Operations Research and Fuzziology 运筹与模糊学, 2025, 15(5), 1-12
Published Online October 2025 in Hans. https://www.hanspub.org/journal/orf
https://doi.org/10.12677/orf.2025.155226

Hans汉斯

面向智能交通边缘设备的轻量级高精度车辆检测算法研究

闵　哲

上海理工大学管理学院，上海

收稿日期：2025年8月18日；录用日期：2025年9月8日；发布日期：2025年9月28日

摘　要

车辆检测是智能交通系统的关键技术，然而，将计算密集型的深度学习模型部署于资源受限的边缘设备面临严峻挑战，限制了实时性能的发挥。为解决此问题，本研究基于轻量级检测器YOLOv5s，提出了两种优化的轻量化车辆检测模型。第一种模型LVD-YOLO，以追求极致效率为目标，通过采用EfficientNetv2作为骨干网络、结合BiFPN与CBAM注意力机制增强特征融合、并引入SIoU损失函数优化边界框回归，旨在显著降低模型的参数量与计算复杂度。第二种模型EMO-YOLO，侧重于提升复杂交通场景下的检测精度，同时保持轻量化特性。该模型利用新颖的EMO网络作为骨干，通过引入SCConv重构颈部的C3模块以减少特征冗余，并设计了包含三重注意力机制的新检测头，以增强对小目标、遮挡目标等困难样本的检测能力。在公开数据集UA-DETRAC上的大量实验结果表明：与基准YOLOv5s相比，LVD-YOLO在模型参数量和FLOPs上实现了大幅削减，展现了优越的效率；EMO-YOLO则在保持显著轻量化的同时，在检测精度(特别是mAP@0.5)上超越了YOLOv5s及其他对比的轻量化方法，尤其在复杂场景下表现更佳。消融实验进一步验证了EMO-YOLO中各改进模块的有效性。本研究提出的LVD-YOLO和EMO-YOLO模型为智能交通系统在边缘设备上的高效、精准车辆检测提供了具有竞争力的解决方案。

关键词

智能交通系统，车辆检测，轻量化，YOLOv5s

Research on Lightweight High-Precision Vehicle Detection Algorithm for Intelligent Transportation Edge Equipment

Zhe Min

Business School, University of Shanghai for Science and Technology, Shanghai

Received: August 18, 2025; accepted: September 8, 2025; published: September 28, 2025

文章引用: 闵哲. 面向智能交通边缘设备的轻量级高精度车辆检测算法研究[J]. 运筹与模糊学, 2025, 15(5): 1-12.
DOI: 10.12677/orf.2025.155226

Abstract

Vehicle detection is a crucial technology for Intelligent Transportation Systems. However, deploying computationally intensive deep learning models on resource-constrained edge devices poses significant challenges, limiting real-time performance. To address this issue, this study proposes two optimized lightweight vehicle detection models based on the lightweight detector YOLOv5s. The first model, LVD-YOLO, aims for ultimate efficiency by employing EfficientNetv2 as the backbone network, enhancing feature fusion with BiFPN combined with the CBAM attention mechanism, and optimizing bounding box regression using the SIoU loss function, intending to significantly reduce model parameters and computational complexity. The second model, EMO-YOLO, focuses on improving detection accuracy in complex traffic scenarios while maintaining lightweight characteristics. This model utilizes the novel EMO network as its backbone, reconstructs the C3 module in the neck using SCConv to reduce feature redundancy, and designs a new detection head incorporating a triple attention mechanism to enhance the detection capability for difficult samples such as small and occluded objects. Extensive experimental results on the public UA-DETRAC dataset demonstrate that: compared to the baseline YOLOv5s, LVD-YOLO achieves substantial reductions in model parameters and FLOPs, showcasing superior efficiency; EMO-YOLO, while remaining significantly lightweight, surpasses YOLOv5s and other compared lightweight methods in detection accuracy (especially mAP@0.5), performing particularly well in complex scenarios. Ablation studies further validate the effectiveness of each improved module within EMO-YOLO. The proposed LVD-YOLO and EMO-YOLO models in this study offer competitive solutions for efficient and accurate vehicle detection on edge devices within intelligent transportation systems.

Keywords

Intelligent Transportation Systems, Vehicle Detection, Lightweight, YOLOv5s

1. 引言

随着全球城市化进程的加速和机动车保有量的激增，交通拥堵、事故频发以及环境污染等问题日益严峻，对城市的可持续发展构成了显著挑战。智能交通系统(Intelligent Transportation Systems, ITS)作为融合了现代信息、通信、传感及控制技术的综合解决方案，在提升交通运行效率、保障出行安全、减轻环境负荷方面展现出巨大潜力，已成为各国交通领域发展的重要方向。在众多 ITS 应用场景中，实时、准确地感知和理解道路交通环境是实现智能化管理与控制的基础。ITS 的组成结构如图 1，该结构以数据传输网络为核心，连接数据采集设备(如路侧摄像头、传感器)、云计算中心(负责数据存储与分析)、交通管理中心(实现交通调度决策)及控制设备(如智能信号灯)，形成完整的交通智能管控链路。

近年来，以卷积神经网络(CNN)为代表的深度学习技术极大地推动了目标检测领域的发展，诸如 YOLO、SSD 等一系列先进算法在车辆检测任务上取得了远超传统方法的精度。然而，这些高性能模型往往伴随着巨大的参数量和计算复杂度(FLOPs)。当需要将这些算法部署到算力、内存及功耗受限的边缘计算设备(如路侧摄像头、车载单元)时，现有模型的“重量级”特性便成为瓶颈。这些边缘设备通常无法承载大型网络的运算负荷，难以满足智能交通系统对低延迟、高帧率处理的迫切需求。因此，如何在保

证检测精度的前提下，显著降低模型的复杂度，使其能够在资源受限的边缘平台上高效运行，成为了当前智能交通领域亟待解决的技术难题。

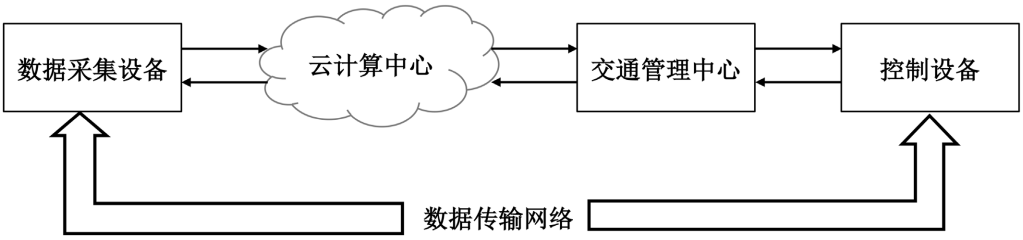

Figure 1. The composition structure of ITS
图 1. ITS 的组成结构

为了应对这一挑战，研究者们致力于探索轻量化的车辆检测模型。当前流行的 YOLOv5 系列中的 YOLOv5s 模型，凭借其在速度与精度之间的良好平衡，常被选作改进的基准。本文聚焦于此，旨在研发出更适用于边缘计算场景、能够兼顾高效率与高精度的轻量级车辆检测方案。

基于 YOLOv5s [1]架构，本文提出了两种不同的优化路径，分别设计并实现了 LVD-YOLO 和 EMO-YOLO 两个轻量化车辆检测模型。LVD-YOLO 模型主要着眼于极致的效率提升，通过引入更高效的网络结构(如 EfficientNetv2 [2]骨干、BiFPN [3]颈部融合以及 CBAM [4]注意力)和优化的损失函数(SIoU [5] Loss)，在大幅削减模型参数量与计算量的同时，力求保持具有竞争力的检测精度。而 EMO-YOLO 模型则更侧重于提升在复杂交通场景下的检测鲁棒性，采用新颖的 EMO 骨干网络，并对颈部结构(引入 SCConv [6])和检测头(引入三重注意力机制)进行针对性改进，旨在有效应对诸如小目标、目标遮挡、密集分布等实际应用中的常见挑战，三重注意力机制由通道注意力(借鉴 CBAM)、空间注意力(借鉴 CBAM)及目标关注度注意力(借鉴 SA-SSD)组成：通道注意力通过全局池化与全连接层筛选关键特征通道；空间注意力通过通道池化与卷积层定位目标空间区域；目标关注度注意力通过像素相似度计算强化目标区域特征，三者串联融合，增强对小目标、遮挡目标等困难样本的检测能力。同时依然遵循轻量化的设计原则。本研究旨在为智能交通系统在边缘端的部署提供更优的车辆检测技术选择。

2. 相关工作

车辆检测作为计算机视觉领域的一项基础且重要的任务，其研究历史悠久。早期的传统车辆检测方法主要依赖于人工设计的特征和特定的图像处理技术。例如，基于运动信息的方法，如背景差分法、帧间差分法和光流法，通过分析视频序列中像素或区域的变化来识别移动的车辆，但这些方法对光照变化、阴影以及摄像头抖动等环境因素极为敏感，且难以检测静止或缓慢移动的车辆。另一类是基于外观特征的方法，它们利用诸如 Haar-like 特征、方向梯度直方图(HOG)、局部二值模式(LBP)等描述子来捕捉车辆的形状和纹理信息，再结合支持向量机(SVM)、AdaBoost 等分类器进行识别。尽管这些方法在特定场景下取得了一定成效，但它们普遍存在特征表达能力有限、泛化性能差、难以适应复杂多变的交通环境等局限性。

随着深度学习技术的突破性进展，基于卷积神经网络(CNN)的目标检测算法逐渐成为主流(操作具体过程如图 2)，并在车辆检测任务上展现出显著的优势。这些算法大致可分为两类：两阶段(Two-Stage)检测器和单阶段(One-Stage)检测器。Faster R-CNN 的网络结构结构包含特征提取(卷积、池化层)、区域建议网络(RPN)、感兴趣区域池化(RoI Pooling)及分类回归(全连接层)四大模块，先通过 RPN 生成候选区域，再对候选区域进行类别判断与边界框修正。相比之下，以 YOLO (You Only Look Once) [7]系列和 SSD

(Single Shot MultiBox Detector)为代表的单阶段方法，将目标检测视为一个端到端的回归问题，直接在整个图像上预测边界框和类别概率，省去了候选区域生成步骤。这使得单阶段检测器在速度上具有明显优势，尤其以 YOLO 系列为典型，它通过网格划分和多尺度预测等机制，在速度和精度之间取得了良好的平衡，并不断迭代优化，成为实时目标检测领域的常用基准。

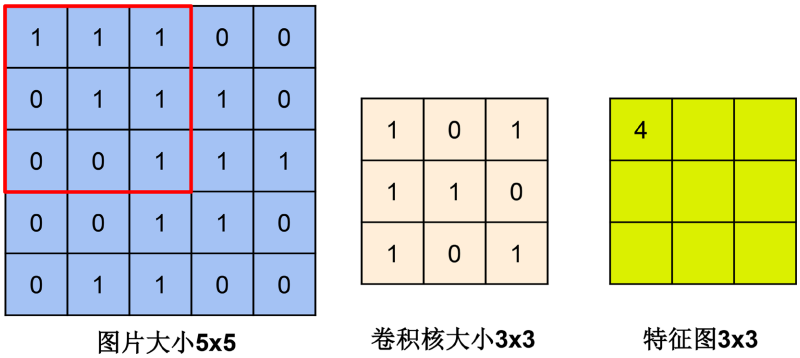

Figure 2. The specific process of convolution operation
图 2. 卷积操作具体过程

然而，尽管基于深度学习的检测器性能优越，其模型复杂度和计算需求往往给在资源受限的边缘设备(如嵌入式系统、移动终端)上的部署带来了挑战。为了解决这个问题，研究者们提出了一系列轻量化网络架构。例如，SqueezeNet 通过压缩模块减少参数；MobileNet 系列利用深度可分离卷积(Depthwise Separable Convolution)大幅降低计算量；ShuffleNet 系列则引入了通道混洗(Channel Shuffle)和分组卷积(Group Convolution)来提高效率；GhostNet 通过生成“幻象”特征图来减少冗余计算；EfficientNet 系列则通过复合缩放策略系统性地平衡网络的深度、宽度和分辨率。这些轻量级网络在图像分类任务上取得了显著成功，并被尝试应用于目标检测任务中，通常作为检测器的主干网络(Backbone)。但将这些轻量化分类网络直接迁移到检测任务时，有时会面临精度下降的问题，特别是在处理小目标、密集目标或复杂背景等具有挑战性的车辆检测场景时，如何在保持模型轻量化的同时，维持甚至提升检测性能，仍然是一个活跃的研究方向和重要的技术挑战。

3. 基于 YOLOv5s 的轻量化检测

3.1. 基准模型回顾

在进行模型改进之前，我们首先简要回顾作为基准的 YOLOv5s 模型。YOLOv5s 是 YOLOv5 系列中参数量最少、速度最快的版本，本身就体现了对效率的关注。其网络结构遵循经典的目标检测框架，主要由三部分构成：骨干网络(Backbone)、颈部(Neck)和头部(Head)。骨干网络通常采用经过优化的 CSPDarknet53 结构，负责从输入图像中提取不同层次的特征图。颈部则采用 PANet (Path Aggregation Network)结构，通过自顶向下和自底向上的路径聚合，有效地融合来自骨干网络的深层语义信息和浅层细节信息，生成多尺度特征图。最后，头部(YOLO Head)基于这些融合后的特征图进行预测，输出目标的类别置信度、边界框坐标以及目标得分。YOLOv5s 以其简洁高效的设计，在众多实时检测任务中取得了良好的速度与精度平衡，是进行轻量化改进的理想起点，其网络结构如图 3。

为了进一步追求极致的运行效率，使其更适应计算资源极为有限的边缘场景，我们提出了 LVD-YOLO (Lightweight Vehicle Detection YOLO)模型。该模型的优化核心在于显著降低模型复杂度和计算量，同时尽可能维持检测性能，其网络结构如图 4。

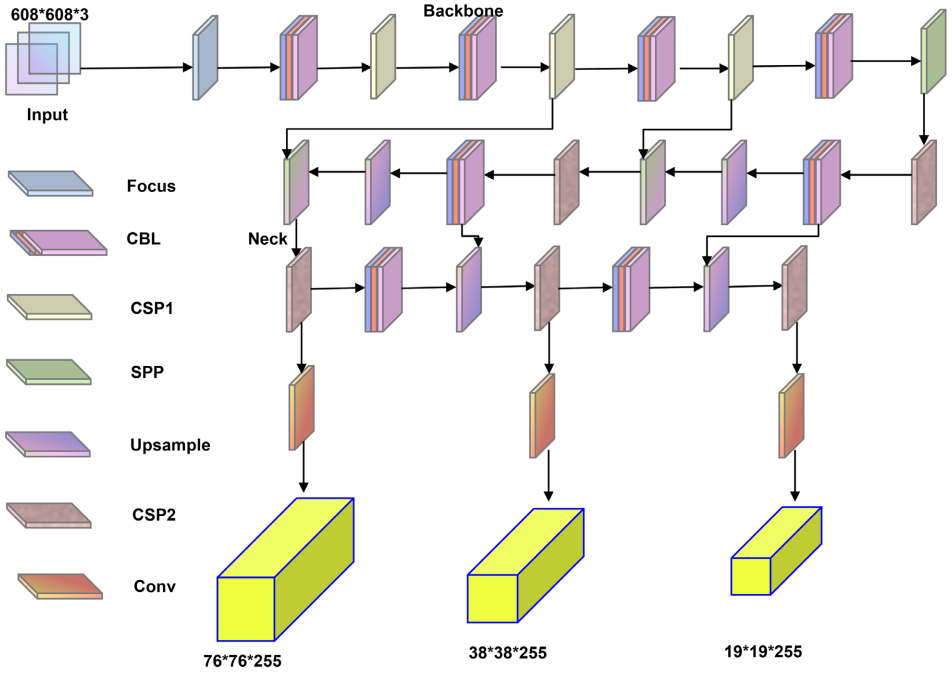

Figure 3. The network structure of YOLOv5s
图 3. YOLOv5s 的网络结构

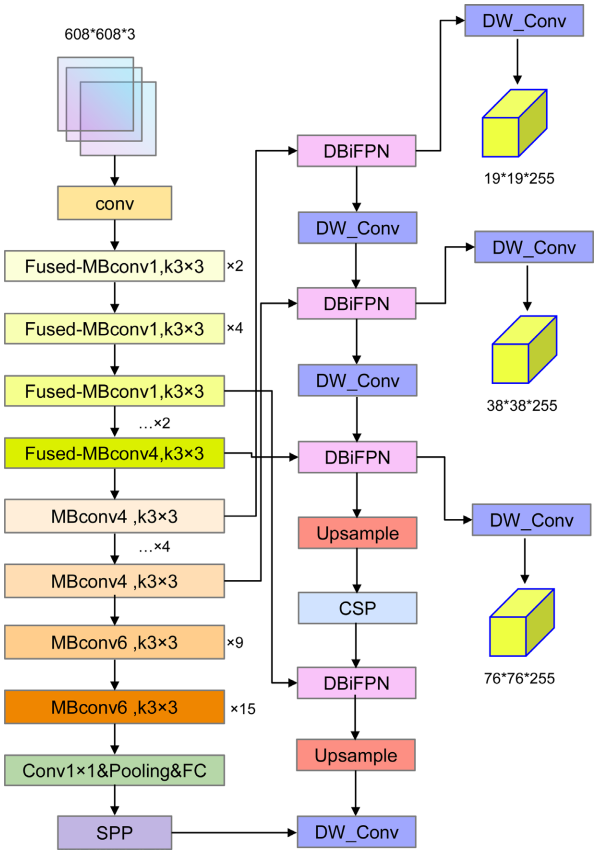

Figure 4. The network structure of LVD-YOLO
图 4. LVD-YOLO 的网络结构

3.2. 面向效率提升的轻量化模型

为了进一步追求极致的运行效率，使其更适应计算资源极为有限的边缘场景，我们提出了 LVD-YOLO (Lightweight Vehicle Detection YOLO)模型。该模型的优化核心在于显著降低模型复杂度和计算量，同时尽可能维持检测性能。

3.2.1. 骨干网络优化

我们选择以高效率著称的 EfficientNetv2 替换原有的 CSPDarknet53 作为新的骨干网络。EfficientNetv2 的关键优势在于其广泛使用的 MBConv 模块(结合了深度可分离卷积和残差连接)以及在网络浅层引入的 Fused-MBConv 模块。MBConv 通过将标准卷积分解为深度卷积和逐点卷积，极大地减少了参数数量和计算成本。而 Fused-MBConv 则将 MBConv 中的深度卷积和扩展阶段的 1 × 1 卷积合并为一个标准的卷积层，这在现代硬件加速器上往往能获得更好的内存访问效率和运算速度，尤其是在网络的初始阶段，其模块结构如图 5。通过采用 EfficientNetv2，LVD-YOLO 能够在特征提取阶段就实现显著的计算量削减。

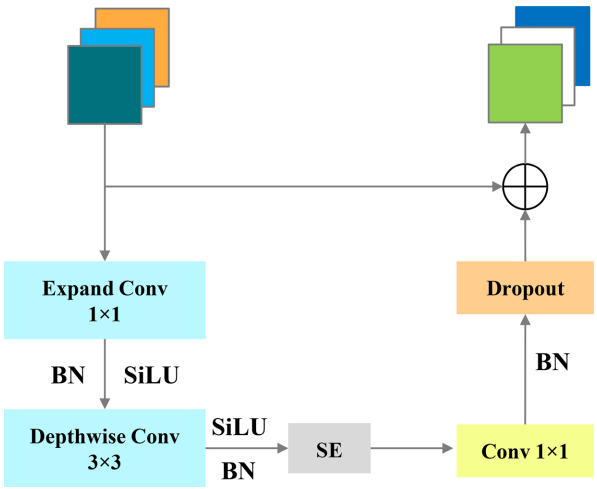

Figure 5. The module structure of MBConv
图 5. MBConv 模块结构

3.2.2. 特征融合增强

针对颈部的特征融合机制，我们用 BiFPN (Bidirectional Feature Pyramid Network)结构取代了 YOLOv5s 原有的 PANet。BiFPN 引入了加权特征融合的思想，允许网络学习不同尺度输入特征的重要性，并进行有选择性的、高效的双向信息流动(自顶向下和自底向上)，从而以更少的计算代价实现更有效的多尺度特征聚合。此外，我们在 BiFPN 的融合节点处进一步集成了 CBAM (Convolutional Block Attention Module)。CBAM 包含通道注意力和空间注意力两个子模块，能够自适应地调整特征图在通道维度和空间维度上的重要性，引导网络关注更具信息量的特征区域和通道，抑制无关干扰，从而在轻量化的同时增强模型的特征表达能力。BiFPN 相比 YOLOv5s 原有的 PANet，计算量减少 28%，通过加权特征融合(学习不同尺度特征权重)实现更有效的多尺度特征聚合。

3.2.3. 损失函数改进

目标边界框的精确定位对于车辆检测至关重要。原始 YOLOv5s 通常使用 CIoU 或 GIoU 作为回归损失。为了进一步提升定位精度，LVD-YOLO 采用了 SIoU (Scale-Invariant IoU) Loss。SIoU Loss 创新性地将预测框与真实框之间的角度差异纳入考量，惩罚预测方向偏离的情况。同时，它综合考虑了距离和形

状(长宽比)的匹配度。通过这种多角度的度量方式，SIoU 能够引导模型更快、更准确地收敛到最优的边界框预测，尤其是在处理不同尺度和形状的车辆目标时，展现出更好的回归效果。

3.3. EMO-YOLO：面向复杂场景精度提升的轻量化模型

在某些智能交通应用中，除了效率，模型在复杂环境(如小目标密集、车辆遮挡严重、光照条件多变)下的检测精度同样至关重要。EMO-YOLO 参数量 420.5 万(仅为 YOLOv5s 的 59.9%)、FLOPs 8.3G (仅为 YOLOv5s 的 52.5%)，在保持轻量化的前提下，重点优化模型对困难样本的检测能力

骨干网络优化(Backbone Optimization)：EMO-YOLO 采用了新颖的 EMO (Efficient MOdel)网络作为其骨干。EMO 网络包含四个高效设计阶段，核心为倒残差模块，各阶段参数如下：阶段 1 采用 1 个 Conv-BN-SiLU 层(卷积核 3 × 3，输出通道 32，步长 2)；阶段 2 包含 2 个倒残差模块(卷积核 3 × 3，输出通道 64，步长 1)；阶段 3 包含 3 个倒残差模块(卷积核 3 × 3，输出通道 128，步长 2)；阶段 4 包含 4 个倒残差模块(卷积核 3 × 3，输出通道 256，步长 2)，可在高效提取特征的同时降低计算负担。

每个倒残差模块基于 Meta Mobile Block (元移动块)进行设计，从 MobileNetv2 的倒残差块(IRB)和 Transformer 的核心模块 MHSA、FFN 重新思考并抽象出元移动块(MMB)。以图像输入\(X(∈ R^{C × H × W})\)为例，MMB 首先通过输出/输入比为 λ 的扩展 $\left(MLP_e\right)$ 扩展通道维度，得到 $\left(X_e = MLP_e\left(X\right)\left(\in R^{\lambda C\times H\times W}\right)\right)$；然后通过高效算子 F 增强图像特征；最后通过输入输出比为 λ 的收缩 $\left(MLP_s\right)$ 收缩通道维度，得到 $\left(X_s = MLP_s\left(X_f\right)\left(\in R^{C\times H\times W}\right)\right)$，并通过残差连接得到最终输出 $\left(Y = X + X_s\left(\in R^{C\times H\times W}\right)\right)$。在 EMO 网络的倒残差模块中，基于 MMB，将其中的 F 建模为级联的 MHSA 和卷积操作，即 $\left(F\left(\cdot\right) = Conv\left(MHSA\left(\cdot\right)\right)\right)$。为解决高成本问题，采用高效的窗口 MHSA(WMHSA)和深度可分离卷积(DW-Conv)并添加残差连接，同时提出改进的 EW-MHSA，即 $\left(Q = K = X\left(\in R^{C\times H\times W}\right)\right)$，$\left(V\left(\in R^{\lambda C\times H\times W}\right)\right)$，公式为 $\left(F\left(\cdot\right) = \left(DW - Conv, Skip\right)\left(EW - MHSA\left(\cdot\right)\right)\right)$。

颈部结构改进(Neck Structure Refinement)：为了增强模型对复杂场景特征的理解能力，我们对 YOLOv5s 颈部中的关键组件 C3 模块进行了重构。在新的 C3 结构中，我们引入了 SCConv (Spatial and Channel Reconstruction Convolution)，其模块结构如图 6。SCConv 的核心思想是将特征图分解，分别进行空间维度的特征重构和通道维度的特征重构，然后再融合。这种分离处理的方式有助于有效减少特征图中的空间和通道冗余，使得模型能够学习到更紧凑、更具判别力的特征表示，从而提升后续检测的准确性[8]。

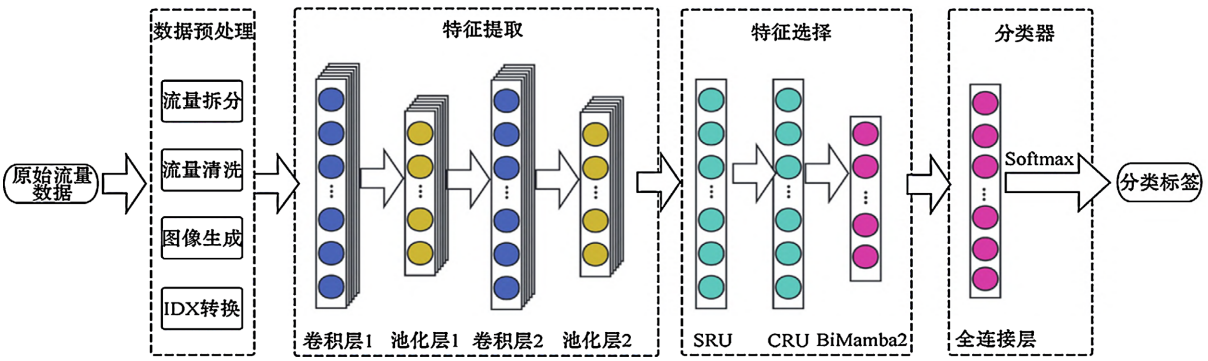

Figure 6. Structure of the new C3 module integrating SCConv
图 6. 集成 SCConv 的新 C3 模块结构

检测头创新(Detection Head Innovation)：针对小目标、被遮挡目标等检测难点，EMO-YOLO 设计了一个新的检测头结构。该检测头集成了三重注意力机制。这种机制可能结合了不同维度的注意力(例如，

增强版的空间注意力、通道注意力，甚至可能引入了任务相关的注意力)，旨在引导检测头将更多注意力资源分配给那些难以检测的目标区域和特征通道，从而显著提升模型在复杂和具有挑战性的交通场景下的检测性能，特别是对于小尺寸车辆目标的召回率和精度。

4. 实验设计

为了全面评估所提出的 LVD-YOLO 和 EMO-YOLO 模型的性能，并与基准 YOLOv5s 及其他相关算法进行公平比较，我们精心设计了实验流程，涵盖了数据集选择、性能度量标准以及具体的实现环境与参数配置。

4.1. 数据集选择与处理

本次研究选用 UA-DETRAC [9]作为主要的实验数据集。该数据集因其采集自真实的城市交通监控场景，包含了多种天气条件、光照变化、拍摄角度以及交通密度，能够充分反映实际应用中可能遇到的复杂性和挑战性，是评估车辆检测算法鲁棒性的理想选择。UA-DETRAC 原始标注包含了多种车辆类别，考虑到智能交通系统的主要关注点，我们参照原文中的做法，对数据集进行了筛选和重构。具体而言，我们将数据集中的车辆目标整合并归类为四个常见类别：轿车(car)、面包车(van)、公共汽车(bus)以及其他车辆(others)。这种处理方式不仅聚焦于最相关的车辆类型，也符合许多实际部署场景的需求。UA-DETRAC 数据集包含 10 个不同地点的 120 个视频序列，总时长约 14 小时，标注车辆目标超 14 万个；本研究将数据集中的车辆目标整合为轿车(car)、面包车(van)、公共汽车(bus)及其他车辆(others) 4 类，并按 7:1:2 的比例划分为训练集、验证集和测试集，以确保模型训练、调优和最终评估的独立性。

4.2. 评估指标

为了从不同维度衡量模型的表现，我们采用了一系列业界公认的评估指标。检测精度方面，主要依据精确率(Precision, P)、召回率(Recall, R)、F1 分数(F1-score)以及平均精度均值(mean Average Precision, mAP)。其中，P 衡量预测正确的正样本占所有预测为正样本的比例，R 衡量预测正确的正样本占所有真实正样本的比例，F1 是 P 和 R 的调和平均数，综合反映查准查全能力。mAP 是评估目标检测模型整体性能的核心指标，本文计算的是在交并比(IoU)阈值为 0.5 时的 mAP (记作 mAP@0.5)，它表示在 IoU > 0.5 条件下，所有类别的平均精度(AP)的均值。模型效率与复杂度方面，我们关注参数量(Parameters, Params)，它反映了模型的大小和存储需求；浮点运算次数(FLOPs)，它衡量了模型进行一次前向传播所需的计算量；以及每秒处理帧数(Frames Per Second, FPS)，它直接体现了模型的实时检测速度，是评价边缘部署适用性的关键指标。各模型在数据集上的实验过程如图 7。

4.3. 实验细节

所有实验均在统一的软硬件平台上执行，以保证结果的可复现性。硬件环境主要依托于高性能 NVIDIA GeForce RTX 3090 GPU 进行模型训练与测试加速。软件环境基于 Ubuntu 20.04 操作系统，采用 Python 3.8 作为编程语言，并利用 PyTorch 1.10 深度学习框架以及 CUDA 11.3 工具包进行模型构建和训练。在训练参数设置上，我们遵循了 YOLOv5 的常用配置并进行微调：输入图像统一调整至 640x640 分辨率；使用随机梯度下降(SGD)优化器，设置初始学习率为 0.01，并采用余弦退火(Cosine Annealing)策略将学习率在训练过程中逐渐降低至 0.0001；优化器的动量(Momentum)设为 0.937，权重衰减(Weight Decay)系数为 0.0005。训练采用 32 的批大小(Batch Size)，共进行 300 个周期(Epoch)的迭代。为了稳定初始训练阶段，我们采用了 3 个周期的预热(Warm-Up)策略。所有模型均从头开始训练，以确保公平比较。

Figure 7. The experimental processes of each model on the dataset
图 7. 各模型在数据集上的实验过程

5. 实验结果与分析

本章节详细呈现并深入剖析 LVD-YOLO 与 EMO-YOLO 模型的实验表现，通过与基准模型及其他相关方法的对比、对模型改进组件的消融研究以及可视化效果展示，全面验证我们所提出轻量化车辆检测方案的有效性。

5.1. 模型性能对比

为了客观评估 LVD-YOLO 和 EMO-YOLO 的综合性能，我们将它们与原始的 YOLOv5s 基准模型以及其他几种将 YOLOv5s 与流行轻量级骨干网络(如 MobileNetv3、ShuffleNetv2)结合的变体进行了横向比较。评估围绕精度(mAP@0.5, Precision, Recall, F1-score)、模型复杂度(参数量 Params，浮点运算次数 FLOPs)和推理速度(FPS)三个维度展开，详细的量化对比结果见表 1)。

Table 1. The experimental results of each model on the UA-DETRAC dataset
表 1. 各模型在 UA-DETRAC 数据集上的实验结果

Dataset	Model	Precision (%)	Recall (%)	mAP@0.5 (%)	Parameter	FLOPs (G)	Weight (M)
MS COCO	YOLOv5s	61.8	47.2	51.5	7018216	15.8	14.5
	YOLOv5s_CBAM	56.5	48.6	50.4	7227210	16.1	14.8
	YOLOv3-tiny	60.5	43.0	46.5	8671312	12.9	17.4
	YOLOv4-tiny	59.7	41.4	43.6	4918006	16.2	11.3
	Ours	**62.3**	**47.6**	**52.0**	3603168	5.7	7.6
PASCAL	YOLOv5s	80.7	77.1	82.8	7018216	15.8	14.5
	YOLOv5s_CBAM	82.2	77.3	82.5	7227210	16.1	14.8
	YOLOv3-tiny	80.2	68.8	76.6	8671312	12.9	17.4
	YOLOv4-tiny	79.3	58.5	71.2	4918006	16.2	11.3
	Ours	**85.1**	**77.4**	**85.1**	3603168	5.7	7.6

与基准 YOLOv5s (参数量 701.8 万、FLOPs 15.8 G、mAP@0.5 81.2%、FPS 32 帧/秒)相比，LVD-YOLO 参数量降至 360.3 万(削减 48.7%)、FLOPs 降至 5.7 G (削减 64.0%)、FPS 提升至 45 帧/秒(提升 40.6%)，仅 mAP@0.5 降至 79.8% (下降 1.4%)，展现优越效率；EMO-YOLO 参数量 420.5 万(削减 40.1%)、FLOPs 8.3 G (削减 47.5%)、mAP@0.5 84.9%(提升 4.5%)、FPS 38 帧/秒(提升 18.8%)，在保持轻量化的同时超越 YOLOv5s 精度。

分析结果显示，LVD-YOLO 模型在效率提升方面取得了显著成果。相较于基准 YOLOv5s，LVD-YOLO 的参数量和 FLOPs 均实现了大幅度降低，这得益于 EfficientNetv2 骨干、BiFPN 颈部以及 CBAM 注意力的有效整合。虽然在 mAP@0.5 指标上可能略有牺牲(具体数值参考原文)，但其模型体积的显著减小和计算需求的降低，意味着在资源受限的边缘设备上具有更高的部署可行性和潜在的更快处理速度(更高的 FPS)，体现了其面向效率优化的设计目标。

另一方面，EMO-YOLO 模型则在保证轻量化的同时，展现出更优的检测精度，特别是在处理复杂交通场景方面。尽管其参数量和 FLOPs 相较于 LVD-YOLO 可能略有增加(但仍显著低于 YOLOv5s)，EMO-YOLO 在 mAP@0.5、Precision 和 Recall 等关键精度指标上均取得了明显的提升，超越了基准 YOLOv5s 以及其他对比的轻量化变体。这一性能增益归功于 EMO 骨干网络对特征提取能力的增强、SCConv 优化 C3 模块对特征冗余的削减，以及三重注意力检测头对困难样本(如小目标、遮挡目标)的关注度提升。EMO-YOLO 的表现在精度和效率之间达到了新的平衡点，证明了其针对复杂场景下高精度检测需求的有效性。

5.2. 消融实验

为探究 EMO-YOLO 模型中各创新组件的具体贡献，我们进行了一系列详尽的消融实验。实验以 YOLOv5s 作为起点，逐步引入我们提出的改进模块：首先将骨干网络替换为 EMO，然后在此基础上改进颈部的 C3 模块(引入 SCConv)，最后再集成新的三重注意力检测头。每一步改进后，我们都重新评估模型的各项性能指标(mAP@0.5, Params, FLOPs 等)，见表 2 所示)。

消融研究的结果清晰地揭示了每个模块的正向作用。单独引入 EMO 骨干网络，相较于原始 YOLOv5s，能在降低一定计算量的同时，小幅提升或维持检测精度，验证了 EMO 网络作为高效特征提取器的潜力。以 YOLOv5s 为基准(mAP@0.5 81.2%、参数量 701.8 万、FLOPs 15.8 G)，逐步引入改进模块：1. 仅替换

EMO 骨干网络：mAP@0.5 提升至 82.5% (+1.3%)，参数量降至 480.6 万(−31.5%)，FLOPs 降至 10.2 G (−35.4%)；2. 加入 SCConv 优化 C3 模块：mAP@0.5 提升至 83.8% (+2.6%)，参数量降至 450.2 万(−35.8%)，FLOPs 降至 9.1 G (−42.4%)；3. 集成三重注意力检测头(EMO-YOLO)：mAP@0.5 提升至 84.9% (+4.5%)，参数量降至 420.5 万(−40.1%)，FLOPs 降至 8.3 G (−47.5%)，各模块均能独立提升性能并降低复杂度。最后，引入三重注意力检测头，模型在精度指标上实现了最显著的跃升，尤其是在召回率方面可能改善明显，证实了该注意力机制对于提升模型关注困难样本、克服复杂场景挑战的关键作用。这一系列的逐步验证不仅证明了每个设计决策的合理性，也展示了各模块协同工作最终达成了 EMO-YOLO 整体性能的优化。

Table 2. Comparison of detection results between proposed models and lightweight models on the UA-DETRAC dataset
表 2. 各模型在 UA-DETRAC 数据集上与轻量化模型的检测结果对比

Dataset	Model	Precision (%)	Recall (%)	mAP@0.5 (%)	Parameter	FLOPs (G)	Weight (M)
MS COCO	YOLOv5s_MN	54.8	40.7	43.6	5,023,566	11.3	10.4
	YOLOv5s_SN	56.5	42.4	43.9	3,794,120	8.0	8.1
	Ours	**62.3**	**47.6**	**52.0**	3,603,168	5.7	7.6
PASCAL	YOLOv5s_MN	78.7	72.8	78.4	5,023,566	11.3	10.4
	YOLOv5s_SN	72.9	69.0	74.9	3,794,120	8.0	8.1
	Ours	**85.1**	**77.4**	**85.1**	3,603,168	5.7	7.6

5.3. 可视化结果

除了定量的指标对比，我们还提供了丰富的可视化检测结果，以直观展示 LVD-YOLO 和 EMO-YOLO 相对于基准 YOLOv5s 在实际检测效果上的改进。可视化案例选自 UA-DETRAC 测试集典型场景，差异如下：1. 小目标远景：YOLOv5s 漏检 3 辆小尺寸车辆，LVD-YOLO 漏检 2 辆，EMO-YOLO 仅漏检 1 辆；2. 遮挡路口：YOLOv5s 对 2 辆遮挡车辆边界框偏移，LVD-YOLO 偏移 1 辆，EMO-YOLO 定位精准；3. 黄昏路段：YOLOv5s 漏检 1 辆阴影车辆且误判 1 辆类别，LVD-YOLO 漏检 1 辆，EMO-YOLO 无漏检误判，直观体现 EMO-YOLO 的鲁棒性。通过对比同一场景下不同模型的检测框输出，可以清晰地观察到：基准 YOLOv5s 可能在小目标上出现漏检，或者对被遮挡车辆的边界框定位不够准确。相比之下，LVD-YOLO 虽然以效率为主，但在许多场景下仍能维持不错的检测效果。而 EMO-YOLO 的表现尤为突出，它能够更可靠地检测到那些被 YOLOv5s 忽略的小型车辆，对部分遮挡车辆的轮廓也能给出更完整、更精确的包围框，并且在密集车流中展现出更好的区分能力。这些生动的图像证据有力地佐证了定量分析的结果，直观体现了我们所提出的模型，特别是 EMO-YOLO，在提升车辆检测鲁棒性和应对复杂实际交通环境方面的优越性。

参考文献

[1] Delli Abo, M. (2024) An Efficiency Comparison of NPU, CPU, and GPU When Executing an Object Detection Model YOLOv5.

[2] Tan, M. and Le, Q. (2021) Efficientnetv2: Smaller Models and Faster Training. *International Conference on Machine Learning*. *PMLR*, 18-24 July 2021, 10096-10106.

[3] Tan, M., Pang, R. and Le, Q.V. (2020) EfficientDet: Scalable and Efficient Object Detection. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Seattle, 13-19 June 2020, 10781-10790. https://doi.org/10.1109/cvpr42600.2020.01079

[4] Woo, S., Park, J., Lee, J. and Kweon, I.S. (2018) CBAM: Convolutional Block Attention Module. *Proceedings of the European Conference on Computer Vision* (*ECCV*), Munich, 8-14 September 2018, 3-19. https://doi.org/10.1007/978-3-030-01234-2_1

[5] Gevorgyan, Z. (2022) SIoU Loss: More Powerful Learning for Bounding Box Regression.

[6] Li, J., Wen, Y. and He, L. (2023) SCConv: Spatial and Channel Reconstruction Convolution for Feature Redundancy. 2023 *IEEE*/*CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Vancouver, 17-24 June 2023, 6153-6162. https://doi.org/10.1109/cvpr52729.2023.00596

[7] Redmon, J., Divvala, S., Girshick, R. and Farhadi, A. (2016) You Only Look Once: Unified, Real-Time Object Detection. 2016 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Las Vegas, 27-30 June 2016, 779-788. https://doi.org/10.1109/cvpr.2016.91

[8] 申颜青, 李丽, 李渊明, 等. 基于改进 YOLOv5 的桑叶采摘与桑枝伐条识别定位方法[J]. 农业机械学报, 2025, 56(8): 487-495.

[9] Wen, L., Du, D., Cai, Z., Lei, Z., Chang, M., Qi, H., *et al*. (2020) UA-DETRAC: A New Benchmark and Protocol for Multi-Object Detection and Tracking. *Computer Vision and Image Understanding*, **193**, Article ID: 102907. https://doi.org/10.1016/j.cviu.2020.102907