSON J, SOLAWETZ J. YOLOv5 is here: State-of-the-art object detection at 140FPS[EB/OL]. (2020-06-10)[2025-02-11]. https://blog.roboflow.com/yolov5-is-here/.

[30] HE K M, ZHANG X Y, REN SH Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2015, 37(9): 1904-1916.

[31] WANG C Y,LIAO H Y M,WU Y H,et al. CSPNet: A new backbone that can enhance learning capability of CNN[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2020:1571-1580.

[32] CHOLLET F. Xception: Deep learning with depthwise separable convolutions[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2017: 1800-1807.

[33] WANG C Y, YEH I H, LIAO H Y M. YOLOv9: Learning what you want to learn using programmable gradient information[J]. ArXiv preprint arXiv: 2402.13616,2024.

[34] WANG C Y, YEH I H, LIAO H Y M. Designing network design strategies through gradient path analysis[J]. ArXiv preprint arXiv: 2211.04800, 2022.

[35] EVERINGHAM M, GOOL L V, WILLIAMS C K I, et al. The pascal visual object classes(VOC) challenge[J]. International Journal of Computer Vision, 2010, 88(2):303-338.

[36] LIN T Y, MAIRE M, BELONGIE S,et al. Microsoft COCO: Common objects in context[C]. European Conference on Computer Vision. Springer International Publishing, 2014:740-755.

[37] VASWANI A, RAMACHANDRAN P, SRINIVAS A, et al. Scaling local self-attention for parameter efficient visual backbones[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), 2021: 12889-12899.

[38] ZHANG H, ZHANG SH J. Shape-IoU: More accurate metric considering bounding box shape and scale[J]. ArXiv preprint arXiv:2312.17663,2023.

[39] HOLLARD L L,MOHIMONT L,GAVEAU N,et al. LeYOLO, new scalable and efficient CNN architecture for object detection[J]. ArXiv preprint arXiv:2406.14239,2024.

[40] MISRA D,NALAMADA T,ARASANIPALAI A U, et al. Rotate to attend: Convolutional triplet attention module[C]. 2021 IEEE Winter Conference on Applications of Computer Vision(WACV), 2021:3138-3147.

[41] 林哲,潘慧琳,陈丹.融合改进 YOLO 和语义分割的遮挡目标抓取方法[J].电子测量与仪器学报,2024,38(12):190-201.
LIN ZH,PAN H L,CHEN D. Grasp method for occlusion method by fusing improved YOLO with semantic segmentation[J]. Journal of Electronic Measurement and Instrumention, 2024, 38(12): 190-201.

作者简介

刘炜,硕士研究生,主要研究方向为计算机视觉、目标检测。

E-mail:363046951@qq.com

皮建勇(通信作者),副教授,博士,主要研究方向为类脑智能、分布式计算。

E-mail:pijianyong@139.com