

基于RT-DETR的轻量化交通标志检测

解浩龙 张孝龙 魏培旭 崔传金

Lightweight traffic sign detection based on RT-DETR

XIE Haolong, ZHANG Xiaolong, WEI Peixu, CUI Chuanjin

引用本文:

解浩龙, 张孝龙, 魏培旭, 等. 基于RT-DETR的轻量化交通标志检测[J]. 应用光学, 2025, 46(2): 300–308. DOI: 10.5768/JAO202546.0202002

XIE Haolong, ZHANG Xiaolong, WEI Peixu, et al. Lightweight traffic sign detection based on RT-DETR[J]. Journal of Applied Optics, 2025, 46(2): 300–308. DOI: 10.5768/JAO202546.0202002

在线阅读 View online: <https://doi.org/10.5768/JAO202546.0202002>

您可能感兴趣的其他文章

Articles you may be interested in

改进YOLOX-S的红外舰船目标检测算法

Infrared ship target detection algorithm based on improved YOLOX-S

应用光学. 2023, 44(5): 1054–1060 <https://doi.org/10.5768/JAO202344.0502006>

基于改进Faster R-CNN的红外目标检测算法

Infrared target detection algorithm based on improved Faster R-CNN

应用光学. 2024, 45(2): 346–353 <https://doi.org/10.5768/JAO202445.0202001>

轻量化铝反射镜拓扑优化与加工试验研究

Topological optimization and machining experiment of lightweight aluminum mirror

应用光学. 2024, 45(5): 1034–1041 <https://doi.org/10.5768/JAO202445.0505001>

基于轻量化校准反射镜的旋转平移干涉绝对检验技术

Absolute test technology for rotation and translation interference based on lightweight calibration mirror

应用光学. 2023, 44(1): 122–127 <https://doi.org/10.5768/JAO202344.0103002>

基于改进YOLOv5s的曲面光学镜片缺陷检测

Defect detection of curved optical lenses based on improved YOLOv5s

应用光学. 2024, 45(4): 781–789 <https://doi.org/10.5768/JAO202445.0403003>

基于机器视觉的手机屏幕表面划痕检测研究

Surface scratch detection of mobile phone screen based on machine vision

应用光学. 2020, 41(5): 984–989 <https://doi.org/10.5768/JAO202041.0502008>



关注微信公众号, 获得更多资讯信息

文章编号: 1002-2082 (2025) 02-0300-09

引用格式: 解浩龙, 张孝龙, 魏培旭, 等. 基于 RT-DETR 的轻量化交通标志检测 [J]. 应用光学, 2025, 46(2): 300-308.

XIE Haolong, ZHANG Xiaolong, WEI Peixu, et al. Lightweight traffic sign detection based on RT-DETR[J]. Journal of Applied Optics, 2025, 46(2): 300-308.



在线阅读

基于 RT-DETR 的轻量化交通标志检测

解浩龙¹, 张孝龙^{1,2}, 魏培旭¹, 崔传金¹

(1. 华北理工大学 电气工程学院, 河北 唐山 063200; 2. 浙江耀恒光电科技有限公司, 浙江 杭州 311600)

摘 要: 交通标志检测是智能车辆行驶过程中的关键, 对于车辆分析路况有着重大的意义。针对现存交通标志检测算法参数较多、精确率不高的问题, 提出一种基于改进 RT-DETR(real-time detection transformer) 的轻量化交通标志检测算法。首先, 将模型主干网络中的 ResNet 网络替换为 VanillaNet 网络, 减少网络的层数与参数量; 其次, 使用加权双向特征金字塔模型 (bi-directional feature pyramid network, BiFPN) 代替 RT-DETR 特征融合模块中的 PAN(path aggregation network) 结构, 增加模型的融合能力, 提取更加丰富的特征信息; 最后, 在特征融合模块中添加 GAM(global attention mechanism) 注意力机制, 加强模型对全局信息的感知, 提升多目标与被遮挡目标的检测性能。该算法在交通标志数据集上进行测试, 结果说明改进后的 RT-DETR 算法的 mAP(mean average precision) 值达到了 87.7%, 相比原始算法提高了 3.6%, 且参数减少 21%, 满足智能车辆设备的部署需要, 证明了改进算法的有效性。

关键词: 机器视觉; DETR 算法; 目标检测; 轻量化; 注意力机制

中图分类号: TN209

文献标志码: A

DOI: 10.5768/JAO202546.0202002

Lightweight traffic sign detection based on RT-DETR

XIE Haolong¹, ZHANG Xiaolong^{1,2}, WEI Peixu¹, CUI Chuanjin¹

(1. College of Electrical Engineering, North China University of Science and Technology, Tangshan 063200, China;

2. Zhejiang Yaoheng Optoelectronic Technology Co., Ltd., Hangzhou 311600, China)

Abstract: Traffic sign detection is the key in the driving process of intelligent vehicles, which is of great significance to vehicle analysis of road conditions. Aiming at the problems of too many parameters and low accuracy of existing traffic sign detection algorithms, a lightweight traffic sign detection algorithm based on improved real-time detection transformer (RT-DETR) was proposed. Firstly, the ResNet network in the backbone network of the model was replaced by VanillaNet network to reduce the number of layers and parameters. Secondly, the bi-directional feature pyramid network (BiFPN) was used to replace the path aggregation network (PAN) structure in the feature fusion module of RT-DETR to increase the fusion ability of the model and extract more abundant feature information. Finally, the global attention mechanism (GAM) was added to the feature fusion module to enhance the model perception of global information and improve the detection performance of multiple targets and obscured targets. The proposed algorithm was tested on traffic sign datasets. The mean average precision (mAP) value of the improved RT-DETR algorithm reaches 87.7%,

收稿日期: 2023-12-21; 修回日期: 2024-04-12

基金项目: 河北省自然科学基金 (E202209110); 河北省人社厅项目 (B2024005010); 河北省教育厅项目 (JQN2023026)

作者简介: 解浩龙 (2000—), 男, 硕士研究生, 主要从事机器视觉与目标识别研究。E-mail: xiehaolong2000@163.com

通信作者: 张孝龙 (1989—), 男, 博士研究生, 讲师, 主要从事图像处理与超声波检测研究。E-mail: zhangxiaolong@ncst.edu.cn

3.6% higher than that of the original algorithm, and the parameters are reduced by 21%, meeting the deployment needs of intelligent vehicle equipment, which proves the effectiveness of the improved algorithm.

Key words: machine vision; detection transformer algorithm; target detection; lightweight; attention mechanism

引言

随着我国交通与汽车行业的快速发展,人们出行的频率激增,交通标志对出行的重要性也日渐凸显出来。交通标志排布的合理性对出行有着至关重要的作用,智能汽车可通过实时获取当前路面的道路标识来综合分析路况,进而合理控制车辆状态。因此,本文通过对交通标志的目标检测来协助计算机对路况进行分析,提高交通出行的安全性。

近年来,国内外众多学者都在使用不同的方法提升识别交通标志的准确性。XU X^[1]等提出了一种基于统计假设检验的自适应颜色阈值分割和形状对称性的交通标志检测方法。戴雪瑞^[2]等使用不同颜色阈值、最大稳定极值区域检测交通标志,发现感兴趣区域后结合形状特征进行判断,并使用支持向量机和定向梯度直方图特征对区域进行分类。WU X^[3]等融合了交通标志颜色信息通道,实现了对比度增强和背景噪声降低,提高了识别的准确性和鲁棒性,但无法检测到形状和颜色相似的交通标志。传统的交通标志检测方法仍有泛化能力差、人工设计的特征算子易受外观背景等因素影响,及不能适用于多种场景检测等缺点。

现如今,计算机视觉与深度学习已成功用于各种目标检测与识别当中,其主流的检测方法主要为单阶段检测器 YOLO(you only look once)^[4]、SSD(single shot multibox detector)^[5]、两阶段检测器 R-CNN(region with convolutional neural network feature)^[6]和端到端检测 DETR(detection transformer)。在交通标志检测方面, PRAKASH A S^[7]等利用 Gabor 滤波器和 Adam 优化器对 LeNet 进行特征提取,提高了模型在图像识别中的处理能力,在德国交通标志 GTSRB 数据集上取得了较高的分类效果。HE Z^[8]等使用简单的卷积和池化操作构建了 10 层架构的轻量级卷积神经网络 TS-CNN,提高了检测的准确率,但缺乏对自然场景和复杂天气条件下图像识别的关注和研究。ZHANG Q^[9]等将 VGG(visual geometry group)的网络结构改为 6 层,引入高层特征,通过块交叉捕捉特征图的上下文信息,提高了对较小交通标志检测的实时性和准确性。伍

锡如^[10]等针对智能驾驶中出现的多目标检测、分割效率低,及鲁棒性差等交通场景,提出一种改进的 Mask R-CNN 交通场景多目标快速检测与分割方法,改进网络能够准确实现多目标的检测与分割。YANG L^[11]等提出了一种基于双目摄像头的车辆多特征检测算法,结合特征金字塔网络实现车牌、标志、灯光 3 个特征的识别,提高了车速检测与测量的鲁棒性。但 YOLO 系列与 R-CNN 系列的算法都需要进行复杂的后处理操作,如它们在检测阶段会生成许多锚点框,但一个物体只需要一个检测框,这样就需要通过非极大值抑制的方法去除冗余的框,在网络训练过程中需要进行复杂的调参。

DETR^[12]是由 Facebook AI 团队于 2020 年推出的一种用于目标检测的模型,其取消了非极大值抑制、anchor 等人工先验模块,在 COCO 数据集上可以达到与 Fast R-CNN 基本相同的效果。但 DETR 作为一种新兴的目标检测模型,还存在着不少问题,例如:训练收敛速度慢、小目标检测能力欠佳和编码器的复杂程度过高等问题。针对以上问题,对其进行优化改进,涌现出很多优秀方法,例如 Deformable DETR^[13]、Conditional DETR^[14]、DAB-DETR(dynamic anchor box-detection transformer)^[15]、RT-DETR(real-time detection transformer)^[16]等。其中 RT-DETR 通过设计了一个混合编码器,使用解耦尺度内交互和跨尺度融合的方法来高效地处理多尺度特征,提高了 DETR 检测器的检测精度与速度。因此,本文在基于 Transformer 的端到端的目标检测方法 RT-DETR 的基础上进行改进,对其主干网络用 VanillaNet^[17]进行替换,并在混合编码器中使用 BiFPN^[18](bi-directional feature pyramid network)进行跨尺度特征融合,同时在其中添加 GAM^[19](global attention mechanism)注意力机制,在不减弱检测准确率的情况下进行轻量化,使其可以部署在智能汽车设备上。

1 RT-DETR 算法原理

RT-DETR 是由百度推出的一种基于 transformer

的一种端到端的目标检测器,无需 NMS(non-maximum suppression)后处理便可直接在模型网络中完成目标检测任务。RT-DETR 主要是由主干网络(backbone)、混合编码器(hybrid encoder)和带预测头的 Transformer 的解码器 3 个部分组成,模型的整体框架如图 1 所示。其主干网络主要是由 ResNet-50 或者 ResNet-101 组成,通过残差神经网络

对图像的特征进行提取,利用主干网络的最后 3 个阶段的输出作为混合编码器的输入,使用混合编码器将其转换为一系列的特征数据。随后,使用 IOU(intersection over union)感知查询选择从编码器输出序列中选择固定数量的图像特征作为解码器的初始对象,带有辅助预测头的解码器迭代优化对象查询以生成方框和置信度分数。

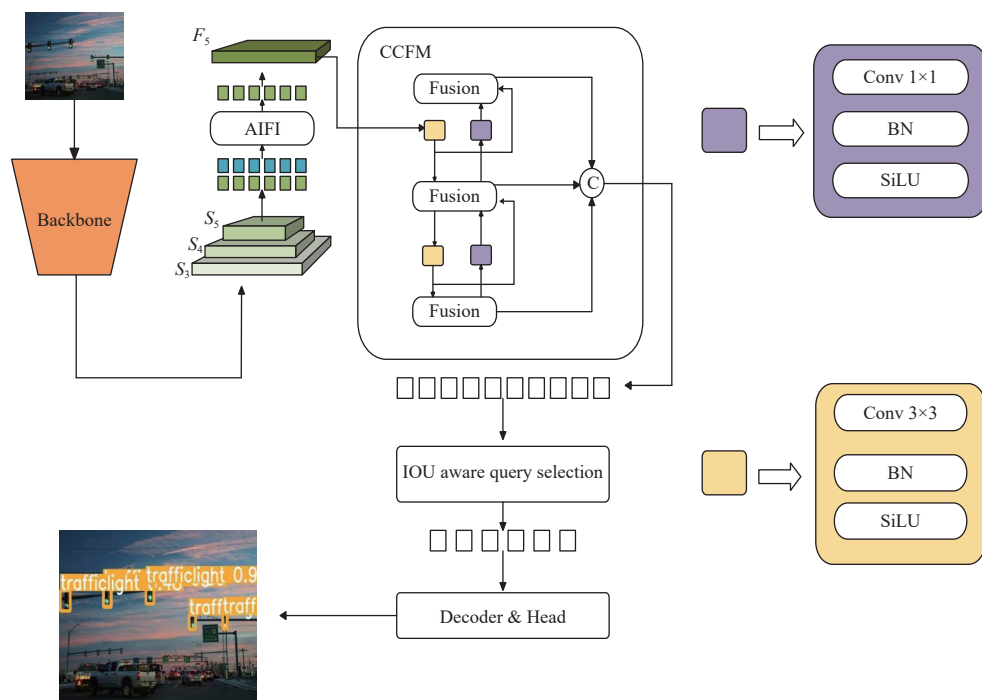


图 1 RT-DETR 框架

Fig. 1 Framework diagram of RT-DETR

与一般 DETR 框架的目标检测算法不同,RT-DETR 设计的混合编码器主要由基于注意力的内部尺度特征交互模块(AIFI)和基于卷积神经网络的跨尺度特征融合(CCFM)模块组成。使用主干网络中最后 3 个阶段提取的特征信息进行特征融合并对 S_5 层使用自注意力,将自注意力机制用于语义概念更加丰富的高级特征,可使用较少的参数量对模型的小目标检测能力进行增强,也能更好地捕获图像中各个实体之间的联系。根据基于注意力的内部尺度特征交互模块,RT-DETR 特别优化了跨尺度特征融合模块,在融合路径中加入了多个卷积层组成融合块,具体的过程可表述为

$$Q = K = V = \text{Flatten}\{S_5\} \quad (1)$$

$$F_5 = \text{Reshape}(\text{Attn}(Q, K, V)) \quad (2)$$

$$\text{OutPut} = \text{CCFM}(\{S_3, S_4, F_5\}) \quad (3)$$

式中: Q 、 K 、 V 是特征图像经过全连接层映射出

的 3 个不同的向量; S_3 、 S_4 、 S_5 为选取的特征图; F_5 是 S_5 经过自注意力机制得到的特征图像; Attn 表示多头自注意力; Reshape 表示将特征的形状恢复为与 S_5 相同的操作; Flatten 是 Reshape 的逆操作。具体模块示意图如图 2。

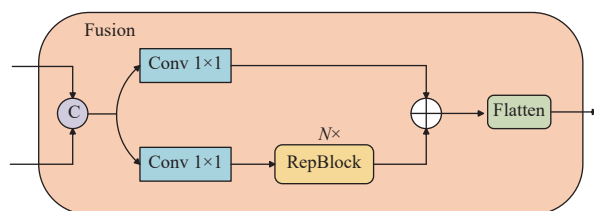


图 2 融合块结构图

Fig. 2 Structure diagram of fusion block

RT-DETR 的预测框主要由 IOU 感知查询模块得到的。因为在进行目标检测时,模型的分类分数与位置置信度有可能不一致,从而导致模型性

能下降, 利用 IOU 感知查询可以约束模型, 对 IOU 分数高的特征产生高分类分数, IOU 分数低的特征产生低分类分数。将解码器的目标优化表述为

$$L(\hat{y}, y) = L_{\text{box}}(\hat{b}, b) + L_{\text{cls}}(\hat{c}, \hat{b}, N_{\text{IOU}}) \quad (4)$$

式中: $\hat{y}$ 和 y 表示预测与基础真值; b 和 c 表示边界框与类别; N_{IOU} 表示置信度分数。

2 改进 RT-DETR 算法

本文主要在原始算法的主干网络、混合编码器和注意力机制上进行优化和改进, 在保持检测精度的情况下对模型进行轻量化操作, 保证其可在较为简易的车载设备上进行检测任务的执行。

首先将主干网络中的 ResNet 替换为 VanillaNet, 通过避免过高的深度和复杂的操作, 从而实现网络的轻量化。此外使用 BiFPN 替换原本跨尺度特征融合模块中的 PAN(path aggregation network) 结构, 引入特征调整与特征选择的操作, 能在保证特征更加充分表达的同时减少参数量, 使其拥有不规则的拓扑结构, 用于不同的环境与任务, 可保持模型在进行轻量化操作时的准确度。最后在网络模型中添加 GAM 注意力机制, 利用其注重三维通道之间信息交互的机制, 减少信息的损失, 提高全局特征之间的交互, 提高深度神经网络的性能, 提升整体的检测精度。具体结构图如图 3 所示。

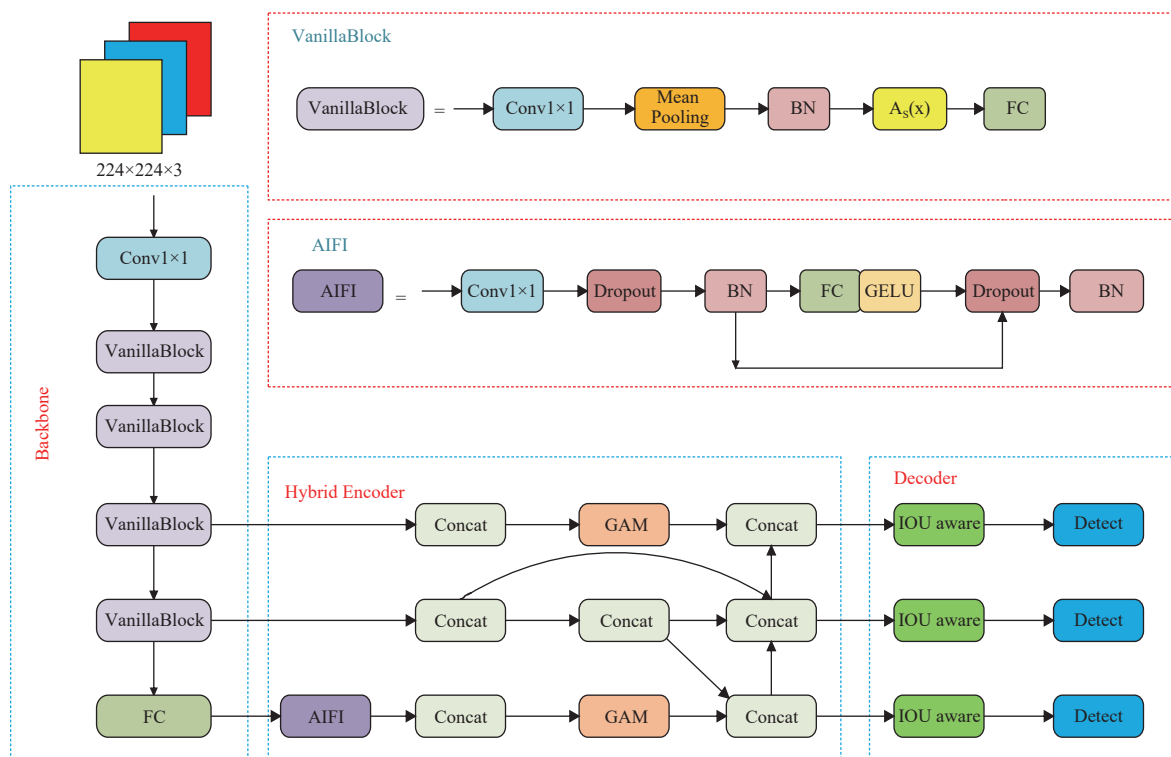


图 3 改进后的 RT-DETR 网络结构图

Fig. 3 Structure diagram of enhanced RT-DETR network

2.1 改进的主干网络

RT-DETR 的主干网络基于 ResNet-50 构建, 引入了残差模块使得卷积神经网络的梯度消失问题得到了缓解, 但因其需要过高层数的堆叠来增强网络的特征提取能力, 网络结构过大且无法对信息过少的目标特征进行提取。普通的车载设备一般是基于嵌入式系统进行开发与设计, 过大的模型无法在这些设备上部署, 本文使用 VanillaNet 来代替 ResNet 在 RT-DETR 在主干网络中的位置, 遵循传统的神经网络结构, 但每一个阶段都只使

用一层, 用以建立一个极为简洁的网络, 减少参数量, 用于车载设备的交通标志检测功能。

本文中采用了一种包含 6 层的 VanillaNet 网络结构, 架构如图 4 所示。在网络的主干部分, 首先应用一个步长为 4 的卷积层, 将 3 通道的特征图像映射到一个具有 512 通道的特征空间。在第 2、3 和 4 层, 引入了步幅为 2 的平均池化层, 以减小特征映射的空间大小, 并在同一时间将通道数增加。在第 5 层, 通道数量保持不变, 同时应用了平均池化层。网络的最后一层是一个全连接层, 负

责输出结果。每个卷积层的核大小都被设定为 1×1 , 这一设计旨在保留特征映射信息的同时最小化每层的计算代价。VanillaNet 利用较少的卷积层达到 ResNet 的性能, 提取所需要的图像特征信息, 同时通过激活函数的堆叠来增强网络的非线性, 改善神经网络的非线性能力。神经网络中输入 x 的单个激活函数为 $A_s(x)$, 它可以是常用的函数, 如 ReLu^[20] 和 Tanh。激活函数 $A_s(x)$ 的并发叠加可表

示为

$$A_s(x) = \sum_{i=1}^n a_i A(x + b_i) \quad (5)$$

式中: n 为激活函数堆叠的次数; a_i 和 b_i 是每一次激活函数的尺度与偏差, 从而避免简单累积。式 (5) 可以看作一个级数, 这种累加方式可以大大增强激活函数的非线性能力, 同时也弥补了 VanillaNet 因为过浅的网络层数导致非线性能力弱的问题。

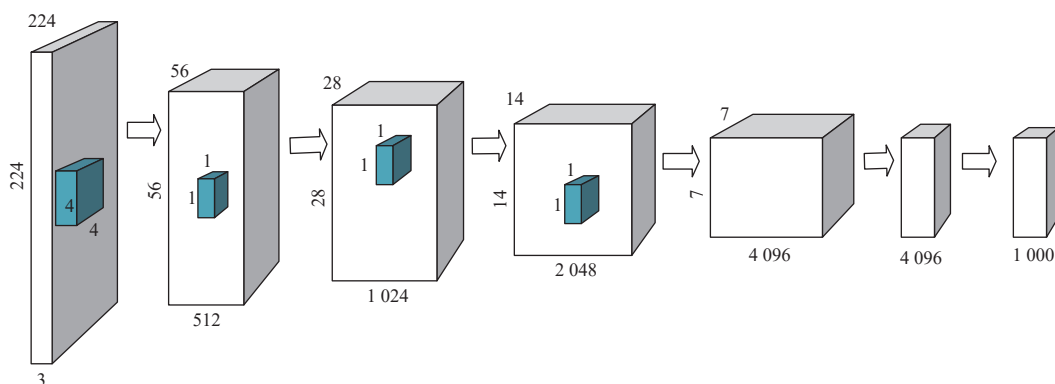


图4 6层 VanillaNet 网络结构

Fig. 4 Structure diagram of 6-layer VanillaNet network

2.2 改进的跨尺度融合模块

在 RT-DETR 混合编码器中, 采用 PAN 与 FPN (feature pyramid networks) 的方法进行跨尺度特征融合, 利用上采样与下采样相结合的方式对得到的各个特征层进行特征融合, 神经网络可以获得不同层级之间的各种图像信息与语义信息, 有利于进行图像的分类与检测。但在对不同分辨率的输入特征进行融合处理时, 不同的输入特征对融合输出特征的贡献度是不同的, 一般的 FPN 结构无法对其进行有效融合。本文使用 BiFPN (bi-directional feature pyramid network) 结构来代替一般的 FPN 结构。BiFPN 被称为加权双向特征金字塔, 是一种既简单又高效的网络架构, 其独特之处在于引入了可学习的权重, 以学习不同输入特征的重要程度。与此同时, 该网络多次应用多尺度特征融合, 进一步增强了对不同尺度信息的捕捉能力。

在多尺度特征融合方面, 传统的自上而下的 FPN 结构受到单向信息流的限制, 之后出现的 PANet 网络结构虽然解决了这个问题, 但是其过大的参数与计算量都不适合其部署于各种微小处理器。BiFPN 网络通过移除那些只有一个输入边的节点来简化网络结构, 这是因为只拥有一个输出

边的节点对整个网络的贡献较小, 简化这一操作有助于构建一个更为精简的双向网络。当原始输入和输出节点位于同一水平时, 在它们之间添加一条额外的边, 将原始输入与输出进行特征融合, 可以解决传统 FPN 结构的单向融合导致对某一特征的过度关注, 而忽略图像特征的内部关联的问题, 在不增加太多参数和计算量的情况下使跨尺度特征融合网络融合更多的特征。与 PANet 不同, BiFPN 将每个双向路径视为一个特征网络层, 并通过多次重复相同层的方式实现更高级的特征融合。具体结构如图 5 所示。

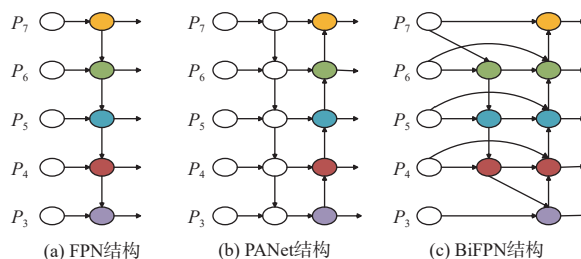


图5 特征融合结构对比

Fig. 5 Comparison of feature fusion structures

考虑到不同分辨率的输入特征在融合时对输出特征的贡献存在差异, BiFPN 采用了一种新的方法, 即为每个输出特征引入了一个额外的权重。

这样一来, 网络可以通过学习这些权重来判断每个特征的重要性, 并在融合过程中对其进行适当的加权处理。以第 6 级的融合特征为例:

$$P_6^{\text{td}} = \text{conv} \left(\frac{w_1 \cdot P_6^{\text{in}} + w_2 \cdot \text{Resize}(P_7^{\text{in}})}{w_1 + w_2 + \varphi} \right) \quad (6)$$

$$P_6^{\text{out}} = \text{conv} \left(\frac{w'_1 \cdot P_6^{\text{in}} + w'_2 \cdot P_6^{\text{td}} + w'_3 \cdot \text{Resize}(P_5^{\text{out}})}{w'_1 + w'_2 + w'_3 + \varphi} \right) \quad (7)$$

式中: conv 代表进行卷积操作; Resize 代表尺寸调整操作; φ 代表归一化参数, 用来避免数值不稳定; w_i 为可学习的权重值, 这个值既可以是标量、矢量, 还可以是多维张量, 又因为权重的值可能是无界的, 所以加入了权重归一化来限定权值的取值范围; P_6^{td} 是自顶向下的路径中第 6 层的中间特征; P_6^{out} 是自底向上的路径中第 6 层的输出特征; P_5^{out} 是第 5 层的输出特征; P_7^{in} 是第 7 层的输入特征。这个方法可以有效提高各个尺度之间的跨尺度融合效率, 且不会增加大量的参数与计算量, 使网络结构轻量化。

2.3 改进的注意力检测模块

交通标志的检测在智能车辆行驶过程中是十分重要的功能, 但是现代道路交通标识种类繁多, 且大部分道路上交通标识重叠堆放, 多数道路上具有大量的其他行驶车辆与遮盖物, 对模型的目标检测任务带来巨大的压力。本文在 RT-DETR 结构中加入全局注意力机制 (global attention mechanism, GAM), 通过简化图像信息与放大全局交互表征来提高神经网络的性能, 将模型的注意力保持在全局信息上来, 在保证不过多地增加参数与计算量的前提下, 提升神经网络的可扩展性与鲁棒性, 可以在保持轻量化的同时增加模型的检测精度。

一般的注意力机制如 CBAM(convolutional block attention module)、BAM(bottleneck attention module)、TAM(time attention module)对通道、空间宽度和空间高度的全部 3 个维度之间的关系关注不足。全局注意力机制 GAM 放大了 3 个维度之间的关系, 且可以捕捉到 3 个维度中的重要特征。该注意力机制利用 CBAM 的顺序通道-空间注意力的基本结构, 重新设计了子模块。给定输入特征为 $F_1 \in R^{c \times h \times w}$, 中间状态 F_2 和输出 F_3 则定义为

$$F_2 = M_c(F_1) \otimes F_1 \quad (8)$$

$$F_3 = M_s(F_2) \otimes F_2 \quad (9)$$

式中: M_c 和 M_s 分别为通道与空间注意力; $\otimes$ 表示各元素进行的乘法运算。具体结构如图 6 所示。

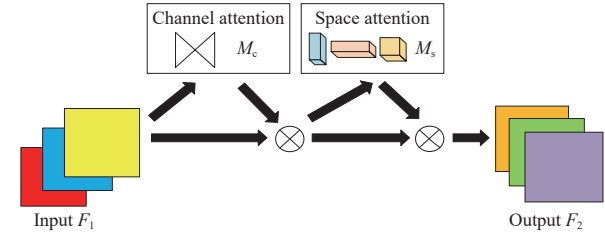


图 6 GAM 注意力结构

Fig. 6 Structure diagram of global attention mechanism

通道注意力模块以三维排列的方式存储横跨 3 个维度的信息。通过采用双层的多层感知器 (MLP) 的方法, 使该模块增强了在跨维度中通道-空间方面的依赖性。在空间注意力子模块中, 为了集中空间信息, 引入两个卷积层以融合空间信息。同时, 由于最大池化操作减少了信息的利用, 产生了不利的影响, 为了更好地保留特征映射, 在这里取消了最大池化操作。

在 RT-DETR 的跨尺度特征融合中对每一个得到的特征输入进行全局注意机制的处理, 使神经网络结构对于被遮挡目标和模糊目标的检测精度提高, 在保持轻量化的情况下可以极大地提升模型检测物体的准确率。

3 实验结果与分析

3.1 实验环境

实验中所使用的硬件环境是在 Windows 10 的操作系统下, CPU 为 Intel(R)Xeon(R)Gold 6330@2.00 GHz, GPU 为 NVIDIA GeForce RTX 3090, GPU 内存为 24 GB, 编程语言是 Python 3.8, Pytorch 的版本是 2.0.0, CUDA 的版本为 11.8, 本文中所有实验都基于本硬件环境进行, 使用相同的数据集。

3.2 实验数据集

本文采用的实验数据集是通过网上搜集、公开数据集抽选和自主拍摄获得。公路交通标志类型错综复杂, 不完全使用同一种数据集而采用多种方式构建数据集, 是因为其可以有效地增加数据集中标志的类型, 丰富数据集的种类。抽选的公共数据集为 RoadSign 交通标志数据集, RoadSign 是 Paddle 官网用于目标检测的公共数据集, 是一款双目相机数据集, 采集于多个城市, 包括了春夏秋冬 4 个季节的街区场景, 是在不同时间、天气条

件、照明条件以及移动模糊下收集的,存在遮挡、模糊等目标信息,可以用于小目标检测。此数据集含有常见的 4 种交通标志种类:人行横道标识 (crosswalk)、速度限制标识 (speedlimit)、停止标识 (stop)、交通灯 (trafficlight),数据集大小为 640×640 像素。通过图像的随机剪切、旋转、缩放等方法,将数据集扩增到 1 600 张,并将其以 8:2 的比例划分为训练集与验证集,部分数据集如图 7 所示。



图 7 交通标志数据集

Fig. 7 Traffic sign datasets

3.3 评价指标

为了评价模型的性能,从不同的角度评价模型改进的有效性,选取精确率 (Precision)、召回率 (Recall)、均值平均精度 mAP (mean average precision)

等评价指标。在轻量化方面,选取参数量 (Params) 和模型大小 (Model size) 来评价模型轻量化程度。

精准率又称查准率,是指被正确预测的样本占所有检出样本的百分比,是针对预测的结果。

$$P = \frac{T_P}{T_P + F_P} \quad (10)$$

式中: P 是指精确率; T_P 是指正确预测且为正样本的样本; F_P 是指错误预测且为正样本的样本。

召回率是指被正确预测为正样本占所有样本的比例,是针对原样本数据集。

$$R = \frac{T_P}{T_P + F_N} \quad (11)$$

式中: R 是指召回率; F_N 是指错误预测为负样本的样本。

mAP 是指所有单个目标类别检测精度的平均值,是将精准率和召回率相结合并能够反映全局性能的指标。

$$P_{\text{mAP}} = \frac{1}{N} \sum_{i=1}^N \int_0^1 P(R) dR \quad (12)$$

式中: N 为样本的类别数。

3.4 消融实验

为了证明改进模块是否有效,本文设计了消融实验进行对比验证。每组模型中仅替换文章中进行改进的模块,在本文的数据集中进行 200 轮的训练。实验结果见表 1。

表 1 消融实验结果

Table 1 Results of ablation experiment

实验	VanillaNet	BiFPN	GAM	Precision	Recall	mAP	Params/ 10^6
1				0.934	0.767	0.841	32.8
2	√			0.920	0.748	0.829	23.2
3		√		0.937	0.755	0.845	30.9
4			√	0.947	0.798	0.874	35.3
5	√	√		0.940	0.795	0.843	23.5
6	√		√	0.945	0.783	0.871	27.3
7		√	√	0.913	0.829	0.879	33.1
8	√	√	√	0.895	0.837	0.877	25.8

通过对 RT-DETR 模型依次添加改进模块进行消融实验,验证每个模块的实验效果,具体结果见表 1。实验 2 在原模型中将主干网络替换为 VanillaNet,模型的大小降低 29.2%,但模型精确度降低 1.5%,mAP 值降低 1.4%,说明网络模型的大小得到减轻,但由于 VanillaNet 的深度相较于一般的 ResNet 较浅,对于一些小目标的检测不够精准,导

致模型精确度与 mAP 值下降。实验 3 将混合编码器中的 PAN 结构替换为 BiFPN 结构,模型的精确率和 mAP 值分别提高 0.3%、0.4%,说明 BiFPN 结构可以利用其中的可学习权重充分混合图像特征,提高模型的识别能力。实验 4 在小目标层上加入 GAM 注意力,模型精确率与 mAP 值分别提高 1.3%、3.9%,通过 GAM 的全局关注力与减弱一些

对图像无关的特征权重, 增强了识别遮挡物与模糊目标的能力。实验 5 和实验 6 通过分别替换主干网络模型中 BiFPN 结构和 GAM 注意力机制, 模型的大小得到一定程度的减小, 精准度也得到了提升, 说明 VanillaNet 结构与 BiFPN 或 GAM 注意力机制相容且不会对模型识别能力得到损害。实验 7 对原始模型同时替换 BiFPN 结构和添加 GAM 注意力机制, 发现模型的 mAP 上升, 说明模型中的 GAM 注意力机制可以使模型学习全局表征、BiFPN 结构融合全局特征, 对模型识别正确目标提供帮助, 但由于 BiFPN 结构融合时移除了一些多余的节点, 导致精准率下降。实验 8 为添加本文中全部改进结构, 模型大小降低了 21.3%, mAP 值提高 4.2%, 相比于实验 7 中的模型, mAP 值基本没有下降, 但模型大小得到有效减少, 可保持模型轻量化的同时总体精确度不变, 使其可以部署在大多数的小型车载设备上, 用于车辆道路检测和交通标志识别。

在同时替换 BiFPN 与添加 GAM 注意力机制时, 模型的精确率有所降低但召回率得到了较大的提升, 使得模型最终 mAP 值得到一定的提升。因为 BiFPN 融合时添加了 GAM 注意力机制, 可能使模型过于追求三维空间的全局交互表征, 忽略了一些专属于某些特征的特点, 但全局交互表征的融合, 有利于增加模型在目标检测时的正确率, 从而提高模型整体的检测精度。

3.5 横向对比实验

为了保证改进 RT-DETR 算法的有效性, 在相同的数据集下, 将改进算法与 YOLO 系列算法、SSD 算法和 Fast R-CNN 算法进行比较, 实验结果见表 2。

表 2 横向对比实验结果

Table 2 Experimental results of transverse comparison					
Model	Precision	Recall	mAP	Params/10 ⁶	Model Size/Mb
Fast-RCNN	0.912	0.797	0.832	78.4	152.9
SSD	0.909	0.784	0.827	51.3	101.7
YOLOv5	0.874	0.662	0.782	46.6	91.3
YOLOv8	0.926	0.812	0.855	43.6	87.2
RT-DETR	0.934	0.767	0.841	32.8	63.1
Improved RT-DETR	0.895	0.837	0.877	25.8	51.6

由表可知, 改进后的 RT-DETR 网络模型在检测性能方面优于当前的主流网络。本文通过交通

场景中选择一些图片进行检测, 得到的检测结果见图 8。



图 8 检测结果图

Fig. 8 Images of test results

4 结论

本文提出了一种轻量化的 RT-DETR 交通标志检测算法。首先在主干网络方面, 将 ResNet 网络替换为更加轻量化的 VanillaNet 网络, 降低了模型的参数量; 其次采用 BiFPN 网络结构替换跨尺度特征融合模块中的 PAN 与 FPN 结构, 使模型在跨尺度特征融合时, 对不同分辨率的不同特征中各添加一个参数, 可以有效提高所需特征在跨尺度特征融合中的比重, 增强模型在模糊环境下的检测性能; 最后再加入 GAM 注意力机制, 可以提高模型对图像特征的全面感知, 在增加少量的参数与计算量的情况下, 增强了神经网络的可扩展性与鲁棒性, 极大地提升了模型的精确度。

实验证明, 本文改进方法相比于原模型, 检测准确率提升 4.2%, 优于当前的主流目标检测算法, 并且参数量下降 21%, 模型大小减少了 11.5 Mb。但是本算法仍存在一些缺陷, 例如对小目标检测精度不足、轻量化程度不够高等。接下来研究的方向是使用剪枝、知识蒸馏的方法进一步压缩模型, 使其可以达到更加轻量化的效果。

参考文献:

[1] XU X, JIN J, ZHANG S, et al. Smart data driven traffic sign detection method based on adaptive color threshold and shape symmetry[J]. Future Generation Computer

- Systems, 2019, 94: 381-391.
- [2] 戴雪瑞, 袁雪, 乐国庆, 等. 复杂场景下基于颜色对和 MSER 的交通标志检测方法[J]. 北京交通大学学报, 2018, 42(5): 107-115.
- DAI Xuerui, YUAN Xue, YUE Guoqing, et al. Detection method of traffic signs based on color pair and MSER in the complex environment[J]. Journal of Beijing Jiaotong University, 2018, 42(5): 107-115.
- [3] WU X, WEI Z, HU Y, et al. Traffic sign detection method using multi-color space fusion[C]//2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA). New York: IEEE, 2020: 314-319.
- [4] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2016: 779-788.
- [5] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//Proceedings of the European Conference on Computer Vision(ECCV). New York: IEEE, 2016: 21-37.
- [6] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2014: 580-587.
- [7] PRAKASH A S, VIGNESHWARAN D, AYYALU R S, et al. Traffic sign recognition using deeplearning for autonomous driverless vehicles[C]//2021 5th International Conference on Computing Methodologies and Communication (ICCMC). New York: IEEE, 2021: 1569-1572.
- [8] HE Z, XIAO Z, YAN Z. Traffic sign recognition based on convolutional neural network model[C]//2020 Chinese Automation Congress (CAC). New York: IEEE, 2020: 155-158.
- [9] ZHANG Q, SHEN Y, YI Z. Video-based traffic sign detection and recognition[J]. SPIE, 2019, 11321: 284-291.
- [10] 伍锡如, 邱涛涛, 王耀南. 改进 mask R-CNN 的交通场景多目标快速检测与分割[J]. 仪器仪表学报, 2021, 42(7): 242-249.
- WU Xiru, QIU Taotao, WANG Yaonan. Fast multi-object detection and segmentation of traffic scene based on improved mask R-CNN[J]. Chinese Journal of Scientific Instrument, 2021, 42(7): 242-249.
- [11] YANG L, LUO J, SONG X, et al. Robust vehicle speed measurement based on feature information fusion for vehicle multi-characteristic detection[J]. Entropy, 2021, 23(7): 910-917.
- [12] 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.
- [13] ZHU X, SU W, LU L, et al. Deformable DETR: deformable transformers for end-to-end object detection[EB/OL]. [2021-03-18]. <http://arxiv.org/pdf/2010.04159-v4>.
- [14] MENG D, CHEN X, FAN Z, et al. Conditional DETR for fast training convergence[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. New York: IEEE, 2021: 3651-3660.
- [15] LIU S, LI F, ZHANG H, et al. DAB-DETR: Dynamic anchor boxes are better queries for DETR[EB/OL]. [2022-01-28]. <http://www.paperreading.club/page?id=107439>.
- [16] ZHAO Y, LV W, XU S, et al. DETRs beat YOLOs on real-time object detection[EB/OL]. [2024-04-03]. <http://arxiv.org/pdf/2304.08069v3>.
- [17] CHEN H, WANG Y, GUO J, et al. VanillaNet: the power of minimalism in deep learning[EB/OL]. [2023-05-23]. arXiv: 2305.12972, 2023. <http://arxiv.org/pdf/2305.12972>.
- [18] TAN M, PANG R, LE Q V. Efficientdet: scalable and efficient object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2020: 10781-10790.
- [19] LIU Y, SHAO Z, HOFFMANN N. Global attention mechanism: retain information to enhance channel-spatial interactions[EB/OL]. [2021-11-10]. <http://arxiv.org/abs/2112.05561?context=cs>.
- [20] GLOROT X, BORDES A, BENGIO Y. Deep sparse rectifier neural networks[J]. Journal of Machine Learning Research, 2011, 15: 315-323.