

基于改进 YOLOv8 的夜间辅助驾驶车辆检测^①



吴湘宁, 潘志鹏, 王梦雪, 涂 雨

(中国地质大学(武汉)计算机学院, 武汉 430078)
通信作者: 涂 雨, E-mail: 912980255@qq.com

摘 要: 夜间辅助驾驶因光线不足导致图像质量下降, 使得车辆检测面临低可见度、对比度降低以及噪声增多等问题. 针对上述问题, 提出了一种改进模型 YOLOv8n-STH (you only look once version 8 nano-SPDConv-triplet attention-HS-FPN), 模型以 YOLOv8n 模型为基础, 针对夜间图像特征提取难、目标小等问题, 在主干网络的部分 C2f (faster implementation of CSP bottleneck with 2 convolutions) 前加入 SPD (space to depth) 模块, 并将部分卷积池化层替换为 SPDConv. 同时, 在 C2f 结构中引入了轻量的三重注意力机制, 使其更加准确地区分目标与背景. 最后, 使用了多尺度选择特征融合模块, 使模型能够高效地筛选出更有效的特征信息. 在两个数据集上的实验结果表明, YOLOv8n-STH 相比于 YOLOv8n 模型, 精确率提升了 2.1%, 模型大小减少了 24.2%, 能部署在资源有限的环境中.
关键词: 夜间辅助驾驶; 车辆检测; YOLOv8; 注意力机制; 特征融合

引用格式: 吴湘宁, 潘志鹏, 王梦雪, 涂雨. 基于改进 YOLOv8 的夜间辅助驾驶车辆检测. 计算机系统应用, 2025, 34(12): 249-259. <http://www.c-s-a.org.cn/1003-3254/10011.html>

Improved-YOLOv8-based Vehicle Detection for Nighttime Assisted Driving

WU Xiang-Ning, PAN Zhi-Peng, WANG Meng-Xue, TU Yu

(School of Computer Science, China University of Geosciences, Wuhan 430078, China)

Abstract: Nighttime assisted driving suffers from image quality degradation due to insufficient light, which causes vehicle detection to face problems such as low visibility, reduced contrast, and increased noise. To address these issues, an improved model, you only look once version 8 nano-SPDConv-triplet attention-HS-FPN (YOLOv8n-STH), is proposed. The model is based on the YOLOv8n and targets problems such as difficult feature extraction and small targets in nighttime images. The space to depth (SPD) module is added in front of part of the faster implementation of CSP bottleneck with 2 convolutions (C2f) structure in the backbone network, and part of the convolutional pooling layers is replaced with SPDConv. Meanwhile, a lightweight triplet attention mechanism is introduced into the C2f structure to enable more accurate distinction between targets and background. Finally, a multi-scale selection feature fusion module is employed to allow the model to efficiently filter more effective feature information. Experimental results on two datasets show that YOLOv8n-STH improves the accuracy rate by 2.1% and reduces the model size by 24.2% compared to the YOLOv8n model, making it suitable for deployment in resource-limited environments.

Key words: nighttime assisted driving; vehicle detection; YOLOv8; attention mechanism; feature fusion

1 引言

在夜间辅助驾驶过程中, 由于光线条件的限制,

辅助驾驶算法常难以准确地辨认前方车辆, 尤其是小型或低亮度的目标, 增加了辅助驾驶的不确定性和风

^① 基金项目: 国家自然科学基金 (U21A2013); 湖北省自然科学基金 (2021CFB506)

收稿时间: 2025-04-29; 修改时间: 2025-05-26; 采用时间: 2025-06-09; csa 在线出版时间: 2025-10-31

CNKI 网络首发时间: 2025-11-03

险. 对于目前大多数的车辆检测识别算法而言, 夜间情况下的图像可见度以及对对比度都较低, 并且在采集图像的过程中很可能出现运动模糊现象, 使得图像整体质量较低, 而距离较远的小目标车辆, 更是增加了检测难度.

车辆目标检测经过多年的发展, 产生了多种检测方式, 主要分为基于运动信息的车辆检测方法^[1]、基于车辆特征的检测方法^[2]和基于深度学习的检测方法.

基于运动信息的车辆检测方法利用车辆在视频序列中的运动特征进行实时检测和跟踪. 通过运动估计、目标提取、目标跟踪和目标检测等步骤, 可以从视频中提取出具有明显运动的车辆目标, 并跟踪它们的运动轨迹. 常见的检测方法有: 帧间差分法^[3]、背景差分法^[4]和光流法^[5]. 舒兆翰等人^[6]提出一种将两者融合的静态背景下运动车辆检测算法, 该融合算法相比于其他传统方法取得了更好的效果.

基于车辆特征的检测方法主要依赖于纹理、颜色、边缘等属性, 在日间环境下效果显著. 然而, 在夜间, 由于光照条件差, 这些特征不如车辆的前照灯和尾灯明显, 因此许多研究转而利用车灯特征进行检测. Tsai 等人^[7]提出了一种基于光衰减分析的夜间车辆检测算法, 通过自动阈值技术提取车灯及其光衰区域, 并结合车灯运动轨迹进行配对优化. Nakane 等人^[8]则利用概率位遗传算法检测尾灯, 结合颜色和几何模型提升准确性. 尽管这些基于车灯的方法在特定场景下有效, 但在城市复杂光照条件下易受干扰.

基于深度学习的检测方法是通过对卷积神经网络 CNN (convolutional neural network)^[9]提取车辆特征信息, 实现对车辆的检测与识别. 目前基于深度学习的检测算法主要分为双阶段检测和单阶段检测算法. 双阶段检测算法, 如 Faster R-CNN^[10]系列和 Mask R-CNN^[11], 通过区域提取和分类两个阶段实现高精度检测. 单阶段检测算法则是通过神经网络直接预测图像中的目标类别和位置信息, 包括 SSD (single shot multibox detector)^[12]、YOLO (you only look once) 系列^[13,14]等. 双阶段检测算法检测精度较高, 但检测速度较低, 单阶段检测算法检测速度快, 但检测精度低^[15].

目前的车辆目标检测算法主要针对白天环境设计, 而夜间检测面临诸多挑战: 光线不足导致图像质量下降, 目标与背景的区别变得困难; 高速行驶中的车辆易

产生运动模糊, 尤其在宽阔道路前方的小目标车辆. 此外, 高精度检测算法通常需要较高的计算资源, 在计算能力受限的车载系统中难以实现. 加之夜间车辆检测数据集的稀缺, 进一步限制了模型泛化能力的培养和对不同夜间环境的适应性.

为此, 提出了基于夜间道路的车辆目标检测模型 YOLOv8n-SPDConv-TA-HSFPN, 简称为 YOLOv8n-STH. 算法在 3 个方面做出了改进: 首先, 在 YOLOv8n 中引入 SPDConv (space to depth convolution) 结构^[16], 以保留更多的图像细节信息, 并使用其中的 SPD 结构来减少后续计算量; 其次, 在 C2f 结构中加入 TA (triplet attention) 注意力机制^[17], 以便更加准确地区分目标与背景; 最后, 将原来的特征融合模块替换为高层筛选-特征融合金字塔 (high-level screening-feature fusion pyramid network)^[18], 在该结构中引入通道注意力 (channel attention, CA) 机制来进行特征筛选, 并引入选择性特征融合 (selective feature fusion, SFF) 模块进行特征融合, 使模型能够筛选出更有效的特征信息, 在降低模型参数量的同时提高了模型的特征表达能力. 多轮消融实验和对比实验的结果表明, YOLOv8n-STH 易于部署, 且检测速度较快, 能满足驾驶场景下的实时要求, 检测精度也有所提升.

2 YOLOv8 算法

YOLOv8 由 Ultralytics 公司在 2023 年 1 月发布, 共有 5 种大小: n、s、m、l、x, 经过不断改进和优化, 检测速度和准确率都更加优秀. YOLOv8 模型如图 1 所示, 与早期 YOLO 版本相比, YOLOv8 作了以下改进.

(1) YOLOv8 向主干网络中添加 C2f 结构, 丰富了模型的梯度流, 让特征信息更丰富的同时, 减少了计算量. 在颈部网络上, 将 C3 模块替换为 C2f 模块. 增强了网络对不同尺寸和形状目标的适应性, 同时保持了运算速度的提升. (2) 早期的 YOLO 网络采用的都是耦合检测头, 而 YOLOv8 采用了解耦检测头, 分类和定位分别由两个独立的网络结构完成, 虽然解耦检测头加大了一点计算成本, 但是模型的收敛速度以及检测精度都有所提升. (3) YOLOv8 采用无锚点 (anchor free) 检测方式, 能够更自然地适应不同尺寸和形状的目标, 同时减少了模型训练和推理过程中的复杂度和计算负担.

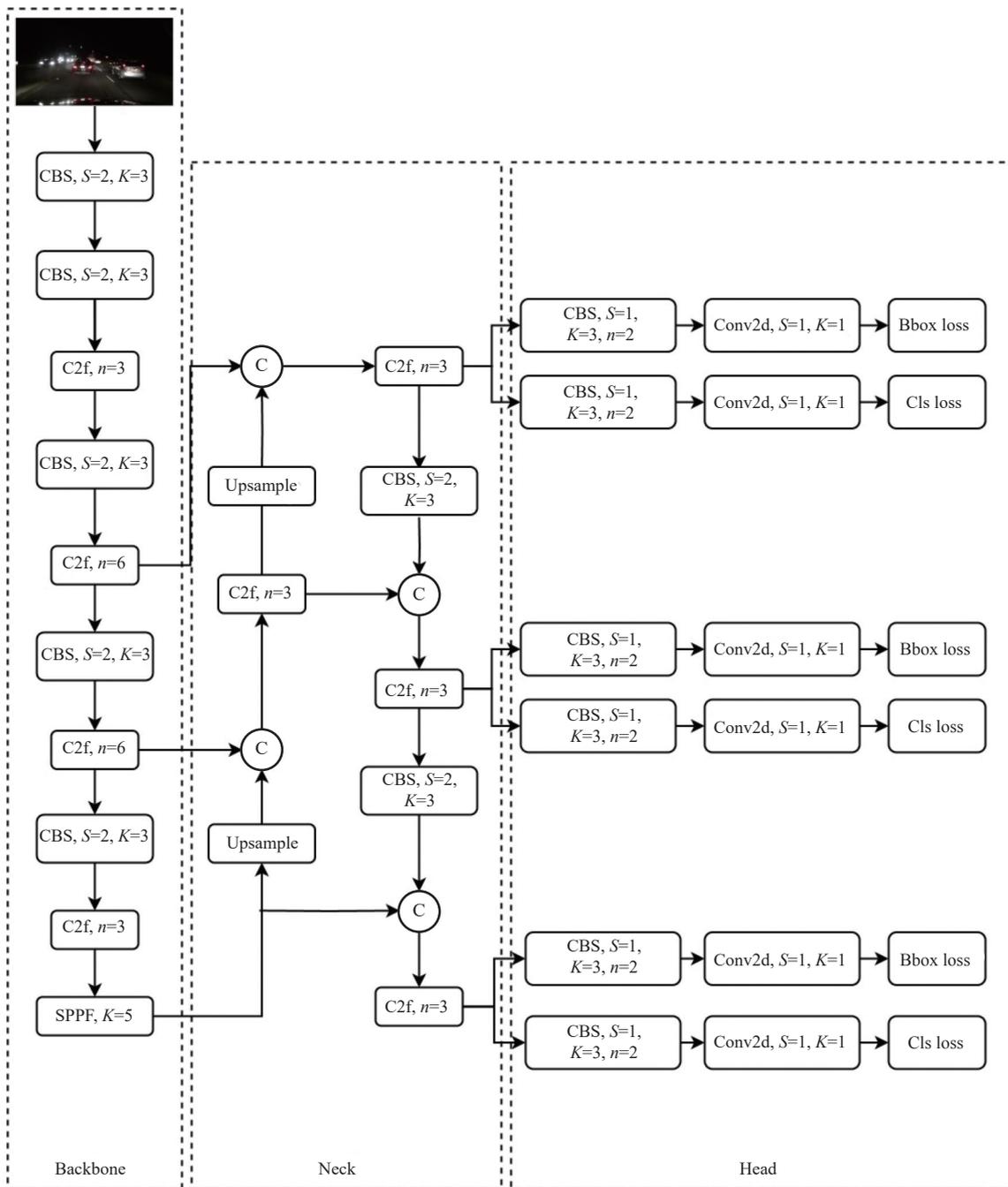


图1 YOLOv8 模型结构图

3 YOLOv8n-STH 模型

3.1 网络总体架构

YOLOv8n-STH 结构如图 2 所示, 包含用于特征提取的主干网络 (Backbone), 用于特征融合的颈部网络 (Neck) 以及执行最后分类和回归任务的检测头 (Head). YOLOv8n-STH 首先向 Backbone 中的部分 C2f 前加入一个空间到深度 (SPD) 层, 将空间信息转换为深度信

息, 以解决夜间图像特征难以提取、存在小目标等问题, 并将部分卷积池化层替换为空间到深度卷积 (SPD-Conv) 层, 在图像质量不高的情况下有效保留了图像的细节信息; 同时在 C2f 结构中引入了轻量的三重注意力 (TA) 机制, 使模型更加聚焦于目标区域, 减少对无用背景的关注, 从而更好地将车辆从背景中检测出来. 在 Neck 中, YOLOv8n-STH 引入高层筛选-特征融合金

字塔 (HS-FPN), 在降低模型参数量的同时提高模型的特征表达能力. HS-FPN 分为特征选择和特征融合两个部分, 在特征选择部分, 引入通道注意力机制 (CA) 使

模型能够筛选出更有效的特征信息; 在特征融合部分, 引入选择性特征融合模块 (SFF) 使模型更精确地聚焦于对当前任务最有用的信息.

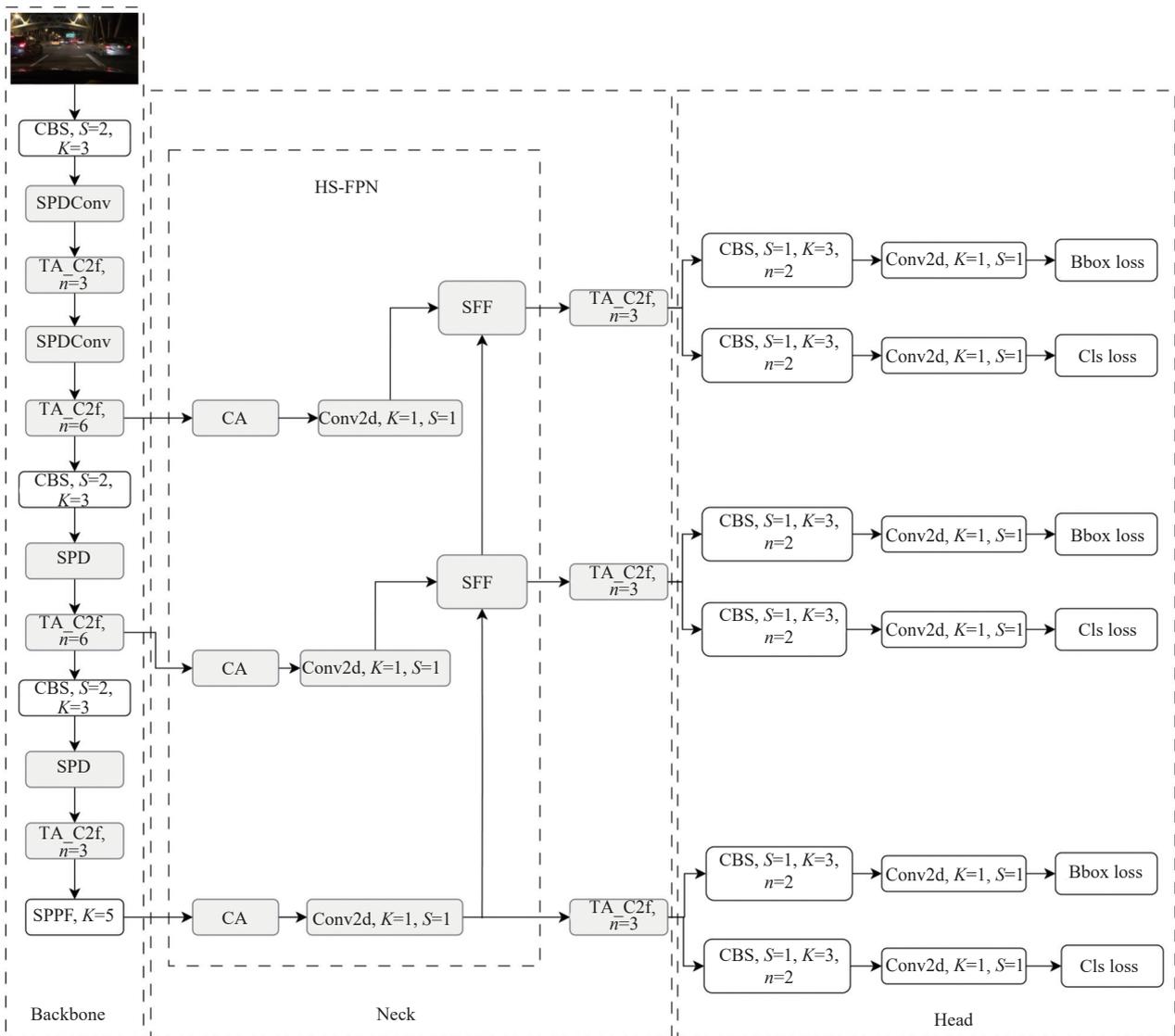


图2 YOLOv8n-STH 模型结构示意图

3.2 SPDCConv 模块

夜晚的光照条件导致背景与前景的对比度不显著, 信息不明显、不充足, 特别是当存在远处较小的车辆时. 为了解决这一问题, 引入了一种不同的卷积网络构件, 即 SPDCConv. 通过替换传统的步长卷积和池化层, SPDCConv 能够保留更多的图像信息, 从而提升模型在学习特征表示方面的能力.

SPDCConv 包含一个将空间维度映射至深度维度的 SPD 层, 以及无步长卷积 (Conv) 层. 在 SPD 层中,

输入特征图的空间尺寸被有效压缩至通道尺寸, 在此过程中确保了通道信息的完整性. 该层将输入的每个像素或特征分配至不同的通道中, 导致空间尺寸的减少和通道数的相应增加. 接着, Conv 层执行标准卷积操作, 与具有步长的卷积不同, 这一层对输入特征图的每个元素进行处理, 无需跳过任何像素. 此设计旨在缓解 SPD 层可能引入的信息损失, 确保更丰富的细节被保留. 通过这种方式, SPDCConv 增强了模型对于夜晚低光条件下保留细节特征以及识别远距离小目

标的能力。

图3为SPDConv的结构图,其中特征图 X 的大小为 (S, S, C_1) ,首先,对 X 执行子特征映射的切片操作,即按照一个比例因子 $scale$ 对 X 进行下采样操作,当 $scale=2$ 时,可以得到4个大小为 $(S/2, S/2, C_1)$ 的子映射 $f_{0,0}$ 、 $f_{0,1}$ 、 $f_{1,0}$ 以及 $f_{1,1}$ 。随后,这些子特征映射沿着通道维度进行拼接,形成了新的特征映射 X' , X' 的大小为 $(S/2, S/2, 4C_1)$ 。将 $scale$ 推广至其他值,即特征图经SPD操作后尺寸会由 (S, S, C_1) 变为 $(S/scale, S/scale, scale^2 \times C_1)$ 。在进行SPD特征转换之后,引入了一个步长为1的卷积层。步长为1意味着会处理输入特征映射的每一个像素点,这种细粒度的处理方式有助于保留更多的图像细节和纹理信息,最大程度地保持图像中的特征信息不被丢弃。通常情况下,超过1的步长虽然可以完成尺寸上的转换,但损失了大量的重要特征信息。

在夜间情况下,为了保留更多的图像信息,使用SPDConv来替换YOLOv8n中Backbone的第1、3层。为了减少后续操作的计算量,在第6、8层的C2f前加入SPD结构,将 W 、 H 的信息集中到通道上,图像尺寸减半,使后续处理的特征图大小减少。

3.3 TA_C2f 模块

在夜间场景中,由于光照条件不稳定以及摄像头传感器的限制,图像中常存在大量的噪声和干扰。并且由于光线昏暗,目标存在低像素,尺寸小,难以与背景区分开等问题。因此,引入了三重注意力(TA)机制。在原有的C2f结构的Bottleneck部分中,加入TA机制,从而形成新的TA_Bottleneck。将新的TA_Bottleneck加入C2f结构中,改进后的结构命名为TA_C2f,其结构如图4所示。这种改进在Backbone中使C2f更加专注于车辆的轮廓和细节特征,同时忽略无关的背景信息。在Neck中也加入TA_C2f模块,进一步优化了特征融合过程,使得模型能够更准确地识别和定位车辆目标。

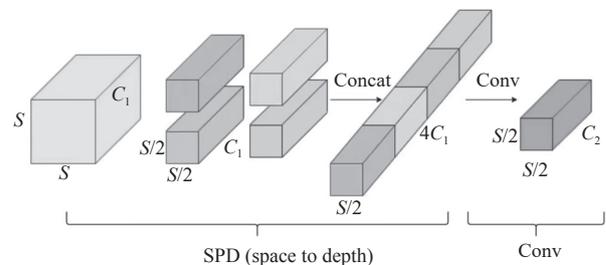


图3 SPDConv 结构图

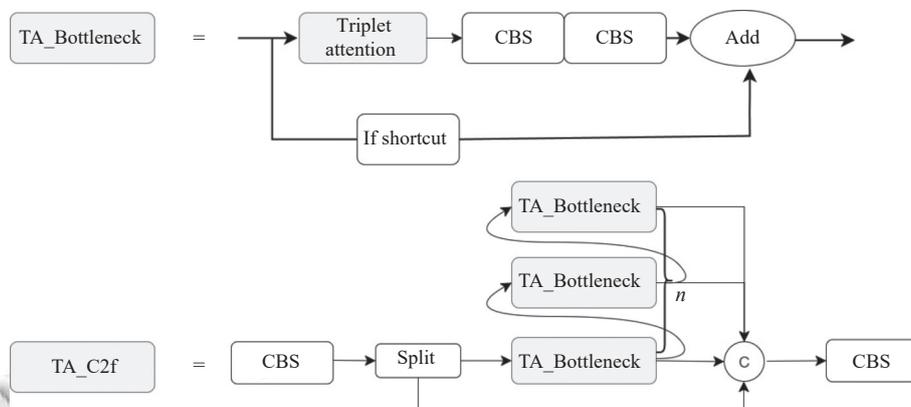


图4 TA_C2f 结构图

TA机制的核心优势在于其能够捕获输入数据的跨维度交互,构建通道与空间位置之间的相互依赖性。这与传统的基于通道和空间的注意力机制(如CBAM^[19])不同,后者的通道注意和空间注意是分离且独立计算的,缺乏有效的交互。相比之下,三重注意力通过3个并行的分支同时关注特征图的通道信息和空间位置信息,且计算成本较低。TA机制结构如图5所示。

通过引入TA机制,模型在特征提取过程中能够更精确地聚焦于关键信息,同时忽略那些无关的背景信息,从而在增加极少参数的情况下,保持了模型的计算效率,并显著提升检测精度。

3.4 HS-FPN 模块

夜间行车时,光线昏暗导致能见度较低,因此采集到的图像比较模糊,有着较高的噪声水平。如果此时直接融合所有层次的特征,会导致大量噪声和无关信息

的混入,降低检测性能.因此需要找到从这些低质量图像中智能筛选和融合重要特征的方法,从而提升检测的准确性.故引入一种高层筛选-特征融合金字塔来实

现多尺度特征的筛选及融合,HS-FPN分为两部分:特征选择(feature selection)和特征融合(select feature fusion),其结构如图6所示.

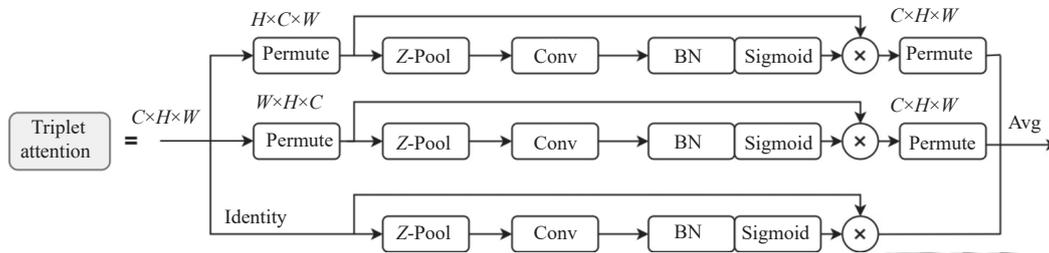


图5 TA机制结构图

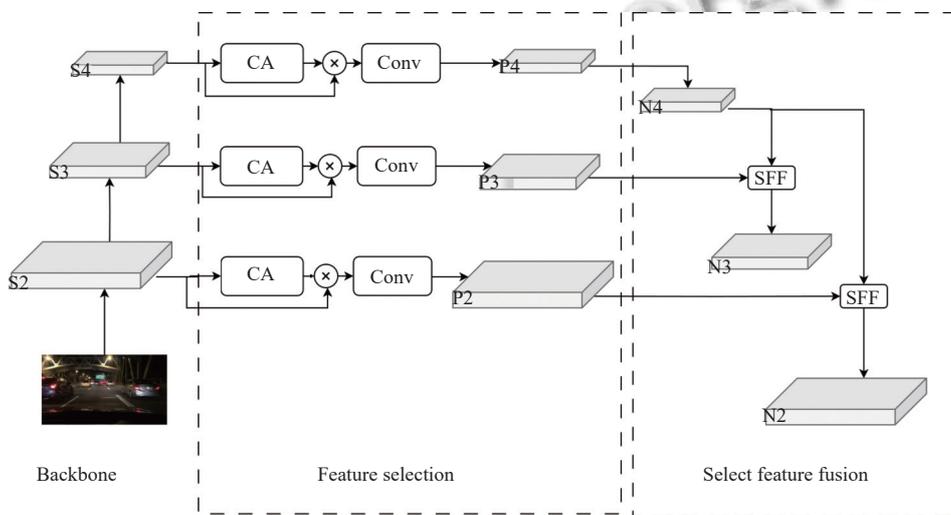


图6 HS-FPN结构示意图

(1)在特征选择阶段,引入通道注意力(CA)机制(如图7所示)以增强特征筛选效率.CA机制通过自适应平均池化(adaptive avg pool)和最大池化(adaptive max pool)聚合特征图的空间信息.平均池化提取通道的全局信息,帮助模型把握整体结构;最大池化则突出显著

特征,如车辆边缘和车灯,使模型聚焦于关键区分性特征.接着,通过卷积和激活函数生成的注意力权重强调重要区域,并抑制次要信息.最后为了统一不同尺度特征图的通道数,应用 1×1 卷积将特征图通道数调整至256,优化特征表示,为之后的特征融合做铺垫.

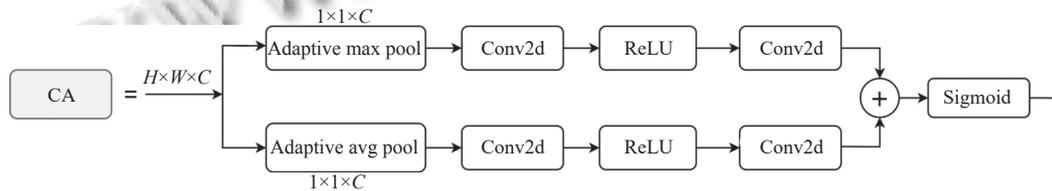


图7 通道注意力机制结构示意图

(2)在卷积神经网络中,深层特征趋向抽象化,捕捉关键对象和全局场景,而浅层特征则富含边缘、纹理等细节.特征融合结合了这些层次的信息,使模型能综合细节与抽象概念,深化对图像的理解,提升识别性能.传统融合方法常通过简单相加或拼接实现,而在夜

间模式下,低层特征可能会受到更多的噪声干扰,为了更精确地聚焦于对当前任务最有用的信息,采用选择性特征融合模块(SFF)来使用高层特征作为权重来筛选低层特征中的重要语义信息.SFF的具体结构如图8所示.

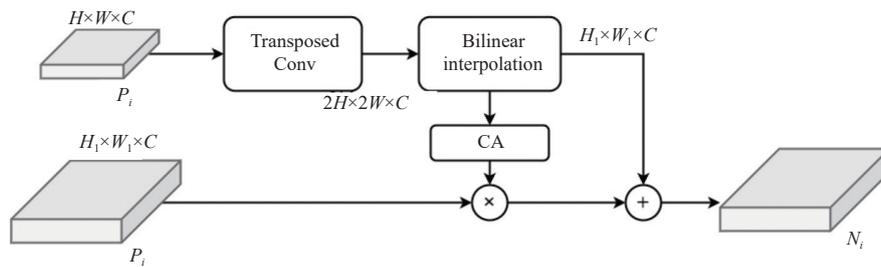


图8 SFF 结构示意图

SFF 通过转置卷积 (transposed convolution) 和双线性插值 (bilinear interpolation) 结合的方式来完成特征图大小的匹配, 双线性插值法简单快速, 但是无法应对复杂的场景, 而转置卷积具有动态学习能力, 并且可以通过调整步长和填充参数来控制输出特征图的尺寸. 转置卷积操作如图 9 所示, 通过“零填充+卷积”实现特征图尺寸放大. SFF 模块首先对输入特征图进行零填充以增加尺寸, 随后应用 3×3 卷积核、步长为 2 的转置卷积进行 2 倍上采样, 将特征图尺寸扩展. 接着, 利用双线性插值对高层特征图进行尺寸调整, 确保与底层特征图一致. 通过通道注意力机制生成的权重与底层特征图逐元素相乘, 筛选并强化重要特征, 最终将调整后的特征图相加, 实现多尺度特征的有效融合. 这一过程优化了特征信息的整合, 提升了模型对关键细节的捕捉能力.

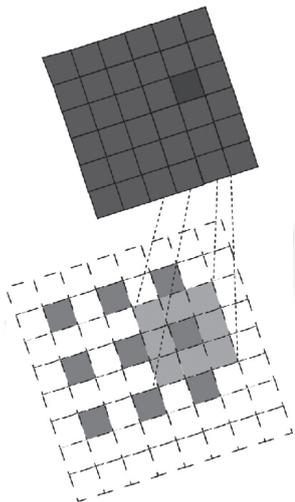


图9 转置卷积操作示意图

4 实验结果与分析

4.1 数据集

(1) BDD100K 数据集

BDD100K (Berkeley deep drive 100K) 由加州大学

伯克利分校开发, 是自动驾驶领域的关键数据集. 它包含超过 10 万张高分辨率图像和 1000 多个视频片段, 覆盖城市街道、高速公路、郊区道路等多样场景, 以及雨、晴、雪等天气条件, 特别包括夜间和白天. 数据集支持物体检测, 标注了车辆、行人、自行车等十几种物体. 针对夜间车辆检测研究, 我们从 BDD100K 中筛选了 21400 张夜间图像, 专注于检测公共汽车 (bus)、卡车 (truck)、汽车 (car) 这 3 类目标.

(2) CarDetection 数据集

CarDetection 数据集由车载摄像头在夜间拍摄, 涵盖高速公路、普通道路、路口和停车场等场景, 共 6515 张图片. 由于原始标注仅限汽车类别且车型与 BDD100K 数据集有差异, 从中选择了 1600 张图片, 用 LabelImg 工具重新标注为这 3 种目标类型, 从而与 BDD100K 数据集保持一致.

将标注后的 CarDetection 数据集与经过挑选后的 BDD100K 数据集合并, 形成 DarkVehicle 数据集, 包含 21400 张夜间道路图片. 为解决类别不平衡, 特别是汽车占比过高 (97.3%) 的问题, 我们对公共汽车和卡车进行了数据增强, 包括旋转、水平翻转和随机裁剪, 并剔除了仅含汽车的图片. 这些措施使数据集更加平衡, 最终包含 25933 张图片, 再将得到的数据集以 8:1:1 的比例划分为训练集、验证集和测试集, 训练集共包含 20747 张图片, 验证集以及测试集分别包含 2593 张图片. DarkVehicle 数据集的部分图像如图 10 所示.

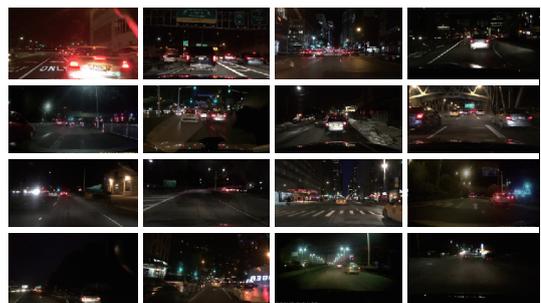


图10 DarkVehicle 数据集示例图

4.2 评估指标

为了评估目标检测模型的性能,使用如下评价指标来衡量其效果.

(1) 精确率和召回率

在深度学习的应用背景下,将样本的实际类别与模型的预测结果对比通常分为4类:真正例(true positive, TP)、假正例(false positive, FP)、真反例(true negative, TN)和假反例(false negative, FN).

精确率(precision, P)表示被正确检测的目标数量与所有被模型预测为正例的样本数量之比.它衡量了模型在识别目标时的准确性.精确率的计算公式如下:

$$P = \frac{TP}{TP + FP} \quad (1)$$

召回率(recall, R)表示被正确检测的目标数量与所有真实正例样本的数量之比,衡量了模型的漏检情况.召回率的计算公式如下:

$$R = \frac{TP}{TP + FN} \quad (2)$$

(2) 平均精度均值

平均精度均值(mean average precision, mAP)能够较全面地评估模型在多类别目标检测任务中的性能.其计算公式如下:

$$mAP = \frac{1}{classCount} \sum_{i=1}^{classCount} AP_i \quad (3)$$

其中, AP (average precision) 是针对单个类别的性能评估,其计算公式如下:

$$AP = \int_0^1 P(R) dR \quad (4)$$

其中, P 表示精确率, R 表示召回率.

(3) FPS

FPS (frames per second) 衡量了模型在处理图像或视频时每秒处理的帧数,反映模型的响应速度和处理效率, FPS 越高,意味着模型处理速度越快,可以实现更实时的目标检测.

(4) 模型大小

模型大小主要指的是模型所占用的存储空间大小以及运行时所需的计算资源,模型大小直接关系到其不同设备上的部署可能性,较小的模型更容易部署在资源有限的设备上.

4.3 实验结果及分析

本文实验环境配置如表1所示,并设置批量大小(batch-size)为8,进行300个epoch迭代.在超参数设置方面,初始学习率(lr0)设置为0.01,权重衰减系数(weight_decay)设置为0.0005.

表1 实验环境配置

环境配置	名称	信息
硬件环境	GPU	NVIDIA GeForce RTX 2080 Ti
	CPU	12th Gen Intel(R) Core(TM) i5-12400F
	内存	32 GB
	显存	16 GB
软件配置	操作系统	Windows 10
	Python	1.8
	CUDA	11.3
	CPU	7.0

4.3.1 消融实验结果与分析

为验证改进的各模块的有效性,本文在DarkVehicle数据集上进行了消融实验.结果如表2所示.

表2 消融实验结果

模块	Precision	Recall	$mAP@0.5$	$mAP@0.5:0.95$
Baseline	0.663	0.522	0.587	0.390
+SPDConv	0.687	0.534	0.606	0.403
+TA	0.663	0.535	0.594	0.394
+HS-FPN	0.668	0.518	0.589	0.388
+SPDConv+TA	0.680	0.541	0.611	0.405
+TA+HS-FPN	0.666	0.538	0.595	0.397
+SPDConv+TA+HS-FPN	0.684	0.549	0.614	0.408

由表2可知,3种改进措施均提升了模型的性能.与YOLOv8n网络相比,加入SPDConv的情况下,模型的精确率提高了2.4%,召回率提高了1.2%, $mAP@0.5$ 提高了1.9%, $mAP@0.5:0.95$ 提高了1.3%.

图11展示了加入SPDConv模块前后的检测效果对比.在黄色框标注区域,可以看到YOLOv8n模型未能检测到的远处小型车辆,而在引入SPDConv模块后(图11(b))被成功识别,同时已检测车辆的置信度也有所提升.这验证了SPDConv在保留图像细节、增强小目标特征表示方面的有效性,显著降低了此类目标的漏检率.图12对比了引入三重注意力(TA)机制前后的检测结果及对应的特征热力图.对比图12(a)和(b),可以观察到YOLOv8n模型在图12(a)黄色框位置存在一个误检目标,而引入TA机制后(图12(b)),该误检被有效抑制.热力图显示,TA机制引导模型更加聚焦于真实的车辆区域(如车灯、轮廓),抑制了背景噪声的干扰.这表明TA机制显著提升了模型在复杂夜间

背景下的目标-背景区分能力,减少了由噪声和光照干扰引起的误检。

随后,引入多尺度选择特征融合模块 HS-FPN,表 3 为引入 HS-FPN 模块前后的模型大小及速度对比,引入 HS-FPN 模块之后,模型参数量下降了约 34.7%,GFLOPs 下降了 14.6%,检测速度每秒也快了 15 帧。在此情况下,检测精度仍有一定的提升,在引入 HS-FPN

模块后, $mAP@0.5$ 提升了 0.2%,虽然在单独使用 HS-FPN 的情况下,模型召回率有所下降,但在与 SPDConv 和三重注意力机制一起使用后,召回率却提升了 2.7%, $mAP@0.5$ 提升了 2.7%, $mAP@0.5:0.95$ 提升了 1.8%,在保留更多细节特征以及更关注目标区域的情况下,模型能更好地利用这些特征进行特征筛选、融合,从而达到更好的检测效果。

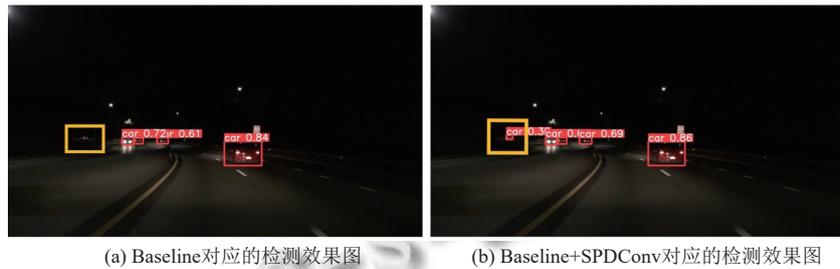


图 11 加入 SPDConv 前后检测效果对比图

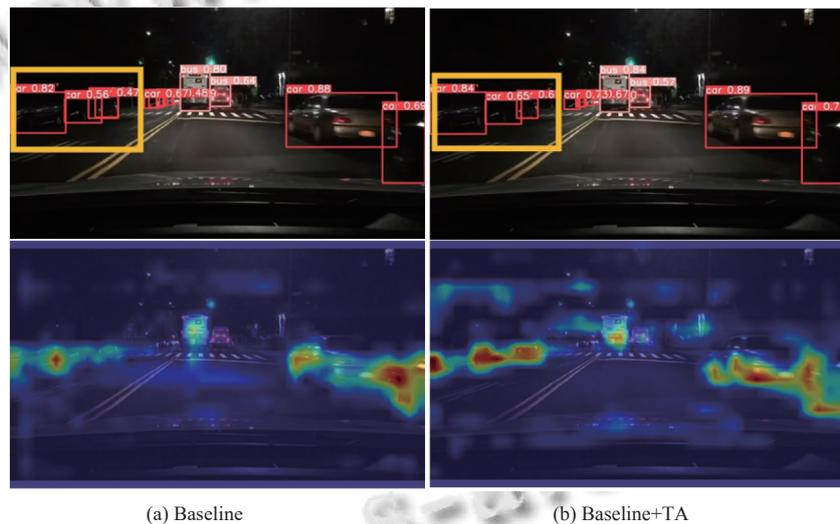


图 12 加入三重注意力机制前后的检测结果及热力图

表 3 引入 HS-FPN 模块前后的模型大小及速度对比

模型	参数量	GFLOPs	FPS (f/s)
YOLOv8n	3 011 433	8.2	110.3
YOLOv8n+HS-FPN	1 966 633	7.0	125.0

4.3.2 对比实验结果与分析

为验证改进模型的性能,将本文模型与目前主流的几种目标检测算法进行对比,包括 Faster R-CNN、SSD、YOLOv5n、YOLOv8s 和 YOLOv8l,在相同的环境配置下,采用 BDD100K 数据集中的夜间部分进行训练和验证。表 4 展示了各模型的检测效果。

根据表 4 的数据,双阶段的检测算法 Faster R-CNN 结构较为复杂,模型较大,且检测速度较慢,相比

于 YOLOv8 系列的算法, Faster R-CNN 在数据集下表现不佳。考虑到夜间车辆检测模型的部署和使用是在车辆上,在移动设备上计算资源通常有限,故需要一个较为轻量且检测速度较快的模型,一阶段的检测算法相对更为合适。一阶段的检测算法 SSD 检测效果最差, $mAP@0.5$ 仅为 53.6%,且模型也较大,比 YOLOv8n-STH 大了约 19 倍。YOLOv5n 的模型大小和 FPS 都相对较优秀,但是其检测效果较差, $mAP@0.5$ 仅有 56%。YOLOv8s 具有更深、更宽的网络结构,故其检测效果较好, $mAP@0.5$ 达到了 60.1%,但由于它包含了更多的网络层和参数,导致其模型比 YOLOv8n 大了 15.3 MB,检测速度也有所降低。作为 YOLOv8 系列中更大的模

型, YOLOv8l 凭借其更复杂的网络结构, 在检测精度上表现最优, 其 $mAP@0.5$ 为 65.9%, $mAP@0.5:0.95$ 为 46.9% 均超过了本文提出的 YOLOv8n-STH 模型. 然而, 这种精度的提升是以显著增加模型大小和降低检测速度为代价的. 而本文的 YOLOv8n-STH 模型仅为 4.7 MB, 易于部署, 且检测速度较快, 能满足驾驶场景下的实时要求, 检测精度也有所提升, 相对于 YOLOv8n, 在模型大小减少了 24.2% 的情况下, $mAP@0.5$ 提升了 2.7%, $mAP@0.5:0.95$ 提升了 1.8%. 尽管在绝对精度上略逊于 YOLOv8l, 但 YOLOv8n-STH 在模型大小和检测速度 (FPS) 方面具有压倒性的优势, 这对于夜间辅助驾驶车辆检测至关重要. 综合来看, YOLOv8n-STH 模型在夜间车辆检测的任务中取得了较为优异的效果, 在提升检测精度的同时还兼顾了检测速度与模型大小.

表 4 各类检测算法的对比实验结果

模型	$mAP@0.5$	$mAP@0.5:0.95$	FPS (f/s)	模型大小 (MB)
Faster R-CNN	0.572	0.367	35.6	108
SSD	0.536	0.311	49.8	95.1
YOLOv5n	0.560	0.332	115.3	5.4
YOLOv8s	0.601	0.393	96.9	21.5
YOLOv8n	0.587	0.390	110.3	6.2
YOLOv8l	0.659	0.469	83.2	83.7
YOLOv8n-STH (本文模型)	0.614	0.408	120.1	4.7

4.3.3 实验结果演示

为了更直观地看到模型效果, 选择了 3 种场景进行可视化效果对比, 图 13 展示了 3 种不同情境下改进模型 YOLOv8n-STH 与基础模型 YOLOv8n 的检测效果对比.

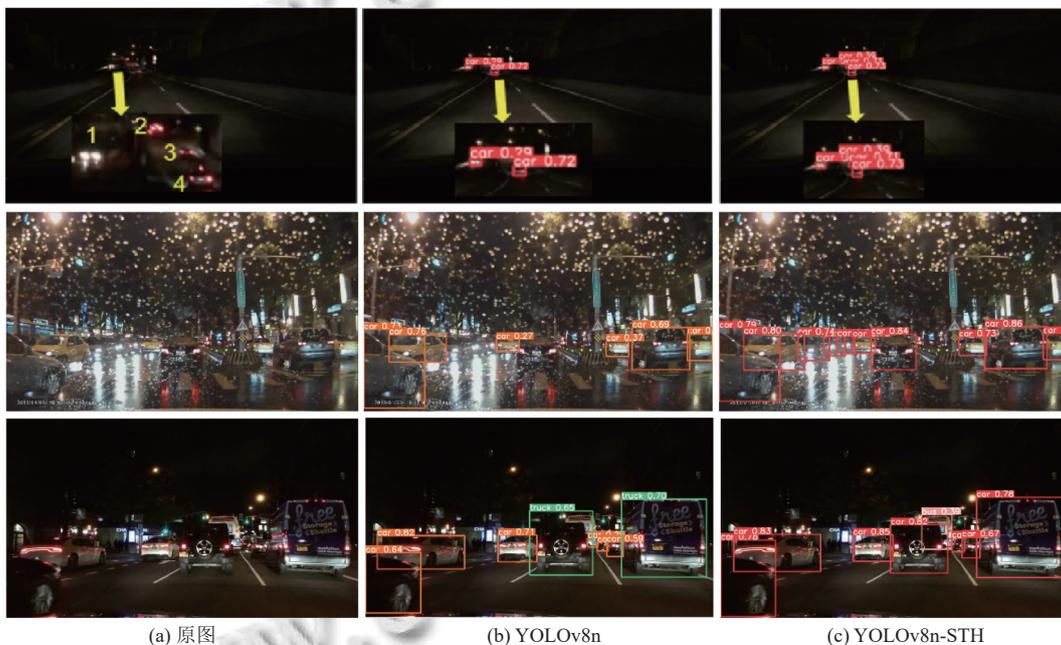


图 13 模型改进前后检测效果对比

第 1 行是远距离小目标车辆检测场景. 图中共有 4 辆车位于视野的远处. YOLOv8n 仅检出两辆车, 而 YOLOv8n-STH 成功检出全部 4 辆, 且置信度更高, 充分体现了 SPDCConv 和特征融合优化对小目标检测的提升; 第 2 行为雨天场景, YOLOv8n 模型出现了漏检, 而 YOLOv8n-STH 在车窗存在大量雨滴噪声干扰的情况下, 依然正确检出所有车辆, 展现了 TA 机制和鲁棒特征融合在噪声环境下的优势; 第 3 行为城市路口, 光照相对较充足, YOLOv8n 不仅将一辆小车和面包车误检为卡车, 还漏检了远处的公交车. YOLOv8n-STH 则

准确识别了所有车辆及其类别, 验证了 HS-FPN 模块通过特征筛选提升类别判别准确性的效果. 通过对比分析, 可以观察到 YOLOv8n-STH 在多种夜间场景下的检测性能均优于基础模型 YOLOv8n, 特别是在小目标检测、噪声鲁棒性和类别准确性方面提升显著.

5 结束语

本文基于 YOLOv8n 提出了一种目标检测算法 YOLOv8n-STH. 首先, 鉴于 YOLOv8n 中的步长卷积可能会跳过一些细节信息, 故将原本的卷积与池化操作

替换为空间到深度卷积 (SPDConv) 层, 并在 C2f 模块前加入空间到深度 (SPD) 层, 使其能够减少后续的计算复杂度, 并能更好地保留特征. 其次, 在 C2f 结构中引入三重注意力 (TA) 机制, 以增强模型对目标区域的聚焦能力, 尤其在目标与背景界限不明显的情况下. 最后对特征融合模块进行改进, 使用一种高层筛选-特征融合金字塔 (HS-FPN), 有效筛选出更具代表性的特征信息, 提高了模型对特征的利用效率, 并通过减少参数量实现了轻量化. 在公开的自动驾驶数据集上的实验表明, 在夜间辅助驾驶场景中, YOLOv8n-STH 模型在对车辆检测的准确率及模型大小方面, 相较于目前主流的目标检测模型有明显的提升.

参考文献

- 1 Feng J, Zeng DN, Jia XP, *et al.* Cross-frame keypoint-based and spatial motion information-guided networks for moving vehicle detection and tracking in satellite videos. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2021, 177: 116–130. [doi: [10.1016/j.isprsjprs.2021.05.005](https://doi.org/10.1016/j.isprsjprs.2021.05.005)]
- 2 乔瑞萍, 董员臣, 王方, 等. 基于颜色通道混合特征的夜间车辆检测仿真. *计算机仿真*, 2020, 37(12): 107–110. [doi: [10.3969/j.issn.1006-9348.2020.12.023](https://doi.org/10.3969/j.issn.1006-9348.2020.12.023)]
- 3 李笑, 杨宇, 徐一鸣. 四帧间差分和改进混合高斯模型对运动目标的检测. *科学技术与工程*, 2020, 20(15): 6141–6150. [doi: [10.3969/j.issn.1671-1815.2020.15.037](https://doi.org/10.3969/j.issn.1671-1815.2020.15.037)]
- 4 康金忠, 王桂周, 何国金, 等. 遥感视频卫星运动车辆目标快速检测. *遥感学报*, 2020, 24(9): 1099–1107.
- 5 陈立潮, 解丹, 曹建芳, 等. 改进光流法和 GMM 融合的车辆实时检测算法研究. *智能系统学报*, 2021, 16(2): 271–278.
- 6 舒兆翰, 李小龙. 融合两帧差分 and 背景差分的车辆检测算法. *计算机工程与设计*, 2023, 44(4): 1174–1180.
- 7 Tsai WK, Chen HJ. High-accuracy vehicle lamp detection for real-time night-time traffic surveillance. *IET Intelligent Transport Systems*, 2020, 14(13): 1923–1934. [doi: [10.1049/iet-its.2020.0063](https://doi.org/10.1049/iet-its.2020.0063)]
- 8 Nakane T, Takeshita T, Tokai S, *et al.* Vehicle rear-lamp detection at nighttime via probabilistic bitwise genetic algorithm. *Proceedings of the 2019 International Conference on Cyberworlds (CW)*. Kyoto: IEEE, 2019. 117–120.
- 9 Lecun Y, Bottou L, Bengio Y, *et al.* Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 1998, 86(11): 2278–2324. [doi: [10.1109/5.726791](https://doi.org/10.1109/5.726791)]
- 10 Ren SQ, He KM, Girshick R, *et al.* Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137–1149. [doi: [10.1109/TPAMI.2016.2577031](https://doi.org/10.1109/TPAMI.2016.2577031)]
- 11 He KM, Gkioxari G, Dollár P, *et al.* Mask R-CNN. *Proceedings of the 2017 IEEE International Conference on Computer Vision*. Venice: IEEE, 2017. 2980–2988.
- 12 Liu W, Anguelov D, Erhan D, *et al.* SSD: Single shot multibox detector. *Proceedings of the 14th European Conference on Computer Vision*. Amsterdam: Springer, 2015. 21–37.
- 13 Redmon J, Divvala S, Girshick R, *et al.* You only look once: Unified, real-time object detection. *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas: IEEE, 2016. 779–788.
- 14 邵延华, 张铎, 楚红雨, 等. 基于深度学习的 YOLO 目标检测综述. *电子与信息学报*, 2022, 44(10): 3697–3708. [doi: [10.11999/JEIT210790](https://doi.org/10.11999/JEIT210790)]
- 15 王婷婷, 戴金龙, 孙振轩, 等. 基于 YOLOv4 算法在车辆检测中的应用. *吉林大学学报 (信息科学版)*, 2023, 41(2): 281–291.
- 16 Sunkara R, Luo T. No more strided convolutions or pooling: A new CNN building block for low-resolution images and small objects. *Proceedings of the 2022 European Conference on Machine Learning and Knowledge Discovery in Databases*. Grenoble: Springer, 2022. 443–459.
- 17 Cai WW, Liu BT, Wei ZG, *et al.* TARDB-Net: Triple-attention guided residual dense and BiLSTM networks for hyperspectral image classification. *Multimedia Tools and Applications*, 2021, 80(7): 11291–11312. [doi: [10.1007/s11042-020-10188-x](https://doi.org/10.1007/s11042-020-10188-x)]
- 18 Chen YF, Zhang CY, Chen B, *et al.* Accurate leukocyte detection based on deformable-DETR and multi-level feature fusion for aiding diagnosis of blood diseases. *Computers in Biology and Medicine*, 2024, 170: 107917. [doi: [10.1016/j.combiomed.2024.107917](https://doi.org/10.1016/j.combiomed.2024.107917)]
- 19 Woo S, Park J, Lee JY, *et al.* CBAM: Convolutional block attention module. *Proceedings of the 15th European Conference on Computer Vision*. Munich: Springer, 2018. 3–19.

(校对责编: 张重毅)