

基于改装电动车的中小目标检测^①

黄峻, 刘涌

(西南科技大学 计算机科学与技术学院, 绵阳 621010)

通信作者: 刘涌, E-mail: 190860881@qq.com



摘要: 由于电动车的普及, 越来越多的电动车进行非法改装雨棚. 然而, 这种改装行为会增加一些安全隐患. 首先, 挡雨板会遮挡骑行者视野, 增加事故的风险. 其次当速度过快时, 挡雨板也会在不经意间擦伤行人, 带来极大地安全隐患, 给交通安全带来了严重威胁. 本文提出了一种改进的 YOLOv7-tiny 算法, 用于电动车非法改装检测. 首先在网络的结构上加入 BiFormer 注意力机制, 不仅可以捕获更多电动车细节的信息, 而且可以使得模型更加关注一些较小的目标信息. 其次将改进的特征金字塔结构和特征融合网络的张量拼接操作进行结合, 提升对中小型目标的检测能力. 最后对框架的 ELAN 模块和 SPPCSPC 模块进行改进和优化, 可以在不增加过多参数数量的同时, 提升对中小目标的检测精度, 增强提取特征的效果.

关键词: 改装电动车; 中小目标检测; YOLOv7-tiny; BiFormer; BiFPN

引用格式: 黄峻, 刘涌. 基于改装电动车的中小目标检测. 计算机系统应用. <http://www.c-s-a.org.cn/1003-3254/9597.html>

Small and Medium-sized Object Detection Based on Modified Electric Vehicles

HUANG Jun, LIU Yong

(School of Computer Science and Technology, Southwest University of Science and Technology, Mianyang 621010, China)

Abstract: Due to the popularity of electric vehicles, more and more electric vehicles are illegally modified with rain shields. However, this modification increases safety hazards. Firstly, rain shields block riders' view, increasing the risk of accidents. Secondly, rain shields can inadvertently scratch pedestrians when the modified vehicles are at excessive speeds, posing a great safety hazard and a serious threat to traffic safety. This study proposes an improved YOLOv7-tiny algorithm for detecting illegally modified electric vehicles. Firstly, a BiFormer attention mechanism is added to the network structure, enabling the model to capture more details of electric vehicles and focus more on smaller target information. Secondly, an improved feature pyramid structure is combined with the tensor concatenation of a feature fusion network to enhance the detection ability of the model for small and medium-sized targets. Finally, the ELAN and SPPCSPC modules of the framework are optimized, which improves the detection accuracy of small and medium-sized targets and enhances the effectiveness of feature extraction without adding too many parameters.

Key words: modified electric vehicles; small and medium-sized object detection; YOLOv7-tiny; BiFormer; BiFPN

电动车的普及与发展已成为现代交通领域的一大趋势. 然而, 随着电动车数量的快速增长, 一些商家也开始利用这一趋势从事违规改装行为, 对社会安全和环境构成了严重威胁. 因此, 建立一种有效的

检测系统来识别违规改装电动车的需求变得迫切. 违规改装电动车是指对电动车进行非法改装或擅自安装其他设备的行为, 例如遮阳伞、挡雨板等. 这些违规改装行为不仅违反法律法规, 还破坏了电动车

^① 基金项目: 科技厅重点研发计划 (2020YFS0307)

收稿时间: 2024-01-26; 修改时间: 2024-02-29, 2024-03-15; 采用时间: 2024-04-10; csa 在线出版时间: 2024-10-31

的正常运行和安全性能. 违规改装电动车的存在给交通管理和执法带来了巨大的挑战, 同时对道路安全和环境保护造成了严重影响. 目前, 传统的违规改装电动车检测方法主要依赖于人工巡查和抽查检测, 效率较低.

近年来, 随着互联网的发展, 卷积神经网络技术在各个领域不断发展, 尤其是在生活方面. 将深度学习应用于电动车违规改装的遮阳伞检测中. 结合深度学习的方法不仅加快了检测速度, 也提高了检测效率, 同时节省了大量的人力成本. 目前深度学习的目标检测主要算法分为两种: 第 1 种是以 RCNN^[1], Faster RCNN^[2], 为代表的二阶段算法, 但是这类算法检测速度比较慢, 在对应研究中不能满足实时性的需求. 第 2 种就是以 YOLO 为代表的一阶段目标检测算法, 这种算法达到实时性的需求, 但是效果不佳. 吴健生等^[3]在 Faster RCNN 进行改进, 引入注意力机制, 多尺度检测模块, 可变型卷积进行识别缺陷. 张光华等^[4]对 YOLOv7-tiny 进行改进, 通过小目标检测层增强了特征提取能力, 改善了对航拍小目标的检测能力. 但是 Faster RCNN 双阶段算法要生成大量的候选框, 导致检测速度过慢, 不符合违规电动车的实时性需求. 而单阶段算法能只提取一次就能完成目标的检测, 速度比较快, 目前对于实

时性方面常用的算法是 YOLO 算法, SSD 算法^[5]等. YOLO 算法是目前比较热门的单阶段检测算法, 包括一些 YOLOv5, YOLOv7^[6]等. 许多的学者在上面进行改进取得了不错的效果. 季娟娟等^[7]设计了一种改进的 YOLOv7, 通过改进的 ELAN 模块和添加注意力机制解决了表面缺陷小导致的漏检率高的问题. 高源等^[8]设计了一种改进的 YOLOv5 的算法, 他通过注意力机制提高特征提取, 数据增强提高模型的检测不同火灾大小的能力, 使模型对火灾高效识别.

目前 YOLOv7-tiny 的检测速度在 YOLO 系列中比较优秀, 所以采用 YOLOv7-tiny 框架对违规电动车的识别进行相关研究.

1 YOLOv7-tiny 模型

YOLOv7-tiny 是一个突出的目标检测模型, 以其高速检测能力而著称. 模型的架构分为 3 个主要部分: 输入模块、主干网络和头部. 如图 1 所示, 主干网络是模型的特征提取核心, 由多个 CBL 模块、MP 层和 ELAN 模块组成. CBL 模块包含了卷积层、标准化层和激活函数, 而 ELAN 层是多个 CBL 层的组合, MP 层则由 CBL 模块和最大池化层构成. 这一层层构建的目的是从图像中抽取出有用的特征信息.

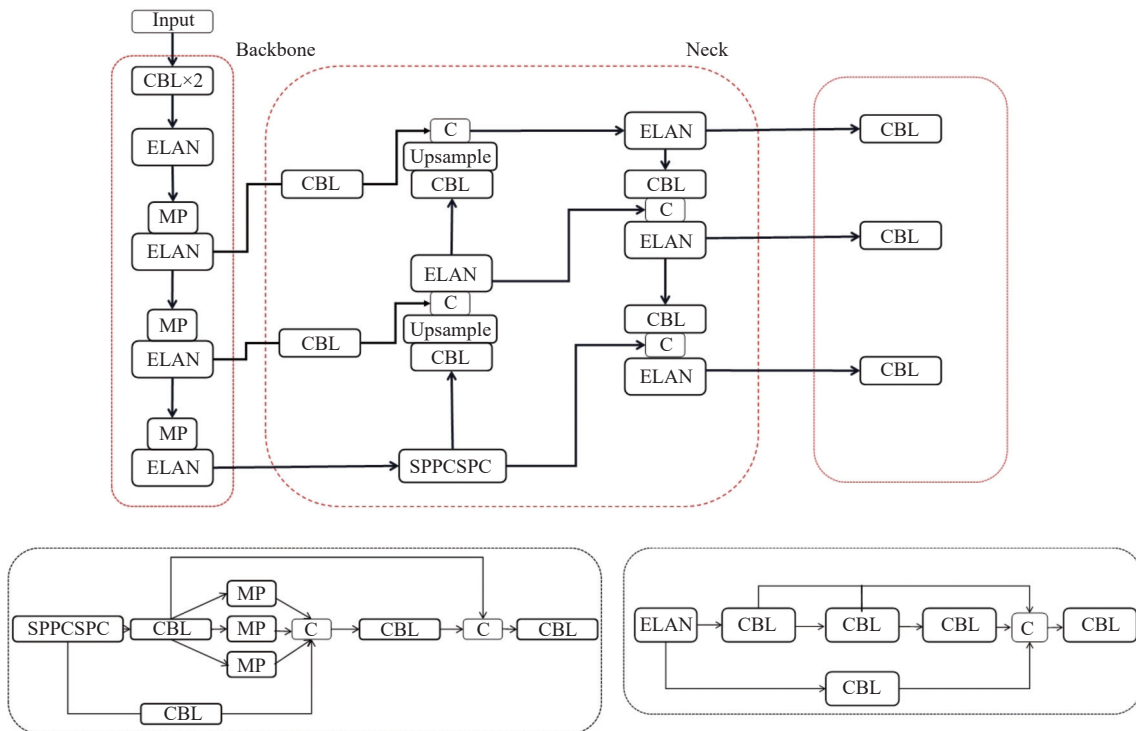


图 1 YOLOv7-tiny 框架图

在主干网络之后是 Neck 部分, 它使用 PANet 结构来融合不同层次的特征, 从而检测各种尺寸的目标. Neck 主要由 CBL 和 ELAN 模块构成. SPPCSPC 模块作为一个连接点, 一方面连接主干网络和 Neck, 另一方面与 ELAN 模块相连, 它有助于增强模型的学习能力, 同时提升处理速度. SPPCSPC 和 ELAN 模块相比于原有的 YOLOv7 框架, 无论是在模型模块数量还是在参数量上都有所减少. Head 部分和 YOLOv7 不同的是采用的是标准卷积进行通道数的调. YOLOv7-tiny 和 YOLOv7 相比, 牺牲了一定的精度, 但是在速度方面具有优势, 参数量和计算量都更少.

2 改进的 YOLOv7-tiny 模型

本文的工作如下: 1) 在 ELAN 模块后引入 BiFormer 注意力机制, 增强对小目标的特征提取. 2) 通过改进的特征金字塔结构和特征融合网络的张量拼接操作进行结合提升对不同大小目标的检测能力. 3) 改进的 ELAN 模块和 SPPCSPC 模块在不引入过多参数的同时, 改善对中小目标的特征提取.

2.1 BiFormer 模块

在实际的交通环境下, 因为复杂的背景而导致漏检和误检的问题. 为了提高检测的准确性, 可以通过引入注意力机制来集中处理信息中的关键部分. 注意力机制的应用^[9-11]在多个领域都显示出了其优越性. 为了解决漏检和误检的问题, 提出了一种新型的注意力机制: BiFormer^[12]. 它具有两种特性. 其一就是动态稀疏注意力, 其二就是金字塔结构. 对于动态稀疏注意力的作用, 它仅关注少量令牌, 避免关注其他无关令牌. 金字塔结构可以能够捕捉更多上下文信息, 增强它对一些密集目标处理能力, 实现了较高的检测效率. BiFormer 是 Transformer^[13]模型的一个变体, 通过引入动态注意力机制优化计算分配, 减少计算量. 这一机制主要通过双层路由来实现, 它也是 BiFormer 的特性之一, 在处理复杂任务时, 首先在宏观层面快速筛选, 然后查询与当前最相关的键值, 排除那些与查询不相关的键值, 此过程在称为“路由区域”的部分完成. 经过宏观层筛选后, 在这些区域内采取更细粒度的 Token-to-Token^[14]注意力机制处理. 该机制的核心在于利用稀疏性, 降低了计算和存储需求.

为了明确每个部分之间的注意力关系, 采取有向图的方式, 用于确定每个区域对应的特定部分. 为了优

化这个步骤, 只保留区域中最重要的几个连接, 一旦确定了连接, 接下来就是通过连接建立一个新的路由索引矩阵.

最后, 为了精确控制信息之间的交流, 区域之间的新路由索引矩阵, 实现了细粒度标记注意力机制. 对于在不同区域的每个查询令牌, 注意力机制将聚焦于索引矩阵指定的特定路由区域. 这一策略有效地改善了计算流程, 提升了处理数据时的效率和准确性. BRA 的计算量有 3 个部分: 投影操作, 构建有向图, 标记注意力. 对于任何的二维特征图 $X \in R^{H \times W \times C}$, 将其划分为 $S \times S$ 个不相关区域, 每个区域就包含 HW/S^2 个特征向量, 其中, k 是需要关注的区域数量, C 为通道数. 总的计算量公式如下:

$$\begin{aligned} FLOPs &= FLOPs_{proj} + FLOPs_{routing} + FLOPs_{att} \\ &= 3HWC^2 + 2(S^2)^2 C + 2HWk \frac{HW}{S^2} C \\ &= 3HWC^2 + C \left(2S^4 + \frac{k(HW)^2}{S^2} + \frac{k(HW)^2}{S^2} \right) \\ &\geq 3HWC^2 + 3C \left(2S^4 \cdot \frac{k(HW)^2}{S^2} \cdot \frac{k(HW)^2}{S^2} \right)^{1/3} \\ &= 3HWC^2 + 3Ck^{2/3} (2HW)^{4/3} \end{aligned} \quad (1)$$

当 $2S^4 = \frac{k(HW)^2}{S^2}$, 式 (1) 成立.

BiFormer 使用 BRA 作为核心, 通过 4 层金字塔进行特征融合. 如图 2 所示. 在初级阶段使用重叠补丁嵌入, 为了降低输入空间分辨率和扩充通道数量, 之后的 3 个阶段, 采用了补丁合并技术. 为了转化特征, 采用使用多个 BiFormer 块, 每个 BiFormer 块的结构如图 3 所示, 采用 3×3 卷积核的深度卷积捕捉局部特征, 然后 BRA 模块和 MLP 模块共同工作来建模跨位置的关系并实现逐个位置嵌入.

在 Neck 部分的 ELAN 模块中引入 BiFormer 注意力机制采用参数共享策略, 参数共享降低了模型的参数量, 使得模型更加轻量化, 可以有效提高中小目标检测精度, 通过表 1 可知, 对比其他注意力机制, 在参数上虽然有所增加, 但是在精度方面的优势比较明显.

2.2 改进的特征金字塔结构

由于在实际的交通道路上, 路况复杂, 会出现中小目标检测不到或者漏检的情况. 框架颈部由特征金字塔网络结构 PANet, 目的是整合多尺度的输入特征. PANet 在特征融合的过程对各尺度特征处理方式相同, 对于实际情况, 往往不同输入会导致最后的结果不同.

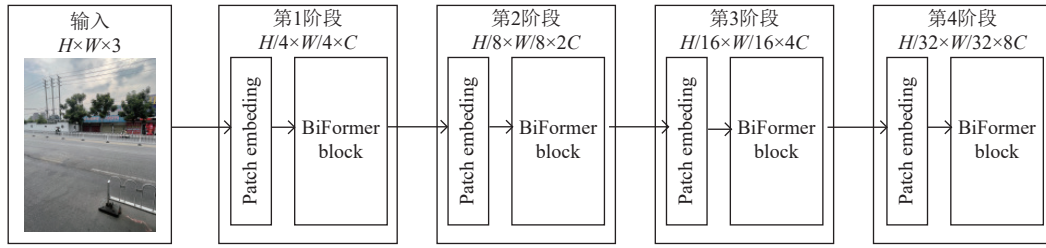


图2 BiFormer 整体框架图

表1 注意力机制对比表

模型	P (%)	$mAP@0.5$ (%)	参数量 (M)
+SE	92.1	93.8	5.76
+CBAM	92.3	94.4	5.76
+SimAM	92.3	94.3	5.75
+BiFormer	92.6	94.6	6.00

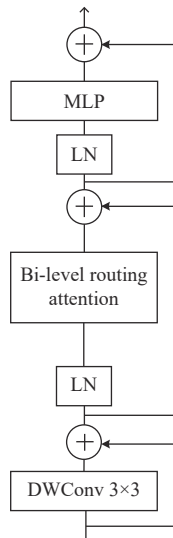


图3 BiFormer 模块详细图

对于一些交通复杂的环境, 速度过快导致目标较小, 出现漏检的情况. 为了实现更高水平的特征融合, 加强多尺度目标特征的融合和传递, 采用 BiFPN^[15] 替换 PANet, BiFPN 删除了两端只有一条边的节点, 在中间的不同层级间引入了双向连接, 这些连接使特征信息在不同层次间流动. 有助于改善不同尺度和层级之间的流动, BiFPN 还引入了训练权重, 用于调整不同输入对特征图的贡献. 原 BiFPN 采用的加权平均融合方式融合特征. 为了更好地捕获和融合高级特征, 将加权平均特征融合的方式改为平均特征融合. 平均特征融合相比加权特征融合更为直接, 减少了模型特征融合中的噪声和过拟合的情况, 有效地强调了对象的重要特征, 将改进的 BiFPN 和原 BiFPN 相结合从而可以适应不同的输入特征, 挖掘深层信息, 提升网络对不同尺

度目标的检测能力. PANet 和 BiFPN 的网络结构如图 4 所示, 图中 P3-P7 代表不同级别的特征融合, 图中包含自上而下的路径, 自下而上的路径, 位于同一级的路径, 通过蓝线可以将高级和低级特征融合, 通过黄线可以实现低级特征的增强, 通过红线可以实现内部信息的交流和补充, 确保信息完整.

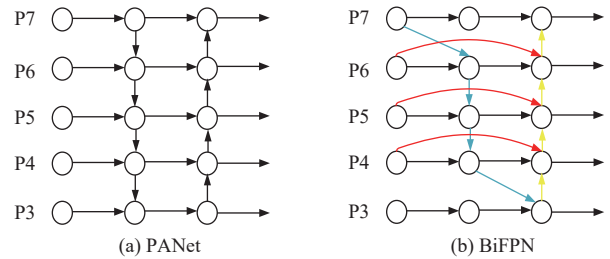


图4 PANet 和 BiFPN 结构图

2.3 改进的 ELAN 模块和 SPPCSPC 模块

高效层聚合网络 (ELAN) 在网络层面上属于梯度路径, 目的就是解决神经网络中的计算效率问题, 它的优点就是使用较少的参数和更少的计算提高性能. 由于原框架的 ELAN 模块每一层的输入是上一层的输出, 每一层接收到上一层输入后, 难以确定哪些是有用的特征, 导致在往后的深层的特征融合效率并不是很高. 为了解决这一个问题, 将原有的 ELAN 改为跳跃连接, 将 ELAN 模块改为跳跃连接的优势在于有助于梯度可以更早地流向网络的早期层, 减少梯度消失的问题, 在参数方面不会有所增加, 但是可以提高参数的使用效率, 从而提升特征提取的能力. 其次在 ELAN 模块中引入全维动态卷积 (ODConv)^[16], 增强网络在复杂背景下检测一些中小型目标的能力^[17]. 它与传统的动态卷积只关注在卷积核维度上的注意力不同, 它在卷积中以并行的方式加入多个维度的互补注意力, 具体的结构图如图 5 所示. 其中 ODConv 对动态卷积的定义为:

$$y = (\alpha_{w1} \odot \alpha_{f1} \odot \alpha_{c1} \odot \alpha_{s1} \odot W_1 + \dots + \alpha_{wn} \odot \alpha_{fn} \odot \alpha_{cn} \odot \alpha_{sn} \odot W_n) * x$$

其中, α_{w1} 为对卷积核 w_i 维度的注意力, α_{c1} 为沿输入通道维度的注意力, α_{s1} 是沿卷积核空间维度的注意力, α_{f1} 为沿输出通道维度的注意力. \odot 表示沿着核空间维度的乘法操作, $*$ 表示卷积运算. x 表示多头注意力模块进行计算, 这 4 类不同的注意力相辅相成, 按照位置, 通道, 过滤器和核的顺序依次乘以卷积核, 使得卷积运算为不同输入 x 的所有空间位置, 输入通道, 过滤器和内核存在差异性, 为捕获丰富的上下文线索提供保证. 改进后的 ELAN 模块和原始 ELAN 模块如图 6 所示.

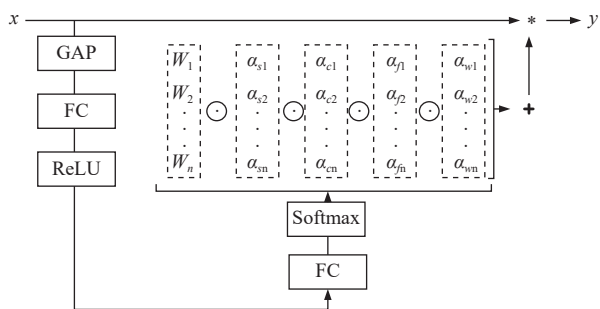


图 5 ODCnv 结构图

空间金字塔池化 SPP 在模型中的作用就是实现不同尺度特征图的融合, 它通过不同的池化操作来聚合特征, 这样可以保留更多的空间信息并增强模型对尺度变化的适应性. CSPNet^[18]主要的作用就是在特征图

的不同阶段引入跨阶段的连接来改善特征重用. 有助于降低成本, 同时保持模型的性能. 结合两者的优点, SPPCSPC 模块在保持轻量化的同时可以增强对多尺度特征的处理能力和信息融合效果, 提高模型检测性能. 图 7 是 SPPCSPC 和改进后的 SPPCSPC 的结构图. 改进后的 SPPCSPC 将原有的并行输出改为了串联式的并行输出, 这样一来, 在学习的过程中细化了对特征的利用, 相比于原结构, 经过多层的最大池化以后, 可以允许网络学习到更多单层学习不到的特征, 同时保留更多的低级和高级特征, 有效减少特征的损失. 改进后的 SPPCSPC 在不引入新参数数量的情况下, 改善了特征提取的能力, 提升检测性能.

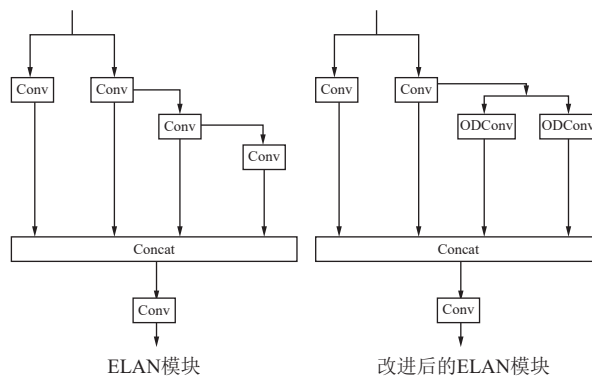


图 6 ELAN 模块改进前后图

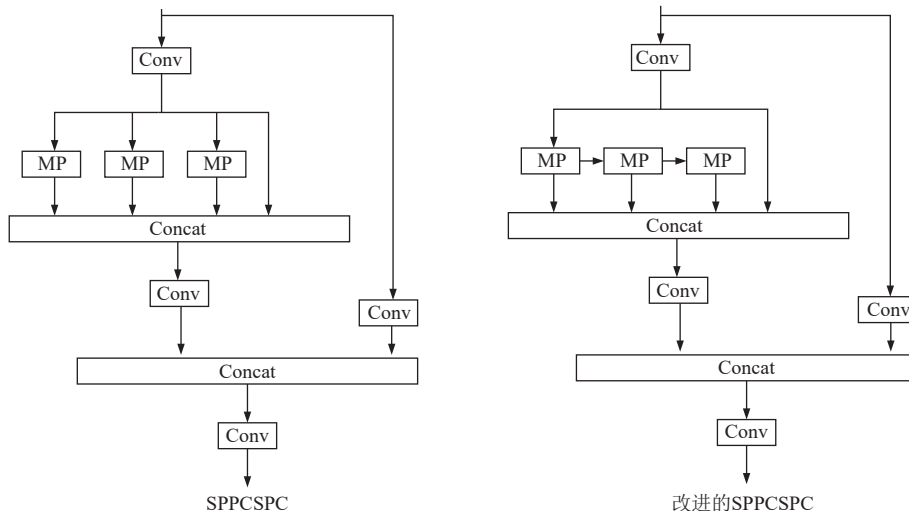


图 7 SPPCSPC 改进前后图

2.4 改进的 YOLOv7-tiny 模型介绍

改进后的 YOLOv7-tiny 模型如图 8 所示, 使用改进的 BiFPN 和原 BiFPN 替换 Neck 中原有的 Concat 模块, 然后采用改进的 SPPCSPC 模块替换原有的

SPPCSPC 模块, 使用改进的 ELAN 模块替换部分的 Neck 部分中的 ELAN 模块, 最后在 Neck 部分中引入注意力机制. 改进后的模型增强了不同背景下对中小违规电动车的识别能力, 提升整体的检测性能.

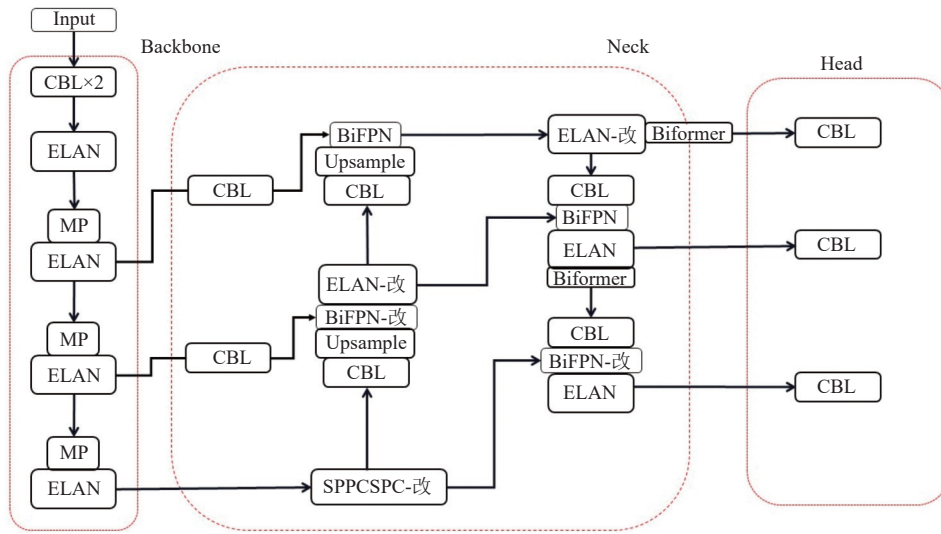


图8 YOLOv7-tiny改进图

3 实验结果与分析

本实验使用 Ubuntu 18.04 操作系统, Inter Core i5-12400 2.5 GHz 处理器, NVIDIA GeForce RTX 3060 12 GB 显卡. 深度学习框架选用 PyTorch 1.13 版本和 CUDA 11.6 版本. 训练参数初始学习率为 0.009 9, 动量值为 0.98, Batchsize 设置为 16, epoch 设置为 200.

3.1 数据集构建

由于目前没有相关数据集, 为了进行相关研究. 本文通过手机进行拍摄电动车违规改装一共 780 张, 通过数据增强将数据集扩展到 4680 张, 并且通过 labeling 进行标注. 数据集以 7:3 的比例生成训练集和验证集. 为了准确评估深度模型在图像上的性能, 本文使用精确率 (precision, P), 平均精度均值 mAP (mean average precision), 参数量, 以及 FPS (frames per second) 用来评估模型检测速度. 其中, P 为预测正样本中真正的正样本数量的比例; mAP 分别为 $mAP@0.5$ 和 $mAP@0.5:0.95$, $mAP@0.5$ 为 IoU 阈值为 0.5 的平均精度, $mAP@0.5:0.95$ 为 IoU 在 0.5–0.95 之间的平均精度. 所有的指标范围均少于 1, 指标可以了解模型的性能, 越高说明模型检测效果越好. 公式如下:

$$P = \frac{TP}{TP + FP} \quad (2)$$

$$R = \frac{TP}{TP + FN} \quad (3)$$

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

$$mAP = \frac{1}{n} \sum_{i=1}^n AP_i \quad (5)$$

其中, P 表示正确识别出来加装雨棚电动车数量占检测出的加装雨棚电动车数量的比值. R 是召回率, 就是被正确检测改装电动车的数量和数据集加装雨棚电动车数量的比值. FP 表示被误检的加装雨棚电动车的数量, FN 表示被误检的非加装雨棚电动车数量. AP_i 表示第 i 个类别的精确率. 该实验只有一个类别.

3.2 实验结果分析

为了更好地验证本文算法的有效性, 在模型中加入了注意力机制, 以及改进后的 BiFPN, SPPCSPC 模块等. 通过以上的改进, 使得模型在精确度, 平均精度 mAP 上有了更好的结果. 图 9 是阈值为 0.5 的交并比结果图. 图 10 是阈值在 0.5–0.95 的结果图, 图 11 就是改进后框架的 $P-R$ 结果图. 通过曲线的对比, 可以知道改进后的框架在平均精度上面已经强于了原有的框架, 体现出良好的检测效果.

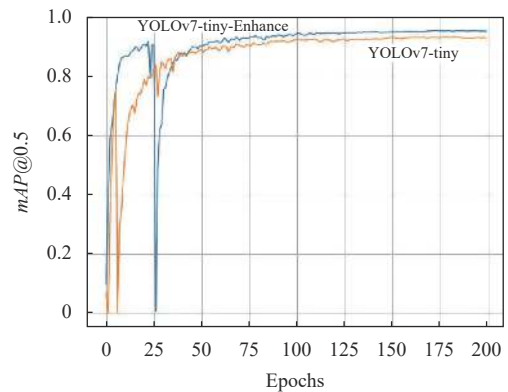


图9 $mAP@0.5$ 对比结果图

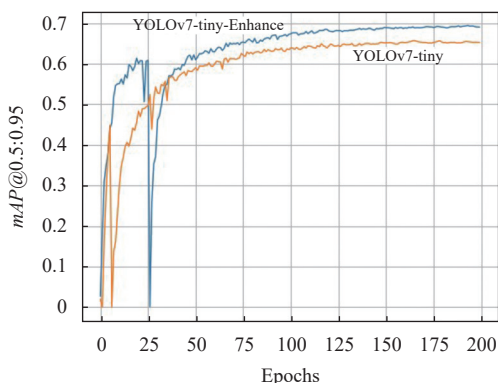


图 10 mAP@0.5:0.95 对比结果图

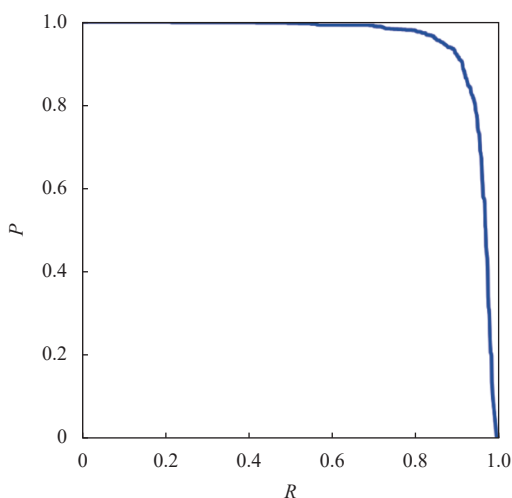


图 11 改进后 P-R 结果图

3.3 检测效果分析

图 12 是原图以及原算法 YOLOv7-tiny 和改进后的 YOLOv7-tiny 算法的效果检测图对比图。

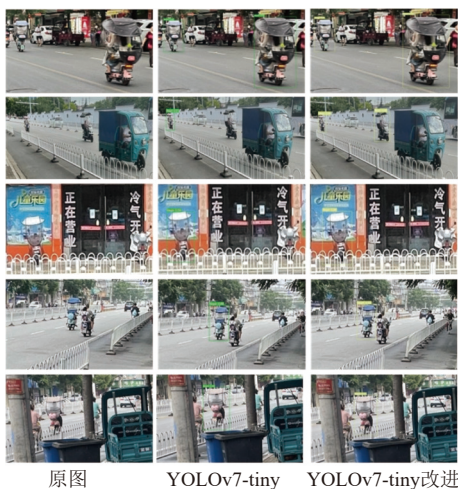


图 12 检测对比结果图

由图 12 可知, 在一些近距离的目标上改进后的 YOLOv7-tiny 算法精度比起原算法拥有更高的精度, 对于一些远距离的中小目标原有的 YOLOv7-tiny 会出现漏检率比较高的情况, 而改进后的 YOLOv7-tiny 算法在对于一些中小目标检测上也有不错的效果, 大大地减少了漏检的概率, 提升了检测性能. 通过对比图可知, 改进后的 YOLOv7-tiny 算法无论在中小目标的检测还是在一些大目标的检测上都要强于原框架.

3.4 对比实验

此外, 本文还将改进后的模型和其他经典的目标检测模型进行比较, 选取精确度、 $mAP@0.5$ 、 $mAP@0.5:0.95$ 、参数量、FPS 进行对比. 数据如表 2 所示. 由表 2 可知, 精确率最高的还是改进后 YOLOv7-tiny 的算法, 改进后的 YOLOv7-tiny 模型的平均精度相比其他所有模型都是最高, 但是检测速度逊色于 YOLOv5 和 YOLOv8. 分析其原因: 加入注意力机制以后, 有助于提高模型检测不同大小的目标. 通过改进 BiFPN 算法以及 BiFormer 模块等影响了对小目标的检测能力, 加强了对复杂背景下目标的识别, 有效提高对目标的识别能力, 但也同时增加了参数量, 导致检测速度下降.

表 2 对比结果实验表

算法	P (%)	mAP@0.5 (%)	mAP@0.5:0.95 (%)	参数量 (M)	FPS
YOLOv7-tiny	92.2	93.2	65.7	5.73	118
YOLOv5s	92.3	94.4	68.9	7.02	104
YOLOv5n	92.2	93.4	65.5	1.71	166
YOLOX-S	90.9	91.0	68.3	8.94	69
SSD	72.5	72.1	52.3	22.6	48
YOLOv8n	90.3	90	64	3.01	163
Ours	94.2	95.2	69.4	6.05	81

3.5 消融实验

为了更好地验证本文算法的有效性, 进行消融实验, 一共 7 组实验. 同样使用上述的数据集进行测试, epoch 设置 200. “√”表示使用该模块, “×”表示未使用该模块. 实验模型的评价如表 3 所示. 在表 3 中可以看出加入注意力机制后精度和平均精度都有提升, 检测速度有所下降, 通过注意力机制, 框架能够提取更多的细节特征, 更好地识别待测物体, 但会引起参数上升. BiFPN 和改进的 BiFPN 相结合能有效提升特征提取能力, 更好地感受周围图像的信息, 提升检测性能, 修改 SPPCSPC 模块以及加入动态卷积以后, 能有效地提高模型拟合的速度, 最终改进的模型虽然在参数量有所上升, 但是检测性能也得到了提升, 也能满足实时检测的需求.

表3 消融结果实验表

No.	BiFormer	BiFPN	SPPCSPC 改进后 和ELAN	BiFPN	P (%)	mAP@ 0.5 (%)	参数量 (M)	FPS
A	×	×	×	×	92.2	93.2	5.73	118
B	√	×	×	×	92.6	94.6	6.00	103
C	×	×	√	×	94.0	94.8	5.91	93
D	×	√	×	√	93.7	95.0	5.73	132
E	√	√	×	×	93.4	94.9	6.00	111
F	√	√	×	√	93.9	95.1	6.00	94
G	√	√	√	√	94.2	95.2	6.05	81

4 结论与展望

本文结合深度学习中的目标检测,结合特定的算法模块替换原有模块,通过 BiFormer 模块增强了原模型的特征提取能力,增强了对小目标的检测能力这有助于提高模型对目标的感知能力和边界的准确性,解决了原算法可能出现未识别到的违规电动车的问题.通过将 BiFPN 和改进的 BiFPN 相互融合,能达到比原有两个更好的检测效果,增强了原模型的泛化性和鲁棒性.最后将改进的 SPPCSPC 模块和 ELAN 模块和原模块相结合,解决了原模型由于卷积损失的特征的问题,提升了模型对中小目标的识别能力.由于注意力机制和动态卷积的引入,导致参数量增加,影响检测速度,今后的研究工作就是,在既能保证精确度的同时,又能保证轻量化^[19].

参考文献

- Girshick R, Donahue J, Darrell T, *et al.* Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014. 580–587.
- Girshick R. Fast R-CNN. Proceedings of the 2015 IEEE International Conference on Computer Vision. Santiago: IEEE, 2015. 1440–1448.
- 吴健生, 王健全, 付美霞, 等. 基于改进 Faster-RCNN 算法的带钢缺陷检测. 鞍钢技术, 2022(6): 23–28, 32.
- 张光华, 李聪发, 李钢硬, 等. 基于改进 YOLOv7-tiny 的无人机航拍图像小目标检测算法. 工程科学与技术, 1–14. <https://doi.org/10.15961/j.jsuese.202300593>. (2023-12-12).
- Liu W, Anguelov D, Erhan D, *et al.* SSD: Single shot multibox detector. Proceedings of the 14th European Conference on Computer Vision. Amsterdam: Springer, 2016. 21–37.
- Wang CY, Bochkovskiy A, Liao HYM. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object

detectors. Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023. 7464–7475.

- 季娟娟, 王佳, 陈亚杰, 等. 基于改进 YOLOv7 的液压阀块表面微小缺陷检测. 计算机工程, 2023, 49(11): 302–310.
- 高源, 杨堃, 曹磊, 等. 基于改进 YOLOv5 的特征抑制多尺度林火检测算法. 林业工程学报, 2023, 8(6): 145–153.
- 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.
- Hou QB, Zhou DQ, Feng JS. Coordinate attention for efficient mobile network design. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 13708–13717.
- Wang QL, Wu BG, Zhu PF, *et al.* ECA-Net: Efficient channel attention for deep convolutional neural networks. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 11531–11539.
- Zhu L, Wang XJ, Ke ZH, *et al.* BiFormer: Vision Transformer with bi-level routing attention. Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023. 10323–10333.
- Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 6000–6010.
- Yuan L, Chen YP, Wang T, *et al.* Tokens-to-token ViT: Training vision Transformers from scratch on ImageNet. Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021. 538–547.
- Tan MX, Pang RM, Le QV. EfficientDet: Scalable and efficient object detection. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 10778–10787.
- Li C, Zhou AJ, Yao AB. Omni-dimensional dynamic convolution. Proceedings of the 10th International Conference on Learning Representations. ICLR, 2022.
- 张梅, 齐至家, 何盼霞. 改进 YOLOv5 的绝缘子缺陷检测算法. 安徽理工大学学报 (自然科学版), 2023, 43(4): 10–18.
- Wang CY, Liao HYM, Wu YH, *et al.* CSPNet: A new backbone that can enhance learning capability of CNN. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Seattle: IEEE, 2020. 1571–1580.
- 苏志威, 黄子涵, 邱发生, 等. 基于改进 YOLOv8 的航空铝合金焊缝缺陷检测方法. 航空动力学报, 2024, 39(6): 116–124.

(校对责编: 孙君艳)