计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
2025,34(1):110−117 [doi: 10.15888/j.cnki.csa.009734] [CSTR: 32024.14.csa.009734]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

复杂条件下交通标识识别①

黄 健, 展 越, 胡 翻

(西安科技大学 通信与信息工程学院, 西安 710600)
通信作者: 展 越, E-mail: zhanyue1126@163.com

摘 要: 该研究旨在深入探究在复杂多变的交通环境下交通标志与信号灯的联合检测问题, 分析并解决恶劣天气、低光照和图像背景干扰等不利因素对检测精度的影响. 为此, 采用了一种改进 RT-DETR 网络的策略. 基于资源有限的运行环境, 并为提高模型对于遮挡以及小目标的检测能力, 提出 PE-ResNet (ResNet with PConv and efficient multi-scale attention) 网络作为主干网络. 为了增强特征融合能力, 提出了 NCFM (new cross-scale feature-fusion module) 模块, 有助于更好地整合图像中的语义信息和细节信息, 对复杂场景的理解更为全面. 最后引入 *MPDIoU* 损失函数, 更精确地衡量目标框之间的位置关系. 改进后的网络相较于基线模型参数量降低了约 14%. 在 CCTSDB 2021 数据集、S2TLD 数据集以及自制的 MTST (multi-scene traffic signs) 数据集上, *mAP*50:95 分别增加了 1.9%、2.2% 和 3.7%. 实验结果表明, 改进之后的 RT-DETR 模型可以有效地改进复杂场景下目标检测精度.

关键词: 目标检测; RT-DETR; 复杂条件; 特征融合; 小目标

引用格式: 黄健,展越,胡翻.复杂条件下交通标识识别.计算机系统应用,2025,34(1):110–117. http://www.c-s-a.org.cn/1003-3254/9734.html

Traffic Sign Recognition Under Complex Conditions

HUANG Jian, ZHAN Yue, HU Fan

(College of Information Engineering, Xi'an University of Science and Technology, Xi'an 710600, China)

Abstract: This study aims to delve into the joint detection of traffic signs and signals under complex and variable traffic conditions, analyzing and resolving the detrimental effects of harsh weather, low lighting, and image background interference on detection accuracy. To this end, an improved RT-DETR network is proposed. Based on a resource-limited operating environment, this study introduces a network, ResNet with PConv and efficient multi-scale attention (PE-ResNet), as the backbone to enhance the model's capability to detect occlusions and small targets. To augment the feature fusion capability, a new cross-scale feature-fusion module (NCFM) is introduced, which facilitates better integration of semantic and detailed information within images, offering a more comprehensive understanding of complex scenes. Additionally, the *MPDIoU* loss function is introduced to more accurately measure the positional relationships among target boxes. The improved network reduces the parameter count by approximately 14% compared to the baseline model. On the CCTSDB 2021 dataset, S2TLD dataset, and the self-developed multi-scene traffic signs (MTST) dataset, the *mAP*50:95 increases by 1.9%, 2.2%, and 3.7%, respectively. Experimental results demonstrate that the enhanced RT-DETR model effectively improves target detection accuracy in complex scenarios.

Key words: object detection; RT-DETR; complex condition; feature fusion; small object

① 基金项目: 陕西省重点研发计划 (2023-YBGY-255); 陕西省科技厅工业攻关 (2022GY-115)
收稿时间: 2024-06-09; 修改时间: 2024-07-10; 采用时间: 2024-07-18; csa 在线出版时间: 2024-11-15
CNKI 网络首发时间: 2024-11-18

在当今社会, 城市化和人口增长导致了交通安全和道路流畅性成为一项重要挑战. 为了解决这些问题, 智能驾驶技术应运而生. 然而, 关于交通标志和信号灯的联合检测的研究和相关数据却相对匮乏, 并且在复杂环境下进行检测与识别也会面临诸多困难. 环境因素, 如雨天、雪天、雾天和低光照环境 (例如夜晚或阴天), 会严重影响图像清晰度, 从而降低了检测的精度. 另一个因素是图像中的目标背景干扰, 包括遮挡、特征干扰和多目标干扰.

近年来, 深度学习和计算机视觉得到极大的发展, 在交通标志检测方面, Wei 等人[1]提出了一种专注于多尺度检测问题的检测器 YOLOF-F (you only look one-level feature fusion), 该方法从单层融合特征中提取多尺度特征信息. 设计了 FFM (特征融合模块) 以融合不同尺度的特征, 并提出了一种新型编码器 CDE (角扩展编码器), 增强特征图中的角点信以及提高位置回归的准确性, 同时保持较快的检测速度. 李禹纬等人[2]在 YOLOv7 算法的基础上进行了创新, 骨干部分采用大核卷积技术以扩大感受野. 在检测部分, 融合了坐标注意力机制和随机池化方法, 这不仅有助于构建通道注意力, 还能精确捕捉对象的位置, 提高网络的泛化性能. 张刚等人[3]开发了一种基于轻量化 SSD 的交通标志检测方法. 使用 MobileNetV3 网络来替换原有的 VGG16 网络, 不仅减少了模型的参数量, 还提高了检测的速率. 通过引入 SE 模块的逆残差结构 B-neck 来替代传统的标准卷积, 增强了低层特征的语义理解. 田鹏等人[4]在面对道路交通场景中交通标志小目标比例高且环境干扰大的问题时, 提出了一种改进的 YOLOv8 交通标志检测算法. 该算法通过引入 BRA (双级路由注意力) 机制来增强网络对小目标的感知能力, 并使用 DCNv3 (可形变卷积 v3) 模块以更好地处理特征图中的不规则形状, 提高对遮挡和重叠目标的检测能力. 此外, 算法还引入了基于辅助边框的 Inner-IOU 损失函数.

在交通信号灯检测方面, Yao 等人[5]提出了一种名为 TL-detector 的轻量级实时交通灯检测模型. 设计了增强主干网络架构, 以生成丰富的信息并降低计算负载. 引入坐标注意力机制以关注位置特征并增强特征提取能力, 采用轻量级的特征融合网络以促进多尺度特征信息的聚合和融合. Liu 等人[6]考虑到实际路境中交通灯尺寸较小且背景复杂, 对 YOLOv5 模型进行了调整, 引入了 P2 层以提取更细致的特征. 受到 MOAT 模型的启发, 他们设计了一种新 C3_MOAT 结构. 同时, 通过在 YOLOv5 检测头部引入结构重参数化的 RepConv 模块, 进一步加强了模型的特征提取能力. 郑岚月等人[7]提出了一种改进的 YOLOv7 算法, 去除了 20×20 的检测尺度, 并增加了 160×160 的尺度, 旨在使模型更轻便的同时, 增加更浅层的特征. 结合 BiFormer 技术中的 BRA (bi-level routing attention) 和坐标轴注意力机制, 针对交通信号灯的位置特性设计了 ABRA (axially-guided bi-level routing attention) 模块. 引入 NWD (normalized Wasserstein distance) 度量方法, 以改进物体定位损失和置信度损失的计算.

以上基于 CNN 的目标检测方法尽管提高了检测精度, 但在复杂环境下的检测性能仍有待提升. 为了解决这一问题, 本文采用基于 Transformer 的 RT-DETR[8] 网络作为检测模型. Transformer 利用其独特的自注意力机制来模拟输入序列中各元素间的相互关系, 从而高效地提取和整合全局上下文信息, 在复杂场景中取得更好的性能. 作为首个实时端到端目标检测模型, RT-DETR 在本文经过进一步的改进, 采用 PE-ResNet 作为轻量化的主干网络, 并设计了特征融合 NCFM (new cross-scale feature-fusion module) 模块, 以增强模型在复杂环境下的适应性. 此外, 本文还对损失函数进行了优化, 采用 *MPDIoU*[9]损失函数. 这些改进使得模型在处理复杂环境时表现更加出色.

1 改进的 RT-DETR 结构

1.1 RT-DETR 结构

RT-DETR 采用 Transformer 架构来建模对象之间的依赖关系, 提供更好的上下文理解能力. RT-DETR 由 3 个关键组件构成: 主干网络、颈部网络和头部网络. 可选择经典的 ResNet 或者可缩放的 HGNetV2 来作为主干网络, 其输出 S3, S4, S5 这 3 种尺寸的特征图. 在颈部网络中, RT-DETR 使用了一层 Transformer 的 efficient hybrid encoder, 其包括两个特征融合模块: attention-based intra-scale feature interaction (AIFI) 和 CNN-based cross-scale feature-fusion module (CCFM). 头部网络为 IoU-aware query selection 和 decoder 两个部分. 根据本文所描述的复杂道路场景的具体需求, 我们对 RT-DETR 网络进行了针对性的改进, 改进后的网络结构如图 1. 图 1 中的 NCFM 是对 CCFM 模块进行改进得到的.

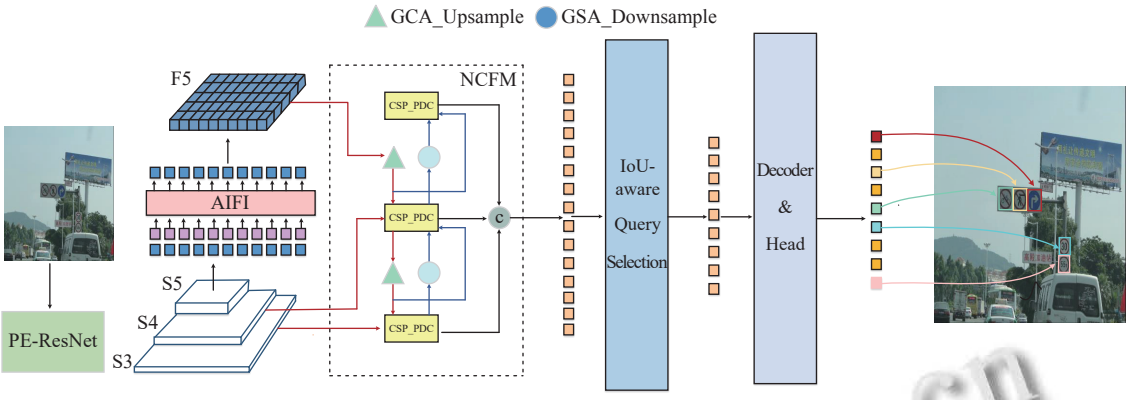


图 1 改进后模型结构图

1.2 PE-ResNet 主干网络

在处理复杂的特征提取任务时, 传统的 ResNet18 网络常常受限于其基本的卷积结构, 这在一定程度上影响了目标检测的精度. 为了克服这一限制, 本文设计了 PE-ResNet 网络, 旨在增强模型应对复杂环境的能力. 改进后的网络结构能够更深入挖掘和利用复杂特征, 从而显著提升了目标检测的整体性能. 如图 2 所示, 该网络通过引入 PConv (partial convolution)[10]模块和 EMA (efficient multi-scale attention)[11]注意力机制, 重新定义了特征提取的过程.

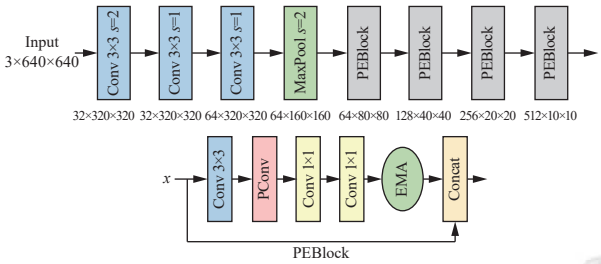


图 2 PE-ResNet 网络结构图

为了更有效地提取特征, 我们使用 3 组 3×3 卷积操作, 并配合最大池化来优化通道数量. 连续使用 4 组 PEBlock 以进一步提取特征图上的信息. 在该网络中, PEBlock 扮演着核心角色. 该模块首先利用一个 3×3 卷积层进行基础的特征抽取, 然后通过 PConv 聚焦于图像的关键区域处理, PConv 只对数据的有效区域进行卷积处理以提取空间特征, 而其他通道保持不变. 在处理遮挡的图像时能有效提取有价值的信息. 为了更精确地细化特征表示, 使用两个 1×1 的卷积层在维持特征图空间维度的同时, 有效调整特征通道并实现特征的变换. 在此基础上, 我们整合了 EMA 注意力机制来增强模型对不同尺度特征的识别, 该机制通过两个 1×1 的卷积分支和一个 3×3 的卷积分支来获取分组特征图中的注意力权重, 实现对多尺度信息的有效整合, 并增强了模型捕捉长距离特征依赖性的能力. 相较于传统的 ResNet 网络, 本研究提出的架构在应对复杂环境任务以及提高特征提取效率方面展现了显著的性能提升.

1.3 NCFM 特征融合模块

为了显著提升网络的特征提取能力, 本文提出了 NCFM 模块. 其由以下几部分组成.

1.3.1 CSP_PDC 融合模块

CSP_PDC 模块结合了 CSP (cross stage partial)[12] Bottleneck 和并行空洞卷积的优势, 在图像处理任务中展现出了良好的性能. CSP_PDC 模块的结构如图 3 所示.

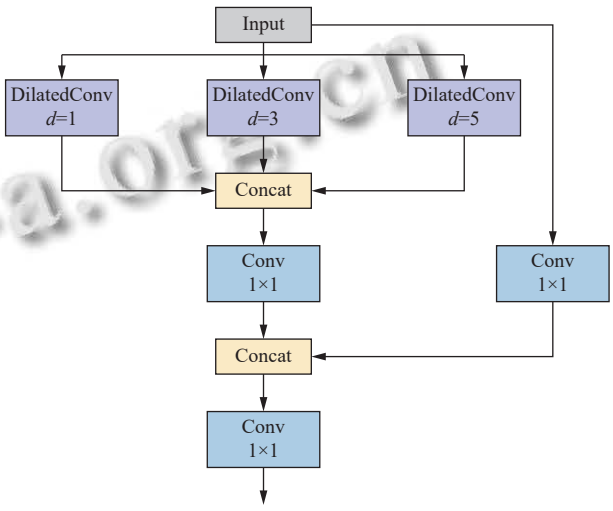


图 3 CSP_PDC 结构图

在该模块中, 我们采用了双分支策略进行特征提取和处理. 第 1 条分支使用并行空洞卷积操作, 接收一个输入张量 x, 并对其进行 3 个不同空洞率的空洞卷积操作, 空洞率依次为 1, 3, 5. 然后将卷积结果在通道维度上进行拼接, 并通过 1×1 的卷积层将通道数调整为初始值, 该操作增强了特征的感受野, 可以捕获更广泛

范围内的特征信息. 第 2 条分支通过一个 1×1 卷积保留了原始特征的部分细节信息. 通过将这两条分支在通道维度上进行拼接, 实现了特征信息的融合, 既保留了原始特征的细节信息, 又通过并行空洞卷积获得了更丰富的特征表示, 从而提高了模块的特征表达能力和网络整体的性能.

1.3.2 GCA_Upsample 上采样模块

在 RT-DETR 网络中, 上采样过程采用步长为 1 的 1×1 卷积, 但受限于卷积核大小, 无法充分捕捉更广泛的空间特征, 导致上采样后的特征缺乏全局一致性. 为了解决这一问题, 本文提出了 GCA_Upsample (gated channel attention upsample) 模块, 如图 4 所示.

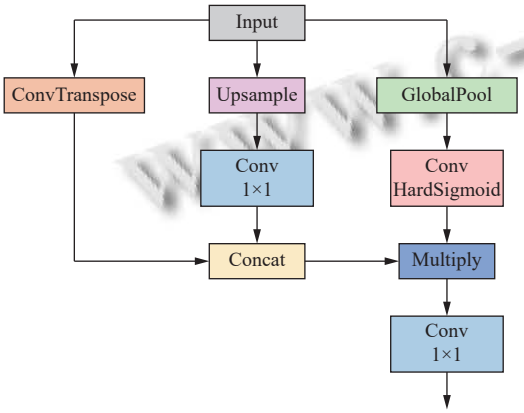


图 4 GCA_Upsample 结构图

通过引入并行上采样分支, 为网络提供多条特征提取路径. 其中一条分支通过转置卷积实现高质量的上采样, 能够捕获更广泛的空间特征, 避免信息丢失, 提升上采样后特征的质量. 另一分支通过 upsample 和 1×1 卷积扩展特征图尺寸, 增强特征表达能力, 减少信息丢失, 提高上采样任务的性能. 两个分支的通道数各为总通道数的一半, 以降低参数量. 通过结合门控通道注意力进行特征选择, 门控通道注意力关注不同通道之间的关系, 有助于网络更好地理解不同特征通道之间的相关性和重要性. 因此, 将其用于上采样过程可以帮助网络更好地捕获和利用高层次的语义信息, 从而提高特征图的分辨率.

1.3.3 GSA_Downsample 下采样模块

GSA_Downsample (gated spatial attention downsample) 模块以优化下采样过程为目标, 与上采样过程类似, 采用并列的特征提取分支. 如图 5 所示在减半通道数的前提下, 其中一条分支采用步长为 2 的 3×3 卷积, 另一条分支则先通过最大池化层对输入特征图进行下采样操作, 从而减小特征图的尺寸. 随后, 经过 1×1 的卷积层 (Conv) 处理, 以调整通道数、优化特征表示并减少参数量. 通过结合门控空间注意力, 网络可以更好地捕获不同位置之间的相关性, 有助于在减少分辨率的同时保留更多的上下文信息, 提高网络对广泛特征的感知能力. GSA_Downsample 模块这一设计旨在弥补传统 3×3 卷积下采样可能带来的细节损失和计算量增加问题, 提高了下采样特征的质量和效率.

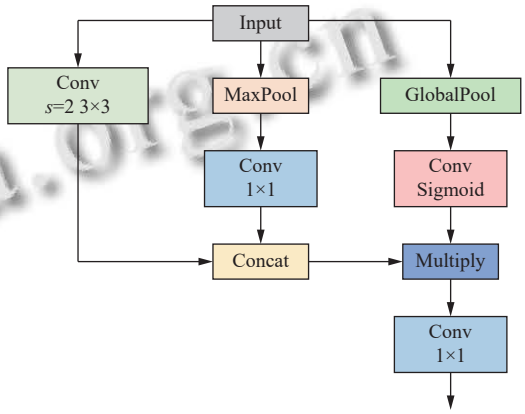


图 5 GSA_Downsample 结构图

NCFM 巧妙地整合了 CSP_PDC 融合模块、GCA_Upsample 上采样模块和 GSA_Downsample 下采样模块, 实现了跨层次特征图的深度融合. 这一融合过程使得高层的抽象特征与低层的细节特征得以相互补充, 共同构建出更为全面和细致的图像表示.

1.4 MPDIoU

RT-DETR 使用了 GIoU (generalized intersection over union) 作为损失函数, GIoU 在检测交通标识时存在对形状适应性较差、对遮挡和重叠情况处理不足以及对定位误差敏感的问题. 并且对每个预测框与真实框均要去计算最小外接矩形, 计算及收敛速度受到限制. 所以本文使用 *MPDIoU* 替代 GIoU 更精准的完成检测. 如图 6 所示, *MPDIoU* 提出了一种新颖的边界框相似性评估方法, 该方法基于最小点对距离进行度量, 直接优化了预测边界框与真实标注边界框之间左上角和右下角点的欧几里得距离. 该度量标准综合了现有损失函数关注的所有关键要素, 包括框的重叠区域、中心点的距离以及宽高的差异, 同时简化了计算过程. 公式如下所示:

$$IoU = (A,B) = \frac{A \cap B}{A \cup B} \tag{1}$$

$$d_1^2 = \left(x_1^{\text{prd}} - x_1^{\text{gt}}\right)^2 + \left(y_1^{\text{prd}} - y_1^{\text{gt}}\right)^2 \tag{2}$$

$$d_2^2=\left(x_2^{\mathrm{prd}}-x_2^{\mathrm{gt}}\right)^2+\left(y_2^{\mathrm{prd}}-y_2^{\mathrm{gt}}\right)^2 \tag{3}$$

$$MPDIoU=IoU-\frac{d_2^1}{w^2+h^2}-\frac{d_2^2}{w^2+h^2} \tag{4}$$

$$\mathcal{L}_{MPDIoU}=1-MPDIoU \tag{5}$$

使用 *MPDIoU* 损失函数能够更准确地衡量目标框之间的位置关系, 提高检测模型在图像目标背景干扰中的鲁棒性和准确性.

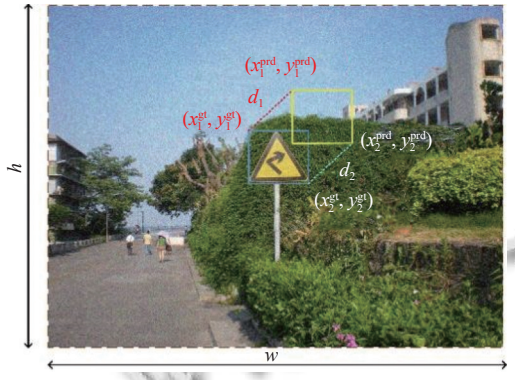


图 6 *MPDIoU* 损失函数

2 实验分析

2.1 实验环境

本实验在 NVIDIA RTX3080 上运行, 深度学习框架使用 PyTorch 1.12.1, 开发环境为 Python 3.8. 输入图片尺寸为 640×640, epoch 为 120, 采用 AdamW 优化器, 初始学习率为 0.000 1, momentum 为 0.9.

2.2 数据集

2.2.1 CCTSDB 2021 数据集

CCTSDB 2021[13]是长沙理工大学团队根据 CCTSDB 2017 进行改进的中国交通标志数据集. 其包含指示标志 (mandatory), 禁止指令 (prohibitory), 警告标志 (warning) 这 3 大种类的交通标志.

2.2.2 S^2LSD 数据集

S^2TLD[14]数据集是由上海交通大学开源的交通信号灯数据集, 其由 4 563 张分辨率为 720×1280 的图片组成. 包括红灯 (red)、黄灯 (yellow)、绿灯 (green)、关闭 (off) 这 4 大种类.

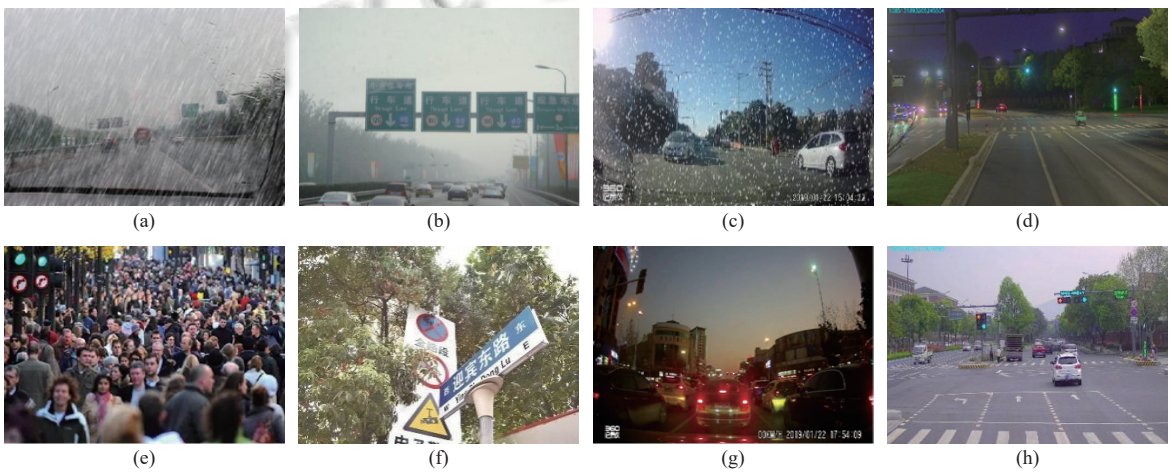
2.2.3 MTST 数据集

为了解决缺乏同时标记交通标志和信号灯的数据集的问题, 我们在交通标志数据集 CCTSDB 2021 的基础上, 利用信号灯数据集 S^2TLD 中的部分图片来增加信号灯的样本. 同时, 采用图像预处理技术增加了各 1 000 张雨天、雪天和雾天的图片. 随后, 我们利用标注工具 labelimg 对扩充后的数据集进行了重新标注, 标注的类别示例如图 7 所示. 经过以上处理, 我们得到了一个完整的数据集. 该数据集不仅包括交通标志和信号灯, 还涵盖了如图 8 所示的包含正常交通环境、恶劣天气、低光照环境以及目标背景干扰场景的图片, 符合本文的检测背景. 将这个数据集命名为 MSTS (multi-scene traffic signs), 即多场景交通标识数据集, 并按照 6:2:2 的比例划分为训练集, 验证集和测试集, 得到训练集 12 801 张, 验证集和测试集各 4 267 张.

(a) 禁止标志 (b) 警告标志 (c) 指示标志

(d) 红灯 (e) 绿灯 (f) 黄灯

图 7 MTST 数据集分类示例



(a) (b) (c) (d)

(e) (f) (g) (h)

图 8 MTST 数据集图片示例

2.3 评价指标

本文采用了多个目标检测评价指标进行验证, 包括准确率 (*P*)、召回率 (*R*)、平均精度均值 (*mAP*50 和 *mAP*50:95)、参数量 (Params)、速度 (FPS).

$$P = \frac{TP}{TP+FP} \tag{6}$$

$$R = \frac{TP}{TP+FN} \tag{7}$$

$$mAP = \frac{\sum AP}{N(\text{class})} \tag{8}$$

AP 值为 *PR* 曲线下面的面积, 可以反映每个类别预测的准确率. *mAP* (mean of average precision) 是对所有类别的 *AP* 值求平均值. *mAP*50 表示在 *IoU* 阈值为 0.5 的情况下 *mAP* 的值. *mAP*50:95 表示在阈值区间 [0.5, 0.95], 步长为 0.05 的平均 *mAP* 值. FPS 在目标检测领域用来衡量模型在实时场景中处理图像的速度.

2.4 实验结果分析

为验证本文模型的泛化性能, 我们在公共数据集 CCTSDB 2021 交通标志数据集和 S^2LTD 信号灯数据集对模型进行测试. 表 1 和表 2 展示了验证集图片在优秀算法, RT-DETR 以及本文提出的模型进行的对比.

表 1 CCTSDB 2021 数据集上的实验结果对比

模型	*P* (%)	*R* (%)	*mAP*50 (%)	*mAP*50:95 (%)	参数量 (M)
文献[2]	93.1	81.9	87.5	—	32
文献[3]	—	—	89.0	—	28.1
RT-DETR	91.2	**82.9**	88.9	55.9	19.8
本文模型	**93.3**	82.7	**89.8**	**57.8**	16.9

注: 加粗数据为最优数据

表 2 S^2TLD 数据集上的实验结果对比

模型	*P* (%)	*R* (%)	*mAP*50 (%)	*mAP*50:95 (%)	参数量 (M)
文献[6]	—	—	95.3	60.1	6.71
文献[7]	—	—	95.5	—	47.5
RT-DETR	96.7	93.7	95.5	65.2	19.8
本文模型	**98.2**	**94.7**	**96.3**	**67.4**	16.9

注: 加粗数据为最优数据

根据表 1 的数据, 在保持参数量最少的情况下, 本文提出的模型在 CCTSDB 2021 数据集上的 *mAP*50 指标相较于对比算法分别提升了 0.8%, 2.3% 和 0.9%. 进一步地, 表 2 所示的实验结果表明, 在 S^2LTD 数据集上, 与基础网络相比, 我们的模型实现了 14% 的参数量减少, 同时在准确率和 *mAP*50 和 *mAP*50:95 指标上分别取得了 1.5%、1.8% 和 2.2% 的提升. 此外, 在文献[7]模型的参数量是本文模型参数量的 2.8 倍的情况下, 本文模型的 *mAP*50 指标上实现了 0.8% 的提升; 而与文献[6]相比较, *mAP*50:95 值提高了 7.3%.

2.5 消融实验

为验证改进方法对 RT-DETR 网络的有效性, 对上述的改进在 MTST 数据集上进行消融实验以证明改进方法的有效性和必要性 (见表 3). √表示在该模型中使用了该方法, 第 1 行表示使用初始 RT-DERT 网络. 结果表明, 每一个模块都可以有效提升检测.

表 3 消融实验表

PE-ResNet	NCFM	*MPD-IoU*	*P* (%)	*R* (%)	*mAP*50 (%)	*mAP*50:95 (%)	参数量 (M)
—	—	—	86.4	86.7	86.8	52.1	19.8
√	—	—	86.8	87.9	87.8	53.8	17.1
√	√	—	88.5	90.0	90.8	55.3	16.9
√	√	√	89.1	90.3	91.0	55.8	16.9

2.6 MTST 数据集对比实验

在相同的环境配置下, 本文选择了基于 CNN 的双阶段网络 Faster R-CNN, 单阶段网络 SSD、YOLOv5、YOLOv7 和 YOLOv8, 以及基于 Transformer 的 DETR、Deformable DETR 和 DINO 网络模型, 在 MTST 数据集上进行对比实验, 结果见表 4. 实验数据表明本文模型在参数量最少的基础上, 召回率 (*R*)、*mAP*50、*mAP*-50:95 以及 FPS 值均高于其他对比网络, 有效地提升了系统在复杂场景下实时且精准地检测交通标识的性能.

表 4 MTST 数据集上对比实验结果

模型	参数量 (M)	FPS (bs=1)	*P* (%)	*R* (%)	*mAP*50 (%)	*mAP*50:95 (%)
Faster R-CNN[15]	108	4	72.3	51.6	69.1	42.3
SDD[16]	91.6	18	73.5	49.7	68.4	39.6
YOLOv5m	21.2	145.1	85.8	84.8	87.3	50.1
YOLOv7[17]	37.2	22	73.3	67.3	69.8	36.4
YOLOv8m	25.9	148.7	**89.3**	88.4	90.1	53.5
DETR[18]	41	—	78.1	76.2	78.4	44.9
Deformable DETR[19]	40	—	82.3	80.6	82.5	47.7
DINO[20]	47	5	84.5	83.2	86.1	49.5
RT-DETR	19.8	179.6	86.4	86.7	86.8	52.1
本文模型	16.9	**195.1**	89.1	**90.3**	**91.0**	**55.8**

注: 加粗数据为最优数据, bs=1代表batch=1

2.7 检测结果对比

为了直观展示改进后模型在应对复杂环境时的性能提升, 本研究选取了 MTST 数据集中的一部分图片, 并结合从网络上采集的样本图片进行了检测比较分析. 图 9 为在恶劣天气下两种模型的检测结果, 图 10 为在低光照环境下的检测结果, 图 11 为在目标背景干扰下的检测结果. 目标背景干扰包括遮挡, 多目标干扰和特

征干扰. 多目标干扰是在图像中同时存在多个无关的目标或特征, 特征干扰指的是图像中不同对象或特征之间的相似性造成的干扰, 例如红灯和汽车尾灯. 检测结果表明本文模型不仅提升了在复杂场景下普通尺寸的目标和小目标的检测精度, 还解决了原网络在图 9(a) 和 (b) 中出现的漏检错检的问题.

(a) 雨天 (b) 雪天 (c) 雾天 (小目标)

图 9 恶劣天气下的检测结果对比

(a) RT-DETR (b) 本文模型

图 10 低光照下小目标检测结果对比

(a) 遮挡 (b) 多目标干扰 (c) 特征干扰 (小目标)

图 11 目标背景干扰情况下的检测结果对比

3 结束语

本文针对复杂条件下交通标志和信号灯联合检测任务, 在 RT-DETR 基线模型的基础上进行改进. 通过替换主干网络为 PE-ResNet 网络, 我们成功将模型在轻量化的基础上提升了在复杂条件下对全局细节信息的提取能力以及对于普通目标和小目标的检测能力.

同时, 我们对跨尺度特征融合模块进行了改进, 构建了 NCFM 模块. 该模块结合添加了门控通道注意力的上采样和门控空间注意力的下采样结构, 并在 fusion 模块中通过使用并列的 3 个空洞卷积来增大网络的感受野, 极大地增加了在复杂道路场景下的性能. 此外, 我们还使用 *MPDIoU* 替代 GIoU 以提高模型的收敛速度和检测框的定位精度. 本文在 MTST 数据集, CCTSDB 2021 数据集以及 S^2TLD 数据集中进行了大量的实验, 结果显示, 我们提出的模型在复杂条件下同时检测交通标志和信号灯取得了显著效果. 具体而言, *mAP*50、*mAP*50:95 和 FPS 均得到了提高, 表明了我们模型在应对复杂场景下的有效性. 在未来的研究中, 我们将进一步优化模型, 旨在保持检测精度的同时减少参数量, 以更好地适用于车载系统等资源受限的场景.

参考文献

1 Wei HY, Zhang QQ, Qin YG, *et al*. YOLOF-F: You only look one-level feature fusion for traffic sign detection. The Visual Computer, 2024, 40(2): 747–760. [doi: 10.1007/s00371-023-02813-1]

2 李禹纬, 付锐, 刘帆. 改进 YOLOv7 的轻量化交通标志检测算法. 太原理工大学学报, 2024, 55(1): 195–203.

3 张刚, 王运明, 彭超亮. 基于轻量化 SSD 的交通标志检测算法. 实验技术与管理, 2024, 41(1): 63–69.

4 田鹏, 毛力. 改进 YOLOv8 的道路交通标志目标检测算法. 计算机工程与应用, 2024, 60(8): 202–212. [doi: 10.3778/j.issn.1002-8331.2309-0415]

5 Yao ZK, Liu Q, Xie Q, *et al*. TL-detector: Lightweight based real-time traffic light detection model for intelligent vehicles. IEEE Transactions on Intelligent Transportation Systems, 2023, 24(9): 9736–9750. [doi: 10.1109/TITS.2023.3267430]

6 Liu PL, Xie ZY, Li TJ. Traffic light detection based on improved YOLOv5. Proceedings of the 8th International Conference on Intelligent Computing and Signal Processing (ICSP). Xi'an: IEEE, 2023. 1891–1895.

7 郑岚月, 张玉洁. 基于改进 YOLOv7 的交通信号灯检测. 系统仿真学报, 1–13. https://doi.org/10.16182/j.issn1004731x.joss.23-1562. (2024-03-08) [2024-04-01].

8 Zhao YA, Lv WY, Xu SL, *et al*. DETRs beat YOLOs on real-time object detection. arXiv:2304.08069, 2023.

9 Ma SL, Xu Y. MPDIoU: A loss for efficient and accurate bounding box regression. arXiv:2307.07662, 2023.

10 Chen JR, Kao SH, He H, *et al*. Run, Don't walk: Chasing higher FLOPS for faster neural networks. Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023. 12021–12031.

11 Ouyang DL, He S, Zhang GZ, *et al*. Efficient multi-scale attention module with cross-spatial learning. Proceedings of the 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Rhodes Island: IEEE, 2023. 1–5.

12 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.

13 Zhang JM, Zou X, Kuang LD, *et al*. CCTSDB 2021: A more comprehensive traffic sign detection benchmark. Human-centric Computing and Information Sciences, 2022, 12: 23.

14 Yang X, Yan JC, Liao WL, *et al*. SCRDet++: Detecting small, cluttered and rotated objects via instance-level feature denoising and rotation loss smoothing. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(2): 2384–2399. [doi: 10.1109/TPAMI.2022.3166956]

15 Ren SQ, He KM, Girshick R, *et al*. Faster R-CNN: Towards real-time object detection with region proposal networks. Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal: ACM, 2015. 91–99.

16 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.

17 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.

18 Carion N, Massa F, Synnaeve G, *et al*. End-to-end object detection with Transformers. Proceedings of the 16th European Conference on Computer Vision. Glasgow: Springer, 2020. 213–229.

19 Zhu XZ, Su WJ, Lu LW, *et al*. Deformable DETR: Deformable Transformers for end-to-end object detection. Proceedings of the 9th International Conference on Learning Representations. ICLR, 2021.

20 Zhang H, Li F, Liu SL, *et al*. DINO: DETR with improved denoising anchor boxes for end-to-end object detection. Proceedings of the 11th International Conference on Learning Representations. Kigali: ICLR, 2023.

(校对责编: 孙君艳)