

DOI:10.19651/j.cnki.emt.2417315

基于高阶空间特征聚合的车型识别算法^{*}

杨潞霞^{1,2} 薛映昭^{1,2} 张红瑞^{1,2} 马永杰³(1. 太原师范学院计算机科学与技术学院 晋中 030619; 2. 智能优化计算与区块链技术山西省重点实验室 晋中 030619;
3. 西北师范大学物理与电子工程学院 兰州 730070)

摘要: 针对复杂交通场景下车型目标密集、遮挡而造成的车型识别精度低的问题,提出基于高阶空间特征聚合的车型识别算法。首先,在特征提取的下采样阶段,设计了HSIDM模块,实现更深层次的特征聚合,减少细小信息损失。其次,在特征融合部分设计了DMFAM模块,动态调整各尺度特征的权重,获取多尺度的上下文信息,以增强模型对多样化特征的适应能力。然后,设计解耦REL-Head检测头,将分类和回归任务拆解,避免任务混杂,增强局部特征的学习能力与抗干扰能力。最后,将本文模型部署到边缘设备进行测试。实验结果显示,本文算法在复杂度交通场景数据集BIT-Vehicle和UA-DETRAC上,mAP相较于原模型分别提升了0.7%和3.9%,并在边缘设备上可以流畅运行,具有较好的识别效果。表明所提出的方法能够有效提高车型识别的精度并应用于受限设备。

关键词: 车型识别;高阶空间交互;动态多尺度特征聚合;解耦检测头

中图分类号: TP391.4; TN0 文献标识码: A 国家标准学科分类代码: 520.20

Vehicle recognition algorithm based on high-order spatial feature aggregation

Yang Luxia^{1,2} Xue Yingzhao^{1,2} Zhang Hongrui^{1,2} Ma Yongjie³(1. School of Computer Science and Technology, Taiyuan Normal University, Jinzhong 030619, China;
2. Shanxi Provincial Key Laboratory of Intelligent Optimization Computing and Blockchain Technology, Jinzhong 030619, China;
3. School of Physics and Electronic Engineering, Northwest Normal University, Lanzhou 730070, China)

Abstract: Addressing the issue of low vehicle recognition accuracy stemming from dense vehicle targets and occlusions in complex traffic scenarios, a vehicle recognition algorithm utilizing high-order spatial feature aggregation is proposed. Initially, during the downsampling phase of feature extraction, the HSIDM module is devised to facilitate deeper feature aggregation and minimize the loss of fine details. Subsequently, within the feature fusion component, the DMFAM module is introduced to dynamically adjust the weights of features across various scales, thereby acquiring multi-scale contextual information and bolstering the model's adaptability to diverse features. Following this, a decoupled REL-Head detector is formulated to disentangle classification and regression tasks, preventing task mixing and enhancing the learning capability and interference resistance of local features. Ultimately, the model presented in this paper is deployed on edge devices for testing. Experimental outcomes reveal that on the complex traffic scene datasets BIT-Vehicle and UA-DETRAC, the mean average precision (mAP) of our algorithm has improved by 0.7% and 3.9% respectively compared to the original model. Additionally, it operates seamlessly on edge devices, demonstrating effective recognition capabilities. This indicates that the proposed approach can effectively enhance the precision of vehicle identification and is suitable for use on constrained devices.

Keywords: vehicle type identification; high-order spatial interaction; dynamic multi-scale feature aggregation; decoupling detect

0 引言

车型识别是指通过计算机视觉技术自动识别和分类车

辆的型号,被广泛应用于监控交通流量、识别违章车辆等方面,对提升城市交通管理效率、城市智能交通系统^[1]建设起着至关重要的作用。

收稿日期:2024-11-08

* 基金项目:国家自然科学基金(62066041)、山西省重点研发计划(202102010101008)、山西省高等学校科技创新项目(2024L295)资助

随着人工智能技术的不断发展,学者们提出了许多基于深度学习的车型识别方法。基于深度学习的车型识别算法一般分为单阶段算法和二阶段算法。对于二阶段算法,桑军等^[2]提出了将 Faster-RCNN 网络与 ZF、VGG-16 以及 ResNet-101 网络结合的方法用于提升识别精度,但模型参数量增加,计算效率低下。江昆鹏等^[3]调整了区域建议网络(RPN)中锚(Anchor)的尺寸大小,提高对车辆目标的感知性。封慧杰等^[4]利用 ShuffleNet 网络对车辆声学信息进行捕捉,识别车型,但噪声对识别精度影响很大。袁公萍等^[5]引入中心损失函数对 Faster-RCNN 中 softmax 损失函数进行优化,使得检测精度提高 0.6%,速度提高 0.29 倍。以上二阶段算法虽提升了识别精度与识别速度,但检测速度较慢且易受到干扰,实时性难以满足复杂交通场景需求。

单阶段算法在图像上直接进行目标检测,不需要生成候选区域,主要包括 YOLO^[6-8] 和 SSD^[9] 等经典算法,国内外学者做了很多研究,李浩等^[10]在 ResNetXt 网络中使用多级注意力机制来提高特征信息的利用率。牟亮等^[11]提出一种基于梯度压缩的 Adam 优化算法(Adam-GC)来提高 YOLOv4 算法的训练速度,提高车型识别的实时性。林艺华等^[12]在 YOLOv7 算法中添加注意力机制加强主干网络对车型深层特征的提取能力及泛化能力。Tas 等^[13]提出了微调预训练模型的方法用于车型识别,进一步提升检测速度。任成汉等^[14]在 RepVGG 网络基础上利用结构重参数化思想融合多分枝网络以提升网络推理速度。SP 等^[15]改进 DETR 的特征提取网络并引入 SPP 结构,提升了车型特征的捕获能力。Kang 等^[16]将模糊注意力引入 YOLO 模型,减少特征图的不确定性,提升检测器对车型目标中心的关注能力。Huang 等^[17]重新设计空间金字塔结构减少模型参数,实现在简单交通场景下的车型识别任务。Song 等^[18]设计轻量级特征提取网络 Mamba_ViT,减少了 40% 的参数量,但检测效果也表现较差。陈红等^[19]利用麻雀搜索算法对检测模型训练超参数进行优化,解决了超参数随机经验设定性问题。以上单阶段算法通过引入注意力机制、改进主干网络和调节训练参数等方法优化识别效果,但对于重叠目标、远距离目标识别效果不佳。

上述车型识别算法中,二阶段算法在车型识别精度方面相较于单阶段有一定优势,但同时面临着识别速度低、参数量大的问题,无法达到实时性要求。单阶段算法在识别速度方面有了较大的提升,使得单阶段算法可以应用于工业生产生活中,但在处理密集目标、遮挡目标时仍然存在漏检、误检等问题。因此,为了解决单阶段算法中车型目标密集、重叠所造成的识别困难的问题,本文在单阶段目标检测模型 YOLOv7 的基础上进行网络设计。首先,提出高阶空间交互下采样模块 (higher-order spatial interaction downsampling module,HSIDM),促进高阶空间特征交互,以减少下采样阶段细小特征损失。然后,构建动态多尺度

特征聚合模块 (dynamic multi-scale feature aggregation module,DMFAM),获取多尺度的上下文特征信息,提高识别的全局性和准确性。最后,设计解耦 REL-Head 检测头,增强局部特征的学习能力,获取更精准的定位效果。

1 YOLOv7 模型

YOLOv7 模型的结构如图 1 所示,包括 BackBone 主干网络和 Head 结构。BackBone 主要包括 ELAN 模块和 MP 模块,ELAN 模块主要用于图像的特征提取,MP 模块用于下采样保持通道数一致。Head 结构主要是对主干部分提取的特征进行特征融合来生成特征图,最后对特征图进行预测输出。

2 基于高阶空间特征聚合的车型识别算法

本文提出的模型由输入端、BackBone(主干网络)和 Head(头部网络)组成,结构如图 2 所示。首先,将模型的主干提取网络中的下采样模块替换为设计的 HSIDM 模块,实现空间特征交互,减少特征损失;其次,在头部网络特征融合部分使用设计的 DFMAM 模块替换普通卷积,实现不同尺度特征的有效融合;最后,使用融合注意力机制的解耦检测头 REL-Head,使网络更加聚焦车辆位置,有助于区分不同车型。

2.1 HSIDM

在获得初始特征图后,需通过下采样降低特征图分辨率,获取不同尺度的目标特征,这一操作减轻了计算负担,加快了处理速度。但是,下采样操作会压缩数据,导致细节特征的丢失,使得模型在捕捉细小特征时变得困难。因此,为了保留更多有价值的特征,减少细节线索的损失,本文设计了一个 HSIDM 模块,其结构如图 3 所示。

首先,将上一层的特征送入 3 条不同的分支网络,第 1、2 条分支包含最大池化以及卷积核为 1、3 的常规卷积模块,这两条分支可以帮助网络捕捉图像的全局特征,在减少空间维度的同时减少分辨率降低了计算复杂度,使得网络在处理时更加高效。该过程可以表示如下:

$$x_1 = f_{Conv}^{1 \times 1}(f_{Maxpool}(x)) \quad (1)$$

$$x_2 = f_{Conv}^{3 \times 3}(f_{Conv}^{1 \times 1}(x)) \quad (2)$$

式中: $f_{Maxpool}$ 表示最大池化, $f_{Conv}^{1 \times 1}$ 表示 1×1 卷积, $f_{Conv}^{3 \times 3}$ 表示 3×3 卷积, x 表示输入特征, x_1, x_2 分别表示第 1、2 条网络分支输出特征。

然后,在第 3 条分支网络,通过 3×3 卷积来提取不同尺度特征,同时缩减通道数以减少计算量,随后通过 GnBlock 生成自加权参数连接特征信息。GnBlock 中 gⁿConv^[20] 利用深度卷积捕捉远距离依赖关系并递归点乘,获取具有更强语义线索的高阶深度空间特征,减少细节线索的损失,提高网络对车型特征的捕获能力,最后通过 laynormal 层加快模型收敛,GnBlock 模块结构如图 4 所示。第 3 条分支操作过程可以表示如下:

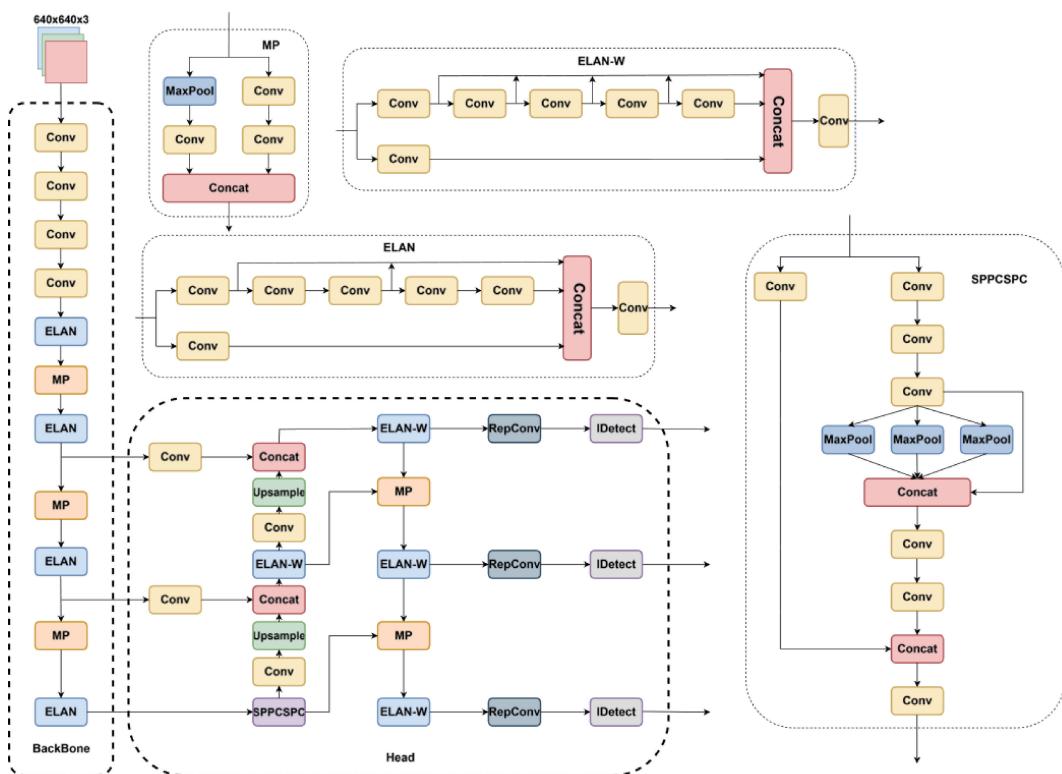


图1 YOLOv7 网络结构图

Fig. 1 Network structure diagram of YOLOv7

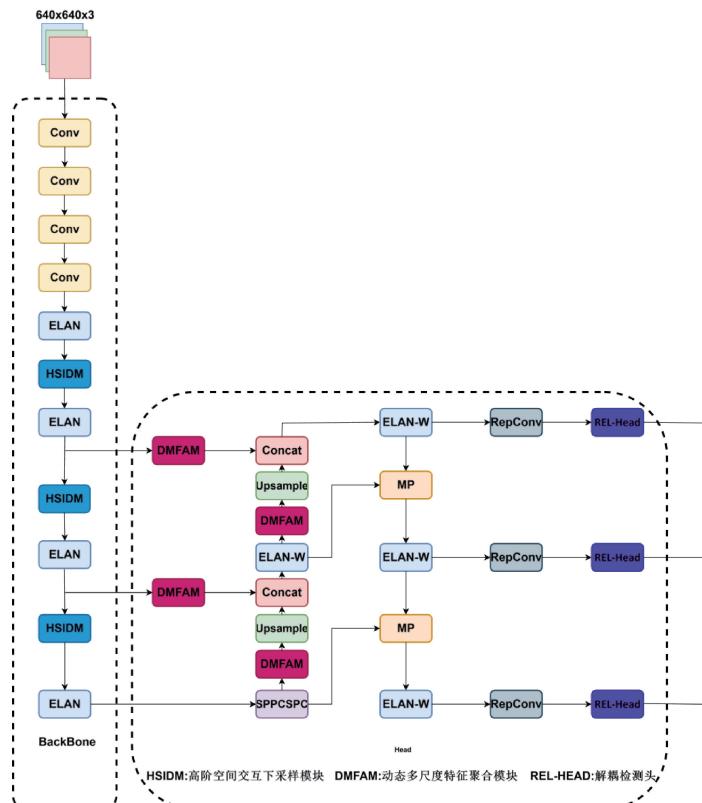


图2 基于高阶空间特征聚合的车型识别算法网络结构图

Fig. 2 Network structure diagram of vehicle recognition algorithm based on high-order spatial feature aggregation

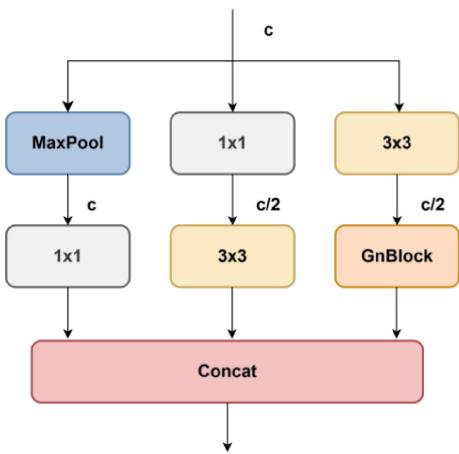


图 3 HSIDM 结构图

Fig. 3 Structure diagram of HSIDM

$$x_3^c = f_{conv}^{3 \times 3}(x)$$
 (3)

$$K_0 = f_{Conv}(x_3^c)$$
 (4)

$$C_0 = \frac{C}{2^{m-1}}$$
 (5)

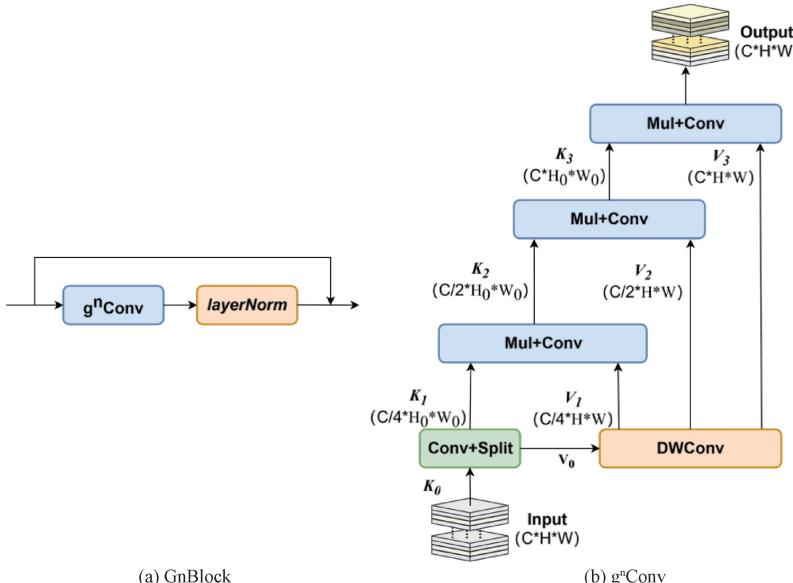
$$\varphi(V_0) = [V_1^{H \times W \times C_0}, V_2^{H \times W \times 2C_0}, \dots, V_m^{H \times W \times C}]$$
 (6)

$$K_i = \begin{cases} K_0^{H \times W \times C_0}, \\ f_{Conv}(K_{i-1} \otimes V_{i-1}), \end{cases}$$
 (7)

$$x_3 = x \oplus f_{LN}(K_m)$$
 (8)

式中: 函数 φ 表示 DW 卷积^[21], m 表示 g^nConv 过程中的空间顺序集, \otimes 表示广播逐元素相乘, \oplus 表示广播逐元素加法, f_{LN} 表示 laynormal 函数, x_3 表示第 3 条分支输出特征。

上述公式首先通过卷积增加维数, 并按照预定比例将其划分为两组特征映射 K_1 和 V_0 。 V_0 经过 DW 卷积后, 输

图 4 GnBlock 模块与 g^nConv Fig. 4 The GnBlock module and g^nConv

出多组子特征映射 V_i ($1 \leq i \leq m$)。然后 K_1 和 V_0 进行点积运算, 通过卷积增加维数得到 K_2 。重复迭代过程, 直到最后一个子特征映射 K_m , 随后与残差边进行逐元素加法, 得到第三条分支网络输出。

最后, 将第 1、2、3 条分支的通道进行拼接操作, 公式如下:

$$y = Cat(x_1, x_2, x_3) \quad (9)$$

式中: Cat 表示 Concat 操作。

2.2 DMFAM

在复杂的现实交通场景场景中, 存在识别对象大小不一、识别对象远离或接近摄像头等问题, 导致同一类别和不同类别识别对象在尺寸与外观方面具有较大差异, 进而对多尺度特征提取以及融合产生负面影响。为了解决上述问

题, 本文提出了 DMFAM 模块, 其网络结构如图 5 所示。

首先, 通过 4 个平行的 1×1 的卷积来调整通道数, 以减少计算量, 该过程可表示为:

$$x_1^{c/4} = f_{Conv}^{1 \times 1}(x^c)$$
 (10)

$$x_2^{c/4} = f_{Conv}^{1 \times 1}(x^c)$$
 (11)

$$x_3^{c/4} = f_{Conv}^{1 \times 1}(x^c)$$
 (12)

$$x_5^{c/4} = f_{Conv}^{1 \times 1}(x^c)$$
 (13)

其次, 通过 3×3 卷积模块以及膨胀率分别为 {1, 3, 5} 的膨胀卷积分支来获得多尺度特征, 可表示如下:

$$y_1^{c/4} = f_{Dconv}^{5 \times 5}(f_{Conv}^{3 \times 3}(x_1^{c/4}))$$
 (14)

$$y_2^{c/4} = f_{Dconv}^{3 \times 3}(f_{Conv}^{3 \times 3}(x_2^{c/4}))$$
 (15)

$$y_3^{c/4} = f_{Dconv}^{1 \times 1}(x_3^{c/4})$$
 (16)

式中: f_{Dconv} 表示不同卷积核大小的空洞卷积。

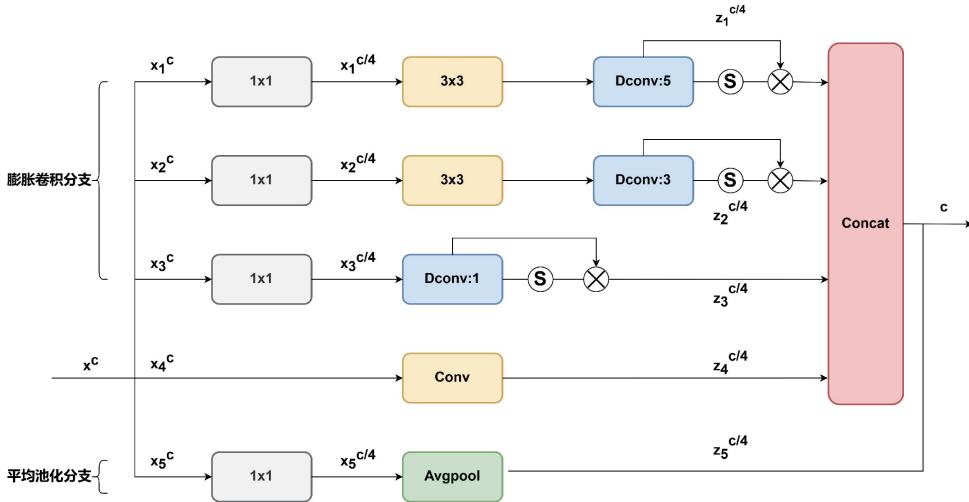


图 5 DMFAM 结构图

Fig. 5 Structure diagram of DMFAM

然后,将膨胀卷积分支通过 $sigmoid$ 函数获取权重系数,以动态调整各个输入特征,帮助模型选择更相关的目标信息,实现多尺度特征感知,如式(17)~(19)所示。

$$z_1^{c/4} = y_1^{c/4} \otimes f_{sigmoid}(y_1^{c/4}) \quad (17)$$

$$z_2^{c/4} = y_2^{c/4} \otimes f_{sigmoid}(y_2^{c/4}) \quad (18)$$

$$z_3^{c/4} = y_3^{c/4} \otimes f_{sigmoid}(y_3^{c/4}) \quad (19)$$

同时,另一条分支使用平均池化对特征图的高度和宽度进行操作,从而提取全局信息,如式(20)所示。

$$z_5^{c/4} = f_{Avgpool}(x_5^{c/2}) \quad (20)$$

最后,将膨胀卷积分支与卷积分支进行拼接,并与平均池化分支进行融合,从而获得了多尺度特征上下文信息,如式(21)~(23)所示。

$$z_4^{c/4} = f_{Conv}(x_4^{c/4}) \quad (21)$$

$$z_6^c = Cat(z_1^{c/4}, z_2^{c/4}, z_3^{c/4}, z_4^{c/4}) \quad (22)$$

$$z^c = z_5^{c/4} \oplus z_6^c \quad (23)$$

2.3 REL-Head 解耦检测头

YOLOv7 模型的检测头结构为耦合头,即图像在经过主干网络和 Head 网络的特征提取和特征融合过程后,通过几个卷积层直接将预测结果输出。然而,分类任务和回归任务所关注的内容不同,直接共同输出结果会对识别效果产生负面影响。

因此,受 YOLOX^[22] 模型的启发,本文将分类任务和回归任务解耦,设计了一个 REL-Head 解耦检测头,避免两个任务混杂,从而提高网络的识别效果。而 YOLOX 的解耦头包含两个分支,存在多个卷积,会造成参数量的增加。基于上述问题,本文设计了残差高效局部注意力机制 (residual efficient local attention, ResELA) 以及局部增强模块 (local enhancement module, LEM),在减少模型参数量的同时提升识别效果,其结构如图 6 所示。

首先,利用 1×1 卷积将通道数调整至 256,减少计算负担,该过程可表示为:

$$x_1^{c/4}, x_2^{c/4} = f_{Conv}^{1 \times 1}(x^c) \quad (24)$$

然后,进入分类分支。分类分支通过一个 ResELA 和 1×1 卷积模块预测分类。ResELA 包含 3×3 卷积以及 ELA^[23] 注意力机制组成的残差边,通过提取水平和垂直方向的特征向量,在每个方向上生成丰富的目标位置特征,增强位置信息,更好地捕获车型特征。ELA 注意力机制的结构如图 7 所示。分类分支的操作过程可以表示如下:

$$Y_1 = f_{Conv}^{3 \times 3}(x_1^{c/4}) \quad (25)$$

$$Y_2 = f_{ELA}(x_1^{c/4}) \quad (26)$$

$$X_1 = f_{Conv}^{1 \times 1}(Y_1 \oplus Y_2) \quad (27)$$

式中: Y_1 表示 ResELA 中 3×3 卷积的输出特征, Y_2 表示 ELA 的输出特征, X_1 表示 ResELA 输出特征。

同时,定位分支通过 ResELA 得到强调特征。计算后特征如式(28)所示。

$$y_2^{c/0} = f_{ResELA}(x_2^{c/0}) \quad (28)$$

式中: f_{ResELA} 表示 ResELA 操作。

随后,通过 LEM 模块以及 1×1 卷积来对位置坐标预测和置信度预测。LEM 模块由残差边和 DW 卷积组成, 1×1 的深度卷积可以进一步调制每个通道的重要性,使模型更加聚焦车型具体特征。 3×3 深度卷积可以学习局部信息并补偿模型的感应缺失,再通过一条残差边减少因为网络层数增多而导致梯度退化与梯度消失问题。操作过程如下:

$$y_1^{c/0} = f_{DWConv}^{1 \times 1}(f_{DWConv}^{1 \times 1}(f_{DWConv}^{1 \times 1}(f_{DWConv}^{1 \times 1}(y_2^{c/0})))) \quad (29)$$

$$X_2, X_3 = f_{Conv}^{1 \times 1}(y_1^{c/0} + y_2^{c/0}) \quad (30)$$

最后,通过特征 X_1 预测出车型类别,特征 X_2 预测 4 个位置坐标, X_3 特征预测置信度。

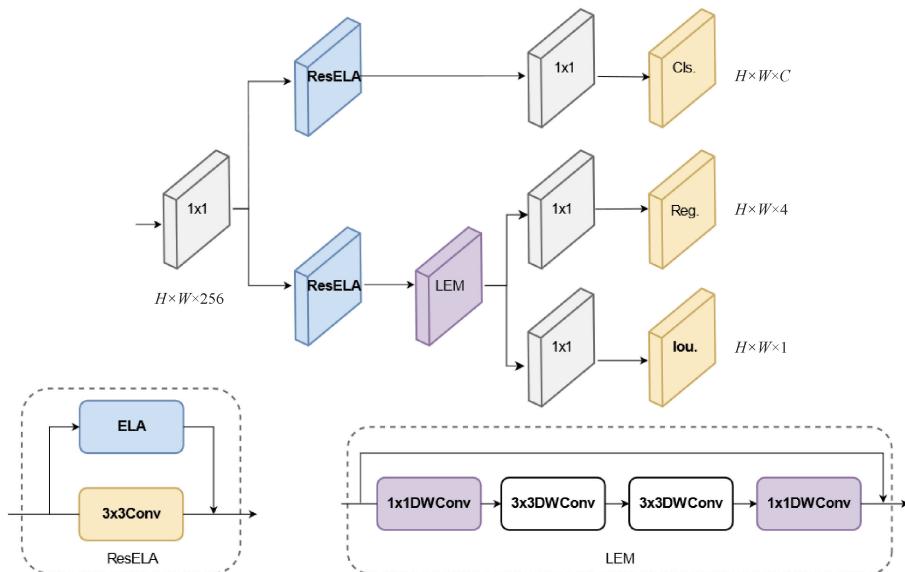


图 6 REL-Head 结构图

Fig. 6 Structure diagram of REL-Head

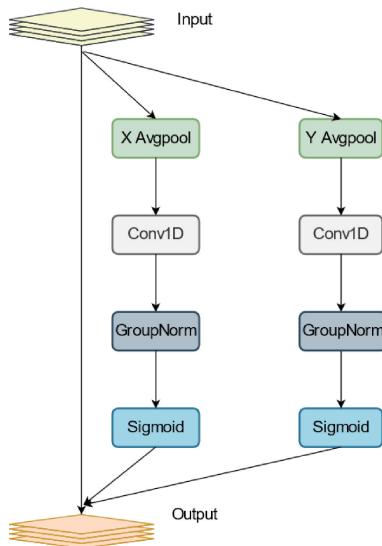


图 7 ELA 注意力机制结构图

Fig. 7 Structure diagram of ELA attention mechanism

3 实验结果与分析

3.1 数据集及实验设置

实验平台的硬件环境为:CPU:Intel(R)Core(TM)i5-12450h, GPU:RTX3050, 显存:6 G, 内存:16 G。软件环境为:windows11+CUDA12.4+python3.9.19+pytorch2.3.1。

为了验证本文所提出算法的有效性,在两个数据集上进行实验分析,分别为北京理工大学开源的车辆车型数据集 BIT-Vehicle Dataset^[24], 交通监控场景下的大型车辆数据集 UA-DETRAC^[25]。BIT-Vehicle 数据集包含 9 850 张车辆图片, 包含 6 种标签, 分别为 SUV、轿车(Sedan)、面包

车(Microbus)、卡车(Truck)、公共汽车(Bus)和货车(Minivan), 如图 8 所示。每类对应的图片数量分别为 1 392、5 922、883、822、558 和 476。

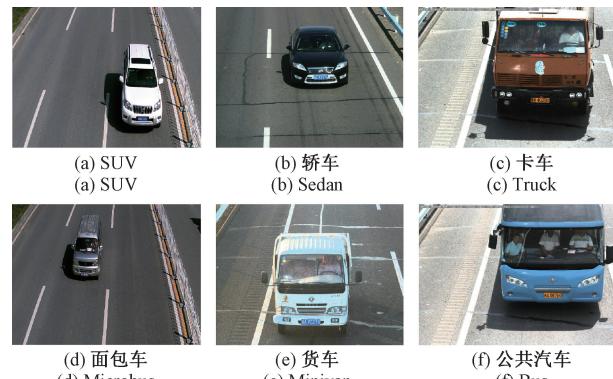


图 8 BIT-Vehicle 数据集

Fig. 8 BIT-Vehicle dataset

由于不同车型数据量差异明显,本文对面包车、卡车、公共汽车和货车类别使用了图像反转、图像拼接等数据增强方法,来扩充数据量。对轿车和 SUV 采取随机抽取策略,来保证每个车型类别数据量平衡。

数据处理完毕后,从 6 个车型类别中分别抽取 1 000 张图片构成训练数据集。同时每个类别抽取 200 张作为测试集,用于评估模型识别性能。

UA-DETRAC 数据集场景复杂,该数据集内容为在中国北京和天津的 24 个不同地点拍摄的 10 h 视频,共有超过 140 000 张图像,总共有 121 万个标记的对象边界框,其中训练集约 82 085 张图片,测试集约 56 167 张图片。车型分为 4 类,分别为轿车、公共汽车、面包车和其他车辆。天气分为 4 种,分别为阴天、晴天、雨天和夜间。

UA-DETRAC 数据集是将拍摄的视频数据的每一帧抽出并组成图片数据集, 这会导致初始数据集中的相邻图片数据几乎相同, 训练时将会浪费资源。为避免相似图片数据输入模型, 基于原始 UA-DETRAC 数据集, 从 UA-DETRAC 训练集每隔 10 张图片选取 1 张组成训练集, 共 8 639 张图片。从初始 UA-DETRAC 测试集每隔 25 张图片选取 1 张组成测试集, 共 2 231 张图片。保证训练集和测试集数据没有重叠, 构成新数据集, 新数据集如图 9 所示。

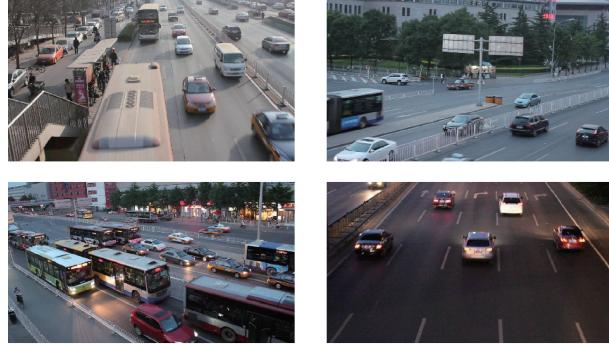


图 9 新建 UA-DETRAC 数据集

Fig. 9 The newly formed UA-DETRAC dataset

3.2 评价指标

在实际应用当中, 车型识别需要满足识别精度和识别速度两方面要求。对此, 本文研究选取 *FPS* 和 *mAP* 作为指标进行实验对比, 以此体现本文模型的实用性和优越性。

FPS 表示每秒钟内可以处理的数据帧数量, 用来评价检测速度的实时性。*FPS* 数值越大, 越能满足实时性要求。

AP 为平均精度, 平均精度是通过精确率和召回率曲线与坐标轴所围成的面积所表示的。如式(31)~(33)所示。

$$p = Precision = \frac{TP}{TP + FP} = \frac{TP}{P} \quad (31)$$

$$r = Recall = \frac{TP}{TP + FN} = \frac{TP}{T} \quad (32)$$

$$\int_0^1 p(r) dr = \sum_{k=1}^N p(k) \Delta r(k) \quad (33)$$

其中, *p* 为精确率, 是实际正样本在所有正样本中占的比例; *TP* 是被检测样本中为正确的正样本; *FP* 是被检测为错误的正样本; *FN* 是正样本却被检测为错误的负样本; *N* 为数据集中的测试集数量。

mAP 通过将需要识别的每个类的 *AP* 值求和后再进行平均获得。计算如式(34)所示, *M* 为识别的类别数。

$$mAP = \frac{\sum_{m=1}^M AP(q)}{M} \quad (34)$$

3.3 消融实验

本文以 YOLOv7 网络为基础框架, 在 BIT-Vehicle 和新建 UA-DETRAC 数据集上进行消融实验, 结果如表 1、2 所示。

表 1 BIT-Vehicle 消融实验

Table 1 BIT-Vehicle ablation experiment

YOLOv7	HSIDM	DMFAM	REL-Head	mAP0.5/%	FPS
✓				97.7	78.7
✓	✓			98.0	78.3
✓		✓		97.9	77.4
✓			✓	98.2	76.2
✓	✓	✓		98.1	76.5
✓	✓		✓	98.0	75.8
✓		✓	✓	98.1	75.6
✓	✓	✓	✓	98.4	75.3

表 2 新建 UA-DETRAC 消融实验

Table 2 The newly formed UA-DETRAC ablation experiment

YOLOv7	HSIDM	DMFAM	REL-Head	mAP0.5/%	FPS
✓				64.6	64.0
✓	✓			65.5	63.0
✓		✓		66.1	63.1
✓			✓	67.5	62.4
✓	✓	✓		66.3	63.1
✓	✓		✓	67.1	62.6
✓		✓	✓	67.8	62.4
✓	✓	✓	✓	68.5	61.0

由表 1 和 2 可以看出, 添加 HSIDM 模块后, 在 BIT-Vehicle 和新组建 UA-DETRAC 数据集上模型精度分别提高了 0.3% 和 0.9%, 且检测速度没有大幅下滑。添加 DMFAM 模块后精度分别提高了 0.2% 和 1.5%。在更换 REL-Head 检测头后, 精度提升较大, 分别提升 0.5% 和 2.9%, 这体现了 ELA 注意力机制的全局特征提取能力以及 LEM 学习局部空间特征的能力。上述结果表明, 改进后的模型与原始模型相比识别精度提升明显, 所设计的模块均对网络模型有效。

3.4 不同场景下的对比试验

1) 简单场景

为了说明本文算法在简单场景下的车型识别效果, 使用 BIT-Vehicle 数据集中的图片作为训练和测试对象。表 3 记录了 Faster R-CNN、CenterNet、Rt-Detr-l、YOLOv3、YOLOv4、YOLOv5m、YOLOv7、YOLOv8s、YOLOv10s、YOLOv11s 和近年来其他学者的改进模型

表 3 各模型在 BIT-Vehicle 数据集上的检测结果对比

Table 3 Comparison of the detection results of each model on the BIT-Vehicle dataset

模型	mAP0.5/%	FPS	AP%					
			SUV	Sedan	Microbus	Truck	Bus	Minivan
Faster R-CNN	95.6	10.5	95.6	97.3	94.6	96.1	96.3	93.6
CenterNet	95.3	29.5	96.6	98.8	88.6	99.6	92.1	96.5
Rt-Detr-l	97.8	89.6	97.5	98.4	97.3	95.4	98.7	99.4
YOLOv3	97.3	78.1	97.7	99.4	95.3	94.0	98.5	98.8
YOLOv5m	94.3	89.3	88.6	98.2	97.2	94.0	99.5	88.5
YOLOv7	97.7	78.7	97.3	99.4	97.3	95.4	98.7	99.4
YOLOv8s	94.8	123.3	88.9	97.7	96.4	93.8	99.5	92.6
YOLOv10s	95.3	155.3	87.8	98.2	97.2	95.8	99.5	93.6
YOLOv11s	95.5	118.4	87.8	98.5	97.6	95.6	99.7	94.1
文献[17]	98.2	106.4	97.4	99.3	98.3	97.0	99.1	97.8
文献[26]	87.7	—	88.2	87.0	86.3	87.2	89.2	88.2
文献[27]	94.87	11.16	96.41	95.84	94.24	97.53	96.41	93.79
本文	98.4	75.3	97.7	99.1	97.7	98.8	99.6	97.5

及本文模型的检测结果,所有模型均不使用预训练权重。

由表 3 可以看出,本文算法的识别性能优于其他算法,相较于官方模型 Faster R-CNN、CenterNet、Rt-Detr-l、YOLOv3、YOLOv5m、YOLOv7、YOLOv8s、YOLOv10s、YOLOv11s 精度分别提升了 2.8%、3.1%、0.6%、1.1%、4.1%、0.5%、3.6% 和 3.1%。相较于国内外改进后的车型识别模型文献[17]、文献[26]、文献[27],精度分别提升了 0.2%、10.7%、3.53%。文献[26]提出一种改进的鸟类群算法来优化卷积神经网络,但仅仅使用卷积神经网络只能有限的捕捉局部信息,使得特征捕获不完全,造成识别效果差,而本文的 HSIDM 结构能够有效的捕获全局上下文信息,减少特征损失。对于车辆检测,不同尺度目标的识别是十分重要的,文献[17]与文献[27]对原模型做出了轻量化操作,但其没有关注不同距离以及不同尺寸的多尺度目标,造成不同尺度目标识别能力不足,而本文设计的 DMFAM 可以动态的适应多尺度目标,REL-Head 结构可以聚焦车型目标中心,有效地提取目标。以上分析表明本文算法在简单场景下的车型识别性能有较好的检测效果,相较于基线模型 YOLOv7,本文算法对 SUV、面包车(Microbus)和卡车(Truck)的识别方面有较大的效果。在 FPS 方面,本文算法相较于 CenterNet 有较大提升,与小参数模型如 YOLOv8s 等相比,FPS 有所降低,但基本满足实时检测的需求。

2) 复杂场景

为了说明本文算法在复杂场景下的车型识别效果以及证明改进的泛化性,使用新建 UA-DETRAC 数据集进行测试,该数据集包含了不同视角、不同路况、不同天气的车辆图像。该数据集满足了车辆场景复杂、数据规模大的数据集要求。表 4 记录了 Faster-RCNN、CenterNet、

Rt-Detr-l、YOLOv3、YOLOv5s、YOLOv6s、YOLOv8s、YOLOv11s、YOLOv7-Tiny、YOLOv7、YOLOv10s 和近些年改进模型的检测结果,所有模型均不使用预训练权重。

表 4 各模型在新建 UA-DETRAC 数据集上的检测结果对比

Table 4 Comparison of the detection results of each model on the new UA-DETRAC dataset

模型	mAP0.5/%	FPS
Faster R-CNN	58.93	9.46
CenterNet	66.10	43.89
Rt-Detr-l	59.60	74.60
YOLOv3	63.57	43.10
YOLOv5s	62.00	55.20
YOLOv6s	59.00	36.00
YOLOv8s	62.60	49.00
YOLOv10s	58.80	93.20
YOLOv11s	61.20	89.60
YOLOv7-Tiny	62.80	69.25
YOLOv7	64.60	64.00
文献[18]	58.80	—
文献[28]	65.04	37.49
本文	68.50	61.00

由表 4 可以看出,复杂场景下本文算法的 mAP 值与 Faster-RCNN、CenterNet、Rt-Detr-l、YOLOv3、YOLOv5s、YOLOv6s、YOLOv8s、YOLOv10s、YOLOv11s、YOLOv7-Tiny、YOLOv7 相比分别提高了 9.57%、2.4%、8.9%、4.93%、6.5%、9.5%、5.9%、9.7%、7.3%、5.7%、3.9%,

并且 FPS 相较于大部分模型具有一定优势。其中,文献[18]引入了新的轻量级提取架构 Mamba_ViT,减少了 40% 参数量的同时获得了更完整的上下文表示,但主干网络的特征提取能力不强,特征损失较多,造成识别的精度不高,本文算法的识别精度相较于文献[18]提升了 9.7%,更加高效的完成了车型识别任务;文献[28]使用膨胀卷积和匈牙利匹配算法有效的提升了对于遮挡目标以及多尺度目标的识别能力,但这也大大增加了参数量和运算量,造成 FPS 较低,而本文算法提升识别精度的同时也保证了较好的识别速度。上述结果表明,所提出的方法在不同天气、路况下这类复杂场景下均有较好的识别效果,同时也

能够满足实时性的要求。

3.5 可视化结果分析

为了更直观的展示本文模型在不同场景下的车型识别能力,选取不同场景、不同光照强度的数据进行可视化结果对比。其中在 BIT-Vehicle 数据集中选取四张不同场景图片,在新建 UA-DETRAC 数据集上选取四张不同天气状况的图片进行对比。图 10(a)为本文算法在 BIT-Vehicle 数据集上的识别效果,图 10(b)为 YOLOv7 检测模型识别效果。图 10(c)、(d)和(e)为最新检测模型识别效果。图 11 为本文算法与 YOLOv7、最新检测模型在新建 UA-DETRAC 数据集上的识别效果。

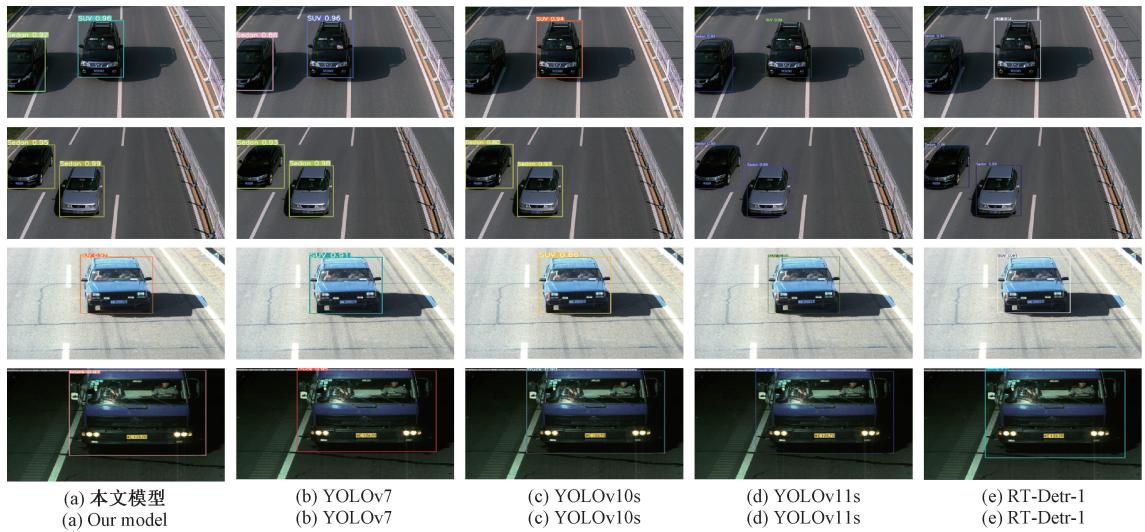


图 10 BIT-Vehicle 数据集识别效果对比

Fig. 10 Comparison of the recognition effect of BIT-Vehicle datasets

从识别效果来看,所提算法对于图像中心的完整车型有着较好的识别效果,如图 10(a)中第 2,3 行所示。同时,在图 10(a)中的 1,2,4 行展示了在多车辆以及黑暗场景下的识别结果,结果表明对于图像边缘的残缺汽车图像、黑夜下的汽车图像,算法相较于基线模型及最新的检测模型都有较大的提升,体现了本文算法在简单场景下优异的车型识别能力。

复杂场景下,YOLOv7 模型改进前后的识别效果如图 11 所示,图 11(a)、(b)、(c)、(d) 分别为模型在新建 UA-DETRAC 数据集的晴天、阴天、雨天、夜间的识别结果。第 1 行识别结果为本文模型,第 2 行识别结果为原始 YOLOv7 模型,3,4,5 行识别结果为最新检测模型。

从图 11(a)可以看出,本文算法在晴天条件下车型识别精度要优于原模型和最新检测模型,其中 YOLOv10s 和 YOLOv11s 均出现了漏检,而 Rt-Detr-l 的识别置信度低于本模型;图 11(b)为阴天条件下的识别结果,可以发现 YOLOv7、YOLOv10s、YOLOv11s 和 Rt-Detr-l 模型出现了错检,且对于远处的小型目标以及重叠目标,YOLOv7、YOLOv10s、YOLOv11s 和 Rt-Detr-l 模型均出现了漏检现

象,而本文算法预测更加稳定,体现了本文算法的抗干扰能力及对密集重叠目标的识别能力。图 11(c)中 YOLOv7、YOLOv10s 和 Rt-Detr-l 模型出现误检,而 YOLOv11s 出现漏检,造成这一结果的原因是误检车型不标准,模型无法做出正确预测,而本文算法可以较好地区分边界,得到正确结果。图 11(d)中 5 个模型都获得了较好的检测效果,但夜间条件下环境黑暗,远处车辆边界与环境难以区分,本文算法设计的新检测头能够聚焦车辆目标中心,提高目标的识别精度。

图 12 为新建 UA-DETRAC 数据集识别热力图,第 1 行为本文算法识别结果,第 2 行为 YOLOv7 模型识别结果。从图 12(a)、(b)可以看出,在光线条件下,本文算法能够更加专注于汽车整体,特别是对于远处车辆。图 12(c)中 YOLOv7 模型出现错误识别,从热力图可以看出,原模型更加关注车厢里的特征,导致识别错误,而本文算法更加关注车头车尾特征,能够更加准确的识别车型。

图 12(d)呈现的夜间环境中,提出的算法相较于 YOLOv7 模型能够更加细致的将车辆边缘于周围环境区分开,体现了所提方法精确的特征提取能力。因此本文算

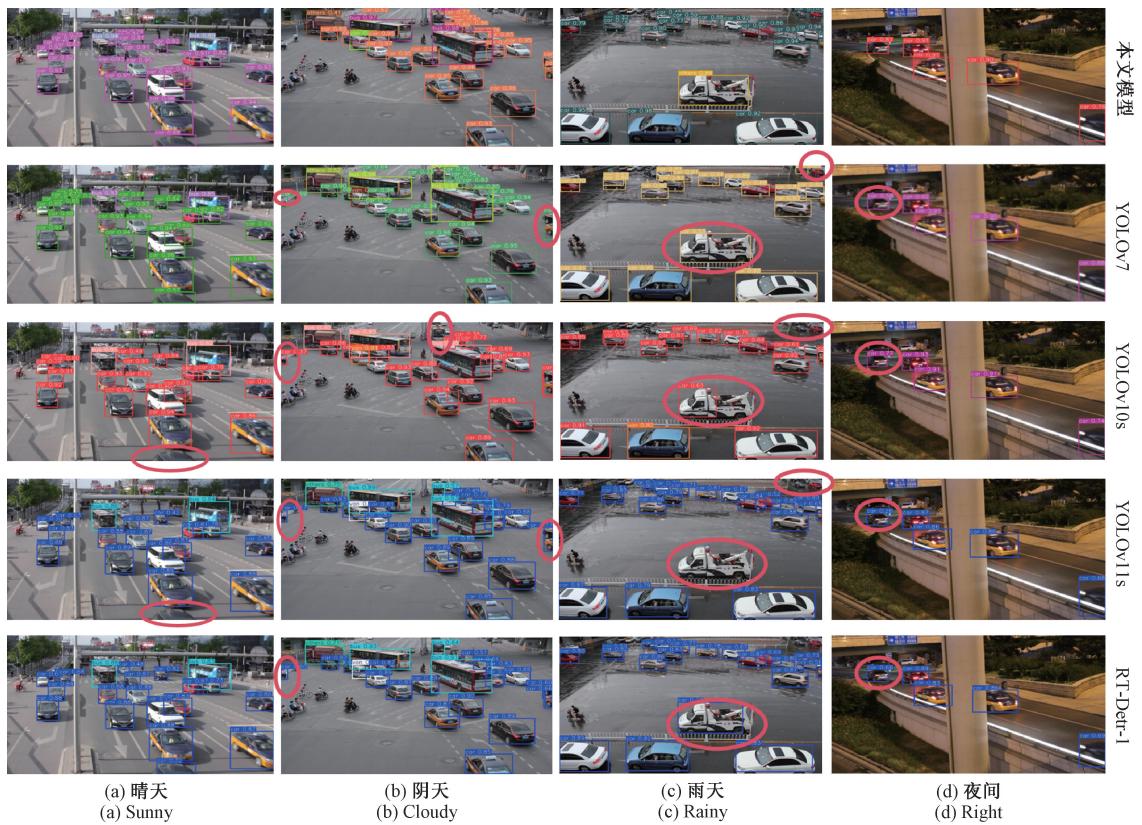


图 11 新建 UA-DETRAC 数据集识别效果对比

Fig. 11 Comparison of the recognition effect of the new UA-DETRAC dataset

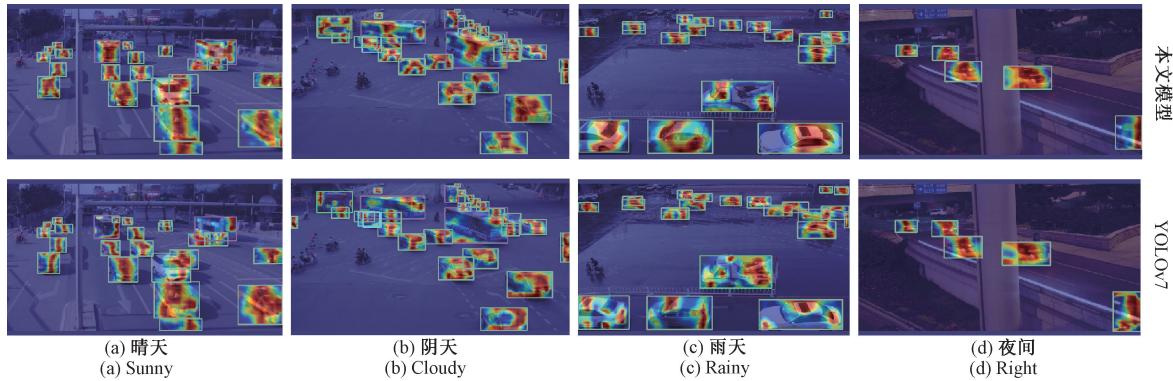


图 12 新建 UA-DETRAC 数据集识别热力图对比

Fig. 12 Comparison of heat maps for Identifying the new UA-DETRAC dataset

法在 BIT-Vehicle 和 UA-DETRAC 数据集上均有良好的识别效果。

3.6 检测头对比分析

为探究本文设计检测头的综合性能,将所设计的解耦检测头与 YOLOX 解耦检测头进行实验对比,分别在新建 UA-DETRAC 数据集上进行测试,结果如表 5 所示。

通过表 5 可知,所设计的检测头相较于原 YOLOX 检测头参数量大大下降,并且识别精度提高 2.2%,证明了本文所设计解耦头的良好的检测性能。

表 5 检测头对比

Table 5 Comparison of detection heads

检测头	P/10 ⁶	mAP0.5/%	FPS
YOLOX	7.7	65.3	63.0
REL-Head	4.4	67.5	62.4

3.7 模型部署

为了进一步验证本文模型的实用性,将所设计模型部署到边缘设备上进行测试。所使用边缘设备为 JETSON

NANO B01,如图 13 所示。



图 13 JETSON NANO B01

本文将模型转换格式并部署到 JETSON NANO B01 边缘设备,使用 UA-DETRAC 数据集进行测试,模型部署及模型测试结果如图 14、15 所示,可以看出本文模型在边缘设备有着较好的识别效果,受限于设备性能,在边缘设备的帧率为 12.4 帧,计算机帧率为 61 帧。以上表明,在资源受限的边缘设备上本模型也可以正常运行。

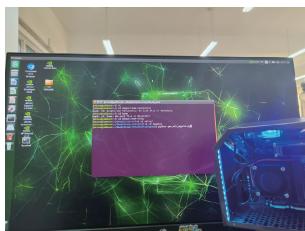


图 14 模型部署

Fig. 14 Model deployment



图 15 模型测试结果

Fig. 15 Model test results

4 结 论

本文提出了一种基于高阶空间特征聚合的车型识别算法。本文算法在特征提取部分设计新的下采样模块 HSIDM,保留车型的细节信息。为了提高模型对不同尺度信息特征的融合能力,设计了 DMFAM,强化网络对不同类型特征的感知能力。此外,为了捕获车型的有效信息,设计解耦 REL-Head 检测头,通过 ResELA 模块获取关键

线索,并利用解耦检测头将分类与回归任务拆分,提升对车型特征的学习能力。在 BIT-Vehicle 和新建 UA-DETRAC 数据集上的实验结果表明,本文模型识别精度优于 YOLOv7 模型。其次,本文算法相较于 Faster-RCNN、CenterNet、YOLOv3、YOLOv5s、YOLOv6s、YOLOv8s、YOLOv7-Tiny、YOLOv10s、YOLOv11s 算法都有更高的精度,具有一定优势且 FPS 指标没有明显下滑,能有效完成不同场景下的车型识别任务。在下一步研究中,将优化模型结构,以此降低模型复杂度和参数量,以便于能够在边缘设备上实现车型识别的部署与应用。

参 考 文 献

- [1] 衣华军,王昆鹏,张书民.城市智能交通系统设计分析及应用研究[J].时代汽车,2022(23):193-195.
- [2] YI H J, WANG K P, ZHANG SH M. Design analysis and application research of urban intelligent transportation system [J]. Auto Time, 2022 (23): 193-195.
- [3] 桑军,郭沛,项志立,等. Faster-RCNN 的车型识别分析[J].重庆大学学报,2017,40(7):32-36.
- [4] SANG J, GUO P, XIANG ZH L, et al. Vehicle detection based on faster-RCNN [J]. Journal of Chongqing University,2017,40(7):32-36.
- [5] 江昆鹏,闫洪涛,杨红卫,等.改进 Mask R-CNN 的细粒度车型识别算法[J].软件,2020,41(3):1-5.
- [6] JIANG K P, YAN H T, YANG H W, et al. Improved Mask R-CNN finegrained car recognition algorithm [J]. Software, 2020, 41(3):1-5.
- [7] 封慧杰,赵红东,于快快,等.基于轻量级卷积神经网络的车辆声学识别[J].传感器与微系统,2024,43(7):136-140.
- [8] FENG H J, ZHAO H D, YU K K, et al. Vehicle acoustic recognition based on lightweight CNN [J]. Transducer and Microsystem Technologies, 2024, 43(7):136-140.
- [9] 袁公萍,汤一平,韩旺明,等.基于深度卷积神经网络的车型识别方法[J].浙江大学学报(工学版),2018,52(4):694-702.
- [10] YUAN G P, TANG Y P, HAN W M, et al. Vehicle category recognition based on deep convolutional neural network [J]. Journal of Zhejiang University (Engineering Science), 2018,52(4):694-702.
- [11] REDMON J. You only look once: Unified, real-time object detection [C]. IEEE Conference on Computer Vision and Pattern Recognition,2016.
- [12] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger [C]. IEEE Conference on Computer Vision and Pattern Recognition,2017: 7263-7271.
- [13] FARHADI A, REDMON J. YOLOv3: An incremental improvement [C]. Computer Vision and

- Pattern Recognition, 2018, 1804: 1-6.
- [9] WEI L, DRAGOMIR A, DUMITRU E, et al. SSD: Single shot multiBox detector[J]. Springer, Cham, 2016, DOI:10.1007/978-3-319-46448-0_2.
- [10] 李浩, 鲍鸿, 詹瑞典. 融合多级注意力机制和信息融合的车型识别[J]. 电子测量技术, 2023, 46(5):164-171.
- LI H, BAO H, ZHAN R D. Research on vehicle type recognition based on multilevel attention mechanism and information fusion [J]. Electronic Measurement Technology, 2023, 46(5):164-171.
- [11] 卞亮, 赵红, 李燕, 等. 基于梯度压缩的YOLOv4 算法车型识别[J]. 工程科学学报, 2022, 44(5):940-950.
- MU L, ZHAO H, LI Y, et al. Vehicle recognition based on gradient compression and YOLO v4 algorithm[J]. Chinese Journal of Engineering, 2022, 44(5):940-950.
- [12] 林艺华, 姜浩, 钟剑, 等. 面向车型识别的YOLOv7 网络优化研究[J]. 公路交通科技, 2023, 40(S2): 361-367.
- LIN Y H, JIANG H, ZHONG J, et al. Optimization on YOLOv7 network for vehicle type identification [J]. Journal of Highway and Transportation Research and Development, 2023, 40(S2):361-367.
- [13] TAS S, SARI O, DALVEREN Y, et al. Deep learning-based vehicle classification for low quality images[J]. Sensors, 2022, 22(13): 4740.
- [14] 任成汉, 黄俊. 基于RepVGG-A0改进的公路车型识别网络[J]. 激光杂志, 2024, 45(1):166-171.
- REN CH H, HUANG J. Improved road vehicle classificcation network based on RepVGG-A0 [J]. Laser Journal, 2024, 45(1):166-171.
- [15] SP K, MOHANDAS P. DETR-SPP: A fine-tuned vehicle detection with transformer [J]. Multimedia Tools and Applications, 2024, 83(9): 25573-25594.
- [16] KANG L, LU Z, MENG L, et al. YOLO-FA: Type-1 fuzzy attention based YOLO detector for vehicle detection [J]. Expert Systems with Applications, 2024, 237: 121209.
- [17] HUANG D, TU Y, ZHANG Z, et al. A lightweight vehicle detection method fusing GSConv and coordinate attention mechanism[J]. Sensors, 2024, 24(8): 2394.
- [18] SONG Z, WANG Y, XU S, et al. Lightweight vehicle detection based on mamba_ViT[J]. Sensors, 2024, 24(22): 7138.
- [19] 陈红, 张乐. 基于麻雀搜索算法改进的YOLOv7-ECA-SSA 模型的车辆检测[J]. 国外电子测量技术, 2024, 43(2):158-164.
- CHEN H, ZHANG L. Vehicle detection based on improved YOLOv7-ECA-SSA model with sparrow search algorithm[J]. Foreign Electronic Measurement Technology, 2024, 43(2):158-164.
- [20] RAO Y, ZHAO W, TANG Y, et al. Hornet: Efficient high-order spatial interactions with recursive gated convolutions[J]. Advances in Neural Information Processing Systems, 2022, 35: 10353-10366.
- [21] CHOLLET F. Xception: Deep learning with depthwise separable convolutions [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1251-1258.
- [22] GE Z, LIU S, WANG F, et al. YOLOX: Exceeding YOLO series in 2021 [J]. ArXiv preprint arXiv: 2107.08430, 2021.
- [23] XU W, WAN Y. ELA: Efficient local attention for deep convolutional neural networks [J]. ArXiv preprint arXiv:2403.01123, 2024.
- [24] DONG Z, WU Y, PEI M, et al. Vehicle type classifycation using a semisupervised convolutional neural network[J]. IEEE Transactions on Intelligent Transportation Systems, 2015, 16(4): 2247-2256.
- [25] WEN L, DU D, CAI Z, et al. UA-DETRAC: A new benchmark and protocol for multi object detection and tracking [J]. Computer Vision and Image Understanding, 2020, 193: 102907.
- [26] CHEN X. Vehicle feature recognition via a convolutional neural network with an improved bird swarm algorithm[J]. Journal of Internet Technology, 2023, 24(2): 421-432.
- [27] 徐浩, 杨德刚, 蒋倩倩, 等. 基于SSD的轻量级车辆检测网络改进[J]. 计算机工程与应用, 2022, 58(12): 209-217.
- XU H, YANG D G, JIANG Q Q, et al. Improvement of lightweight vehicle detection network based on SSD[J]. Computer Engineering and Applications, 2022, 58(12):209-217.
- [28] WANG K, LIU M. YOLOv3-MT: A YOLOv3 using multi-target tracking for vehicle visual detection[J]. Applied Intelligence, 2022, 52(2): 2070-2091.

作者简介

杨潞霞, 博士, 教授, 主要研究方向为智能信息处理。

E-mail:ylxfyz328@163.com

张红瑞(通信作者), 博士, 讲师, 主要研究方向为智能交通系统下的机器视觉任务。

E-mail:zhanghongrui@tynu.edu.cn

薛昭映, 硕士研究生, 主要研究方向为机器视觉。

E-mail:1429984629@qq.com

马永杰, 博士, 教授, 主要研究方向为智能交通图像处理、进化算法。

E-mail:ylxfyz328@163.com