面向边缘端的实时道路障碍检测研究

宋明欣[1] 罗中泽[1] 郑奇志[1] 董光辉[1]
SONG Mingxin LUO Zhongze ZHENG Qizhi DONG Guanghui

摘 要 针对出行者在复杂路况下因情绪紧张而容易误判道路障碍的问题，为了实现道路障碍的实时目标检测，减少事故的发生概率，提出了一种基于 YOLOv9 的道路障碍检测模型。首先，采用 Gold-YOLO 模块作为 Neck 网络来代替传统 FPN 结构，提高模型对于不同层级特征的融合能力的同时实现了检测准确性的提升；然后，采用 CIoU 边界回归损失函数作为模型训练的损失函数，保证了模型的定位准确性；最后，在 WOTR 数据集上进行实验并部署到边缘计算设备测试。实验结果表明，改进后的模型 mAP@0.5 和 mAP@0.5:0.95 分别达到了 87% 和 61.8%，在不显著增加模型计算量的情况下，目标检测的精度得到了普遍提升。

关键词 YOLOv9；道路障碍检测；Gold-YOLO；边缘计算

doi：10.3969/j.issn.1672-9528.2025.01.028

0 引言

近年来，在经济蓬勃发展的有力推动下，道路体系得以持续完善，进而使得人们的出行方式变得愈发多样。多样化的出行方式给人们提供便利的同时，也使得道路交通状况变得更加复杂，由此引发的交通安全问题愈发受到关注。因此，实现对道路障碍精准和迅速的检测具有重要的研究意义。

基于深度学习的目标检测方法能够自动提取复杂数据中的特征，大幅提升目标检测的准确性和实时性。根据是否生成候选框，深度学习方法可以分为两阶段的目标检测算法（Two-stage）及一阶段的目标检测算法（One-stage）。Two-stage 算法对输入样本进行候选区域选取，然后通过卷积神经网络对候选区域进行目标分类和定位。其典型算法包括 R-CNN[1]、Fast R-CNN[2] 和 Faster R-CNN[3] 等，尽管这类算法有较高的检测精度，但受限于计算量大、运行速度慢以及实时性较差等问题，难以部署到边缘端设备上。而 One-stage 算法无须生成候选框，因此具有较快的检测速度，适合需要实时性的应用场景，其中的 YOLO[4] 系列算法具有准确率高、泛化能力强、网络结构简单等优点，能够满足道路目标检测实时性和轻量化的要求。因此，本文采用 YOLO 算法对道路障碍进行目标检测识别。本文的主要贡献如下：

（1）为了提高实时道路障碍检测的精确度和速度，本文采用 Gold-YOLO 模块作为 Neck 网络来代替传统 FPN 结构，提升了模型对不同尺寸物体的检测能力。

（2）针对道路障碍检测场景复杂，易出现漏检错检的问题，在模型训练时采用 CIoU 损失函数，增强了模型对于不同目标检测的鲁棒性和定位准确性。

（3）使用边缘计算设备进行部署，成功实现边缘端对于道路障碍的目标检测识别。

1 实验方法

1.1 YOLOv9 算法

2024 年 2 月，Wang 等人 [5] 发布了新一代 YOLO 系列目标检测版本，即 YOLOv9。与 YOLO 系列的其他版本一样，其网络结构主要分为三部分：Backbone、Neak 和 Head。Backbone 用于提取图像中的特征信息；Neak 负责进一步处理和融合特征信息；Head 负责输出目标的类别、位置以及置信度等信息。YOLOv9 在 YOLOv7[6] 的基础上，针对深度网络传输中的数据丢失问题，提出了可编程梯度信息 PGI 的概念，并设计了一种基于梯度路径规划的新型轻量级网络架构，即广义高效层聚合网络 GELAN，极大地提高了目标检测的性能。基于以上改进，YOLOv9 在性能和灵活性上得到了极大的提升，在准确性、速度和整体性能方面位于现有的目标检测算法前列。因此本文采用 YOLOv9C 作为道路障碍识别的基础模型进行研究。

1.2 Gold-YOLO

在模型中，特征融合对于提高检测的准确性和鲁棒性至关重要。特征融合使模型能够结合来自网络不同层次的信息，

1. 黑龙江省哈尔滨市东北林业大学 黑龙江哈尔滨 150036
[基金项目] 国家级大学生创新创业计划项目（202410225053）

这些层次分别捕获了图像的不同方面的特征。YOLOv9 模型的 Neck 网络采用了传统的 FPN 结构进行特征融合，其结构图如图 1 所示。这种结构主要融合相邻层级的特征，对远距离层级的信息融合不足，整体上信息融合的有效性受到限制。为了更好地进行多尺度特征融合，将 YOLOv9 的 Neck 网络替换为 Gold-YOLO。

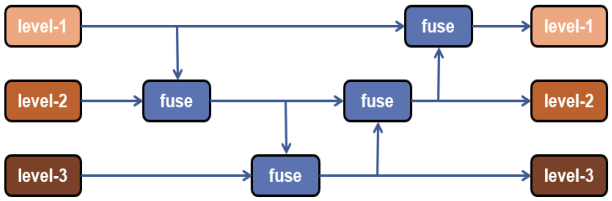


图 1 传统的 FN 结构图

2023 年，Wang 等人[7]提出 Gold-YOLO 这一改进方法，其架构如图 2 所示。Gold-YOLO 采用了一种新的 GD 机制，通过一个统一模块收集所有层级的信息并有效融合，然后分发到各个层级。

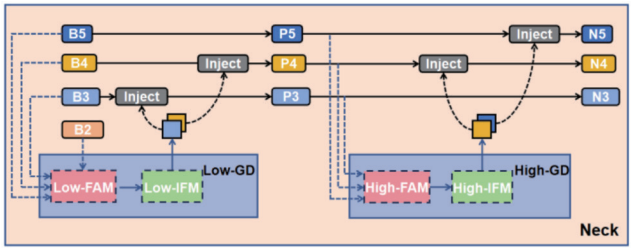


图 2 Gold-YOLO 架构图

其中 Low-GD 将主干网络输出的特征进行融合，以获取小目标的高分辨率特征。在低级全局特征对齐模块 Low-FAM 中，采用 AvgPool 操作对输入特征进行下采样，然后将特征调整为组中最小的特征大小，从而得到对齐后的特征。这样既保证了信息的有效聚合，又通过变换器模块减少了后续处理的计算负担。在低阶信息融合模块 Low-IFM 中，多层重参数化卷积块 RepBlock 输出的特征在通道维度上分裂，这些特征会与其他层级的特征融合。Inject 信息注入模块接收当前层特征和全局注入信息。并通过两个卷积层处理，进一步提取和融合信息。这样的设计在确保特征间的有效融合的同时也优化了计算效率。High-GD 将 Low-GD 生成的特征进行融合。高级特征对齐模块 High-FAM 同样通过 AvgPool 操作简化输入特征的尺寸为组中的最小大小，之后的高级信息融合模块 High-IFM 用于不同层级特征的融合。

1.3 损失函数

在目标检测中，边界框回归的损失函数对于精确定位至关重要。它通过衡量预测边界框与真实边界框之间的差异来优化模型参数，从而显著提高模型的定位准确性。目前常用的边界框回归损失函数包括 IoU[8] 损失函数、CIoU[9] 损失函数、EIoU[10] 损失函数、MPDIoU[11] 损失函数、Inner-ShapeIoU[12] 损失函数和 Focal-CIoU 损失函数等。

2 实验分析和可视化

2.1 实验环境与数据集

本实验平台采用 Intel(R) Xeon(R) Silver 4310 CPU @ 2.10 GHz，NVIDIA A100 80 GB PCIe GPU 和 Ubuntu 20.04.2 操作系统。深度学习框架是 PyTorch 2.0.1，Python 的版本为 3.8.0, CUDA 的版本是 11.7。本实验的模型训练参数设置如下：批量大小为 32，训练周期为 200，图像大小为 640，初始学习率为 0.01，权重衰减系数为 0.000 5，动量为 0.937。使用 SGD 作为迭代训练的优化器。

本实验采用 Xia 等人[13]提出的 WOTR（walk on the road）公开数据集作为道路目标检测数据集。WOTR 数据集共包含 13 928 张图像，其中 9056 张用于训练集，2534 张用于测试集，其余 2338 张用于验证。该数据集中包括 15 种常见障碍物和 5 种道路判断对象，涵盖了道路行走和过马路两种主要场景，该数据集中的对象类别如图 3 所示。

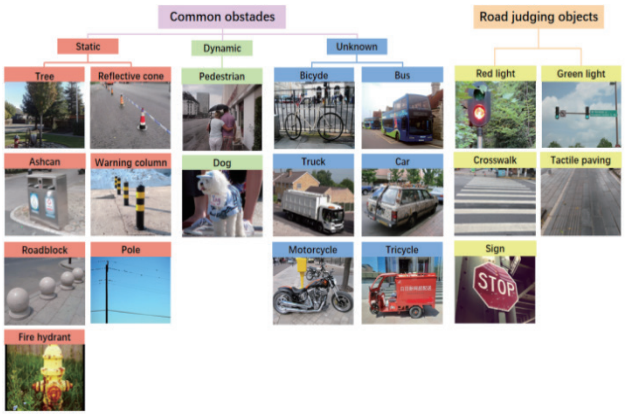


图 3 WOTR 数据集中的对象类别

2.2 评价指标

在本文中，采取 P、R、AP、mAP、GFLOPs（千兆浮点运算）以及 Params 作为评价指标来评价模型性能。其中 P 表示模型正确预测为正类别的对象占所有预测为正类别的对象的比例，R 表示模型正确预测为正类别的对象占所有实际正类别的对象的比例，均被用于评估模型在分类准确率和错误率方面的性能。AP 是一种用于衡量目标检测算法性能的常用指标，AP 越高，模型对于该目标的检测性能越好。对于多个类别的物体检测任务，对每个类别的 AP 进行平均得到 mAP，其计算公式为：

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

$$\mathrm{mAP}=\frac{1}{n}\sum_{i=1}^{n}\mathrm{AP}_i \quad (2)$$

GFLOPs 是一个关键的性能指标，用于评估模型的计算效率和实时性能。GFLOPs 越低，模型的计算效率就越高。Params 是指模型中需要训练的参数总数，用于评估模型的复杂性。Params 越小，模型所需计算资源越少。

2.3 实验结果及分析

在提出 WOTR 数据集的同时，Xia 等人还使用多种目标检测模型对该数据集进行测试，给出了详细的测试结果。其中，YOLOv7 模型超越了较为先进的 YOLOv8 模型，表现出了最佳的检测效果。因此，Xia 等人以 YOLOv7 为基础提出了一种新颖的 PC-YOLO 模型，提升了道路障碍检测速度，其测试结果如表 1 所示。

表 1 PC-YOLO 模型检测结果对比

Model	AP/%	mAP@0.5/%	mAP@0.75/%	GFLOPs	Params/10^6
Faster-R-CNN-fpn	53.4	82.6	58.3	134.7	41.8
YOLOv4-csp-s-mish	52.8	81.1	56.7	20.7	8.1
YOLOv5-L	60.7	85.6	66.9	109.1	46.5
YOLOv7	63.5	88.6	70.2	105.4	37.3
YOLOv8-L	61.3	85.9	67.3	165.2	43.7
YOLOv7 + PC-YOLO	60.5	86.4	66.3	236	37.4

本文使用 YOLOv9C 模型对该数据集进行测试，并发现 YOLOv9C 模型相比于 PC-YOLO 模型表现出了更好的检测精度和效率。因此，本文以 YOLOv9C 作为实验的基础模型，并对其进行改进。

2.3.1 颈部网络对比实验

本实验采用 Gold-YOLO 架构来代替原模型的颈部网络，以增强模型的特征信息融合能力。根据测试结果表 2 显示，改进后模型的 mAP@0.5 和 mAP@0.5:0.95 分别达到了 87% 和 61.8%，相比于原 YOLOv9C 模型均提升了 0.4 个百分点。同时 mAP@0.5 相比于 Xia 等人改进后的 PC-YOLO 模型提升了 0.6 个百分点。在不显著增加模型计算量的情况下，目标检测的精度得到了普遍提升。

2.3.2 损失函数对比实验

为了比较不同损失函数对模型训练效果的影响，选取了几种常见的边界框回归损失函数进行对比实验。从实验结果表 3 中可以看出，CIoU 损失函数的 mAP@0.5 上达到了 87%，在所有损失函数中达到了最好的效果。因此选择 CIoU 损失函数作为模型训练的损失函数。

表 3 损失函数对比

Model	IoU/%	CIoU/%	EIoU/%	Focal-CIoU/%	Inner-ShapeIoU/%	MPD-IoU/%	SIoU/%
map@0.5	86.6	87	85.4	85.6	85.4	85.6	86.7

2.3.3 实验结果对比

为了分析本模型在复杂道路障碍场景下的检测情况，本实验采用 YOLOv9C 和本模型对多个代表性样本进行检测，测试结果图 4 所示。对于（a）中的垃圾箱和电线杆，YOLOv9C 的检测边界框相较于真实边界框出现了较大偏移，而本模型检测边界框的位置更接近于真实边界框。观察（b）（c）可以看出，YOLOv9C 检测结果中会出现将一个目标识别为同一类型的多个目标重叠的情况，本模型更好地避免了这种问题，定位准确性进一步增强。并且对于图 4 中的部分目标，本模型的检测结果表现出了更高的置信度，检测的可靠性得到了提升。

图 4 YOLOv9C 与本模型检测结果对比

2.3.4 边缘端测试结果

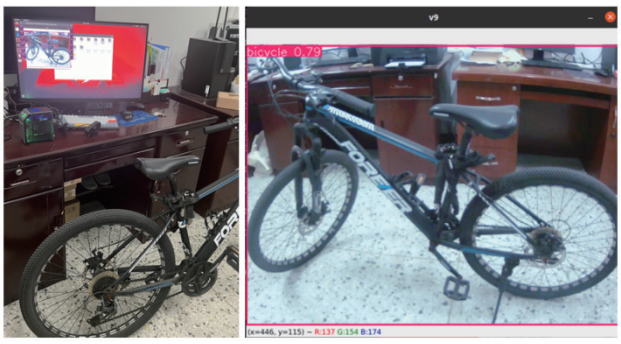
本文中，先在树莓派上进行模型移植，在移植过程中，树莓派出现了检测效率低、时间延迟长和识别准确率低等问题，不能满足复杂道路场景应用的要求。最终本决定将模型移植到 Jetson Orin Nano 上，这是一款强大的嵌入式人工智能开发板，采用 NVIDIA Ampere 架构的 GPU 和 6 核 ARMCortex-

表 2 YOLOv9C 与 Gold-YOLO 检测结果对比

Model	*P*/%	*R*/%	mAP@0.5/%	mAP@0.5:0.95/%	GFLOPs	Params/10^6
Baseline (YOLOv9C)	87.2	79.4	86.6	61.4	236.9	50.7
Baseline+Gold-YOLO	86.7	79.1	87	61.8	265.6	57.4

A78AE CPU，能够处理复杂的计算任务。

本实验采用 NVIDIA Jeston Orin Nano 8 GB 对该模型进行移植，实验硬件如图 5（a）所示，测试结果如图 5（b）所示。可以看出 NVIDIA Jeston Orin Nano 8 GB 表现出了优越的性能。



(a) (b)

图 5 硬件测试结果

3 结论

为了提高实时道路障碍检测的精确度和速度，本文提出了一种基于改进 YOLOv9 的目标检测模型。该模型采用 Gold-YOLO 模块作为 Neck 网络，避免了传统 FPN 结构的信息丢失问题，提升了模型对不同尺寸物体的检测能力。在模型训练时采用 CIoU 损失函数，增强了模型对于不同目标检测的鲁棒性和定位准确性。与 YOLOv9C 相比，改进模型的性能在各方面得到了极大提升，并且成功部署到边缘计算设备上进行测试。

参考文献：

[1] HE K M, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C/OL]//2017 IEEE International Conference on Computer Vision (ICCV). Piscataway:IEEE,2017[2024-03-16].https://ieeexplore.ieee.org/document/8237584.

[2] GIRSHICK R. Fast R-CNN [DB/OL].(2015-09-21)[2024-02-24].https://arxiv.org/abs/1504.08083.

[3] REN S Q, HE K, 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. Piscataway:IEEE,2016, 39(6): 1137-1149.

[4] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2016:779-788.

[5] WANG C Y, YEH I H, LIAO H Y M . YOLOv9: learning what you want to learn using programmable gradient information [DB/OL].(2024-02-29)[2024-03-11].https://doi.org/10.48550/arXiv.2402.13616.

[6] 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/OL]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2023[2024-01-15].https://ieeexplore.ieee.org/document/10204762.

[7] WANG C C, HE W, NIE Y, et al. Gold-YOLO: efficient object detector via gather-and-distribute mechanism [DB/OL]. (2023-10-23)[2024-02-24].https://doi.org/10.48550/arXiv.2309.11331.

[8] YU J H, JIANG Y N, WANG Z Y, et al. Unitbox: an advanced object detection network[DB/OL].(2016-08-04)[2024-04-11].https://doi.org/10.48550/arXiv.1608.01471.

[9] ZHENG Z H, WANG P, LIU W, et al. Distance-IoU loss: faster and better learning for bounding box regression[J]Proceedings of the AAAI conference on artificial intelligence, 2020, 34(7):7.

[10] ZHANG Y F, REN W Q, ZHANG Z, et al. Focal and efficient IOU loss for accurate bounding box regression [J]. Neurocomputing, 2022, 506(9): 146-157.

[11] MA S L, XU Y.Mpdiou: a loss for efficient and accurate bounding box regression [DB/OL].(2023-07-14)[2024-04-12].https://doi.org/10.48550/arXiv.2307.07662.

[12] ZHANG H, XU C, ZHANG S J. Inner-IoU: more effective intersection over union loss with auxiliary bounding box [DB/OL].(2023-11-14)[2024-01-19].https://doi.org/10.48550/arXiv.2311.02877.

[13] XIA H Y, YAO C, TAN Y M, et al. A dataset for the visually impaired walk on the road [J]. Displays, 2023, 79(9): 102486.

【作者简介】

宋明欣（2004—），女，河北承德人，本科，研究方向：深度学习。

罗中泽（2003—），男，江西赣州人，本科，研究方向：深度学习、人工智能。

郑奇志（1998—），女，山西朔州人，硕士，研究方向：嵌入式、模式识别。

董光辉（1978—），男，黑龙江安达人，博士，副教授、副主任，研究方向：植物物理信息检测、图像处理及模式识别。

（收稿日期：2024-09-29）