基于 YOLOv11 改进的驾驶员视角下前方车头与车尾检测

许牛琦[1] 王 金[1] 付 迪[1] 郝 韵[1]
XU Niuqi WANG Jin FU Di HAO Yun

摘 要 随着智能驾驶和自动驾驶技术的快速发展，车辆对环境感知的需求日益增长。车头与车尾的目标检测技术能够提供车辆的位置和运动状态信息，对于提高驾驶的安全性至关重要。文章基于 YOLOv11s 网络，关注车头与车尾的空间上下文信息，通过添加注意力机制，融合全局特征，提出 YOLOv11AD 网络结构，提高识别效果，优化检测准确率。在自定义的车头、车尾数据集中采用了各种 YOLO 模型进行测试，包括 YOLOv5s、YOLOv8s、YOLOv9s、YOLOv10s、YOLOv11s 和改进的 YOLOv11AD 网络模型，以确保在复杂环境中进行有效识别。其中改进的 YOLOv11AD 网络模型在 YOLOv11 的基础上增加 AFGCAttention 和 DBB 模块，Precision 提高了 0.4%，Recall 提高了 3.7%，mAP50 提高了 0.2%，mAP50-95 提高了 0.8%。

关键词 YOLOv11；车辆检测；深度学习；智能驾驶

doi：10.3969/j.issn.1672-9528.2025.11.025

0 引言

随着自动驾驶技术的不断发展，驾驶员的角色逐渐从主要操控者转变为监督者和潜在接管者。在自动驾驶阶段，驾驶员仍需保持一定的警觉性，及时监控车辆，以备随时进行接管。然而，研究表明，车内非驾驶相关次任务可能会引起驾驶员分心，增加认知负荷，从而削弱其接管能力 [1]。特别是在需要高度集中注意力和快速反应的情况下，如会车、跟车场景，这种影响尤为显著 [2]。

双车道公路上的会车与跟车是两种典型的驾驶场景，对驾驶员所产生的负荷程度并不相同。会车通常会在短时间内产生较高的驾驶负荷，因为要求驾驶员在短时间内进行高度集中的注意力和快速的反应；而跟车虽然单次负荷较低，但其持续性和累积效应也不容忽视，尤其是在长时间驾驶或交通繁忙的情况下。因此，准确识别并区分不同驾驶场景（会车、跟车）对于确保行车安全至关重要。

本文旨在探讨如何利用目标检测算法来实现会车与跟车场景的精准鉴别，进而优化自动驾驶系统的性能，保障行车安全。道路车辆检测在计算机视觉中的发展可分为两个重要阶段：基于手工特征设计的传统方法和基于深度学习的现代方法。在传统方法阶段，研究者主要从车辆的颜色、纹理和形状等先验知识出发，采用 Haar 小波特征 [3]、方向梯度直方图 [4]（HOG）等人工设计的特征进行识别。这类方法存在明显的局限性：首先，特征设计过程高度依赖领域经验，需要耗费大量时间进行参数调优；其次，预设特征的泛化能力较弱，当面对光照变化、视角转换等复杂场景时，检测性能显著下降；最后，针对不同应用场景（如城市道路与高速公路）需要重新设计特征工程，严重制约了方法的实际应用价值。

随着机器学习技术的突破，基于深度学习的车辆检测方法通过数据驱动的方式实现了质的飞跃 [5]。这类方法能够从海量标注数据中自动学习多层次的特征表达，不仅有效克服了人工设计特征的局限性，还能更精准地捕捉数据的内在分布规律。当前主流算法根据检测机制可分为两大类别：两阶段检测器（如 Faster R-CNN 系列）首先生成候选区域再进行精细分类与回归，在检测精度上具有优势；而一阶段检测器（以 YOLO 系列为代表）采用端到端架构直接在特征图上完成边界框预测和分类，显著提升了检测速度。这种技术演进使车辆检测系统在复杂动态场景下的鲁棒性和实时性都得到了全面提升。YOLO 系列模型以其快速准确的检测性能，在实时车辆检测方面取得了很多成功应用 [6-12]。

YOLOv5 是由 Ultralytics 推出的开源框架，首创轻量化（n/s/m/l/x）多版本架构设计，通过动态学习率调整和模块化结构显著提升训练效率与推理速度。其优势在于部署便捷性与工业级应用适配能力，但存在小目标检测精度不足及密集场景误检率偏高的局限。

YOLOv8 由 Ultralytics 公司于 2023 年发布，在复杂场景

1. 北京工业大学城市交通学院 北京 100124
[基金项目] 北京市自然科学基金（8232005、L221026）

下表现优异，引入新的注意力机制和数据增强策略，支持全方位的视觉 AI 任务，使得用户可以在各个应用和领域中利用 YOLOv8 的功能。尽管进行了优化，但高性能的模型仍需要较大的计算资源。复杂的网络结构和多个模块增加了模型的复杂度和训练难度。

YOLOv9 于 2024 年 2 月正式提出。为了解决深度网络梯度传输过程中的数据丢失问题，YOLOv9 引入了可编程梯度信息 PGI 的概念。PGI 可以为目标任务提供完整的输入信息来计算目标函数，从而获得可靠的梯度信息来更新网络权重，有效解决传统深度网络中因信息逐层压缩导致的特征稀释与梯度弥散问题。主要由主分支、辅助可逆分支、多级辅助信息三部分组成。主分支用于推理，辅助可逆分支解决了网络深度增加导致的信息瓶颈问题。多级辅助分支解决了深度监管带来的误差积累问题。但 AutoML 的引入增加了模型开发的时间成本，对于极端复杂或遮挡严重的目标，检测效果可能受到影响。

YOLOv10 于 2024 年由清华大学团队开源，针对超大规模模型进行了优化，突破性地提出 NMSfree（非极大值抑制）训练的一致双分配，实现了高效的端到端检测。引入了整体效率精度驱动的模型设计策略。虽然进行了轻量化设计，但相对于一些更简单的模型来说，YOLOv10 仍面临模型复杂度与边缘设备适配的平衡挑战。

YOLOv11 是 YOLO 系列目标检测模型中的最新版本。通过 C3k2 模块增强浅层特征捕获能力，结合 SPPF 金字塔池化与 C2PSA 跨尺度注意力机制，构建多粒度特征融合网络。延续多规格模型（n/s/m/l/x）传统，规模从小到大，适合不同的应用场景。YOLOv11s 更好地兼顾了检测速度和精度，有较强的鲁棒性。能够在动态变化的道路环境中快速准确地检测和分类物体。

在驾驶负荷研究中，驾驶员需关注视线前方会车（车头）和跟车（车尾）的信息。YOLOv11s 通过深度卷积层逐层提取特征，能够识别车头和车尾结构及轮廓差异，但仍存在复杂情境中车头、车尾漏检误检现象，为提高车头、车尾检测精度，本文基于 YOLOv11s 网络，添加注意力机制，提出 YOLOv11AD 网络结构。利用车头与车尾的空间上下文信息，结合全局特征融合技术，提高识别效果，优化检测准确率。此外，构建专门针对车头和车尾的标注数据集，使 YOLOv11AD 网络模型更好地学习这些特定特征，可以实现对驾驶员视线前方车辆信息的快速且精准识别，进而支撑驾驶负荷分析，为自动驾驶系统提供可靠的决策支持。

1 改进的 YOLOv11 网络结构

本文在轻量级 YOLOv11s 模型的基础上进行改进与测试。YOLOv11s 网络结构中 Backbone 部分负责特征提取，采用了一系列卷积和反卷积层，同时使用了残差连接和瓶颈结构来减小网络的大小并提高性能。通过在 backbone 中加入 AFGCAttention。在 Head 中 YOLOv11 特有的 C3k2 模块中引入 DBB 结构。

通过引入 AFGCAttention 和 DBB，模型能够在复杂的场景中更精准地提取和利用重要特征，同时动态调整卷积核的表征能力。AFGCAttention 利用注意力机制使模型专注于关键特征区域，减少不相关或冗余信息的干扰，从而提高特征表示的质量，使得后续检测头能够基于更高质量的特征进行边界框预测和类别分类，特别是在处理复杂背景或目标时，显著提升了模型的鲁棒性和准确性。DBB（dynamic block branching）则通过对训练时的模块进行结构重参数化，在测试时优化 $K\times K$ 卷积的学习过程，引入先验知识指导卷积核学习不同尺度下的表征能力，而非仅限于固定大小的卷积核，其更大的输出通道数能够捕捉到更丰富的特征信息，进一步增强了模型对多尺度特征的学习能力。两者共同作用提升了模型在目标检测和分类任务中的整体性能。至此完成基于 YOLOv11s 改进的 YOLOv11AD 网络，其网络结构图如图 1 所示。

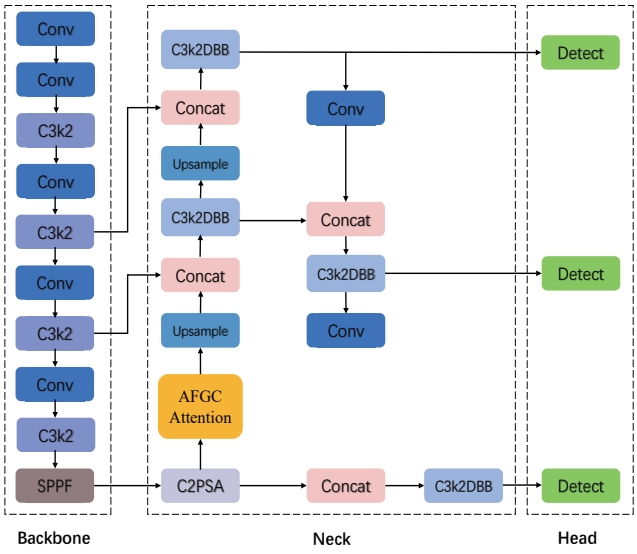

图 1 YOLOv11AD 网络结构图

2 实验与分析

2.1 实验数据集

会车车辆（车头）从另一车道向试验车靠近，形成会车，而在同一车道内，试验车会跟随前方车辆（车尾），形成跟车。为了研究前方会车和跟车对驾驶负荷的影响，前方车辆信息来自 Tobii 眼动仪采集数据时自动录制的视频数据，该眼动仪记录各项眼动数据的同时，可以类似于行车记录仪，录制驾驶员可见的视野范围内的视频。

关于自制数据集，从眼动仪录制视频中，使用 OpenCV 库自动捕捉视频中的连续帧。以每 12 帧的间隔截取图像（视频帧率为 24 帧 /s，即每秒提取 2 张图像）。考虑行车视距值，剔除空标签的样本，从连续帧中筛选可能对驾驶负荷产生影响的驾驶员视角下的前方车辆图像如图 2 所示，分为车头和车尾两类标签。最终获得了包含车头和车尾的 1 500 张图像。其中训练集数据 1 200 张，测试集数据 300 张。

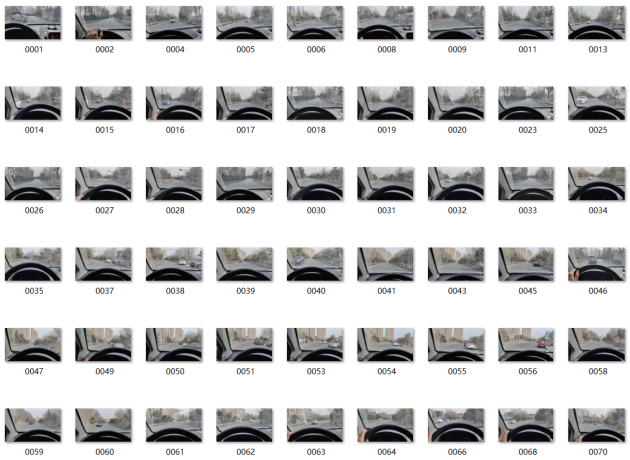

图 2 驾驶员视角下的前方车辆图像

2.2 实验环境与参数指标

YOLOv11AD 网络在操作系统 Windows11 下进行，GPU 为 GeForceRTX3060，主机内存为 32 GB，编程语言为 Python3.9，使用 CUDA version 11.8 对 GPU 进行加速，使用深度学习框架 PyTorch2.0.0 进行训练，训练参数配置如表 1 所示。

表 1 训练参数配置

参数名称	配置	参数名称	配置
epochs	300	optimizer	Sgd
batchsize	32	close_mosaic	10
workers	0	momentum	0.937
imgsz	640	lr0	0.01

在目标检测任务中，模型性能通常以平均精度均值（mAP）（公式 1、公式 2）为核心评估标准，其通过整合精确度（Precision）（公式 3）与召回率（Recall）（公式 4）构建综合评价维度，Precision 量化检测结果的可靠性，侧重降低误检率；Recall 评估目标覆盖完整性，旨在减少漏检风险。此外，mAP@0.5 以交并比（IoU）阈值 0.5 衡量基础检测能力，而 mAP@0.5:0.95 通过 0.5 至 0.95 多阈值平均反映精细化定位水平。目标检测任务中，mAP 值越高，模型效果越好。

$$AP = \int_0^1 P(y)dy \tag{1}$$

$$mAP = \frac{1}{n}\sum_{i=0}^{n} AP_i \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

式中：TP 表示真正例（正确检测到的目标数）；FP 表示假正例（错误检测出的目标数）；FN 表示假反例，即实际存在但未被检测到的目标数。

2.3 实验结果分析

将眼动仪录制视频送入 YOLO 系列模型，对有无跟车和有无会车进行分类。YOLOv11AD 的 loss 随着训练轮数的增加而逐渐减小并趋于稳定，如图 3 所示，这意味着模型在训练过程中逐渐学习到车头车尾数据的特征，且正在收敛。其车头、车尾检测可视化，如图 4 所示。

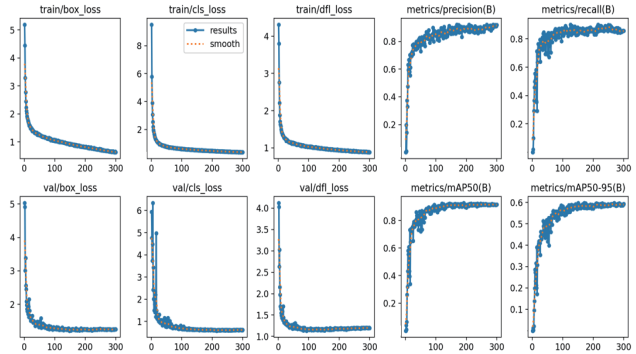

图 3 YOLOv11AD 训练结果

图 4 YOLOv11s 检测结果

YOLOv5s、YOLOv8s、YOLOv9s、YOLOv10s、YOLOv11s、YOLOv11AD 的检测结果如表 2、表 3、表 4 所示。其中 YOLOv11AD 的综合表现最佳，车头检测的 Precision、Recall、mAP50、mAP 50-95 为 0.951、0.903、0.958、0.632。车尾检测的 Precision、Recall、mAP50、mAP 50-95 为 0.894、0.864、0.903、0.580。相对于 YOLOv11s，YOLOv11AD 的平均 Recall 提高了 3.7%，

mAP50 提高了 0.2%。此外，相对于车尾，车头检测精度更高，是因为驾驶员视线中，会车可以获取更近距离的车头数据，而跟车处于相对较远的位置，车头数据集质量更高。此外，车头部分通常具有更明显的视觉特征，例如结构、轮廓、格栅等，这些特征在图像中较为突出。总的来说，该精度基本可以满足驾驶负荷分析。

表 2 对比实验：车头检测结果

车头	Precision	Recall	mAP50	mAP50-95
YOLOv5s	0.910	0.878	0.937	0.627
YOLOv8s	0.922	0.882	0.952	0.634
YOLOv9s	0.943	0.883	0.952	0.630
YOLOv10s	0.935	0.880	0.945	0.635
YOLOv11s	0.950	0.883	0.961	0.631
YOLOv11AD	0.951	0.903	0.958	0.632

表 3 对比实验：车尾检测结果

车尾	Precision	Recall	mAP50	mAP50-95
YOLOv5s	0.835	0.821	0.864	0.504
YOLOv8s	0.815	0.848	0.877	0.56
YOLOv9s	0.879	0.826	0.899	0.583
YOLOv10s	0.854	0.839	0.887	0.571
YOLOv11s	0.888	0.811	0.903	0.573
YOLOv11AD	0.894	0.864	0.903	0.580

表 4 对比实验：平均检测结果

平均	Precision	Recall	mAP50	mAP50-95
YOLOv5s	0.872	0.843	0.904	0.589
YOLOv8s	0.868	0.865	0.914	0.597
YOLOv9s	0.911	0.854	0.925	0.606
YOLOv10s	0.894	0.860	0.916	0.603
YOLOv11s	0.917	0.847	0.931	0.597
YOLOv11AD	0.921	0.884	0.933	0.605

3 结论

本文提出的 YOLOv11AD 网络可以精确检测会车与跟车情景，相比于 YOLOv11s 有更高的精度和鲁棒性，确保在各种复杂路况下稳定检测车头与车尾。综上所述，本文提出的 YOLOv11AD 网络在自动驾驶安全中有应用潜力，通过场景理解以减少认知负荷，进一步提升驾驶安全性，可以为未来的智能交通系统提供一定理论和技术支持。

参考文献：

[1] 马艳丽 , 卢俊 , 朱洁玉 , 等 . 不同认知负荷非驾驶任务下高度自动化驾驶接管绩效预测 [J]. 汽车工程 , 2023, 45(12): 2330-2337.

[2] 胡汇 . 山区公路弯道会车过程驾驶人心生理及行为耦合特性分析 [J]. 科技和产业 ,2023,23(15):169-174.

[3] 李琳 , 靳志鑫 , 俞晓磊 , 等 .Haar 小波下采样优化 YOLOv9 的道路车辆和行人检测 [J]. 计算机工程与应用 , 2024, 60(20): 207-214.

[4] 曲爱妍 , 张正 , 梁颖红 , 等 . 基于方向梯度直方图特征的车脸识别方法研究 [J]. 软件工程 ,2021,24(9):38-43.

[5] 李明熹 , 林正奎 , 曲毅 . 计算机视觉下的车辆目标检测算法综述 [J]. 计算机工程与应用 ,2019,55(24):20-28.

[6] BAUTISTA C M, DY C A, MANALAC M I,et al.Convolutional neural network for vehicle detection in low resolution traffic videos[C]// 2016 IEEE Region 10 Symposium (TENSYMP). Piscataway:IEEE,2015:277-281.

[7] KANG L, LU Z W, MENG L Y, et al.YOLO-FA: type-1 fuzzy attention based YOLO detector for vehicle detection[J]. Expertsystems with applications, 2024, 237:121209.

[8] WANG J F, CHEN Y, DONG Z K,et al.Improved YOLOv5 network for real-time multi-scale traffic sign detection[J]. Neural computing and applications, 2021, 35:7853-7865.

[9] LI C Y, LI L L, JIANG H L,et al. YOLOv6: a single-stage object detection framework for industrial applications[EB/OL].(2022-09-07)[2024-05-26].https://doi.org/10.48550/arXiv.2209.02976.

[10] WANG C-Y, BOCHKOVSKIY A, LIAO H-Y M, et al. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).Piscataway:IEEE,2023:7464–7475.

[11] WANG C-Y, YEH I-H, LIAO H-Y M. YOLOv9: learning what you want to learn using programmable gradient information[EB/OL].(2024-02-29)[2024-12-11].https://doi.org/10.48550/arXiv.2402.13616.

[12] WANG A, CHEN H, LIU L H, et al. YOLOv10: real-time end-to-end object detection[EB/OL]. (2024-10-30)[2024-12-23].https://doi.org/10.48550/arXiv.2405.14458.

【作者简介】

许牛琦（1997—），男，安徽合肥人，硕士研究生，研究方向：基于点云和图像的道路安全分析，email: 1347594942@qq.com。

（收稿日期：2025-01-24 修回日期：2025-07-18）