面向车载视频的违停自动检测方法

龚 航[1] 范立新[2]
GONG Hang FAN Lixin

摘 要 自动、高效地对违停车辆进行治理，可以降低人工巡查的工作量并提高执法效率。基于此，文章提出了一种面向车载视频的违停自动检测方法。该方法基于 YOLOv7-tiny 进行设计，首先通过引入 SPPFCSPC 模块来降低模型的复杂度，提高推理速度；然后结合融合通道注意力的上采样模块来增强模型多尺度特征融合能力，提高检测准确率；其次，采用不规则四边形预测框来拟合目标的实际区域，提升定位准确率；最后，设计一种跨域协同训练策略来保护各方训练数据隐私，提高协同训练速度。实验结果表明，所提方法在验证集上的平均准确率（mAP@50）达到了 98.9%，相比原始 YOLOv7-tiny 提升了 6.34%，能有效满足车载视频的违停检测需求。

关键词 人工智能；智能交通；违停检测；YOLO

doi：10.3969/j.issn.1672-9528.2026.06.042

0 引言

2024 年，我国机动车保有量达 4.53 亿辆，较 2023 年增长 4.1%[1]。在机动车数量快速增长的同时，违法停车的现象也日益突出，成为了加剧交通拥堵、诱发安全事故等重要原因之一。传统违停执法主要依赖固定点位监控与人工巡查的方式，存在覆盖范围窄、执法效率低等问题；而采用无人机巡检 [2] 的方式，虽然能提高检测范围，但存在维护成本高、操作困难等问题，不利于大面积推广使用。在当前警车与铁骑巡逻成为交警日常勤务的背景下，探索移动化、自动化的违停检测技术，对提高道路交通治理效能具有重要意义。

随着人工智能技术的发展，特别是深度学习在计算机视觉领域的广泛应用，为面向车载视频的违停检测技术的实现提供了可能。特别是以 YOLO[3] 为代表的单阶段目标检测算法，凭借其优异的实时性与准确性，可以快速地从复杂的视频流中识别 [4] 和定位违法车辆。与此同时，“联邦学习”技术范式 [5] 的提出，也为解决智能交通领域的数据隐私 [6] 问题提供了新路径。

然而，为实现在移动车载视频环境下的违停车辆实时与精准识别，仍存在两个技术难题。一是算力局限性难题，当前车载的计算终端计算能力有限，要求实现上必须采用参数规模小、模型复杂度低的检测算法，可能会损失一部分检测精度；二是干扰因素多样性难题，在车辆运动过程中，道路环境的变化、视频画面的抖动、相机拍摄的角度等，都会对算法的检测造成偏差。

为解决上述技术难题，本文提出了一种面向车载视频的违停自动检测方法。主要贡献可归纳为以下四个方面：

（1）优化了 YOLOv7-tiny[7] 网络结构，引入 SPPFCSPC 模块 [8] 以降低参数模型复杂度，在保持检测精度的同时提升推理速度；

（2）提出了一种融合通道注意力机制 [9] 的上采样结构，增强模型在多尺度目标下的特征提取能力，提升检测精度；

（3）改进了 YOLOv7-tiny 输出层与损失函数，以不规则四边形检测 [10] 提升对倾斜停车线等目标的定位精度；

（4）设计了一种跨域协同训练策略，在保障数据安全的前提下，利用双方算力和数据完成了检测模型的协同训练。

1 违停自动检测

1.1 实验数据集

目前国内外尚未有开源的面向车载视频的违停车辆数据集。为此，本研究通过自主采集与协作获取两种方式构建实验数据集。

自主采集的方式是通过在摩托车或汽车上加装高清车载摄像头，在实际道路环境中进行采集。经人工筛选与标注后获得有效图像 5 930 张，部分数据如图 1 所示。协作获取的方式是通过与交警部门合作，获取内部执法设备拍摄的违停车辆图片数据。由于数据保密要求，该部分数据仅支持在交警部门内部研究使用。经人工筛选与标注后获得有效图像 5 000 张。

1. 韶关市数据产业研究院 广东韶关 512029
2. 上海长城汽车科技有限公司 上海嘉定 201805

图 1 数据集样例

本文采用 Labelme 工具对图像进行标注，所有标注任务均由专业人员完成。标注过程采用了四点标注法，支持描绘任意四边形边界框。与传统 YOLO 矩形框相比，任意四边形更能贴合停车位、车牌等目标的真实轮廓，能有效减少无关背景因素的干扰。

最终标注完成后的数据集组成如下：自主采集数据包含训练集 4 535 张、验证集 1 345 张；协作获取数据包含训练集 4 000 张、验证集 1 000 张。

1.2 检测算法设计

本文采用 YOLOv7-tiny 作为违停检测算法的研究框架。该算法在目标检测精度与速度之间取得了较好的平衡，其轻量化设计尤其适合在边缘计算终端上部署。然而，标准 YOLOv7-tiny 模型在车载视频环境下，违停车辆的检测准确率和目标的定位精度并不高，需要做进一步优化。

1.2.1 改进空间金字塔池化

YOLOv7-tiny 采用了 SPPCSPC 空间金字塔池化模块（如图 2 所示），相较于 SPP 模块[11]，虽然性能表现更优，但引入了更多的参数，使模型推理效率下降。因此本文对 YOLOv7-tiny 的 SPPCPSC 模块进行了改进，将原有的 SPPCSPC 模块升级为 SPPFCSPC 模块。该改进的核心在于用快速空间金字塔池化（SPPF）替代传统的空间金字塔池化（SPP）组件。

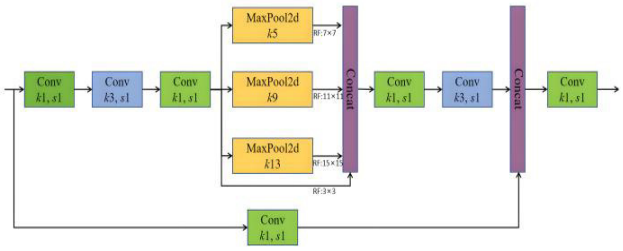

图 2 SPPCSPC 模块结构

如图 3 所示，SPPFCSPC 模块采用了串行堆叠池化层的方式，通过连续使用多个大小为 5×5 的最大池化层来完成多尺度信息的融合。与原有 SPPCSPC 模块的并行多尺度池化相比，这种串行设计既可以保持模型相同的感受野，使模型能够捕获丰富的上下文信息，同时又可以利用参数共享的方式降低模型的计算复杂度，从而提高模型的推理速度。

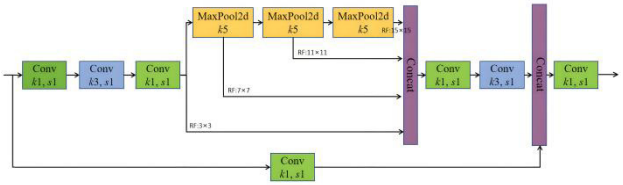

图 3 SPPFCSPC 模块结构

实验表明，改进后的空间金字塔池化模块能更好地满足违停检测算法的实时性要求，增强模型在资源受限环境中的部署适用性。

1.2.2 加强多尺度特征融合

在目标检测任务中，YOLOv7-tiny 采用了多尺度特征融合的方式来提高目标的检测率，但在不同尺度特征融合问题上，YOLOv7-tiny 只是简单地将深层特征图经过上采样后与浅层特征图进行拼接运算，融合效率不高。如何高效地融合不同尺度的特征图，是提高目标检测网络性能的一个关键手段。

为此，本文提出了一种基于通道注意力机制的上采样模块（channel attention upsample, CAU），通过生成深层网络的通道注意力特征图来指导浅层网络选择目标的定位细节，使融合后的特征图包含更多的目标位置信息。

如图 4 所示，CAU 模块首先对深层特征图执行全局平均池化操作，得到一个和深层特征图通道数一样的特征向量。然后利用两个 1×1 卷积（不含 BN 层）对生成的特征向量进行压缩和激励，得到深层特征图每个通道对应的权值。接着将该权值向量与浅层特征图进行相乘后，通过一个 1×1 卷积进行降维和跨通道信息整合。最后将该加权后的浅层特征图与经过上采样操作之后的深层特征图进行拼接，完成不同尺度的特征融合。

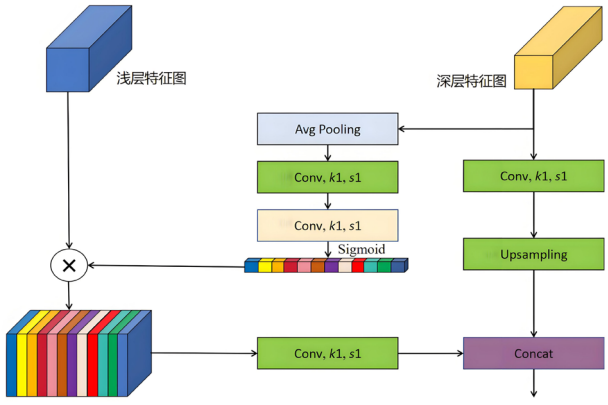

图 4 CAU 模块结构

与原 YOLOv7-tiny 简单的上采样相比，本文提出的 CAU 模块充分利用了深层特征图的语义信息，不仅能更加有效地加强不同尺度下的特征融合，而且还能通过一种简单的

方式为浅层网络的特征映射提供指导信息。

1.2.3 采用不规则四边形检测

现有的 YOLO 系列目标检测算法通常采用矩形边界框来锚定目标区域（如图 5（a）所示），但面对倾斜停车位、车牌和道路标志线等非规则目标时，采用矩形框标注不可避免的会引入更多的背景噪声，使模型无法聚焦目标实际特征区域进行训练，导致模型的检测准确率和定位精度不高。因此，本文对 YOLOv7-tiny 边框预测部分进行了改进，采用不规则四边形来拟合目标的实际区域（如图 5（b）所示）。

（a）采用矩形边界框来锚定目标区域　（b）采用不规则四边形来拟合目标的实际区域

图 5 标注方式对比

改进后的 YOLOv7-tiny 模型输出层的边界框预测由原来的中心点、长度 4 参数表示法，调整为 8 参数的不规则四边形顶点表示法。

在损失函数设计方面，本文设计了一种双重损失函数。该损失函数由几何约束损失与坐标回归损失两部分构成。其中，几何约束损失通过 6 个空间关系，确保顶点坐标按照左上、右上、右下与左下的顺序依次输出，使四边形保持合理的凸多边形结构；坐标回归损失则采用 Smooth L_1 函数分别计算每个顶点的坐标误差。

改进后的边框损失函数可表示为：

$$\begin{aligned} L_{\text{box}} = \frac{1}{\lambda}[&\max(0, y_1 - y_3)^2 + \max(0, y_2 - y_3)^2 \\ &+ \max(0, y_1 - y_4)^2 + \max(0, y_2 - y_4)^2 \\ &+ \max(0, x_1 - x_2)^2 + \max(0, x_4 - x_3)^2] \\ &+ \text{Smooth } L_1(p_{\text{pred}}, p_{\text{truth}}) \end{aligned} \tag{1}$$

$$\text{Smooth } L1 = \begin{cases} \dfrac{0.5 \cdot (x-y)^2}{\text{beta}}, & \text{if } |x-y| < \text{beta} \\ |x-y| - 0.5 \cdot \text{beta}, & \text{other} \end{cases} \tag{2}$$

式中：p_{pred} 表示预测的四边形坐标，p_{truth} 表示真实四边形坐标，λ 表示平均惩罚系数（本文设定为：1/6），beta 表示 L_1 与 L_2 损失之间的平滑过渡点（本文设定为 0.12）。

1.3 模型跨域协同训练

为了保护交警部门的数据隐私，本文采用了一种基于数据并行的跨域协同训练策略。该策略在不直接共享训练图像数据的前提下，实现了算法模型的分布式联合优化。

传统的数据并行训练策略 [12]，通常每迭代一轮数据，需要进行一次参数传递。对于低带宽网络环境，这种频繁的数据交互方式，不仅会占用大量的网络资源，而且也会降低算法的训练效率，从而导致模型完成训练时间大幅度增加。

针对上述问题，本文对该协同训练策略进行了改进，允许用户自定义参数同步频率。训练流程如图 6 所示，开始训练时每个节点使用各自数据集训练自己的模型副本；当每个节点都迭代完成 10（N=10）轮数据之后，各节点之间开始同步梯度信息并更新各自的模型副本；当参数同步之后各节点损失值开始收敛，停止训练并保存主节点模型副本作为最终结果。

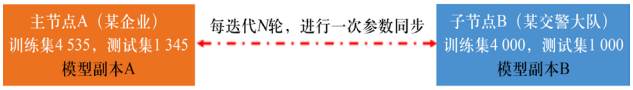

图 6 协同训练示意图

实验结果表明，本文改进的模型跨域协同训练策略，能高效地完成跨区域之间的模型训练，降低约 90% 的网络传输负载。而且，由于模型训练过程仅传输模型参数，并不涉及数据集传输，有效地保护了各参与方的数据隐私安全。

1.4 实验结果分析

本文采用双节点协同训练的方式对改进算法进行训练。其中，一个训练节点 A 部署于江苏省某企业，硬件配置为：Intel Core i5-14400F、NVIDIA GeForce RTX 4060 显卡、32 GB 内存；另一个节点 B 部署于江苏省某交警大队，硬件配置为：Intel Core i5-12450H、NVIDIA GeForce RTX 4050 Laptop 显卡、16 GB 内存。测试设备为瑞芯微 RK3588 开发板，内存为 8 GB。

1.4.1 算法性能分析

如图 7 所示，算法在训练 50 个 epoch 后，训练节点 A 的模型副本损失值开始收敛；训练至 100 个 epoch 时，模型的平均准确率（mAP@0.5）均已稳定在 90% 以上。

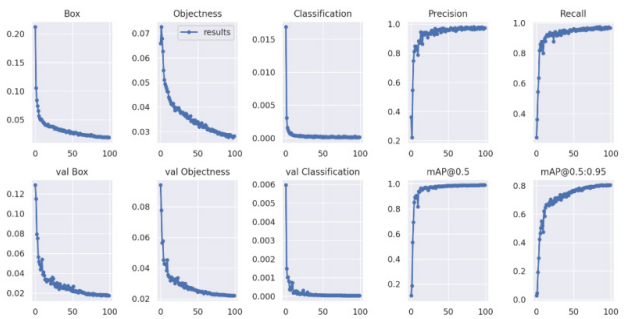

图 7 损失与性能指标可视化曲线图

为进一步验证上述改进方法的优越性，本文对改进算法的 3 个关键部分进行了消融实验，分别是：改进空间金字塔池化模块（SPPFCSPC）、加强多尺度特征融合的基于通道注意力机制的上采样模块（CAU）与改进 YOLOv7-tiny 输出

层模块（Polygon）。消融实验结果如表 1 所示，其中 MTTD 为模型在测试设备上的单张图片检测时间。

表 1 消融实验结果

Net	*P*/%	*R*/%	mAP@0.5/%	Size/MB	MTTD /ms
YOLOv7-tiny	**93.54**	**93.12**	**92.56**	**12.3**	**36**
YOLOv7-tiny + SPPFCSPC	93.54	92.88	92.55	12.3	34
YOLOv7-tiny + CAU	97.42	94.28	96.42	12.8	38
YOLOv7-tiny + Polygon	96.50	95.24	96.70	12.3	37
YOLOv7-tiny + SPPFCSPC + CAU + Polygon	**97.38**	**96.55**	**98.9**	**12.8**	**39**

根据上述消融实验结果，可以清晰地看出：

（1）采用 SPPFCSPC 模块替换原有的空间金字塔池化层，并不会对模型的准确率造成影响，反而在实际部署环境中可以提升单帧 2 ms 的检测速度。

（2）引入 CAU 模块后，模型的准确率和召回率分别提高了 3.88% 和 1.16%，平均准确率（mAP@0.5）提高了 3.86%，说明本文提出的 CAU 模块能有效地加强网络多尺度的特征融合和提升模型对目标的检测准确率。

（3）采用 Polygon 模块替换原有的矩形边框输出模块后，模型的准确率和召回率分别提高了 2.96% 和 2.12%，平均准确率（mAP@0.5）提高了 4.14%，说明不规则四边形预测能提高模型对倾斜目标的检测准确率。

（4）结合三种改进策略后，最终模型的平均准确率（mAP@0.5）达到了 98.9%，检测效果得到了较大提升。

1.4.2 测试效果分析

本文开发的违停检测模型最终部署在搭载瑞芯微 RK3588 芯片的工业级平板上（内嵌北斗定位模块），设备安装示意图如图 8 所示。

图 8 设备安装示意图

如图 8 所示，系统所用平板、摄像头均安装在交警铁骑上，其中平板通过防震支架固定在仪表盘上方；两个摄像头（电子防抖）呈 120° 夹角固定在警灯柱上，用于检测右侧违停车辆。在实际抓拍过程中，首先由执法人员手动点击平板，开启违停抓拍功能；接着系统进入检测状态，开始加载违停检测模型分析视频流。若视频中车辆未停在停车线区域内，并且车辆位于马路崖左侧，即可认为该车辆属于违法停车；最后生成车辆违停证据图片（如图 9 所示），并通过 4G 网络上传至后台服务器。

目前基于本文改进算法设计的铁骑移动智能抓拍系统，已经通过了公安部交通管理科学研究所的测试，在全国多地都有安装应用。测试数据显示，本文提出的面向车载视频的违停自动检测方法具有 90% 以上的违停车辆抓拍率。

图 9 实际抓拍效果图

2 结语

本文提出的面向车载视频的违停检测方法，虽然取得了不错的检测效果，但仍然存在一些不足。首先，系统在抓拍过程中无法判断车辆是否在缓慢行驶，这可能会导致误抓拍；其次，临停路段通常需间隔一段时间进行二次抓拍，但在第二次抓拍过程中，系统无法判断车辆中途是否短暂离开过；最后，在雨天环境下，系统会存在停车线检测率低和车牌识别效果差等问题。针对上述不足，现阶段暂未有较好的解决办法，只能依靠人工干预的方式进行处理，即执法人员手动操控平板开启或关闭违停检测系统。

未来，本文将继续对该检测方法进行优化。一方面，对现有的检测算法进行迭代升级，使系统的实际违停车辆抓拍率达到 95% 以上；另一方面，通过引入新的检测算法，解决当前无法判别车辆缓慢行驶的问题。此外，本文计划在现有算法基础上，继续探索更多的机动车违法行为抓拍，例如闯红灯、未按规定车道行驶等，使系统功能更加强大。

参考文献：

[1] 吴博峰 . 全国机动车保有量达 4.53 亿辆 [N]. 中国消费者报，2025-01-21.

[2] 梁定康 , 钱瑞 , 陈义豪 , 等 . 基于视觉的无人机巡检违章违停系统设计与实现 [J]. 电子科技 ,2018,31(5):76-77.

[3]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).

融合梯度与注意力机制的可解释性在线协作讨论交互文本分类方法

邵 峥[1*] 李金鹏[1] 王 勇[1] 何常乐[1]
SHAO Zheng LI Jinpeng WANG Yong HE Changle

摘 要 深度学习模型在在线协作讨论交互文本分类任务中取得了显著性能，但其决策过程的不透明性严重制约了该类模型在教育领域中的可信应用。基于此，文章提出了一种融合梯度与注意力机制的可解释性在线协作讨论交互文本分类方法。以 CNN-BiGRU 网络为基础分类模型，通过将卷积层与循环层得到的梯度重要性与注意力权重进行非线性融合，以得到词级重要性分数，从而在保证分类精度的同时提供词粒度的可解释依据。从分类性能与解释质量两个维度进行评估，实验结果显示所提出的可解释性分类方法在 Macro-F_1 上达到 87.47%，优于基线模型。在可解释性上，该方法的 AOPC 值和 PAAC 值分别为 44.16% 和 38.79%，均优于其他可解释性方法。

关键词 可解释性；注意力机制；梯度；非线性融合

doi：10.3969/j.issn.1672-9528.2026.06.043

0 引言

近年来，深度学习在自然语言处理（NLP）领域取得了突破性进展，比如在文本分类[1]、情感分析[2]和关系抽取等任务中都展现出了超越传统机器学习方法的性能。然而，随着这些模型的规模不断扩大以及神经网络深度的增加，其表现出的“黑箱”行为也日益明显，内部决策机制难以被理解和解释，严重限制了其在高信任场景中的应用与推广。例如在协作讨论场景下，教师不仅需要了解模型根据学习者发布的协作文本类型所做出的决策，更要了解决策的依据，以便

1. 信阳学院计算机与人工智能学院 河南信阳 464000
[基金项目] 信阳学院校级科研项目资助（2025-XJLYB-004）

Piscataway:IEEE,2016:779-788.
[4] 龚航 , 刘培顺 . 夜间行驶车辆远光灯检测方法 [J]. 计算机科学 ,2021,48(12):256-263.
[5] 周国娟 . 联邦学习发展现状与展望 [J]. 网络安全技术与应用 , 2025(10):61-64.
[6] 陈丹霞 , 曾鹏 . 交通数据隐私保护技术研究及其在智能交通系统中的应用 [J]. 中国高新科技 ,2024(9):69-71.
[7]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]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway:IEEE,2003:7464-7475.
[8]Li Chuyi, Li Lulu, Jiang Hongliang, et al.YOLOv6: a single-stage object detection framework for industrial applications[PP/OL].(2022-09-07)[2026-05-28].https://doi.org/10.48550/arXiv.2209.02976.
[9]Hu Jie, Shen Li, Sun Gang. Squeeze-and-excitation networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City, UT, USA: IEEE, 2018: 7132-7141.
[10] 周秋红 . 基于改进 SSD 模型的道路病害检测研究 [J]. 黑龙江交通科技 , 2023, 46(4): 30-32.
[11]He Kaiming, Zhang Xiangyu, Ren Shaoqing, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.
[12] 金昊婴 . 面向分布式机器学习训练的通信优化技术探析 [J]. 长江信息通信 , 2025, 38(8): 52-54.

【作者简介】
龚航（1995—），男，广东韶关人，硕士研究生，研究方向：人工智能算法与应用。

（收稿日期：2025-11-07 修回日期：2026-06-08）