



DOI:10.12404/j.issn.1671-1815.2402441

引用格式:龙小羽,南新元.智能驾驶场景下的中小型障碍物检测方法[J].科学技术与工程,2025,25(9):3778-3787.

Long Xiaoyu, Nan Xinyuan. Small and medium-sized obstacle detection methods in intelligent driving scenarios[J]. Science Technology and Engineering, 2025, 25(9): 3778-3787.

# 智能驾驶场景下的中小型障碍物检测方法

龙小羽,南新元\*

(新疆大学电气工程学院,乌鲁木齐 830017)

**摘要** 针对智能驾驶场景下路面中小型障碍物易发生漏检、小目标障碍物难检测、模型参数量大等问题,提出了改进YOLOv8n的障碍物目标检测算法。在主干网络中融入分布移位卷积(distribution shifting convolution, DSCConv),将浮点运算替换为整数运算,减少了冗余计算量,通过量化和分布移位的方式模仿原始卷积层,维持了准确率;通过添加小目标检测层,更好地捕捉小目标的特征信息,适配小目标的尺度特征;结合SimAM无参数注意力机制,引入SPPF-SimAM模块,提高特征表示的质量与多样性,在不增加参数量,的情况下实现了检测精度的提升;通过组合鬼影混洗卷积(ghost-shuffle convolution, GSCConv)和VoV-GSCSP模块的方式轻量化颈部特征融合网络,降低了模型的参数量和计算量。实验结果表明,改进后模型的准确率、召回率、平均精度均值相较于原始模型分别提升了1.6%、8.0%、6.2%,参数量降低了6.7%,所提算法有效提升了智能驾驶场景下中小型障碍物的检测精度,并且在检测性能与模型轻量化之间达到较好的平衡。

**关键词** 障碍物检测;YOLOv8n;智能驾驶;小目标检测;注意力机制

中图分类号 TP391; 文献标志码 A

## Small and Medium-sized Obstacle Detection Methods in Intelligent Driving Scenarios

LONG Xiao-yu, NAN Xin-yuan\*

(School of Electrical Engineering, Xinjiang University, Urumqi 830017, China)

**[Abstract]** Aiming at the problems such as small and medium-sized obstacles on the road are prone to miss detection, small target obstacles are difficult to detect, and the number of model parameters is large in smart driving scenarios, the obstacle target detection algorithm with improved YOLOv8n was proposed. Distribution shifting convolution (DSCConv) was used in the backbone network to replace floating point operation with integer operation, reducing the amount of redundant computation, and maintaining the accuracy by imitating the original convolution layer by quantization and distribution shifting. By adding small target detection layer, the feature information of small target can be captured better and the scale characteristics of small target can be adapted. Combined with SimAM parameterless attention mechanism, SPPF-SimAM module was introduced to improve the quality and diversity of feature representation, and the detection accuracy was improved without increasing the number of parameters. By combining ghost-shuffle convolution (GSCConv) and VoV-GSCSP modules, the neck feature fusion network was lightweight, reducing the number of parameters and calculation of the model. The experimental results show that the accuracy, recall, and mean average precision of the improved model are improved by 1.6%, 8.0%, and 6.2%, respectively. The number of parameters is reduced by 6.7% compared with the original model, and the proposed algorithm effectively improves the detection accuracy of small and medium-sized obstacles in smart driving scenarios, and achieves a better balance between the detection performance and the model lightweighting.

**[Keywords]** obstacle detection; YOLOv8n; intelligent driving; small target detection; attention mechanism

近年来,随着数字技术的不断发展与成熟,智能驾驶技术在汽车工业中备受关注,人们对智能驾驶的接受程度也在不断提升,具有智能驾驶功能的汽车在市场上的需求在快速增长。其中路面障碍物的检测是实现更加安全可靠自动驾驶的关键步

骤之一。特别是对于中小型障碍物的准确识别与检测,对于提升自动驾驶汽车的安全性和稳定性具有至关重要的意义。

目前智能驾驶系统中的目标检测任务主要是检测出道路上的车辆、行人、交通标志等信息来保

收稿日期:2024-04-06 修订日期:2024-12-06

基金项目:国家自然科学基金(62303394);新疆维吾尔自治区自然科学基金(2022D01C694)

第一作者:龙小羽(2000—),女,汉族,新疆和静人,硕士研究生。研究方向:深度学习、目标检测。E-mail:874480489@qq.com。

\*通信作者:南新元(1967—),男,汉族,新疆乌鲁木齐人,硕士,教授。研究方向:智能控制系统开发、机器视觉。E-mail:xynan@xju.edu.cn。

证自动驾驶车辆行驶的安全性<sup>[1]</sup>。随着科学技术的不断发展,乘客对自动驾驶车辆的可靠性和舒适性要求大大提升。在车辆行驶过程中,可能会出现一些异常障碍物,例如路锥、废弃轮胎、石墩、前方掉落的货物等,这些障碍物日常情况下较少见且体积相对较小,车辆需要及时发现问题并进行避让,这种“罕见又突发”的情况,在智能驾驶行业被称为“边角情况”(corner case, CC)<sup>[2]</sup>,在处理这些情况时,智能驾驶系统必须具备高度的感知能力和快速的响应性,以确保车辆和乘客的安全。

随着深度学习的不断发展,越来越多的学者开始关注智能驾驶领域的异常障碍物检测问题,引发了大量学者的深入研究。文献[3]提出了一种用于小型障碍物检测的 MergeNet 深度网络架构,通过多阶段的训练过程,每个阶段学习互补的特征,融合这些特征来预测场景的分割图,能够在公共数据集上对障碍物进行高保真度分割,但该方法的场景泛化性不足,且运行速度较慢。文献[4]采用深度学习方法,以检测到的道路消失点为导向,设计了一种基于全局特征的道路消失点检测模型,以确定图像中障碍物的位置区域,并在该区域上进行障碍物检测。文献[5]提出了一种端到端的实时障碍物检测,创造了 RODSNet 联合学习来自立体 RGB 图像的语义分割和视差图,利用两种高效的单任务网络架构和多任务学习方案中的模块来识别道路上异常的小型障碍物。文献[6]在 SSD 模型基础上,提出一种感受野增强和特征融合的小目标检测模型,兼顾检测效率的同时,有效提高了对道路上小目标的检测精度。文献[7]针对高速公路存在抛洒物常出现漏检和误检问题,以 YOLOv5s 为主体,提出了一种基于特征提纯和上下文信息增强的 CIEFRNet 抛洒物检测算法,提升了检测效果。

现有的路面异常障碍物检测已经取得了一定的进展,但是其准确性、响应速度、模型轻量化程度及泛化性能均有待提高,且部分研究依靠激光雷达等设备,成本较为高昂。为了能够更容易部署到移动设备以贴合实际,需要在更低成本的计算资源条件下实现更高的检测精度和响应性。针对上述存在的问题,现以 YOLOv8n 为基准模型,在此基础上构建一种中小型障碍物检测算法。融合 DSConv 卷积<sup>[8]</sup>优化主干网络结构,将浮点运算替换为整数运算,减少模型计算量的同时通过量化和分布移位维持检测性能。智能驾驶场景中包含大量的小目标障碍物,添加特征图为  $160 \times 160$  的小目标检测层提升小目标的检测精度。为了提高目标特征表示的质量,对 SPPF 模块进行优化,将无参数注意力机制 Si-

mAM<sup>[9]</sup>与 SPPF 模块相结合构成 SPPF-SimAM 模块,通过特征加权、空间关系建模和全局信息整合等方式,更高效地提取到目标的关键信息。由于添加的小目标检测层造成模型复杂程度增加,结合 Slim-neck 模块<sup>[10]</sup>将颈部网络的卷积替换为轻量级卷积 GSConv,部分 C2f 模块替换为 VoV-GSCSP 模块,降低模型参数量和计算量。以期在检测性能与模型轻量化之间达到较好的平衡,同时为智能驾驶过程中的安全问题提供新的解决思路。

## 1 YOLOv8 网络模型概述

YOLOv8 是一种单阶段目标检测算法,具有广泛的应用领域,还可用于目标跟踪、姿态估计、图像分割等任务。YOLOv8 的架构主要包含主干(Backbone)特征提取网络、颈部(Neck)特征融合网络和检测头(Head)这 3 个部分。Backbone 负责网络的特征提取功能,采用了一系列卷积和 C2f 模块。Neck 部分采用 PANet 结构,利用上采样和通道融合,增强特征的代表能力,最终将 PANet 的 3 个输出分支输入 Head 中。Head 部分负责产生最后的检测结果,采用目前较为主流的解耦头结构,包括一个检测头和一个分类头,有助于加快模型的收敛速度。

YOLOv8 包括 5 个模型:YOLOv8-n、YOLOv8-s、YOLOv8-m、YOLOv8-l、YOLOv8-x,模型的参数量和计算量随着模型深度和宽度的增大而增加。YOLOv8n 是目前模型最小、速度最快且易于在低成本、低算力设备上部署的版本,适用于智能驾驶场景中的障碍物检测任务。为了满足智能驾驶场景中对障碍物检测实时性的需求,并在保持检测准确率的同时降低成本和算力消耗,故选用 YOLOv8n 作为实验基础模型。

## 2 改进 YOLOv8n 的中小型障碍物检测算法

基于 YOLOv8n 模型进行改进,着眼于实现低延迟、高精度、轻量化的网络模型,旨在提高模型对小目标的感知能力和检测准确率,同时减小模型的体积和计算量,以适应实际应用的需求。在 YOLOv8n 的基础上,通过在 Backbone 中引入 DSConv 替换原有的传统卷积,减少冗余计算量的同时提升检测精度;增加小目标检测层提升了对小目标障碍物的检测精度;引入一种 SPPF-SimAM 模块,通过无参数注意力机制更高效地提取关键特征;采用 GSConv 和 VoV-GSCSP 模块轻量化 Neck 部分,总体实现网络模型的轻量化和精确度的提升。改进后的 YOLOv8n 网络结构如图 1 所示。



第  $j$  块卷积核的 softmax 值;  $i$  为卷积层数,  $j$  为当前层的第  $j$  块卷积核;  $\hat{\xi}$  为偏置, 初始值为 0。通过梯度下降最小化损失得到  $\xi$ ,  $\xi$  为张量中每个参数的值, 表达式为

$$\xi = \min_{\hat{\xi}} \sum_j T_j \ln \left( \frac{T_j}{I_j} \right), \forall (1, \text{BLK}, 1, 1) \text{ slice} \quad (3)$$

DSConv 相较于传统卷积的优势在于计算速度快, 占用内存更少。因此, 在主干特征提取网络中引入 DSConv, 能够在维持精度的同时有效降低模型的数量和计算量。

### 2.2 添加小目标检测层

数据集中存在着大量的小目标障碍物, 这些小目标障碍物与常见的车辆、行人相比, 所占尺寸相对较小, 形状和外观也更加多样化。小目标通常具有更小的像素区域和更低的分辨率, 细节与边界信息不明显, 且目标的特征更加微弱、模糊。输入图像尺寸对小目标障碍物的识别也产生了干扰, 原 YOLOv8n 的检测层会输出  $20 \times 20$ 、 $40 \times 40$ 、 $80 \times 80$  三种不同尺寸的特征图, 对原始图像进行缩放会使小目标的识别更加困难。

基于此, 对 YOLOv8n 网络进行优化, 添加一个专门处理小目标的检测层, 引入更多的上下文信息和特征表示, 从而提高模型对小目标检测的鲁棒性以及识别的准确性。当输入为尺寸  $640 \times 640$  的特征图时, 在原来的网络中增加一个  $160 \times 160$  特征图的检测层, 从主干网络的第六层提取特征信息, 与颈部网络上采样后的信息进行特征融合, 能够对驾驶视角中距离较远的障碍物进行更加准确的检测, 一定程度上降低了各种尺度上对小目标障碍物的漏检与误检, 但模型的复杂程度也因此而增大。

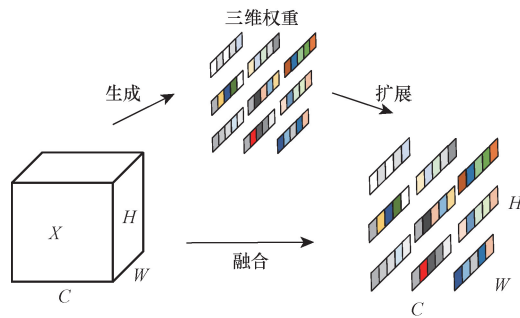
### 2.3 SPPF-SimAM 模块

面向驾驶场景的障碍物检测通常具有复杂的环境背景, 这需要模型提升对目标障碍物关键特征的提取能力。如图 3 所示, SimAM 注意力机制作为一种具有全三维权重的无参数注意力机制, 可以通过测量一个目标神经元和其他神经元之间的线性可分离性来估计各个神经元的重要性实现其功能, 最小能量函数  $e_i^*$  为

$$e_i^* = \frac{4(\hat{\sigma}^2 + \lambda)}{(t - \hat{\mu})^2 + 2\hat{\sigma}^2 + 2\lambda} \quad (4)$$

$$\hat{\mu} = \frac{1}{M} \sum_{i=1}^M x_i \quad (5)$$

$$\hat{\sigma}^2 = \frac{1}{M} \sum_{i=1}^M (x_i - \hat{\mu})^2 \quad (6)$$



$X$  为特征图;  $C, H, W$  分别为特征图的通道数、高、宽

图 3 SimAM 模块结构

Fig. 3 SimAM module structure

式中:  $\lambda$  为超参数, 设置为 0.000 1;  $t$  为输入特征的值;  $\hat{\mu}$  和  $\hat{\sigma}^2$  分别为通道中单个神经元的平均值和方差;  $x$  为其余神经元;  $i$  为空间上的维度;  $M$  为通道中神经元数量。由式(4)可知,  $e_i^*$  越低神经元的区别越大, 即重要性越高。整个 SimAM 模块通过 Sigmoid 优化, 可以表示为

$$\hat{X} = \text{Sigmoid} \left( \frac{1}{E} \right) \odot X \quad (7)$$

式(7)中:  $E$  为神经元的最小能量函数集合;  $X$  为输入特征;  $\odot$  表示矩阵元素相乘。Sigmoid 可以在不影响神经元之间相对重要性的前提下抑制  $E$  的值不能过大。

SPPF 通过 3 次连续的池化操作, 结合不同尺度的输出进行融合, 将 SimAM 引入 SPPF 模块中, 能够更好的整合不同池化层的特征信息, 提高关键特征信息的表征能力, 进而提升特征融合的效果, 保证特征提取的信息更加丰富, 改进的 SPPF-SimAM 结构如图 4 所示。

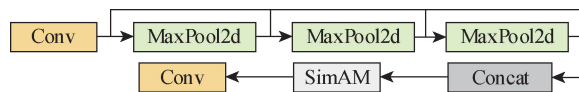


图 4 SPPF-SimAM 模块结构

Fig. 4 SPPF-SimAM module structure

### 2.4 轻量化颈部特征融合网络

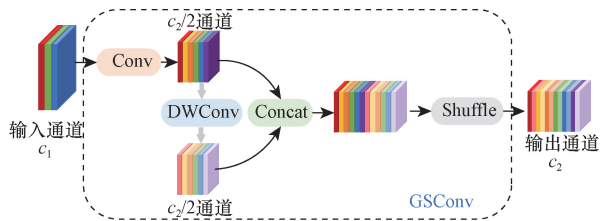
在模型中添加小目标检测层时增加了额外的网络层数, 这导致模型的计算量增加, 从而降低整体的性能和速度。为了在维持精度的同时减少模型的计算资源, 利用 Slim-Neck 方法中的轻量级卷积 GSConv 和 VoV-GSCSP 模块实现颈部特征融合网络的轻量化, 可以有效缓解高昂的计算成本。

#### 2.4.1 GSConv 卷积

GSConv 是一种新型的轻量化卷积模块, 由标准卷积 (standard convolution, SC)、深度可分离卷积 (depth-wise separable convolution, DSC) 和 Shuffle 组成, 使用 Shuffle 将 SC 生成的信息渗透到 DSC 生成

的信息的每个部分中。Shuffle 是一种统一的混合策略,该方法通过在不同通道上统一交换局部特征信息,可以将来自 SC 的信息完全混合到 DSC 的输出中,而无需任何附加功能,使得卷积计算的输出尽可能接近 SC 的输出,降低了计算成本,计算量约为 SC 的 50%。其模块结构如图 5 所示。

GSConv 更适合于在颈部网络使用,因为在主干网络中使用会导致模型的网络层数更深,会加剧对数据流的阻力,增加推理时间,特征图传送到颈部网络时,其通道尺寸已达到最大,宽高维度最小,不再需要变换。而在颈部网络使用时冗余信息更少,不需要压缩。



$c_1$ 、 $c_2$  分别为输入和输出通道数

图 5 GSConv 模块结构

Fig. 5 GSConv module structure

#### 2.4.2 跨阶段部分网络模块 VoV-GSCSP

以 GSConv 模块为基础引入 GS bottleneck 结构,使用一次性聚合方法设计跨阶段部分网络模块 VoV-GSCSP,能够有效降低网络结构的复杂性,模型结构如图 6 所示。

通过组合 GSConv 和 VoV-GSCSP 的方式能够轻量化颈部特征融合网络,有效降低模型的参数量和计算量,从而使模型在不牺牲检测精度的条件下,加快模型的推理速度。

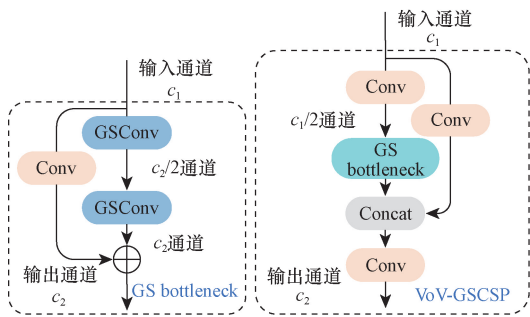


图 6 GS bottleneck 和 VoV-GSCSP 模块结构

Fig. 6 GS bottleneck and VoV-GSCSP module structure

## 3 实验与结果分析

### 3.1 数据集的建立

#### 3.1.1 数据来源

智能驾驶过程中出现的中小型障碍物通常存在日常情况下“罕见又突发”的特点,在单一数据集

上训练容易导致模型出现过拟合等问题,为了提高模型的泛化能力,应对不同的驾驶场景,需要制作场景丰富的数据集。

因此,结合公开数据集、网络资源、实地采集等方式建立障碍物联合数据集。首先,从 Lost And Found<sup>[11]</sup>、Small Obstacle<sup>[12]</sup>、Street Obstacle Sequences<sup>[13]</sup>、CODA<sup>[14]</sup>这 4 个公开数据集中筛选出 2 564 张图像,Lost And Found 包含 13 个驾驶场景,是一个专注于小型道路危险检测的数据集;Small Obstacle 数据集的研究对象主要为施工活动、货物掉落等场景出现的小障碍物;Street Obstacle Sequences 包含 20 个真实场景的视频序列,涵盖 13 个不同的异常障碍物类型;CODA 数据集具有丰富的驾驶场景,各个场景至少包含一个对智能驾驶车辆或周围环境有危险性或较为新奇的物体。为了在不同场景下获得更丰富可靠的数据样本,通过网络引擎搜索、真实道路场景自行采集的方式获得 1 437 张图像。为了保证自行采集数据的可用性,每个图像中需要包含至少一个障碍物类别,且在多个时段、多种角度和距离进行采集。通过上述方式累计获得共 4 001 张图像,联合多种场景的数据集包含了多样的光照情况及天气条件,不仅仅局限于单一的数据场景,有助于提升模型的泛化能力。

由于驾驶视角的特殊性,仅采取镜像翻转扩充的方式数据增强至 8 002 张构成障碍物联合数据集。按照 8:1:1 的比例划分为训练集、验证集和测试集供本实验使用。

#### 3.1.2 数据集标注

选用 14 类边角情况下的异常中小型障碍物构建数据集,包括:路锥、石墩、废弃轮胎、小动物、球、箱式货物、水瓶、水马、施工标志牌、防撞桶、路桩、石块、塑料袋、三角警示牌。对联合数据集进行标签及图片格式的统一,使用 LabelImg 进行标注,以障碍物名称作为标签类别,共计 19 878 个障碍物标签,各集合标签类别及其数量如表 1 所示。

### 3.2 实验环境

实验均在 Windows11 操作系统中完成,选择 Pytorch2.0.1 深度学习框架,CUDA 版本为 11.7,编程语言为 Python3.8。CPU 型号为 Inter(R) Core(TM) i7-12700H,运行内存 16 GB, GPU 采用 NVIDIA GeForce RTX 4060 Laptop,8 G 显存。

为消除参数变化的影响,在模型训练过程中均设置相同的实验参数,初始学习率为 0.01, Batchsize 设置为 8, Workers 设置为 4, Epoch 设置为 100。

### 3.3 评价指标

采用准确率(precision)、召回率(recall)、平均精

表1 数据集标签统计  
Table 1 Statistics on dataset labels

障碍物类别	标签数量			
	训练集	验证集	测试集	合计
路锥	2 311	279	350	2 940
石墩	1 788	186	182	2 156
废弃轮胎	835	82	95	1 012
小动物	1 160	129	151	1 440
球类	489	57	54	600
箱式货物	2 173	277	292	2 742
水瓶	692	100	130	922
水马	1 157	112	105	1 374
施工标志牌	579	58	81	718
防撞桶	983	136	117	1 236
路桩	1 548	148	228	1 924
石块	1 483	169	160	1 812
塑料袋	565	48	63	676
三角警示牌	270	30	26	326
合计	16 033	1 811	2 034	19 878

度均值(mean average precision, mAP)、参数量(parameters)、浮点运算量(giga floating-point operations per second, GFLOPs)、每秒帧数(frames per second, FPS)指标对模型性能进行评价。

准确率、召回率、平均精度均值用于衡量模型的检测准确程度,参数量用于衡量模型的复杂程度和规模大小,浮点运算量用于评估模型的计算资源消耗,模型的检测速度由每秒检测帧数来表示。

### 3.4 实验结果分析

#### 3.4.1 不同卷积对主干网络的影响

智能驾驶场景下的障碍物检测需要满足高精度、实时性、低计算量的要求,这就需要在满足模型轻量化的同时维持检测精度。为了验证在主干网络引入 DSConv 卷积的有效性,与常见的卷积模块 ODConv<sup>[15]</sup>、DWConv<sup>[16]</sup>、GhostConv<sup>[17]</sup> 进行对比实验,实验结果如表 2 所示。

从表 2 可以看出,4 种卷积模块都能够降低基准模型的 GFLOPs,一定程度上能够减少模型的计算资源。在检测速度方面,仅 DWConv、DSConv 有所提升。但是除了 DSConv 外,其他 3 种卷积方式的 mAP 都呈现小幅度的降低,未达到轻量化的同时维持检测精度的需求,其中 ODConv 降低了 0.9%,

表2 引入不同卷积的对比结果  
Table 2 Comparison result of introducing different convolutions

卷积	mAP/%	Parameters/10 <sup>6</sup>	GFLOPs	FPS/(帧·s <sup>-1</sup> )
Conv	84.6	3.0	8.1	60
ODConv	83.7	3.0	7.1	60
DWConv	84.5	2.6	7.2	67
GhostConv	84.3	2.8	7.7	52
DSConv	84.8	3.0	7.1	73

DWConv 降低了 0.1%, GhostConv 降低了 0.3%, 而 DSConv 提升了 0.2%。

由此可知,DSConv 模块更具有优越性,能够取代原模型中传统的卷积模块,在降低模型计算量的同时提升了检测精度,实现了更高的检测速度。

#### 3.4.2 不同注意力机制与 SPPF 融合对比实验

为提升模型对障碍物关键特征的提取能力,优化 SPPF 结构能够更有效地进行多尺度融合,在 SPPF 模块的相同位置引入不同注意力机制,包括 Triplet Attention<sup>[18]</sup>、Shuffle Attention<sup>[19]</sup>、LSKA<sup>[20]</sup>、CoT Attention<sup>[21]</sup> 进行对比试验,结果如表 3 所示。可以看出,在 LSKA、CoT 增加了一定的参数量和计算量,但检测精度的提升幅度较低,在同等参数量与计算量的情况下,采用的 SimAM 模块在 mAP 指标表现较突出,在不增加计算负担的前提下有效提升了检测性能。

表3 SPPF 与不同注意力机制结合的对比结果  
Table 3 Comparison result of SPPF combined with different attentional mechanisms

算法	mAP/%	Parameters/ 10 <sup>6</sup>	GFLOPs	FPS/ (帧·s <sup>-1</sup> )
SPPF	84.6	3.0	8.1	60
SPPF-Triplet Attention	85.1	3.0	8.1	64
SPPF-Shuffle Attention	84.6	3.0	8.1	67
SPPF-LSKA	84.6	3.3	8.3	67
SPPF-CoT Attention	84.9	5.3	9.9	55
SPPF-SimAM	85.3	3.0	8.1	66

#### 3.4.3 障碍物检测结果

为验证所提出的中小型障碍物检测算法相较于原模型而言对不同类别障碍物识别精度的优越性,将模型改进前后的障碍物检测结果进行比较,对比结果如表 4 所示。

原 YOLOv8n 模型对石墩、水瓶、路桩、石块的 mAP 值较低,均位于 80% 以下,分析其原因可知这四类障碍物的共同特点是数据集中包含大量的小目标,并且路桩会出现极端长宽比的情况,同时石墩与石块的形态颜色相似,易发生混淆。本文模型对石墩、水瓶、路桩、石块四类障碍物的 mAP 分别提升了 9.0%、9.9%、13.1%、18.5%,说明本文所提模型对小目标的检测能力有所提高,且对其余障碍物类别的 mAP 均在 80% 以上,改进前后模型的检测效果如图 7 所示,可以看出改进后的检测算法对于小目标障碍物具备更高的特征提取能力,对于较远的目标也可以检测到,与原始模型相比,误检漏检的情况有明显改善。

表 4 不同障碍物类别平均精度均值对比结果  
Table 4 Comparison result of mAP of different obstacle classes

障碍物类别	准确率/%		召回率/%		mAP/%	
	YOLOv8n	本文算法	YOLOv8n	本文算法	YOLOv8n	本文算法
路锥	88.4	92.5	74.0	77.4	81.0	87.0
石墩	82.5	92.8	70.9	78.2	79.1	88.1
废弃轮胎	88.2	95.0	88.4	99.2	94.8	98.2
小动物	95.7	79.8	90.7	84.6	94.7	84.3
球类	97.1	98.1	72.2	92.6	91.8	97.7
箱式货物	97.8	98.2	92.4	95.4	96.9	98.6
水瓶	79.9	77.8	63.1	75.4	74.0	83.9
水马	81.1	81.4	75.2	74.3	80.5	85.9
施工标志牌	90.6	88.7	77.8	82.7	82.7	88.1
防撞桶	86.8	91.0	72.8	77.9	82.9	89.1
路桩	83.1	89.2	55.3	69.3	71.8	84.9
石块	88.1	93.0	70.0	94.4	79.2	97.7
塑料袋	88.8	91.7	77.8	81.0	85.3	90.4
三角警示牌	95.6	79.8	84.0	84.6	89.0	84.3
平均值	88.8	90.4	76.0	84.0	84.6	90.8

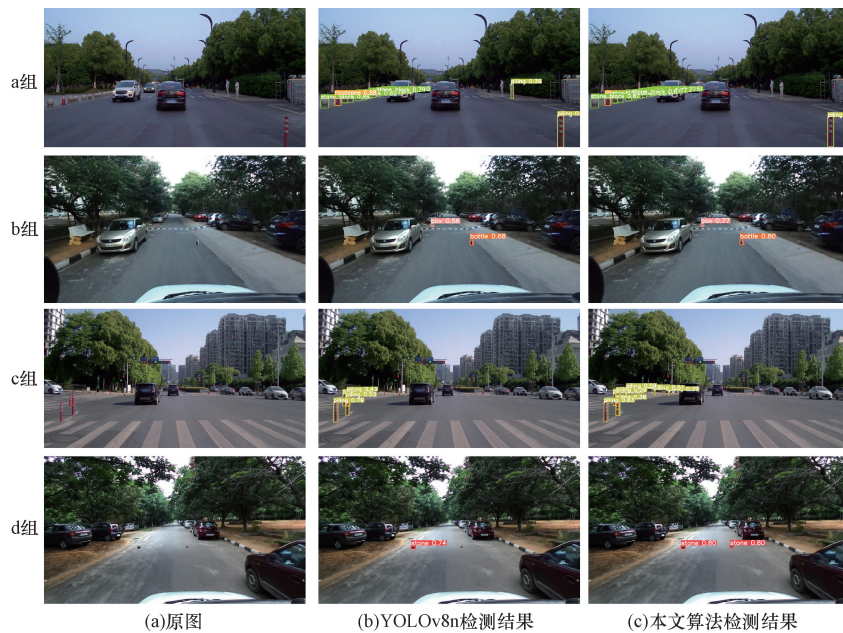


图 7 检测结果对比图

Fig. 7 Comparison plot of the detection results

为了更直观地体现本文模型相对于原模型的优化效果,将原模型与本文模型训练过程中的 mAP 值变化幅度可视化,如图 8 所示,可以看出从 80 轮开始模型 mAP 近趋于平稳,且改进后模型的平均精度均值收敛速度更快,精度更高。

3.4.4 消融实验

为了验证各个模块对 YOLOv8n 模型改进的有效性,以 YOLOv8n 为基准模型,通过不同的模块组合方式设计了编号 A ~ H 共 8 组消融实验,实验结果如表 5 所示。

由表 5 中结果可知,提出的 4 种改进方式均对基准模型产生了优化效果。如实验 B 所示,将主干

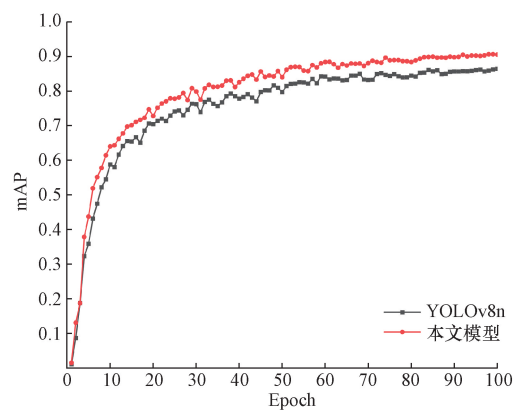


图 8 改进前后平均精度均值曲线图

Fig. 8 Mean average precision curve before and after improvement

网络中的标准卷积替换为 DSCConv 卷积后,在维持检测精度与速度的同时降低了模型的计算量,且 FPS 每秒提升 13 帧,说明 DSCConv 通过量化和分布移位维持了和原始卷积相似的输出;实验 C 在原模型中加入小目标检测层后,检测精度提升了 5.5%,说明小目标检测层能够关注到数据集中更多的特征信息,检测精度提升的幅度较大,但由于网络模型的层数的扩增,增加了模型的计算量;实验 D 引入 SPPF-SimAM 模块,通过 SimAM 注意力机制在不增加参数量的前提下提升了检测精度;实验 E 通过轻量化颈部网络,将颈部网络中的全部卷积与部分 C2f 模块替换为 GSCConv 与 VoV-GSCSP 模块,在 mAP、参数量、GFLOPs、FPS 四个评价指标上均有优化效果,其参数量和 GFLOPs 均有所下降,在模型轻量化的同时提升了 0.1% 的 mAP,表明轻量化颈部网络能够在模型性能与轻量化之间达到较好的平衡。

由实验 F 可知,在主干网络中替换 DSCConv 的同时加入小目标检测层,模型的 mAP 和 FPS 得到大幅优化,相比原模型来看,mAP 增大了 6.1%,FPS 每秒提升 8 帧,同时由于 DSCConv 计算量更小,减少了仅加入小目标检测层时的计算量,维持了检测性能;实验 G 在 F 的基础上引入 SPPF-SimAM 模块,充分发挥注意力机制对关键信息提取的作用,提升了模型的检测精度。实验 H 为本文模型,在 G 的基础上引入轻量化模块,mAP 相较于原模型提升 6.2%,参数量降低了 6.7%,但是 GFLOPs 稍有提升。消融实验结果表明,改进后的模型能够更有效地检测小目标障碍物,改善了误检和漏检情况,并且基本满足了轻量化模型和实时性要求。

### 3.4.5 模型对比实验

为了进一步评估改进的 YOLOv8n 在障碍物检测研究中的性能,考虑到智能驾驶领域对目标检测精度和实时性的双重需求,仅将改进后的模型与部分

主流的单阶段目标检测算法:SSD<sup>[22]</sup>、CenterNet<sup>[23]</sup>、YOLOv3<sup>[24]</sup>、YOLOv5、YOLOv7<sup>[25]</sup>、YOLOX<sup>[26]</sup>、YOLOv9<sup>[27]</sup> 进行比较,对比实验结果如表 6 所示。

相对于表 6 中其他主流模型,本文模型的参数量和计算量是最小的,同时检测精度高于其他检测算法。本文模型的检测精度和速度相较于 SSD 和 CenterNet 而言有较大的提升,mAP 分别提升了 31.4% 和 35.0%,FPS 也有较大幅度提升。与 YOLO 系列其他版本的模型相比,本文模型的 mAP 值有较为显著的提升,其中提升幅度最大的是相对于 YOLOX-tiny 提高了 27.5%,但 FPS 值略低;YOLOv8s 与本文模型的 mAP 值最接近,但其模型参数量和计算量较大,未满足轻量化需求,而本文模型的参数量和计算量仅仅相当于 YOLOv8s 的 25.2% 和 36.8%;对于 YOLOv9s 模型,虽然与本文模型的 mAP 值较为接近,但计算复杂度相对较高,导致检测速度变慢。因此本文模型更易部署到低成本、低算力的移动设备,兼顾了检测性能和模型的轻量化。

表 6 主流模型对比实验结果

Table 6 Results of the mainstream model comparison experiments

算法	mAP/%	Parameters/ 10 <sup>6</sup>	GFLOPs	FPS/ (帧·s <sup>-1</sup> )
SSD	59.4	25.3	62.1	15
CenterNet	55.8	32.7	109.7	33
YOLOv3-tiny	66.6	8.7	12.9	63
YOLOv5s	74.2	7.1	16.0	56
YOLOv7	80.2	36.5	103.4	46
YOLOv7-tiny	70.6	6.0	13.1	64
YOLOX-s	68.9	8.9	26.8	69
YOLOX-tiny	63.3	5.0	15.3	70
YOLOv8s	89.5	11.1	28.5	55
YOLOv9s	89.3	9.6	38.8	19
本文算法	90.8	2.8	10.5	53

表 5 消融实验结果

Table 5 Results of ablation experiments

编号	DSCConv	小目标 检测层	SPPF-SimAM	GSCConv + VoV- GSCSP	准确率/%	召回率/%	mAP/%	Parameters/10 <sup>6</sup>	GFLOPs	FPS/(帧·s <sup>-1</sup> )
A	×	×	×	×	88.8	76.0	84.6	3.0	8.1	60
B	√	×	×	×	89.3	75.0	84.8	3.0	7.1	73
C	×	√	×	×	90.2	84.7	90.1	3.0	12.5	68
D	×	×	√	×	91.6	75.1	85.3	3.0	8.1	66
E	×	×	×	√	88.9	74.3	84.7	2.8	7.5	62
F	√	√	×	×	91.7	82.8	90.7	3.0	11.5	68
G	√	√	√	×	92.7	84.2	90.9	3.0	11.5	51
H	√	√	√	√	90.4	84.0	90.8	2.8	10.5	53

注:“√”表示引入该方法;“×”表示未引入该方法。

### 3.4.6 泛化性能实验

为了验证改进后模型的泛化能力,选择在目标检测数据集 PASCAL VOC 2012<sup>[28]</sup>和自动驾驶领域数据集 KITTI<sup>[29]</sup>上进行实验与分析。本次实验中使用的 PASCAL VOC 2012 共包含 17 125 张图片,其中训练集和测试集分别包含 13 870 张和 3 255 张;KITTI 共包含 7 481 张图片,训练集和测试集分别包含 6 058 张和 1 423 张。实验结果如表 7、表 8 所示。

由表 7、表 8 中结果可知,改进后模型的 mAP 在 PASCAL VOC2012 和 KITTI 上分别相较于原始模型提升了 0.6% 和 1.3%,速度上仍满足实时性需求。在自动驾驶领域的 KITTI 数据集上涨幅相对较高,由此可说明,本文模型具有较好的泛化能力,验证了本文模型的有效性。

表 7 改进前后模型在 PASCAL VOC 2012 的对比实验结果  
Table 7 Comparison result of the model before and after improvement on PASCAL VOC 2012

算法	mAP/%	Parameters/ 10 <sup>6</sup>	GFLOPs	FPS/ (帧·s <sup>-1</sup> )
YOLOv8n	56.4	3.0	8.1	88
本文算法	57.0	2.8	10.5	61

表 8 改进前后模型在 KITTI 的对比实验结果  
Table 8 Comparison result of the model before and after improvement on KITTI

算法	mAP/%	Parameters/ 10 <sup>6</sup>	GFLOPs	FPS/ (帧·s <sup>-1</sup> )
YOLOv8n	76.9	3.0	8.1	75
本文算法	78.2	2.8	10.5	53

## 4 结论

为了提高智能驾驶汽车对路面中小型障碍物的检测精度,提出一种改进 YOLOv8n 的目标检测算法。在 Backbone 网络中融入 DSCConv 减少了冗余计算量,针对数据集中存在大量的小目标障碍物这一问题,添加 160×160 特征图的小目标检测层加强对小目标的特征提取能力,提出一种 SPPF-SimAM 模块更加关注目标的重要特征,最后采用 GSCConv 与 VoV-GSCSP 的组合模块轻量化 Neck 网络。实验结果表明,本文模型的 mAP 相较于原 YOLOv8n 模型提升了 6.2%,参数量降低了 6.7%,在检测性能与模型轻量化之间达到较好的平衡。面对大量的小目标障碍物,本文模型能够实现更好的检测效果,且具有较好的泛化能力。

对比其他主流模型,所提模型的综合性能更强,但是在检测速度上仍有提升的空间。由于实际场景中的障碍物目标与驾驶路况更为复杂多变,未

来研究将优先从提升数据集质量出发,扩充障碍物类别数量,解决样本不均衡问题,提升模型应用的泛化能力与应用价值。

### 参考文献

- [1] 茅智慧,朱佳利,吴鑫,等.基于YOLO的自动驾驶目标检测研究综述[J].计算机工程与应用,2022,58(15):68-77.  
Mao Zhihui, Zhu Jiali, Wu Xin, et al. Review of YOLO based target detection for autonomous driving[J]. Computer Engineering and Applications, 2022, 58(15): 68-77.
- [2] Breitenstein J, Termöhlen J A, Lipinski D, et al. Corner cases for visual perception in automated driving: some guidance on detection approaches[J]. ArXiv Preprint, 2021: 2102.05897.
- [3] Gupta K, Javed S A, Gandhi V, et al. Mergenet: a deep net architecture for small obstacle discovery [J]. ArXiv Preprint, 2018: 1803.06508.
- [4] 俞骏威,张黎明,陈凯,等.基于道路消失点的远距离路面微小障碍物检测[J].同济大学学报(自然科学版),2019,47(S1):213-216.  
Yu Junwei, Zhang Liming, Chen Kai, et al. Long-distance small road obstacles detection based on road vanishing point[J]. Journal of Tongji University(Natural Science), 2019, 47(S1): 213-216.
- [5] Song T J, Jeong J, Kim J H. End-to-end real-time obstacle detection network for safe self-driving via multi-task learning[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(9): 16318-16329.
- [6] 吴晨曦,应保胜,许小伟,等.基于改进单步多框目标检测的道路小目标检测算法[J].科学技术与工程,2023,23(5):2051-2058.  
Wu Chenxi, Ying Baosheng, Xu Xiaowei, et al. Road small target detection algorithm based on improved single shot multibox detector [J]. Science Technology and Engineering, 2023, 23(5): 2051-2058.
- [7] 李旭,宋焕生,史勤,等.CIEFRNet:面向高速公路的抛洒物检测算法[J].计算机工程与应用,2024,60(5):336-346.  
Li Xu, Song Huansheng, Shi Qin, et al. CIEFRNet: abandoned objects detection algorithm for highway[J]. Computer Engineering and Applications, 2024, 60(5): 336-346.
- [8] Nascimento M G, Fawcett R, Prisacariu V A. DSCConv: efficient convolution operator[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019: 5148-5157.
- [9] Yang L, Zhang R Y, Li L, et al. SimAM: a simple, parameter-free attention module for convolutional neural networks[C]//International Conference on Machine Learning. Online: PMLR, 2021: 11863-11874.
- [10] Li H, Li J, Wei H, et al. Slim-neck by GSCConv: a better design paradigm of detector architectures for autonomous vehicles [J]. ArXiv Preprint, 2022: 2206.02424.
- [11] Pinggera P, Ramos S, Gehrig S, et al. Lost and found: detecting small road hazards for self-driving vehicles[C]//2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Daejeon: IEEE, 2016: 1099-1106.
- [12] Singh A, Kamireddypalli A, Gandhi V, et al. Lidar guided small obstacle segmentation[C]//2020 IEEE/RSJ International Confer-

- ence on Intelligent Robots and Systems (IROS). Las Vegas; IEEE, 2020: 8513-8520.
- [13] Maag K, Chan R, Uhlemeyer S, et al. Two video data sets for tracking and retrieval of out of distribution objects[C]//Proceedings of the Asian Conference on Computer Vision. Macau; Springer, 2022: 3776-3794.
- [14] Li K, Chen K, Wang H, et al. Coda: a real-world road corner case dataset for object detection in autonomous driving[C]//European Conference on Computer Vision. Tel Aviv; Springer, 2022: 406-423.
- [15] Li C, Zhou A, Yao A. Omni-dimensional dynamic convolution [J]. ArXiv Preprint, 2022: 2209.07947.
- [16] Chollet F. Xception: deep learning with depthwise separable convolutions[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu; IEEE, 2017: 1251-1258.
- [17] Han K, Wang Y, Tian Q, et al. Ghostnet: more features from cheap operations[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Seattle; IEEE, 2020: 1580-1589.
- [18] Misra D, Nalamada T, Arasanipalai A, et al. Rotate to attend: convolutional triplet attention module [C]//IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa; IEEE, 2021: 3139-3148.
- [19] Zhang Q L, Yang Y B. SA-Net: shuffle attention for deep convolutional neural networks [C] //ICASSP-IEEE International Conference on Acoustics. Toronto; IEEE, 2021: 2235-9.
- [20] Lau K W, Po L M, Rehman Y A U. Large separable kernel attention: rethinking the large kernel attention design in CNN[J]. Expert Systems with Applications, 2024, 236: 121352.
- [21] Li Y H, Yao T, Pan Y W, et al. Contextual transformer networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(2): 1489-1500.
- [22] Liu W, Anguelov D, Erhan D, et al. SSD: single shot multibox detector[C]//European Conference on Computer Vision. Amsterdam; Springer, 2016: 21-37.
- [23] Zhou X, Wang D, Krähenbühl P. Objects as points[J]. ArXiv Preprint, 2019: 1904.07850.
- [24] Redmon J, Farhadi A. YOLOv3: an incremental improvement [J]. ArXiv Preprint, 2018: 1804.02767.
- [25] 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]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver; IEEE, 2023: 7464-7475.
- [26] Ge Z, Liu S, Wang F, et al. YOLOX: exceeding YOLO series in 2021[J]. ArXiv Preprint, 2021: 2107.08430.
- [27] Wang C Y, Yeh I H, Liao H Y M. YOLOv9: learning what you want to learn using programmable gradient information[J]. ArXiv Preprint, 2024: 2402.13616.
- [28] Everingham M, Eslami S A, Van Gool L, et al. The pascal visual object classes challenge: a retrospective[J]. International Journal of Computer Vision, 2015, 111: 98-136.
- [29] Geiger A, Lenz P, Stiller C, et al. Vision meets robotics: the KITTI dataset [J]. The International Journal of Robotics Research, 2013, 32(11): 1231-1237.