

基于改进 RT-DETR 的公路路面交通标识检测算法

孙海青¹ 杨传颖¹ 敖乐根²

(1. 内蒙古工业大学信息工程学院 呼和浩特 010080; 2. 内蒙古灵奕高科技(集团)有限责任公司 呼和浩特 010010)

摘 要: 为了降低复杂天气、光照变化和交通标识污损等环境因素干扰和算法自身复杂度高导致模型性能的下降,提出一种以 RT-DETR 为基准模型的公路路面交通标识检测算法。首先,使用设计的轻量化模块 CSP-PMSFA 作为算法的主干网络,降低模型的计算量和参数量,提高模型的表达能力。然后,针对原有算法尺度内特征交互模块计算量大、模型容量小且计算效率有限等问题,引入了级联分组注意力 CGA 进行改进。最后,设计了跨尺度特征融合模块 EMBSFPN,用于解决感受野适应性不足,特征信息处理有限的问题;采用 EUCB 上采样模块的解码机制在保证准确度的情况下保留和融合特征信息,优化性能的同时提高模型的鲁棒性。实验结果表明,改进算法在 ROAD MARK 路面标识检测数据集上与原有算法相比 mAP50 提高了 2.6%、FPS 提高了 13.6 帧,计算量和参数量分别下降了 29.7% 和 39.6%,整体优于其他改进算法,在进行轻量化的同时提高了检测精度和速度,具有实用性。

关键词: 交通标识;RT-DETR;深度学习;PartialConv 高效卷积;注意力机制;特征融合;EUCB 上采样

中图分类号: TP391.4; TN919.8 **文献标识码:** A **国家标准学科分类代码:** 520.6040

Highway traffic sign detection algorithm based on improved RT-DETR

Sun Haiqing¹ Yang Chuanying¹ Ao Legen²

(1. School of Information Engineering, Inner Mongolia University of Technology, Hohhot 010080, China; 2. Inner Mongolia Lingyi High Tech (Group) Co., Ltd., Hohhot 010010, China)

Abstract: In order to reduce the interference of environmental factors such as complex weather, lighting changes, and traffic sign fouling, as well as the degradation of model performance caused by the high complexity of the algorithm itself, a highway road traffic sign detection algorithm based on RT-DETR as the benchmark model is proposed. Firstly, the designed lightweight module CSP-PMSFA is used as the backbone network of the algorithm to reduce the computational and parameter complexity of the model and improve its expressive power. Then, to address the issues of high computational complexity, small model capacity, and limited computational efficiency in the feature interaction module within the original algorithm scale, cascaded group attention CGA was introduced for improvement. Finally, a cross scale feature fusion module EMBSFPN was designed to address the issues of insufficient receptive field adaptability and limited feature information processing; the decoding mechanism using EUCB upsampling module preserves and fuses feature information while ensuring accuracy, optimizing performance and improving the robustness of the model. The experimental results show that the improved algorithm has improved mAP50 by 2.6% and FPS by 13.6 frames compared to the original algorithm on the ROAD MARK road marking detection dataset. The computational and parameter requirements have decreased by 29.7% and 39.6%, respectively. Overall, it outperforms other improved algorithms and improves detection accuracy and speed while being lightweight, demonstrating practicality.

Keywords: traffic signs; RT-DETR; deep learning; PartialConv efficient convolution; attention mechanism; feature fusion; EUCB upsampling

0 引 言

随着现代交通技术的不断升级和自动驾驶技术的飞速

发展,公路路面交通标识检测精度的高低和速度的快慢对保障行车安全和提高交通效率至关重要。然而,在实际交通环境中,受雨雪天气的影响、光照及车流量变化等外界因

素干扰往往会导致路面交通标识发生变形和模糊,影响算法的实时性和准确性。因此,对公路路面交通标识进行有效检测和评估变得更为重要^[1]。

目前,公路路面交通标识检测算法实现了从传统检测算法到深度学习检测算法的过渡。传统检测算法较为简单、直观、但容易被光照、噪声等因素干扰,在面对路面交通标识变形、遮挡、褪色等干扰时存在检测的限制^[2]。深度学习技术的发展,使得基于卷积神经网络(CNNs)和基于Transformer的检测算法^[3]逐渐占据主导地位。基于卷积神经网络的检测算法又分为以YOLO(you only look once)^[4]、SSD(single shot multibox detector)^[5]为代表的一阶段检测算法和以Fast R-CNN(fast region convolutional neural network)^[6]和Faster R-CNN^[7]为代表的两阶段检测算法。而基于Transformer的检测算法主要以DETR(detection transformer)^[8]及其改进版本为主。在进行目标检测任务时,基于卷积神经网络的检测算法需要进行后处理操作,容易造成漏检和误检的问题,而基于Transformer的检测算法因具有实时端到端检测的优势得到广泛关注。尤其是开源深度学习平台百度飞桨在2023年提出的RT-DETR(real time detection transformer)^[9],一种基于Transformer的实时端到端检测器,为解决因消除非极大值抑制(non-maximum suppression, NMS)^[10]造成计算成本高,实用性低的问题提供了具体方案。

近年来的交通标识检测研究中,闫世洋等^[11]为解决交通标志检测存在的准确率低、参数量大等问题,提出了SC-YOLOv8交通标志检测算法。使用Adown改进下采样并设计了C2f_SC模块降低模型参数,同时使用Wise-MPDIoU替换原损失函数,虽然改进模型的整体性能得到提升,但在密集交通和敏感变换环境下其检测效果不佳。为了解决现有算法对车门内拉手表面小目标、多尺度、易反光等缺陷检测时精度低和速度慢的问题,徐仟祥等^[12]通过将并行膨胀卷积与CA注意力机制结合的方式应用于RT-DETR模型的主干网络中,同时添加额外的检测层并将改进的BIFPN应用于多尺度特征融合。虽然改进模型的精度和速度得到提升,但仍需大量的内存和计算资源,这使得其部署起来较为困难。针对背景干扰对路面缺陷检测影响较大的情况,周建新等^[13]等提出YOLOv8 DRM模型,设计了C2f-Dysnake模块和RDFPN模块,用于提高特征提取能力和小目标检测精度,同时,在主干网络引入MPCA注意力机制,用于捕获不同尺度的特征信息。改进模型虽然提升了精度,但轻量化程度不高,实用价值有限。Xia等^[14]提出了用于交通标识多尺度检测的DSRA-DETR模型,将膨胀空间金字塔模型思想和多尺度特征残差聚合模块融入到DSRA-DETR中,用于聚合各种尺度特征并减少特征噪声和低级特征的损失。但由于DETR的计算成本较高,该模型仍需占用较大的内存和计算资源,导致扩展性受限,运行成本增加。许小伟等^[15]通过在RT-DETR主

干网络使用SPDRSFE模块进行特征提取,引入渐进特征金字塔结构实现特征融合并利用Conv3XCC3模块进行特征再提取。该模型的精确率、召回率和mAP均有提高,但检测速度较慢,实时检测效果一般。沈烨等^[16]提出的FSSD-DETR实时检测算法,利用FADC模块优化特征提取,引入小目标检测层并基于SSFF和TFE模块改进颈部网络,提升检测准确性。最后采用DySample上采样算子改善最近邻插值法。虽然该模型相较于其他主流算法在实时性和检测精度方面表现良好,但仍存在计算量、参数量较大的问题,对系统硬件要求较高。

针对上述模型在交通标识检测时存在的计算量、参数量较大,检测不同大小和尺度交通标识的精度低,受环境因素干扰后的特征信息处理有限等情况,本文提出一种基于改进RT-DETR的公路路面交通标识检测模型。改进的工作为:

1)在主干网络设计了轻量化的CSP-PMSFA(cross stage partial-partial multi-scale feature aggregation)模块,用于降低模型的计算量,提高交通标识的检测效率和表达能力。

2)在尺度内特征交互(attention-based intrascale feature interaction, AIFI)模块引入CGA(cascaded group attention)^[17]注意力机制,采用级联分组的方式解决注意力头冗余导致计算效率下降的问题,并通过增加网络深度的方式提升模型容量,保留更多交通标识的特征信息。

3)设计了EMBSFPN(efficient multi-branch scale feature pyramid network)作为模型的跨尺度特征融合模块(cross-scale feature fusion module, CCFM),用于实现不同大小、尺寸交通标识的特征融合,并在FPN阶段选择卷积核逐步获得多尺度感受场信息,提高检测精度。

4)改进模型的上采样,使用EUCB(efficient up-convolution block)模块减少交通标识因道路类型、交通流量和天气条件因素变化造成适应性不足的问题,提高模型的鲁棒性。

1 RT-DETR 介绍

RT-DETR模型是一种基于Transformer架构的实时目标检测、嵌入和跟踪模型,其将Transformer的高效性与DETR的灵活性进行结合,解决了传统目标检测方法在计算效率和实时性能方面的不足。RT-DETR的核心思想在于对目标检测和目标跟踪两个传统独立的任务进行统一建模。在目标检测阶段,通过卷积神经网络对输入图像进行特征表示,然后利用Transformer编码器将特征生成一组包含目标位置和类别的嵌入向量。在目标跟踪阶段,为将当前帧的目标嵌入与前一帧中的目标嵌入进行匹配,RT-DETR采用匈牙利算法以实现目标的连续跟踪。该模型的架构由骨干网络、混合编码器、Transformer解码器及检测头等部分组成。为便于提取输入图像的特征表示,骨干网

络采用经典的卷积神经网络 ResNet。混合编码器为全新高效编码器,它的设计结合了尺度内特征交互模块 AIFI 和跨尺度特征融合模块 CCFM,便于将多尺度特征转化为图像特征序列。此外,检测头包含目标的边界框和置信度得分,用于将解码器的输出转换为最终的检测结果。RT-DETR 模型的解码器来源于 DINO 解码器的设计,其会直接预测一定数量的预测框,在模型训练时进行预测框与真实框的一一对比。然后按照一定的规则对每个真实框和预

测框进行适配,再计算损失函数^[18]。

2 改进 RT-DETR 的整体架构

虽然 DETR 系列拥有良好的性能,但由于其基于 Transformer 架构占用了较多的计算资源,会产生计算量大、推理速度较慢等现象。因此,本文以 ResNet-18 的网络结构作为 RT-DETR 的基准模型进行改进,改进 RT-DETR 的整体架构如图 1 所示。

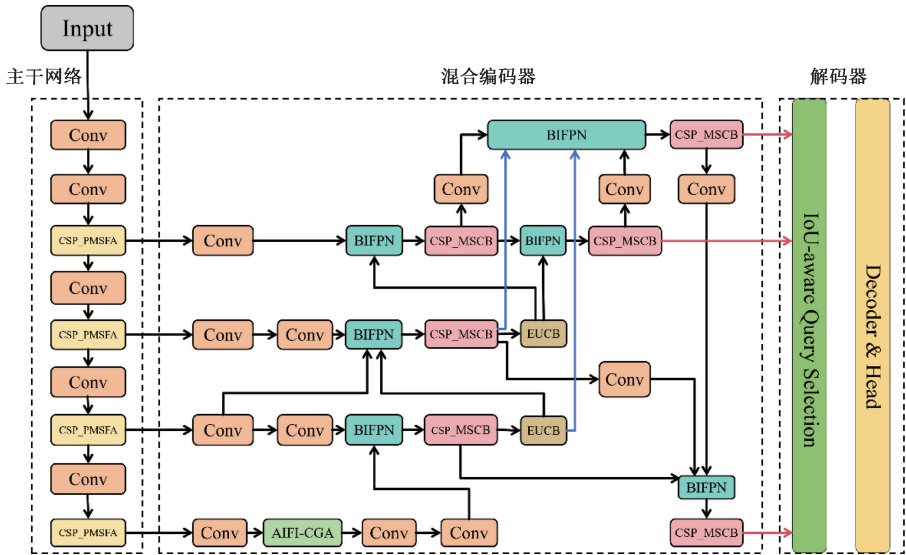


图 1 改进 RT-DETR 整体架构

Fig. 1 Improve the overall architecture of RT-DETR

2.1 CSP-PMSFA 模块的设计

在进行交通标识检测时,由于基准模型 RT-DETR 主干网络易受主流 CNN 计算的中间特征映射影响,一方面会在处理全局信息时更加关注局部特征,造成检测精度低的问题。另一方面,模型会随着网络层数的增加堆积冗余信息,产生较高的计算量和参数量,限制模型对不同交通环境的泛化能力。Han 等^[19]为解决在有限内存和计算资源的嵌入式设备中部署卷积神经网络困难的问题,提出了运用低成本的线性运算获取特征图信息的 Ghost 模型,达到轻量化的同时提升检测性能。Chen 等^[20]避免实现更快的神经网络来追求减少 FLOPs,而是引入了一种简单但快速、有效的卷积 PConv,试验结果表明,推出的 FasterNet 在 GPU、CPU 和 ARM 处理器等多种设备上的运行效果较好,具有高速性和有效性。基于此,根据 GhostNet 轻量化和 FasterNet 有效卷积的思想本文设计了 CSP-PMSFA 主干网络,CSP-PMSFA 结构如图 2 所示。

CSP-PMSFA 模块包含一个 3×3 、 5×5 和 7×7 的 PartialConv 卷积、一个 1×1 卷积。PartialConv 卷积会在一半的通道上对输入图像进行多尺度特征信息的提取,最后的 1×1 卷积会将不同尺度的特征融合在一起并使用残差连接把输入特征与处理后的特征相加。这样设计既

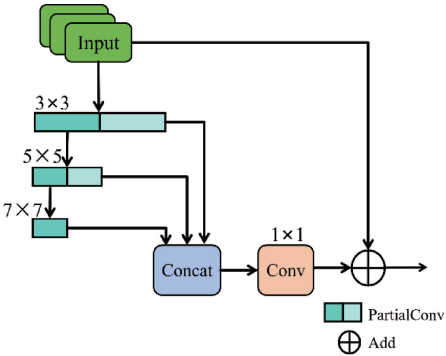


图 2 CSP-PMSFA 结构

Fig. 2 CSP-PMSFA structure

可以在多尺度特征提取下保留交通标识每个尺寸的原始信息,将每个不同尺度的特征信息进行交互,又可以利用 CSP 与残差块结合,优化网络结构,降低梯度消失和爆炸的风险。

2.2 级联分组注意力 CGA 的引入

为了提高模型对不同大小、形状交通标识的检测能力,本文利用级联分组注意力 CGA 改进了尺度内特征交互模块 AIFI 的自注意力机制。如图 3 所示,该模块会将注意力计算分解到不同的头部中,对交通标识图像进行头

部与完整特征不同切分的输入处理,利用更小的 Q 、 K 通道维数在不引入额外参数的情况下,增加网络深度,提高模型容量和计算效率。

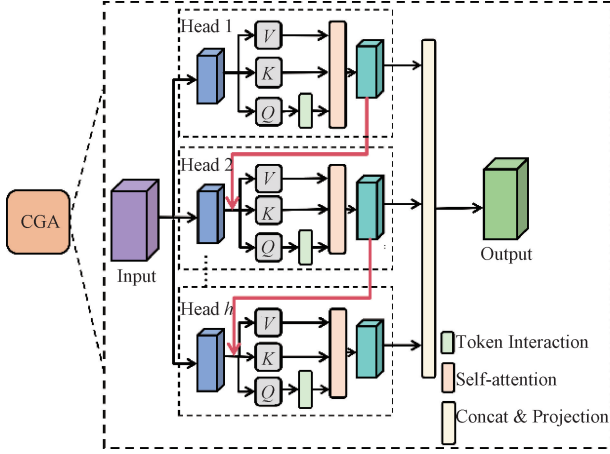


图 3 CGA 结构

Fig. 3 CGA structure

级联分组注意力可表示为:

$$\tilde{X}_{ij} = \text{Attn}(X_{ij} W_{ij}^Q, X_{ij} W_{ij}^K, X_{ij} W_{ij}^V) \quad (1)$$

$$\tilde{X}_{i+1} = \text{Concat}[\tilde{X}_{ij}]_{j=1,h} W_i^P \quad (2)$$

其中, X_{ij} 是输入特征 X_i 的第 j 个分割, 即 $X_i = [X_{i1}, X_{i2}, \dots, X_{ih}]$ 和 $1 \leq j \leq h$; h 是头的总数; $W_{ij}^Q, W_{ij}^K, W_{ij}^V$ 为将分割输入特征映射到不同空间的投影层; W_i^P 是将级联的输出特征投影会与输入一致维度的线性层。如式(3)所示,级联分组会将每个头部的输出添加到后续头部中,以逐步细化特征表示。

$$X'_{ij} = X_{ij} + \tilde{X}_{i(j-1)}, 1 < j < h, \quad (3)$$

其中, X'_{ij} 是通过式(3)计算的第 j 个输出分割 X_{ij} 和

第 $j-1$ 个头部输出 $\tilde{X}_{i(j-1)}$ 。在计算自我注意力时,它会将 X_{ij} 替换为第 j 个头的新输入特征。此外, Q 投影之后会应用另一个令牌交互层,这进一步增强了特征表示。使自我注意力能够共同捕获局部和全局关系。

2.3 EMBSFPN 模块的设计

1) BiFPN 的应用与 CSP_MSCB 模块的设计

由于受到交通流量、雨雪天气和光照变化等环境因素的干扰,公路路面交通标识存在不少尺寸较小或形变较大的目标,而 RT-DETR 的跨尺度特征融合模块 CCFM,在处理多尺度交通标识时却对目标的捕获和识别能力有限,易出现感受野适应性不足,某些特征信息被忽略或过度融合的现象,造成检测精度低、速度慢等问题。因此,本文引入 BiFPN (bi-directional feature pyramid network)^[21] 对多尺度特征信息进行动态选择加权融合并设计了 CSP_MSCB (cross stage partial multi-scale convolution block) 模块,通过在多个尺度上进行深度卷积

操作,增强级联扩展路径生成的特征,并逐步获得多尺度感知场信息。该模块首先将输入特征图卷积并拆分,一部分流向 Bottleneck 和 MSDC^[22] 模块,另一部分经不同卷积层处理后合并再卷积输出。该模块一方面保留原有的 CSP 思想,拥有较好的扩展性。另一方面,其会在多个尺度上执行深度卷积,增强级联扩展路径生成的特征并逐步获得多尺度感知场信息。CSP_MSCB 结构如图 4 所示,首先,输入特征图经过一个卷积层进行变换,拆分成两部分,一部分直接进入 $n-1$ 个 Bottleneck 流向 MSDC 模块,另一部分经过不同的卷积层处理,最终两部分合并再卷积得到输出。MSDC 结构使用点向 (1×1) 卷积层 PWC_1 扩展通道数量,到达批归一化层 BN 和 ReLU6 激活层 R6,然后使用多尺度深度卷积 MSDC 捕获多尺度和多分辨率上下文,同时为了避免深度卷积忽略通道关系带来的损失,利用 channel_shuffle 操作对通道关系进行合并。

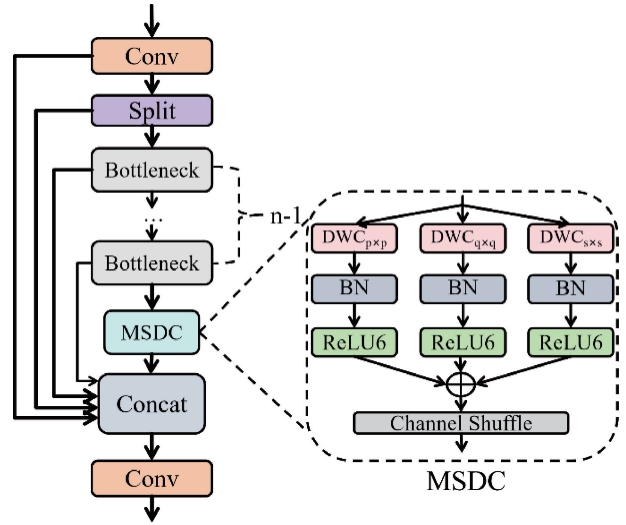


图 4 CSP_MSCB 结构

Fig. 4 CSP_MSCB structure

具体过程如下:

$$\text{MSDC}(x) = \sum_{ks \in KS} \text{DWCB}_{ks}(x) \quad (4)$$

$$\text{DWCB}_{ks}(x) = \text{R6}(\text{BN}(\text{DWC}_{ks}(x))) \quad (5)$$

其中, DWCB_{ks} 是与核大小 ks 的深度卷积; BN 和 R6 分别为批归一化和 ReLU6 激活。同时, MSDC 为优化正则化; x 为递归更新的输入; 输入 x 与之前的 DWCB_{ks} 进行残差连接, 如式(6)所示。

$$x = x + \text{DWCB}_{ks}(x) \quad (6)$$

其中, 本文为增强算法对复杂交通场景的适应能力, 针对不同的特征图尺寸, 采用了不同的卷积核进行处理^[23]。在 P3 层引入 1、3、5 的卷积核到 MSDC 的 DWC_{ks} 中, P4 层引入 3、5、7 的卷积核到 MSDC 的 DWC_{ks} 中, P5 层引入 5、7、9 的卷积核到 MSDC 的 DWC_{ks} 中分别进行深度卷积以适应不同分辨率交通标识图像的需求。

2)EUCB 上采样模块的引入

在深度学习的特征融合阶段,为了与原始图像或其他阶段的特征图进行匹配,通常会采用上采样来逐步恢复特征图的分辨率。由于 RT-DETR 结合了 Transformer 架构,传统的上采样会依赖手工特征工程,对交通标识中的边缘或轮廓等细粒度任务存在局限性,在计算效率和精度方面存在权衡问题。对此,本文引入了具有深度卷积的上采样模块 EUCB^[22]。如图 5 所示,首先对输入内容进行比例因子为 2 的 UpSampling Up 操作来升级特征图,然后再经过一个 3×3 深度卷积 DWC 后,应用 BN 和 ReLU 激活对特征图进行放大增强。最后,为与下阶段进行匹配,对特征图使用 1×1 的卷积操作来减少通道。

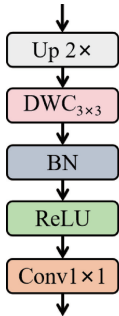


图 5 EUCB 上采样模块
Fig. 5 EUCB upsampling module

3 实验设计与结果分析

3.1 数据集

本实验使用 ROAD MARK 路面标识检测数据集^[24]作为实验数据集,该数据集由 Parisa Karimi Darabi 团队收集并创建针对公路路面交通标识检测任务的专用数据集,其涵盖了乡村公路、城市街道等多个场景下的各类道路标记,如车道线、箭头、人行横道、限速标志、交通信号灯及其他相关标识等共计 3 892 张图像。但为提高数据集的质量,首先,对数据集中存在的重复、模糊和质量低下的图像进行清洗。然后,将清洗后的数据进行旋转、缩放和翻转操作增加数据的多样性。同时,自行采集部分公路路面交通标识检测数据集,使用 LabelImg 软件进行人工标注,与处理后的数据集进行归一化操作,调整数据格式,最终得到共约 6 187 张图片,并以 7:2:1 的比例将改进后的数据集划分为训练集、验证集和测试集。同时,为了验证改进模型的泛化能力,选取了 BDD100K^[25]数据集中的部分数据集进行验证,在进行筛选和数据增强后得到 8 000 张数据集,并按照 7:2:1 的比例重新划分为训练集、验证集和测试集。

3.2 实验环境及参数配置

本实验在 Ubuntu 20.04. x LTS 操作系统下进行, GPU 为 NVIDIA GeForce GTX 3090,24 GB 显存,CPU 为 Intel(R) Xeon(R) Gold 6136 CPU @ 3.00 GHz,深度学习

架构为 Pytorch2. 4. 1,编程语言为 Python3. 9,加速计算架构为 CUDA12. 1。具体实验参数设置如表 1 所示。

表 1 实验环境配置及参数
Table 1 Experimental environment configuration and parameters

参数名称	配置
optimizer	AdamW
epochs	350
Lr0	0.000 1
batchsize	4
workers	4
imgsz	640

3.3 评价指标

为了检验改进算法的有效性,本文选用调和平均数(F1)、精确度(precision, P)、召回率(recall, R)、平均精度(mean average precision, mAP)、每秒传输帧速(FPS)、Giga 浮点运算每秒(GFLOPs)和参数量(parameters)来评价改进模型的检测效果。相关计算公式为:

$$F1 = \frac{2TP}{2TP + FP + FN}$$
 (7)

$$P = \frac{TP}{TP + FP}$$
 (8)

$$R = \frac{TP}{TP + FN}$$
 (9)

$$mAP = \frac{\sum_1^N \int_0^1 p(r) dr}{N}$$
 (10)

其中,TP (true positive)为模型预测正确正样本的数量,FP (false positive)为模型错误预测正样本的数量,FN (false positive)为模型错误预测为负例的正样本数量。N 为类别的数量,p(r) 表示精确率和召回率曲线。

3.4 不同注意力机制对比实验

为验证本文在 AIFI 模块引入级联分组注意力 CGA 的有效性和优越性,本文在基础模型 RT-DETR 上添加了 CGA 注意力机制,并与 DA (deformable attention transformer)^[26]注意力模块、MSMHSA (multi-scale multi-head self-attention)^[27]注意力模块、EAA (efficient additive attention)^[28]注意力模块进行对比,对比实验结果如表 2 所示。根据实验结果可知,在基准模型 RT-DETR 中引入 CGA 注意力模块比引入 DA 注意力模块、MSMHSA 注意力模块、EAA 注意力模块在 mAP50 上分别高出 1.3%、0.7%、0.5%,虽然较 EAA 注意力模块的 mAP50:95 低 0.1%,但 CGA 注意力模块的 FPS 达到了 83.6 帧,是所有注意力模块中最高的且计算量最低,相比其他注意力机制,其综合性能最好,验证了级联分组注意力 CGA 的有效性和优越性。

表 2 注意力机制对比

Table 2 Comparison of attention mechanisms

模型	mAP50	mAP50:95	FPS	GFLOPs	Params
	/ %	/ %			
RT-DETR	91.7	77.0	71.3	57.2	19 964 740
+DA	92.4	77.0	60.4	57.0	20 019 524
+MSMHA	93.0	77.5	70.2	57.2	19 924 036
+EAA	93.2	77.8	67.7	57.2	19 888 708
+CGA	93.7	77.7	83.6	57.1	19 718 600

3.5 不同检测模型的对比实验

为了综合评估改进算法的优越性,考虑到任务实时性需求,本文选取了 SSD^[5]、Faster-RCNN^[7]、RT-DETR^[9]、YOLO 系列多个版本^[4]和目前主流的算法与本文的改进

模型进行横向对比实验。所有模型均在实验环境和参数设置保持一致的前提下,在同一数据集上进行训练和测试,对比实验结果如表 3 所示。根据实验结果可知,改进 RT-DETR 的 F1 分数达到 0.91,与 FSSD-DETR 持平且略高于其他模型。在 mAP50 指标上,改进模型以 94.3% 的准确率先,在更为严格的 mAP50:95 指标上改进模型以 77.4% 的优势验证了其高 IoU 阈值下的鲁棒性。同时,改进 RT-DETR 在 FPS_{bs=4} 的测试环境下达到了 84.9 fps,优于大多数对比模型,仅次于 SC-YOLOv8 和 YOLOv8 DRM,但这两者在精度上远低于改进 RT-DETR。此外,改进模型拥有更低的计算量和参数量,表明其在保持高精度的同时,也实现了较好的计算效率和模型复杂度控制,满足实时检测的要求。

表 3 与其他模型的对比实验

Table 3 Comparative experiments with other models

模型	F1	mAP50/ %	mAP50:95/ %	FPS _{bs=4}	GFLOPs	Params
Faster-RCNN	0.47	50.5	30.2	16.9	170.2	41 124 788
SSD	0.56	60.5	53.2	34.9	121.0	25 334 517
YOLOv5m	0.89	91.1	74.3	52.6	64.2	25 052 743
YOLOv7	0.88	92.8	73.9	30.0	103.4	36 546 500
YOLOv8m	0.90	92.8	75.5	50.7	78.9	25 847 287
YOLOv9m	0.90	93.1	77.5	53.1	76.8	20 022 391
YOLOv10m	0.88	92.5	75.7	71.4	63.5	16 465 438
SC-YOLOv8 ^[12]	0.80	80.6	63.9	94.7	15.2	4 294 712
YOLOv8 DRM ^[13]	0.78	78.9	60.2	99.6	8.4	2 083 471
RT-DETR	0.90	91.7	77.0	71.3	57.2	19 964 740
DSRA-DETR ^[14]	0.90	92.3	76.8	74.5	56.1	18 372 972
FSSD-DETR ^[17]	0.91	93.1	77.2	72.9	58.5	21 348 917
改进模型	0.91	94.3	77.4	84.9	40.2	12 044 633

3.6 改进模型消融实验

为系统分析本文引入的改进模块对模型性能的影响及有效性,本文以 RT-DETR 为基准模型,在此基础上依次添加改进模块,进行消融实验,实验结果如表 4 所示。根据实验结果可知,当使用轻量化模块 CSP-PMSFA 对主干部分进行改进后,mAP50 提高了 1.9%,计算量和参数量分别下降了 16.6%和 29%。在此基础上继续使用 CGA

级联分组注意力机制,虽然 mAP50 略有下降,计算量稍有上升,但 FPS 又提高了 9.4 帧。最后,引入设计的 EMBSFPN 模块后,模型的 mAP50 提高了 2.6%,达到 94.3%,FPS 提高了 13.6 帧,计算量和参数量分别下降了 29.7%和 39.6%。总体而言,各部分改进模块均可以综合提高模型在复杂的公路路面交通标识环境中的学习和泛化能力,同时,检测精度与速度的提升具有较高的实用价值。

表 4 改进模型消融实验

Table 4 Improve the model ablation experiment

CSP-PMSFA	CGA	EMBSFPN	mAP50/ %	mAP50:95/ %	GFLOPs	FPS	Params
			91.7	77.0	57.2	71.3	19 964 740
✓			93.8	77.4	47.7	72.4	14 107 204
	✓		93.7	77.7	57.1	83.6	19 718 600
		✓	93.5	77.0	48.7	67.8	17 954 773
✓	✓		93.3	77.4	47.8	81.8	13 937 352
✓	✓	✓	94.3	77.4	40.2	84.9	12 044 633

3.7 泛化实验

为了验证改进模型的泛化能力,本文在 BDD100K 数据集上进行了不同改进模型的性能对比实验。实验选取了 DSRA-DETR、FSSD-DETR、RT-DETR 以及改进后的模型作为对比对象,泛化实验结果如表 5 所示。根据实验结果可知,改进后的模型在 mAP50/% 指标上达到了 45.6%,相较于 DSRA-DETR 的 42.7%、FSSD-DETR 的 44.2%和 RT-DETR 的 41.2%,有了明显的提升。各类交通元素的检测中,改进模型也表现出优异的性能。例如,在 Person 和 Car 类别的检测上,改进模型分别达到了 61.8%和 77.5%的准确率,均高于其他 3 个模型。此外,在 FPS 指标上,改进模型达到 81.9 帧,远高于 DSRA-

DETR 的 78.3 帧、FSSD-DETR 的 66.7 帧和 RT-DETR 的 67.3 帧,表明改进模型在保持高准确率的同时具备更快的检测速度。

3.8 热力图可视化分析

为了更加直观地展示改进模型在交通标识检测任务中的性能提升,本文利用 GradCAM (gradient-weighted class activation mapping)^[29]对基准模型和改进模型进行输出层的可视化,通过比较两者对交通标识关注区域的轮廓和细节进一步验证改进模型的性能表现。深红色区域为模型关注的焦点区域,交通标识检测任务中不同类别标识的深红色区域越多表示模型精度越高、性能更好。部分热力图可视化效果如图 6 所示。

表 5 不同改进模型的泛化实验结果
Table 5 Generalization experimental results of different improved models

模型	mAP50/%	Person	Rider	Car	Bus	Truck	Bike	Motor	Traffic Light	Traffic Sign	Train	FPS _{bs=4}
DSRA-DETR ^[14]	42.7	55.7	31.9	70.2	35.1	42.7	31.8	32.5	39.6	43.7	15.0	78.3
FSSD-DETR ^[17]	44.2	57.9	28.5	76.7	30.2	43.1	30.5	33.9	39.7	44.2	16.3	66.7
RT-DETR	41.2	54.1	30.4	75.8	36.9	41.6	29.7	30.1	38.7	42.4	13.2	67.3
改进模型	45.6	61.8	32.7	77.5	33.1	43.8	33.1	35.7	40.2	44.6	17.5	81.9

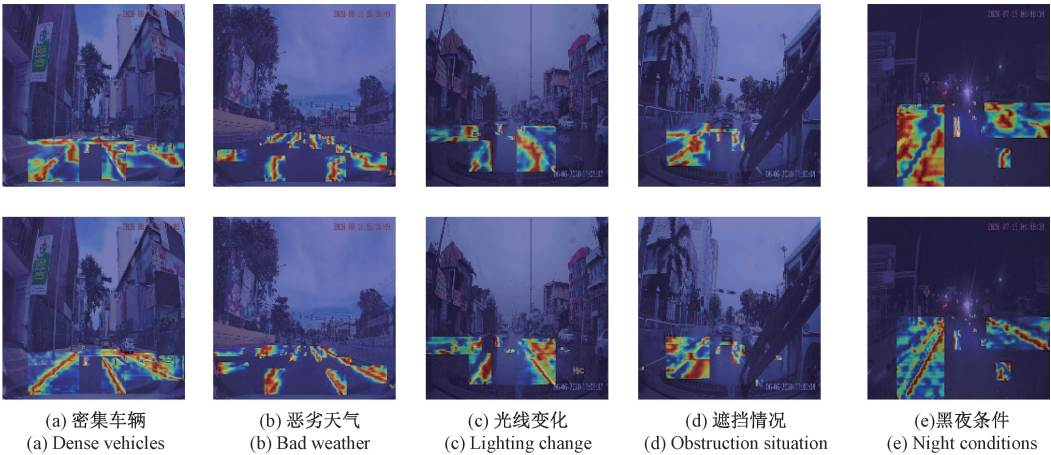


图 6 改进前后模型热力图可视化结果

Fig. 6 Visualization results of thermal maps before and after improvement of the model

根据热力图的结果可知,基准模型往往会被光照强弱、交通标识形变、背景遮挡和天气变化等外界因素的干扰存在局限性,会更加关注交通标识以外的区域,易发生误检和漏检的现象。而改进模型由于具备动态调整感受野大小的能力,会对不同尺度的特征层进行卷积核的选择,所以更善于捕捉交通标识的轮廓和细节,在热力图中能够呈现出更为清晰的标识区域。图 6(a)列展现了公路上车辆密集行驶的情景。虽然行驶车辆的驾驶速度不同给路面中的交通标识检测带来了一定的困难,但改进模型因检测速度较快,可快速提取不同分辨率的交通标识。图 6(b)、(c)列分别展现了恶劣天气和光线变化导致路面标

识类别难以区分的情景。尽管这使得路面标识类别的区分变得更加困难,但改进模型 EMBSFPN 模块的设计弥补了这一缺点,其通过自适应选择不同尺度特征进行加权融合,并利用不同卷积核来获得多尺度的公路路面交通标识,检测效果较为突出。图 6(d)列展现了车辆快速行驶到有树木遮挡时的情景。虽然由于遮挡、光线较暗导致路面中的交通标识特征信息不明显,但改进模型主干网络的 CSP-PMSFA 模块特征提取能力更强,能够准确提取低分辨率的小目标。图 6(e)列展现了夜间车辆行驶在公路上的情景,由于路面相对行驶的车辆光照较强、夜间路面光线较差,导致公路路面交通标识对比度不明显,但改进模

型的级联分组注意力 CGA 保留更多局部信息,上采样模块引入了 EUCB,深度卷积的应用增强了鲁棒性,优化了模型的结构,更加适应不同尺度的交通标识。

4 结 论

针对以往交通标识检测模型会因天气条件、光照变化等环境因素和模型自身计算量、参数量较大造成检测精度低、速度慢的问题,本文提出了一种基于改进 RT-DETR 的公路路面交通标识检测算法。首先,利用设计的主干网络 CSP-PMSFA 来降低计算量、参数量,提取交通标识图像的多尺度特征信息。其次,在模型的尺度内特征交互部分引入 CGA 注意力机制提升模型的检测性能。最后,利用设计的 EMBSFPN 模块进行多尺度特征加权融合并动态调整感受野获取感知场信息。实验结果显示,改进 RT-DETR 算法在数据集上的性能表现验证了改进模型的实用性,能够在降低参数量、计算量的同时提升检测平均精度和速度,满足实时检测的需求。但本文算法仍存在待改进的地方,接下来的研究工作一方面将聚焦于如何优化网络结构和参数设置,在达到更好的实时性的同时提升效率;另一方面,降低模型对硬件设备的成本要求,搭建具有功能完善的公路路面交通标识检测系统。

参考文献

- [1] 霍爱清, 南思媛, 胥静蓉. 改进 YOLOX 的弱光线路交通标志检测[J]. 电子测量技术, 2023, 46(6): 62-67.
HUO AI Q, NAN S Y, XU J R. Improving YOLOX's weak light road traffic sign detection[J]. Electronic Measurement Technology, 2023, 46(6): 62-67.
- [2] YAO J N, HUANG B Q, YANG S, et al. Traffic sign detection and recognition under low illumination[J]. Machine Vision and Applications, 2023, 34(5): 75.
- [3] 李清格, 杨小冈, 卢瑞涛, 等. 计算机视觉中的 Transformer 发展综述[J]. 小型微型计算机系统, 2023, 44(4): 850-861.
LI Q G, YANG X G, LU R T, et al. Overview of transformer development in computer vision[J]. Small Microcomputer Systems, 2023, 44(4): 850-861.
- [4] 毛少华, 王文东. 基于深度学习的 YOLO 系列物体检测算法研究综述[J]. 延安大学学报(自然科学版), 2024, 43(2): 88-95.
MAO SH H, WANG W D. A review of YOLO series object detection algorithms based on deep learning[J]. Journal of Yan'an University (Natural Science Edition), 2024, 43(2): 88-95.
- [5] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]. Computer Vision-ECCV 2016: 14th European Conference, 2016: 21-37.
- [6] GIRSHICK R. Fast R-CNN[C]. IEEE International Conference on Computer Vision, 2015: 1440-1448.
- [7] REN SH Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 39(6): 1137-1149.
- [8] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers [C]. European Conference on Computer Vision, Cham: Springer International Publishing, 2020: 213-229.
- [9] ZHAO Y AN, LYU W Y, XU SH L, et al. DETRs beat YOLOs on real-time object detection[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 16965-16974.
- [10] BODLA N, SINGH B, CHELLAPPA R, et al. Soft-NMS-improving object detection with one line of code[C]. IEEE International Conference on Computer Vision, 2017: 5561-5569.
- [11] 闫世洋, 罗素云. 基于 SC-YOLOv8 的交通标志检测算法研究[J]. 电子测量技术, 2024, 47(15): 117-124.
YAN SH Y, LUO S Y. Research on traffic sign detection algorithm based on SC-YOLOv8 [J]. Electronic Measurement Technology, 2024, 47(15): 117-124.
- [12] 徐仟祥, 曾勇, 卢倩, 等. 基于改进 RT-DETR 的车门内拉手表面缺陷检测方法[J]. 电子测量技术, 2024, 47(18): 172-181.
XU Q X, ZENG Y, LU Q, et al. Surface defect detection method for door handle based on improved RT-DETR[J]. Electronic Measurement Technology, 2024, 47(18): 172-181.
- [13] 周建新, 张媛, 贾梓涵, 等. 改进 YOLOv8 的路面缺陷检测算法[J]. 电子测量技术, 2024, 47(19): 146-154.
ZHOU J X, ZHANG Y, JIA Z H, et al. Improved YOLOv8 road defect detection algorithm [J]. Electronic Measurement Technology, 2024, 47(19): 146-154.
- [14] XIA J A, LI M J, LIU W K, et al. DSRA-DETR: An improved DETR for multiscale traffic sign detection[J]. Sustainability, 2023, 15(14): 10862.
- [15] 许小伟, 陈燕玲, 占柳, 等. 基于改进 RT-DETR 的路面坑槽检测模型[J]. 武汉科技大学学报, 2024, 47(6): 457-467.
XU X W, CHEN Y L, ZHAN L, et al. A road pit detection model based on improved RT-DETR[J]. Journal of Wuhan University of Science and

- Technology, 2024, 47(6): 457-467.
- [16] 沈烨, 王泉, 吴昊. FSSD-DETR 自动驾驶场景实时目标检测算法[J]. 电子测量技术, 2024, 47(14): 88-95.
SHEN Y, WANG Q, WU H. FSSD-DETR real time object detection algorithm for autonomous driving scenarios [J]. Electronic Measurement Technology, 2024, 47(14): 88-95.
- [17] LIU X Y, PENG H W, ZHENG N X, et al. Efficientvit: Memory efficient vision transformer with cascaded group attention[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 14420-14430.
- [18] 蒋智臣, 胡俐蕊. 基于改进 RT-DETR 的浅水海洋生物识别方法 [J]. 电子测量技术, 2024, 47(18): 155-163.
JIANG ZH CH, HU L R. Shallow water marine biological identification method based on improved RT-DETR [J]. Electronic Measurement Technology, 2024, 47(18): 155-163.
- [19] HAN K, WANG Y H, TIAN Q, et al. GhostNet: More features from cheap operations[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 1580-1589.
- [20] CHEN J R, KAO S H, HE H, et al. Run, don't walk: Chasing higher FLOPS for faster neural networks[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 12021-12031.
- [21] TAN M X, PANG R M, LE Q V. Efficientdet: Scalable and efficient object detection[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 10781-10790.
- [22] RAHMAN M M, MUNIR M, MARCULESCU R. EMCAD: Efficient multi-scale convolutional attention decoding for medical image segmentation[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 11769-11779.
- [23] YANG ZH Q, GUAN Q, ZHAO K, et al. Multi-branch auxiliary fusion YOLO with re-parameterization heterogeneous convolutional for accurate object detection[C]. Chinese Conference on Pattern Recognition and Computer Vision (PRCV). Singapore: Springer Nature Singapore, 2024: 492-505.
- [24] TDX L. ROAD MARK Dataset[EB/OL]. [2024-12-18]. <https://universe.roboflow.com/lr-tdx/road-mark>.
- [25] YU F, CHEN H F, WANG X, et al. Bdd100k: A diverse driving dataset for heterogeneous multitask learning [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 2636-2645.
- [26] XIA ZH F, PAN X, SONG S J, et al. Vision transformer with deformable attention [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 4794-4803.
- [27] WU H L, HUANG P, ZHANG M, et al. CMTFNet: CNN and multiscale transformer fusion network for remote sensing image semantic segmentation[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 1-12.
- [28] SHAKER A, MAAZ M, RASHEED H, et al. Swiftformer: Efficient additive attention for transformer-based real-time mobile vision applications[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 17425-17436.
- [29] SELVARAJU R R, CONGSWELL M, DAS A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization [J]. International Journal of Computer Vision, 2020, 128(2): 336-359.

作者简介

孙海青, 硕士研究生, 主要研究方向为机器学习与图像处理。

E-mail: 2171298792@qq.com

杨传颖(通信作者), 副教授, 主要研究方向为机器学习与图像处理。

E-mail: ycy@imut.edu.cn

敖乐根, 正高级工程师, 主要研究方向为信息化与智能化。

E-mail: 648165126@qq.com