DOI:10.11918/202402020

基于注意力变形和动态查询机制的交通小目标检测

李建新[1,2],朱进玉[1],乔鸿政[3],石浩楠[1]

(1.长安大学 电控学院,西安 710064;2.中汽零部件技术(天津)有限公司,天津 300300;
3.厦门大学 航空航天学院,福建 厦门 361005)

摘 要:深度学习推动了交通目标检测发展,但复杂交通场景下密集遮挡环境中的小目标检测精度仍不足。针对上述问题提出一种注意力变形和动态查询机制的交通小目标检测算法CDAQ-DDETR,在Deformable DETR的基础上,通过引入CBAM注意力双塔机制和DCNv2可变形卷积重构原始残差网络,增强算法对密集区域交通小目标的语义获取能力;借助AFN网络思想添加低层特征,同时构建注意力感知融合金字塔模块,提高算法对多尺度中小交通目标的检测效果;依靠在原解码器前向集成动态查询机制模块结合输入图像匹配目标特性,以构建最佳查询向量提升算法对多样化背景干扰的适应泛化能力。在VisDrone2019数据集上进行实验,结果表明:CDAQ-DDETR算法在平均精确率(mAP@0.5:0.95)上已达到37.9%,在平均召回率(mAR@0.5:0.95)上已达到57.4%,相比现阶段主流SOTA算法在检测精度上提升5.5%,召回率提升8.0%,尤其针对于小目标检测精度提升6.9%,召回率提升了10.0%,同时利用可视化实验分析其更加适用于密集场景下交通小目标检测的实际应用。

关键词:交通目标检测;密集场景;小目标检测;Deformable DETR;Transformer算法
中图分类号:TP391 **文献标志码**:A **文章编号**:0367-6234(2025)07-0081-15

Traffic small object detection based on attention deformation and dynamic query mechanism

LI Jianxin[1,2], ZHU Jinyu[1], QIAO Hongzheng[3], SHI Haonan[1]

(1. School of Electrical Control, Chang'an University, Xi'an 710064, China;
2. CATARC Component Technology (Tianjin) Co., Ltd., Tianjin 300300, China;
3. School of Aeronautics and Astronautics, Xiamen University, Xiamen 361005, Fujian, China)

Abstract: While deep learning has advanced traffic object detection, accurately detecting small objects in complex traffic scenes with dense occlusion remains challenging. To address these issues, this paper proposes a novel small traffic object detection algorithm, CDAQ-DDETR, which incorporates an attention-based deformation and dynamic querying mechanism. Building upon Deformable DETR, the algorithm introduces the CBAM attention-based dual-tower mechanism and DCNv2 Deformable convolutions to reconstruct the original residual network, thereby enhancing the semantic acquisition capabilities for small traffic objects in dense areas. By leveraging the AFN network concept to add lower-level features and constructing an attention-aware fusion pyramid module, the algorithm improves detection performance for multi-scale small and medium traffic objects. Additionally, by integrating a dynamic query mechanism module before the original decoder, combined with matching input image characteristics, it constructs optimal query vectors, enhancing the algorithm's adaptability and generalization ability against diverse background interferences. Experiments conducted on the VisDrone2019 dataset show that the CDAQ-DDETR algorithm has achieved a mean Average Precision (mAP@0.5:0.95) of 37.9% and a mean Average Recall (mAR@0.5:0.95) of 57.4%. Compared to the current state-of-the-art (SOTA) algorithms, there is an improvement of 5.5% in detection precision and 8.0% in recall rate, particularly, an increase of 6.9% in precision and 10.0% in recall rate for detecting small objects. Visualization experiments further demonstrate its practical applicability and superior performance in detecting small traffic objects in dense scenes.

Keywords: traffic object detection; dense scenes; small object detection; Deformable DETR; Transformer algorithm

收稿日期:2024-02-26;**录用日期**:2024-04-18;**网络首发日期**:2025-07-08
网络首发地址:https://link.cnki.net/urlid/23.1235.T.20250708.1143.006
基金项目:国家自然基金重点项目(52232015);陕西省科技发展计划项目"两链"融合重点专项(2023KXJ-297)
作者简介:李建新(1999—),女,硕士研究生;朱进玉(1991—),男,硕士,工程师
通信作者:朱进玉,jyzhu@chd.edu.cn

交通目标检测作为一种利用计算机视觉科技定位识别图像或视频中特定交通目标的感知技术,广泛应用于智能监控、自动驾驶、交通疏导等关键性研究领域[1]。然而由于交通目标多处在复杂密集多样化场景中,外观易受遮挡、重叠、小尺度等多因素影响,导致交通目标检测难度不容小觑,因此在密集场景下的交通小目标高效精准检测是当前重要的研究课题。

传统的交通目标检测方法普遍依靠人工检测、特征检测[2]、机器学习检测[3],易在高密集遮挡区域导致小目标检测不稳定,且泛化能力差,常出现漏检误检。随着人工智能图像识别技术日益完善,深度学习目标检测算法凭借端到端训练,无需人工干预,可实现多样化场景下的特定目标检测,为交通密集场景小目标检测提供现实理论基础。目前,现有目标检测技术按照特征提取不同主要分为两类算法:一种为精度占优的基于候选框检测的 Two-stage 算法(如 Faster-RCNN[4]、Cascade-RCNN[5]),另一种为速度占优的基于回归方法检测的 One-stage 算法(如 GFL[6]、YOLOv8[7])。

鉴于目标检测算法更卓越的特征提取能力及可扩展性,较多研究学者将其应用到密集多样化背景干扰遮挡场景下交通小目标检测任务中。黄继鹏等[8]证明了 Faster-RCNN 中区域框建议模块对于小目标特征候选的有效性,同时在其后向通过上下采样构建特征金字塔,用以增强特征获取效率及抑制密集小目标遮挡信息干扰,但是其应用的主体网络 VGG16[9]本质上因卷积结构限制,依然无法获取更高层次的语义信息且增加了超参数冗余运算。针对密集场景下小目标检测效果不佳问题,杨彪等[10]通过在 YOLOv5 中添加长短时记忆 LSTM[11]神经网络,利用异构图学习目标间交互,增强小目标感知能力,但 YOLOv5 仍采用预选框特征提取,存在局部感受野受限问题,多层堆叠获取全局信息易致信息量衰竭,使特征注意力仅集中在个别区域。自然语言处理领域内的 Transformer 模型[12]依靠长距离建模和并行运算,能够融合长距离上下文信息解决长序列的遗忘现象[13],Dosovitskiy 等[14]将 Transformer 应用至图像分类提出 ViT 模型,通过注意力机制构建长距离依赖解决局部感受野受限问题,获得了相较传统分类算法更好的效果,Carion 等[15]运用 Transformer 至目标检测提出了 DETR 模型,通过集合预测思想将 Transformer 集成在检测通道中,这种真正的端到端检测思路摒弃了传统分类算法所需的预选框及 NMS 后处理,解决了传统候选框目标检测算法超参冗余及感受野受限问题。

为了进一步提高小目标检测及模型训练效率,Zhu 等[16]提出 Deformable DETR 模型,利用可变形注意力模块高效筛选出关键采样点,并无需 FPN[17]帮助便可扩展聚合多尺度特征,实现了高效注意力机制与多尺度特征融合。但是其使用主体网络 ResNet-50[18]进行特征提取,由于此网络深度较浅且具有多层下采样,对于交通高密集场景下多尺度小目标提取特征能力较弱,同时编码器前向的查询向量为随机初始化,这种固定尺寸且静态初始化的目标查询机制限制了算法对多样化交通图像内容的适应性。针对上述问题,本文提出一种改进 Deformable DETR 的多样化场景内交通密集小目标检测算法 CDAQ-DDETR,研究通过构建注意力变形特征提取模块、创建注意力感知融合金字塔和集成动态查询机制,分别提升了交通小目标的语义获取能力、缓解了多尺度目标漏检误检问题,并增强了算法对复杂背景的适应性。

1 Deformable DETR

Deformable DETR 结合可变形卷积机制高效关注关键采样点动态捕捉采样以实现局部稀释特征提取。Deformable DETR 由骨干网络 ResNet-50、编码器和解码器 3 个部分组成。输入图像经骨干网络 ResNet-50 深度特征提取后生成 4 个层级的多尺度特征C_2,C_3,C_4,C_5。C_2层级因包含较少的特征语义信息被舍弃,C_5层级进行下采样以及通道数调整以获得包含更大感受野映射的高层特征层C_6,接着将已获得的$C_3 \sim C_6$每个层级进行扁平化 Flatten 以及 mask 掩码处理,同时对每个尺度的特征图添加位置编码得到多尺度一维特征序列$X_3 \sim X_6$,最后将所有特征序列拼接为 X 并输入进下一阶段编码器中。

在 Deformable DETR 中,编码器通过 6 个堆叠编码块对特征序列进行增强,融合全局与局部信息以支持目标检测。每个编码块包含多尺度可变形自注意力层、残差连接与层归一化组合、以及前馈神经网络配以另一组残差连接和层归一化,数据依次流经这些组件完成特征处理。其中 MSDeformAttn 是 Deformable DETR 算法中的关键性创新,首先对于前向所拼接的一维特征序列 X 中的每个元素位置,算法模型会根据特征内容动态自适应选取关注采样点,这些采样点是由查询(query)点的二维位置和自学习偏移量动态跨越图像关键位置选取的,公式为

$$p_k = x + \Delta p_k \tag{1}$$

式中:x 表示当前查询点的二维位置,Δp_k表示算法自学习获取的偏移量,p_k表示更新获得的第 k 个采样点。

接着利用 fatt 函数计算每个注意力头 h，在每一个特征尺度 l 上，元素查询向量 $\boldsymbol{q}$ 与所对应的采样点 p_{klh} 之间的相对位置及特征相似度，进而获得查询点与采样点之间的注意力得分，最后运用 softmax 函数进行归一化以获得注意力权重 $A_{klh}(\boldsymbol{q},p_{klh})$，计算公式为

$$A_{klh}(\boldsymbol{q},p_{klh}) = \mathrm{softmax}(f_{\mathrm{att}}(\boldsymbol{q},p_{klh})) \tag{2}$$

最后通过对每个注意力权重针对多尺度采样点特征进行点乘加权融合，最终构建充分融合了多维尺度不同关键采样点位置信息的特征表示，即为多尺度可变形自注意力 MSDeformAttn 的输出，计算公式为

$$\mathrm{DeformAttn}_q = \sum_{l=1}^{L}\sum_{k=1}^{N_l}\sum_{h=1}^{H} A_{klh}(\boldsymbol{q},p_{klh}) \cdot \boldsymbol{W}_{lh} \cdot \boldsymbol{x}_{pklh} \tag{3}$$

式中：$A_{klh}(\boldsymbol{q},p_{klh})$ 表示查询点 $\boldsymbol{q}$ 在第 h 个注意力头第 l 层尺度上对于第 k 个采样点 p_{klh} 的注意力权重，$\boldsymbol{W}_{lh}$ 表示在 l 层尺度上第 h 个注意力头的线性权重矩阵，$\boldsymbol{x}_{pklh}$ 表示更新获得的第 k 个采样点 p_k 在第 h 个注意力头第 l 层尺度上特征向量。

解码器借助查询向量将前向编码器提供的特征表达转换为目标检测结果，内部由 6 个解码块串联。每个解码块含多头可变形自注意力、多尺度可变形交叉注意力、前馈神经网络及残差连接和层归一化。多头可变形自注意力对全部位置特征序列进行注意力计算，多尺度可变形交叉注意力关注多尺度特征向量中的关键序列。经多次迭代精细化处理后，最终通过专门预测头进行目标预测和定位。

2　CDAQ-DDETR 算法

为了减少中小目标关键细节丢失，同时增强密集区域交通目标语义信息关注，弱化动态查询特征序列关键区域的多样化背景干扰，本文提出一种基于 Deformable DETR 改进的结合注意力变形和动态查询机制的密集多样化场景下的交通小目标检测算法，其网络结构如图 1 所示。

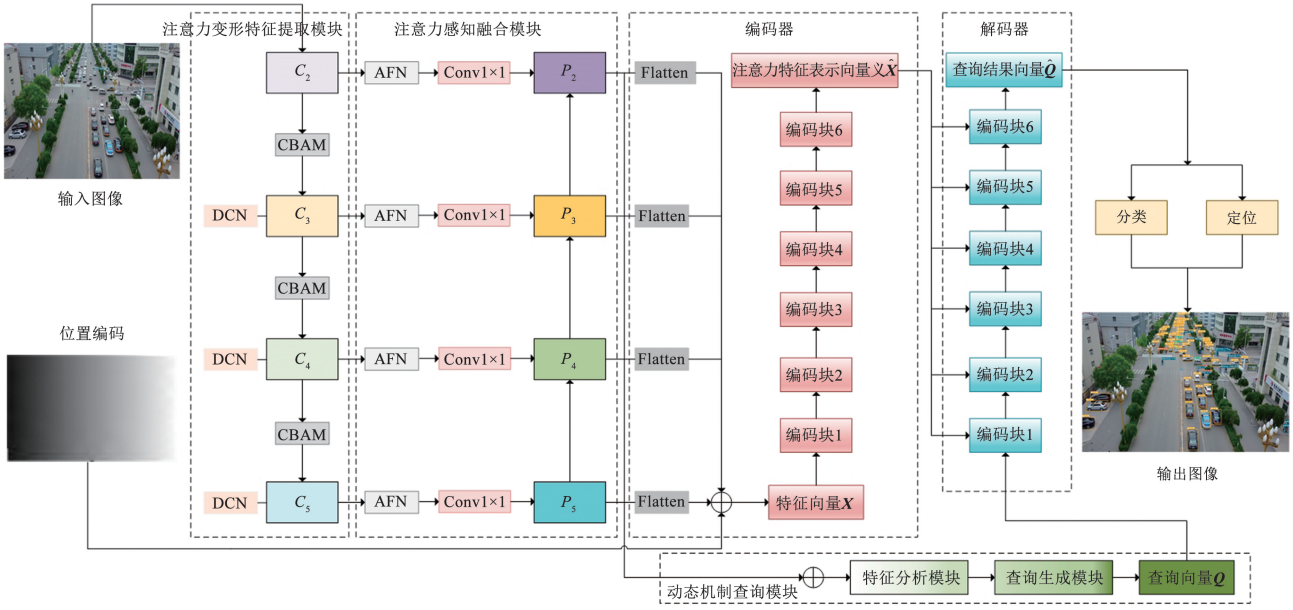


图 1　CDAQ-DDETR 算法网络结构

Fig. 1　CDAQ-DDETR algorithm network structure

首先将 CBAM 双塔注意力模块与 DCNv2 可变形模块嵌入原始 ResNet-50 残差网络重新构建注意力变形特征提取模块，输入图像经此模块提取出多尺度特征 $\mathrm{ConvBA}_2 \sim \mathrm{ConvBA}_5$，接着依靠 AFN 网络思想添加低层特征，同时创建注意力感知融合金字塔模块，将前向包含低层细节特征的 ConvBA_2 与包含高层丰富语义特征的 $\mathrm{ConvBA}_3 \sim \mathrm{ConvBA}_5$ 都输入至注意力感知融合金字塔模块进行注意力特征融合更新，接着将获得的最佳多尺度特征金字塔 $P_2 \sim P_5$ 进行扁平化 Flatten 以及 mask 掩码操作，获得的 $X_2 \sim X_5$ 向量序列，再加入位置编码并拼接得到输入特征向量 $\boldsymbol{X}$，在编码器中，特征向量 $\boldsymbol{X}$ 经 6 个编码块得到包含图像丰富语义信息和空间信息的注意力特征表示向量 $\hat{\boldsymbol{X}}$，为了使算法更加灵活地适应多样化交通图像目标内容，在解码器前向集成动态查询机制模块，将最佳多尺度金字塔输出的 $P_2 \sim P_5$ 经过预处理及简单的元素相加策略直接融合作为输入，通过其内部特征分析模块和查询生成模块动态初始化生成查询向量 $\boldsymbol{Q}$，接着并行输入前向注意力特征表示向量 $\hat{\boldsymbol{X}}$ 至解码器中，经过6 个解码块串联输出得到对象查询结果向量 $\hat{\boldsymbol{Q}}$，最后预测阶段将根据对象查

询结果向量$\hat{\boldsymbol{Q}}$完成目标定位及分类。

2.1 注意力变形特征提取模块

2.1.1 CBAM 双塔注意力模块

为了强化原始特征提取网络对通道及空间内关键特征信息的自适应学习能力,提高算法对密集区域交通小目标的语义信息获取能力,本文将 CBAM(convolutional block attention module)双塔注意力模块加入至 Deformable DETR 原始骨干网络 ResNet-50 每个残差卷积块后,使算法在卷积特征提取过程中同时捕获空间和通道双注意力信息。CBAM 通过两个顺序的子模块对输入特征图进行精细化重标定,即通道注意力模块(channel attention module,CAM)与空间注意力模块(spatial attention module,SAM),形成的双塔注意力结构如图 2 所示。

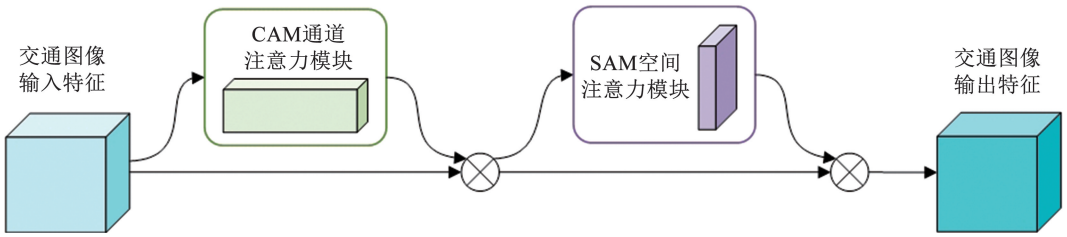


图 2 CBAM 双塔注意力模块结构

Fig. 2 CBAM dual-tower attention module structure

通道注意力模块 CAM 的主要作用是让模型学习到不同通道内有用特征的重要性,其结构如图 3 所示,具体实现:接收前向残差块的输出特征图 $F\in\mathbf{R}^{C\times H\times W}$,沿着空间维度分别通过一个全局平均池化层(GAP)和一个全局最大池化(GMP)生成两个包含全局统计信息的描述符。其中,全局平均池化公式为

$$F_{\text{avg}}^{c}=\frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}F(c,i,j) \tag{4}$$

式中 $F(c,i,j)$ 表示在 c 通道、i 高度、j 宽度上的特征值。之后进行全局最大池化,计算公式为

$$F_{\max}^{c}=\max_{i=1,j=1}^{H,W}F(c,i,j) \tag{5}$$

接着对两个全局池化后的特征图分别经过多层感知机进行一维卷积,然后相加并结合 Sigmoid 激活函数计算得到每个通道的权重W_c,计算公式为

$$W_c=\sigma[\text{MLP}(\text{AvgPool}(F))+\text{MLP}(\text{MaxPool}(F))] \tag{6}$$

式中:σ 表示 Sigmoid 激活函数,MLP 表示多层感知机。

最后,将得到的权重W_c与原始特征图 F 相乘,即得到不同通道精细化重标定后的通道注意力特征 F',计算公式为

$$F'=W_c\cdot F \tag{7}$$

空间注意力模块 SAM 的主要作用是让模型更加聚焦特征图中特定重要的空间区域,其结构如图 4所示。

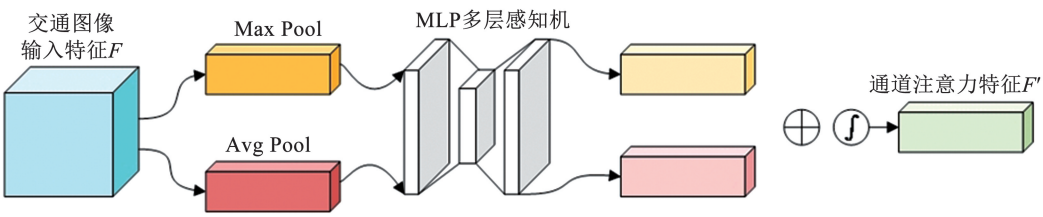


图 3 CAM 通道注意力模块结构

Fig. 3 CAM channel attention module structure

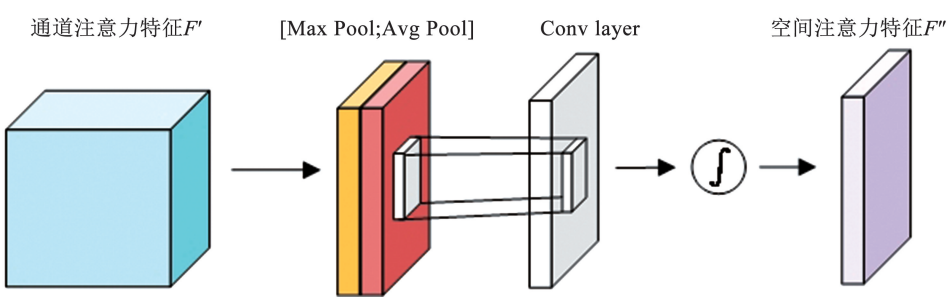


图 4 SAM 空间注意力模块结构

Fig. 4 SAM spatial attention module structure

具体实现：接收前向 CAM 输出的通道注意力特征图 $F' \in \mathbf{R}^{C\times H\times W}$，对特征图沿通道维度进行最大池化与平均池化操作，将获得的两个二维特征图再次堆叠以捕捉空间位置不同的统计信息。其中，平均池化计算公式为

$$F_{\text{avg}}^{\text{2D}} = \frac{1}{C}\sum_{c=1}^{C} F'(c,i,j) \tag{8}$$

为了完善空间注意力模块中对特征图空间信息的处理机制，最大池化最大池化公式为

$$F_{\max}^{\text{2D}} = \max_{c=1}^{C} F'(c,i,j) \tag{9}$$

接着将堆叠后的特征图通过一个卷积层结合 Sigmoid 激活函数计算得到每个空间的权重 W_s，计算公式为

$$W_s = \sigma(f^{7\times 7}([F_{\max}^{\text{2D}}; F_{\text{avg}}^{\text{2D}}])) \tag{10}$$

式中：$f^{7\times 7}$ 表示卷积核为 7 ×7 的卷积层，$[\cdot;\cdot]$ 表示沿通道维度的堆叠。

最后，将得到的 W_s 权重值与通道注意力调整后的特征图 F' 相乘，即得到不同空间区域精细化重标定后的空间注意力图 F''。

2.1.2　DCNv2 可变形卷积模块

为了提高原始特征提取网络对图像内目标的几何形变的适应性，增强算法对复杂场景下形态各异交通小目标的细节特征表示，综合计算效率和特征表达的丰富性，本文将 DCNv2(deformable convolutional networks version 2)可变形卷积模块集成到 formable DETR 原始骨干网络 ResNet-50 残差卷积 $C_3 \sim C_5$ 层，用以使算法网络高效灵活地提取小目标不同形状和尺寸的目标。DCNv2 的核心特性是将标准卷积替换为可变形卷积，可变形卷积机制通过引入额外的可学习偏移量来动态调整卷积采样点的位置，使网络卷积动态适应图像内目标的结构形变，可变形卷积与普通卷积对比如图 5 所示。普通卷积是在输入特征图上利用卷积核滑动对局部区域进行加权求和完成特征提取的，具体公式为

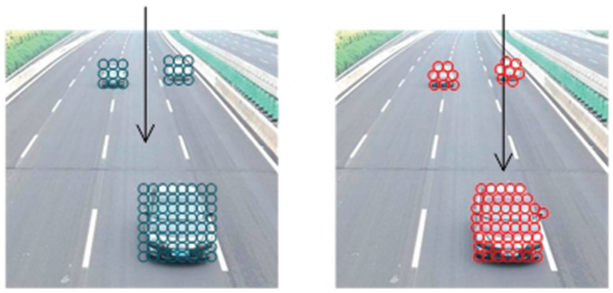
$$Y(p_0) = \sum_{p_n\in \mathbf{R}} W(p_n)\cdot X(p_0 + p_n) \tag{11}$$

式中：$X(p_0 + p_n)$ 表示输入特征图在 p_0 位置周围的像素值，p_0 表示实际位置，p_n 表示周围位置偏移，R 表示卷积核覆盖的区域，$W(p_n)$ 表示位置偏移 p_n 的卷积核权重，$Y(p_0)$ 表示卷积后特征图在 p_0 位置的像素值。

DCNv2 可变形卷积是在普通卷积核采样点位置引入额外可学习偏移量 Δp_n，使得卷积核动态变形以匹配输入特征，计算公式为

$$Y(p_0) = \sum_{p_n\in \mathbf{R}} W(p_n)\cdot X(p_0 + p_n + \Delta p_n)\cdot \Delta m_n \tag{12}$$

式中：Δp_n 表示通过主网络共同训练学习到的偏移量，Δm_n 表示学习到的调制因子。



(a) 普通卷积　　(b) 可变形卷积

图 5　可变形卷积与普通卷积对比

Fig. 5　Comparison between deformable convolution and standard convolution

2.2　注意力感知融合金字塔模块

为了增强原始颈部网络对于多尺度目标关键信息的特征表达，提高算法针对多样化特征内部重要性的动态调整能力，本文将注意力感知融合模块引入原始 neck 颈部中，用以对前向多尺度卷积特征进行关注度加工和动态特征融合。注意力感知融合模块的核心特性是利用 AFN(attention-aware fusion network)注意力机制属性对多尺度输入特征图内的关注重要性进行自适应调整，这种关键点均衡策略使算法在小目标检测和大目标检测之间实现更好的平衡，同时经 AFN 模块后的多尺度特征图更加聚焦于有用的特征，接着结合 FPN 特征融合金字塔模块捕获从粗略到细粒度的信息，最终构建最佳特征融合金字塔。具体实现如图 6 所示，首先针对密集多样化场景下的交通中小目标特点，将前向包含低层细节特征的 ConvBA2 与包含高层丰富语义特征的 ConvBA3 ~ ConvBA5 都输入至注意力感知融合金字塔模块，接着在每一尺度特征图平行加入一个 AFN 模块，在 AFN 模块中通过全局平均池化 GAP 提取每个尺度通道的全局信息，设单个尺度输入特征图为 $X \in \mathbf{R}^{C\times H\times W}$，其中 C，H，W 分别表示输入特征图的通道、高度及宽度，经全局平均池化操作计算得到的每个通道的全局特征表示即为 $F_{\text{global}} \in \mathbf{R}^{C}$，计算公式为

$$F_{\text{global}} = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W} X_{:,i,j} \tag{13}$$

接着通过一个多层感知机 MLP(包含 Sigmoid 激活函数)计算可得注意力权重 $W \in \mathbf{R}^{C}$，最后将注意力权重应用至原始特征图中即得加权后的输出特征图 $A \in \mathbf{R}^{C\times H\times W}$，计算公式为

$$W = \sigma(\text{MLP}(F_{\text{global}})) \tag{14}$$

$$A = W\cdot X \tag{15}$$

经过 AFN 模块的输出特征图被表示为$A_2 \sim A_5$，接着 FPN 特征融合金字塔模块将对这些多尺度特征图进行上采样及加法操作重构形成最佳特征融合金字塔$P_2 \sim P_5$，具体实现：将 AFN 模块输出的最深层特征图A_5作为金字塔重建的起始点，通过一个 1×1的横向卷积层调整A_5的通道数以匹配其他层级，这一步骤的计算公式为

$$P_5 = \mathrm{Conv}_{1\times1}(A_5) \tag{16}$$

接着将金字塔顶层特征图P_5自顶向下进行上采样操作，同时与卷积层A_4经 1×1 横向卷积调整通道数后的特征图相加得到P_4，这一步骤的计算公式为

$$P_4 = \mathrm{UpSample}(P_5) + \mathrm{Conv}_{1\times1}(A_4) \tag{17}$$

这一步骤重复进行直至所有特征图完成融合，对于更高层次的特征图P_i融合操作可以概括为

$$P_i = \mathrm{UpSample}(P_{i+1}) + \mathrm{Conv}_{1\times1}(A_i) \tag{18}$$

式中：UpSample(·)表示上采样操作，$\mathrm{Conv}_{1\times1}(\cdot)$表示横向 1×1 卷积操作，$i=4,3,2$。

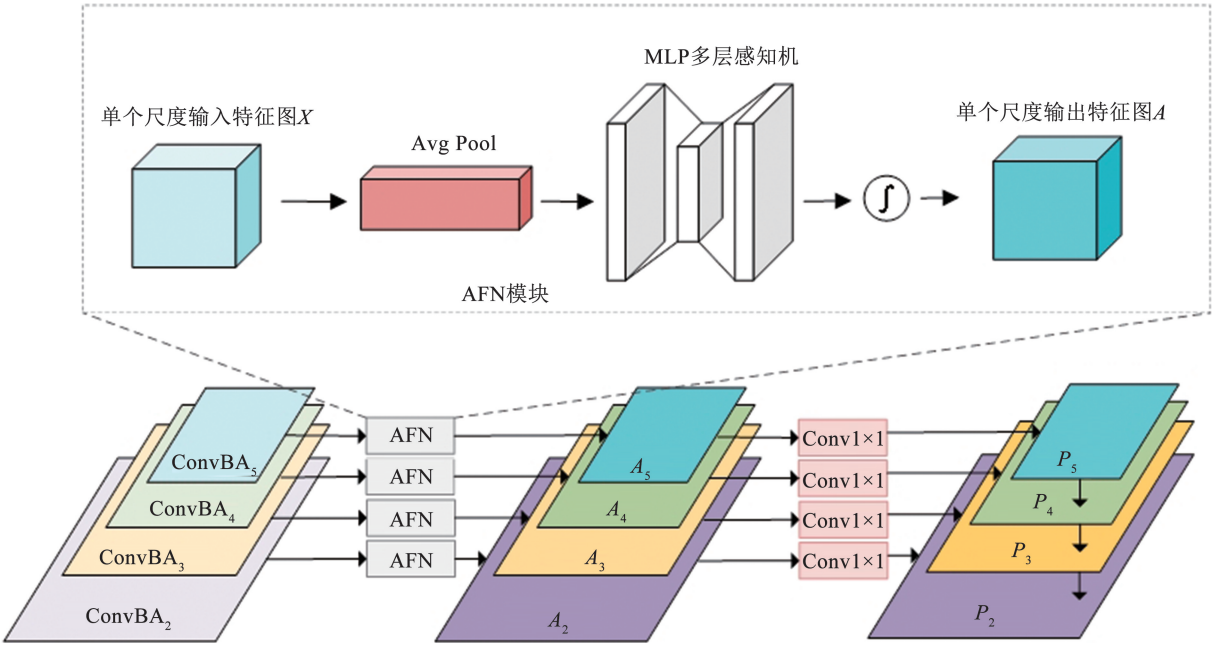


图 6 注意力感知融合金字塔模块结构

Fig. 6 Attention-aware fusion pyramid module structure

2.3 动态查询机制模块

在 Deformable DETR 等基于 Transformer 的目标检测算法里，解码器用于目标定位的关键目标查询向量多为静态初始化，不依赖于输入图像内容。然而实际交通场景中小目标有诸多特点，致固定数量和配置的查询向量难捕获潜在目标细节，限制了模型对多样化场景中交通目标的适应性。为了解决这一局限性，本文设计一种动态目标查询机制模块嵌入注意力感知融合金字塔模块与解码器之间，通过图像卷积特征图中的具体信息生成与之相匹配的目标查询向量，从而提升算法对多样化场景的适应性和整体模型的检测精度。具体实现如图 7 所示，所设计的动态查询机制模块内部包含特征分析模块和查询生成模块两个子模块，特征分析模块的主要目标是从输入图像特征图中提取出对目标检测最有价值的信息，这包括捕获全局上下文信息和识别局部细节信息，全局信息有助于算法理解图像的整体场景，如交通场景内的城市街道或高速公路，局部信息用于算法识别图像内的细粒度目标，如交通场景内的不同类别车辆或行人，具体实现：首先将前向注意力感知融合金字塔模块输出的多尺度融合特征图$P_2 \sim P_5$预处理为相同的空间尺寸$P'_2 \sim P'_5$，并将其通过简单的元素相加策略直接融合为F_{fused}，具体实现公式为

$$F'_i = \mathrm{PreprocessFeatures}(F_i),\ i \in \{P_2, P_3, P_4, P_5\} \tag{19}$$

$$F_{\mathrm{fused}} = F'_{P_2} + F'_{P_3} + F'_{P_4} + F'_{P_5} \tag{20}$$

式中：F_i表示原始特征图，F'_i表示预处理后的特征图，PreprocessFeatures 包括必要的上下采样操作以及可能的 1×1 卷积来调整通道数。

接着本文通过两个并行的路径提取F_{fused}内的全局上下文信息和局部细节信息，其中一分支使用全局平均池化(GAP)层提取全局上下文信息 C，具体实现公式为

$$C = \mathrm{GAP}(F_{\mathrm{fused}}) = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W} F_{i,j} \tag{21}$$

式中：H 和 W 分别表示特征图的高度和宽度，$F_{i,j}$表示位置(i,j)上的特征向量。

另一分支使用一个 3 × 3 卷积结合 ReLU 函数以捕捉局部细节信息 L，具体实现公式为

$$L = \mathrm{ReLU}(\mathrm{Conv}_{3\times3}(F_{\mathrm{fused}})) \tag{22}$$

查询生成模块的主要目标是通过特征分析结果动态定制生成查询向量的数量以及初始状态，以便后续解码器的具体目标识别及定位。具体实现：将全局上下文信息 C 和局部细节信息 L 输入至一个简单的全连接层（fully connected layer，FC）以实现融合，并在全连接层中根据输入信息动态调整并生成每个查询的特征向量 $\boldsymbol{Q}$，具体实现公式为

$$\boldsymbol{Q} = \boldsymbol{Q}G(C, L) \tag{23}$$

式中 $\boldsymbol{Q}G(\cdot)$ 表示全连接层，根据输入的全局和局部信息动态调整并生成每个查询的特征向量。

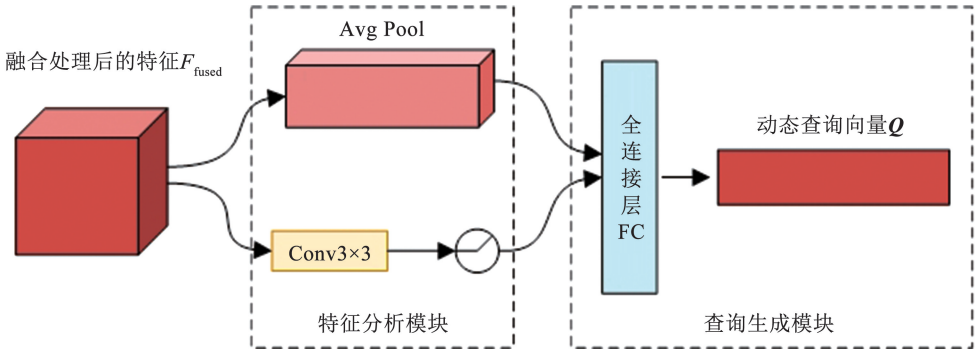


图 7　动态查询机制模块结构

Fig. 7　Dynamic query mechanism module structure

3　实验过程与分析

3.1　数据集与实验环境

本文所用数据集基于天津大学 AISKYEYE 团队开源的 VisDrone2019-DET[19] 无人机交通小目标数据集，通过对该数据集进行分析，其内部涵盖了城市交通、高速公路、乡村农田等 14 种不同的交通场景，图像采集跨越了白天 6:00 至夜晚 18:00 和夜晚 18:00 到次日 6:00 不同光照条件下的全天候不同时间段，同时该数据集还捕捉了包含晴天、雨天等多种气象条件下的实际场景，数据类别包含行人（pedestrain）、人（person）、自行车（bicycle）、汽车（car）、面包车（van）、卡车（truck）、三轮车（tricycle）、遮阳三轮车（awning-tricycle）、公交车（bus）、摩托车（motor）共计 10 个类别目标，数据集内包含标签的图片有 7 019 张（训练集 6 471 张、验证集 548 张），图像内物体共计 381 964 个标注框，数据预处理采用默认参数，包括图像尺寸调整为 800 × 1 333 像素，数据标准差为［0.229，0.224，0.225］。图 8（a）展示了完整数据集各标记类别目标数量分布不均衡，汽车、行人因城市交通监控的重要性及目标特性，高频出现且处于高密集遮挡区域，三轮车、遮阳三轮车、公交车数量少可能源于实际交通场景中出现频率低，但数据集仍涵盖多种目标，具有多样性和复杂性。图 8（b）、8（c）显示数据集目标以中小尺度为主，对应多尺度目标检测任务。

实验平台为 Windows server 2012 服务器，硬件配置采用 Intel（R）Xeon（R）Gold 5117 CPU 处理器、NVIDIA Tesla V100-32G GPU 处理器，软件配置采用 Python3.8、CUDA10.2、CUDNN7.6.5，采用 Pycharm 及 Anaconda 构建深度学习虚拟运行环境，结合基于 Pytorch 的开源目标检测工具箱 MMDetection 作为算法基本的开发框架。

3.2　评估指标

在本研究中，为了全面评估提出的改进算法模型性能，本文采用了目标检测领域广泛认可的 COCO API 评估协议。该协议提供了包括精确率 P，单个类别平均的精确率 AP、模型平均识别精度 mAP、召回率 R、单个类别平均的召回率 AR，以及模型平均召回率 mAR 在内的一套全面评价指标，同时这套指标内部还包含了对不同尺度大小（small、medium、large）和多种重叠阈值（IoU = 0.5:0.95、IoU = 0.5、IoU = 0.75）的全套检测精度和召回率输出，其中不同尺度大小以像素面积划定，面积小于 32 × 32 像素为小尺寸、大于 96 × 96 像素为大尺寸，介于中间值即为中尺寸目标，同时 IoU = 0.5:0.95 是指从 0.5 到 0.95 以 0.05 为步进，用于综合评定不同 IoU 阈值下的平均精度，IoU = 0.5 和 IoU = 0.75常用于评估模型对目标边界框定位的宽松准确性。以下本节将详细介绍这些评估指标以及其在本研究中的具体应用。

1）精确率（P）为算法检测正确的正样本数量与检测出的正样本总数的比例，计算公式为

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{24}$$

式中：TP 表示正确检测为正样本的数量，FP 表示错

误检测为正样本的数量。

2)召回率(R)是算法检测正确的正样本数量与实际正样本总数的比例,计算公式为

$$R = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{25}$$

式中:FN 表示未被检测为正样本但实际上是正样本的数量。

3)单个类别平均精确率(AP)是衡量算法在单个类别上检测精度的指标,其计算方法是基于模型在各个召回率水平下精确率的平均值,计算公式为

$$\mathrm{AP} = \int P(r)\,\mathrm{d}r \tag{26}$$

4)单个类别平均召回率(AR)是衡量算法在单个类别上实际检测正样本的指标,其计算方法是基于模型在各个精确率水平下召回率的平均值,计算公式为

$$\mathrm{AR} = \int R(p)\,\mathrm{d}p \tag{27}$$

5)模型平均精确率(mAP)是衡量算法在整体类别上检测精度的指标,计算公式为

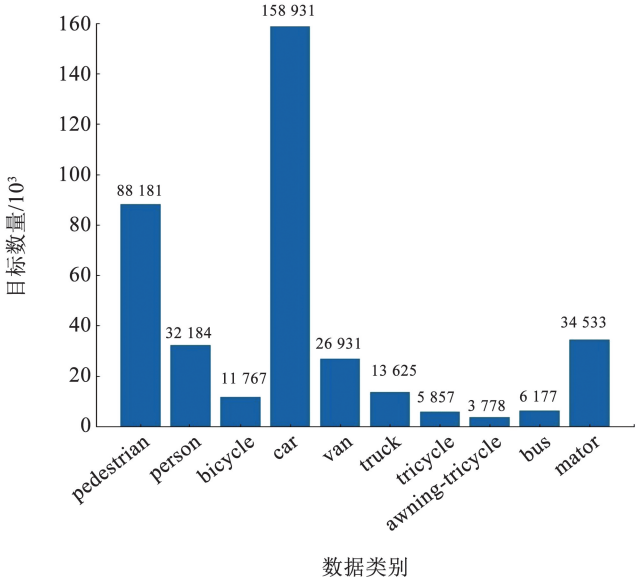
$$\mathrm{mAP} = \frac{1}{C}\sum_{i=1}^{C} \mathrm{AP}_i \tag{28}$$

式中:C 表示细分类个数,AP_i表示当前类别精确率。

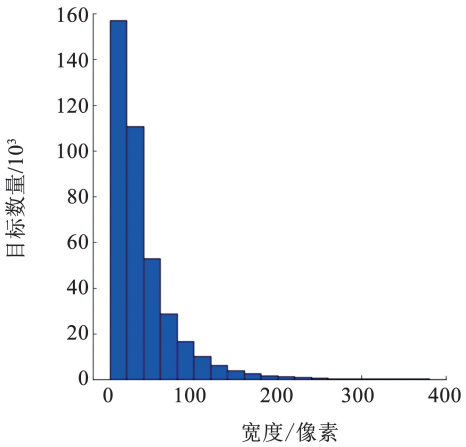
6)模型平均召回率(mAR)是衡量算法在整体类别上实际检测正样本的指标,计算公式为

$$\mathrm{mAR} = \frac{1}{C}\sum_{i=1}^{C} \mathrm{AR}_i \tag{29}$$

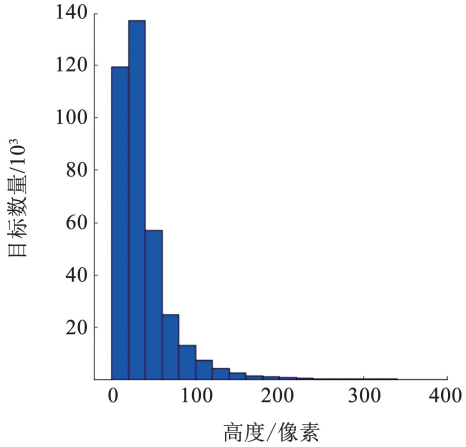
式中:C 表示细分类个数,AR_i表示当前类别召回率。



(a) 各标记类别目标数量分布直方图



(b) 所有标记目标尺寸(宽度)分布直方图



(c) 所有标记目标尺寸(高度)分布直方图

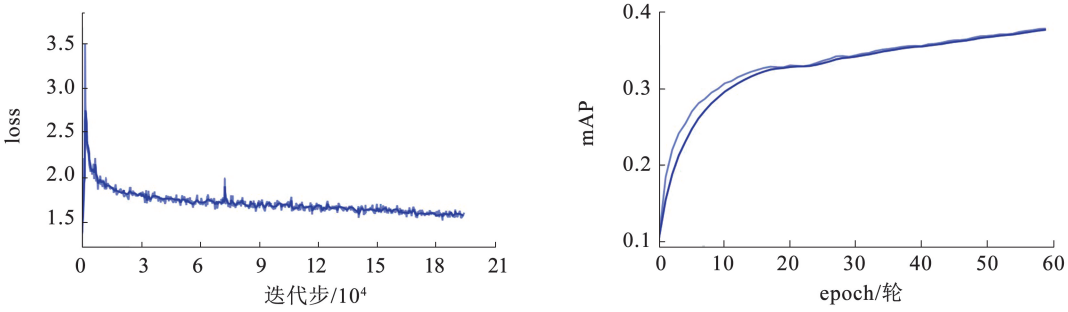
图 8 数据集分析

Fig. 8 Dataset analysis

3.3 模型训练及性能评估

本文改进算法 CDAQ-DDETR 在训练过程中,为了使模型拟合收敛加速,实验中借鉴迁移训练思想将原始 Deformable DETR 算法在 COCO 数据集上的

预训练权重引入训练开始点，初始化 CDAQ-DDETR 算法模型的参数：最大迭代训练周期 epoch 为 60 轮，默认批迭代系数 batch_size 为 2，每个 epoch 会结合批迭代系数对训练集进行迭代学习，即每个 epoch 进行 3 235 轮共计 194 100 迭代步训练，默认学习率为 0.000 2，权重衰减系数为0.000 1，采用 AdamW 优化器开始训练。模型评估方式遵循目标检测领域的标准评估协议 COCOAPI，每个 epoch 结束后均会利用验证集（548 张）进行评估计算，同时绘制实时的训练迭代损失 loss 曲线及算法模型精度均值 mAP 曲线如图 9 所示。



(a) 训练迭代损失loss曲线　　(b) 算法模型精度均值mAP曲线

图 9　模型训练分析

Fig. 9　Model training analysis

在算法迭代训练过程中，每个 epoch 学习结束后都会结合验证集生成当前阶段的 model 权重，同时模型算法会将最新生成的 model 权重与前向所有权重进行对比，用以获得最佳模型权重 best_model，在模型评估阶段，利用最佳模型权重 best_model 对 548 张验证集图像进行性能评估绘制单个类别 *P-R* 曲线，如图 10 所示，以下 *P-R* 曲线反映了改进算法内各类别目标在不同召回率水平下查准率的变化，揭示了模型算法在识别正样本的准确性与全面性之间的权衡关系，可见行人、人、汽车、面包车、公交车、摩托车的 *P-R* 曲线靠近图表右上角，这表明改进算法在这些类别上能够在保持高召回率的同时实现高准确率，同时对于自行车、三轮车这些样本数据分布极少的类别，改进算法在维持较高召回率的同时也尽量减少了误检的情况，这在一定程度上减轻了后续处理的负担，卡车类别的 *P-R* 曲线倾向于左上区域，表明改进后的算法在确认检测到的卡车时变得更加准确，尽管改进算法在追求高准确率时牺牲了召回率，但其优势在于显著减少误判，提高了模型算法可靠性。通过对遮阳三轮车 *P-R* 曲线的分析，尽管面临遮阳三轮车目标的复杂性和多样性、遮挡和角度变化的挑战，以及数据集中表示不足的问题，改进算法仍展现出了其在处理这一特定且复杂目标上的明显优势。

3.4　实验结果对比

3.4.1　消融实验

为了验证本文所提每种改进模块的有效性，利用 VisDrone2019 无人机交通小目标数据集对基准模型 Deformable DETR 进行消融实验，评估增添不同改进模块在相同实验参数环境下算法的检测性能，主要评估对象有 IoU 交互比阈值在 0.5 到 0.95 均值条件下每个细分类目标的单个类别精准率 AP@0.5:0.95 与单个类别召回率 AR@0.5:0.95，如表 1、2 所示。在不同 IoU 交互比阈值下的算法模型精准率 mAP(0.5:0.95)、mAP(0.5)、mAP(0.75) 和召回率 mAR(0.5:0.95)，以及针对小、中、大不同尺度目标对象在 IoU 交互比阈值 0.5 到 0.95 均值条件下的算法模型精准率 mAP(small)、mAP(medium)、mAP(large) 和召回率 mAR(small)、mAR(medium)、mAR(large)，如表 3 所示。

实验 1 为基线模型 Deformable DETR，实验 2 添加了注意力变形特征提取模块，有助于提高算法对密集交通场景内多姿态目标的检测性能，尤其在汽车、公交车、摩托车 3 个类别中体现，其类别精准率 AP 分别提升 4.2%、6.4%、4.3%，其类别召回率 AR 分别提升 3.6%、7.4%、3.2%，在对于不同尺度目标检测性能方面，实验 2 着重于提高大尺度目标的检测性能，即 mAP(large) 提高了 11.3%、mAR(large) 提高了 12.2%，但在中小目标检测方面，mAP(small) 提高 2.7%、mAP(medium) 提高 4.9%、mAR(small) 提高 2.9%、mAR(medium) 提高 4.3%，这样的提升性能仍需改进，因此，针对中小目标识别特点实验 3 在实验 2 的基础上集成注意力感知融合金字塔模块，动态调整不同尺度特征的注意力关注融合权重，进而缓解多尺度中小目标误检漏检问题，相较于实验2，实验3在小目标检测精准度mAP

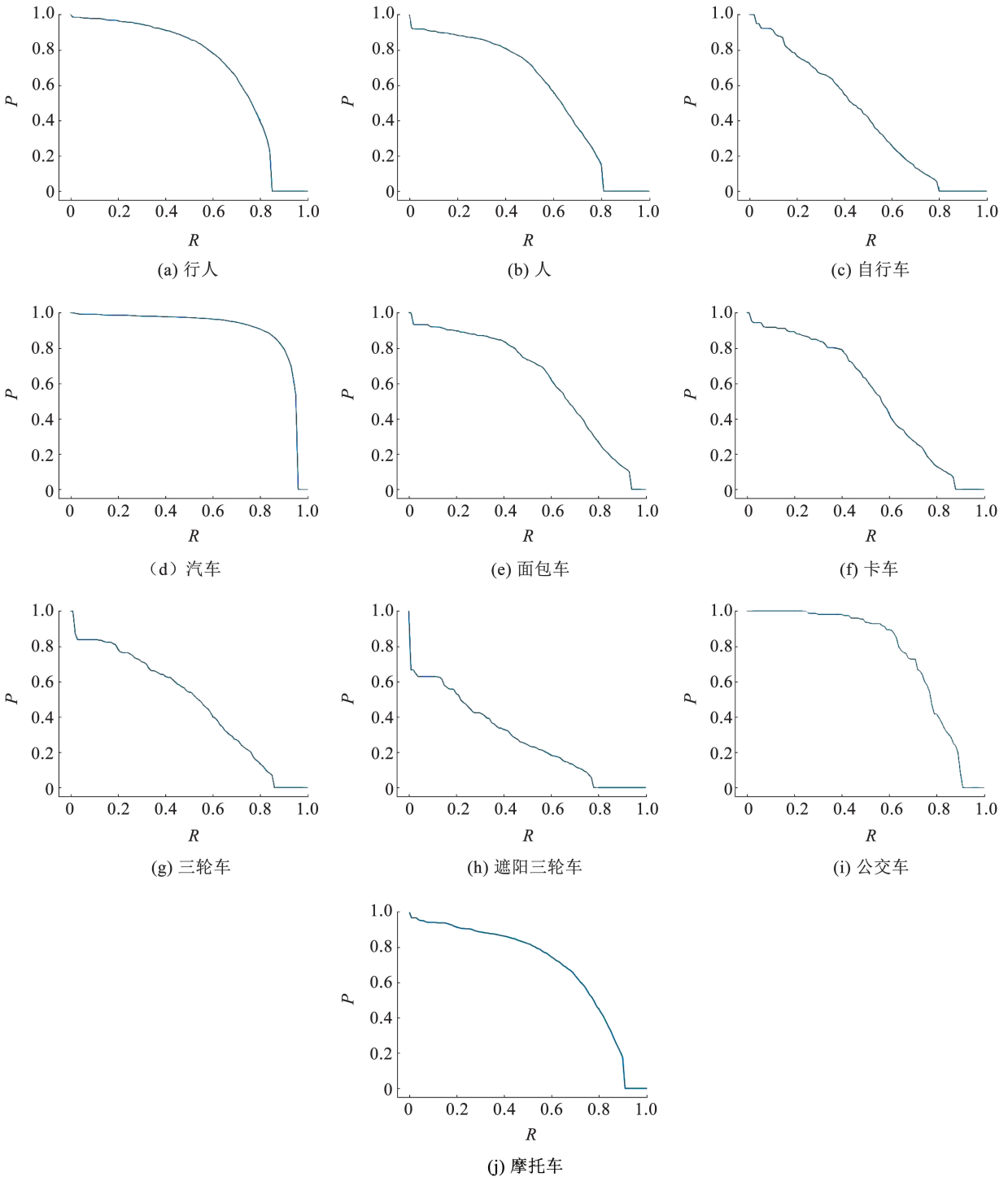


图 10 模型评估各交通目标类别 *P-R* 曲线

Fig. 10 Model evaluation of *P-R* curves for each traffic object category

表 1 算法在各细分类目标精准率 AP@0.5:0.95

Tab. 1 Algorithm precision for each fine-grained category: AP@0.5:0.95

实验序号	注意力变形提取	感知融合金字塔	动态查询机制	AP@0.5:0.95/%									
				行人	人	自行车	汽车	面包车	卡车	三轮车	遮阳三轮车	公交车	摩托车
1				16.4	10.3	5.5	42.0	22.4	18.5	9.3	5.2	21.2	14.7
2	√			20.4	13.0	8.3	46.2	25.5	21.8	12.4	6.8	27.6	19.0
3	√	√		25.6	17.6	10.6	61.4	39.4	33.6	23.2	13.4	50.7	28.3
4	√	√	√	36.6	25.7	22.1	66.9	46.4	38.9	29.0	19.3	57.7	36.0

表 2　算法在各细分类目标召回率 AR@0.5:0.95

Tab. 2　Algorithm recall for each fine-grained category: AR@0.5:0.95

实验序号	注意力变形提取	感知融合金字塔	动态查询机制	AR@0.5:0.95/%									
				行人	人	自行车	汽车	面包车	卡车	三轮车	遮阳三轮车	公交车	摩托车
1				23.7	18.6	13.3	49.2	32.5	28.2	19.1	14.3	24.4	24.9
2	√			27.3	21.5	15.3	52.8	36.2	31.0	22.3	17.6	31.8	28.1
3	√	√		35.6	31.2	29.4	67.5	64.9	54.7	47.5	46.3	60.5	43.4
4	√	√	√	48.7	41.7	44.0	73.5	71.9	63.5	55.3	52.9	70.5	52.2

表 3　算法在不同阈值和多尺度目标上的预测性能

Tab. 3　Predictive performance of the algorithm across different thresholds and multi-scale objects

实验序号	注意力变形提取	感知融合金字塔	动态查询机制	mAP (0.5:0.95) /%	mAP (0.5) /%	mAP (0.75) /%	mAP (small) /%	mAP (medium) /%	mAP (large) /%	mAR (0.5:0.95) /%	mAR (small) /%	mAR (medium) /%	mAR (large) /%
1				16.6	29.8	16.3	8.7	26.2	37.0	24.8	15.5	37.7	47.3
2	√			20.1	35.0	19.7	11.4	31.1	48.3	28.4	18.4	42.0	59.5
3	√	√		30.4	49.8	31.0	20.7	42.5	50.4	48.1	39.1	61.4	73.0
4	√	√	√	37.9	59.3	40.0	30.2	48.5	44.2	57.4	50.7	67.9	67.8

(small)提升了 9.3%，在中目标检测精准度 mAP(medium)提高了 11.4%，在小目标及中目标召回率评估指标中分别提升了 20.7% 和 19.4%，极大验证了改进策略的有效性，接着针对实际多样化交通场景内，俯拍角度目标之间、背景与目标之间多存在相互遮挡干扰问题，实验 4 在实验 3 的基础上加入动态查询机制模块，该模块通过对输入图像内潜在细节信息进行捕获学习，综合提高算法对实际复杂交通场景中多尺度中小目标的识别和定位能力，实验 4 最终改进算法 CDAQ-DDETR 与实验 1 原始基准模型 Deformable-DTER 的检测结果对比如图 11 所示，本文从中标记出几处有代表性的区域验证各改进模块的有效性。引入注意力变形特征提取模块后，改进算法对密集远端多姿态交通目标更敏感；集成注意力感知融合金字塔模块后，对多尺度中小交通目标细分类更准确；加入动态查询机制模块后，对背景融合度高、相互遮挡的交通目标能精准定位。

3.4.2　主流算法对比

为了证明本文所提改进算法 CDAQ-DDETR 的优越性及有效性，结合验证集将本文改进算法 CDAQ-DDETR 与现阶段主流算法模型进行对比实验，评估在数据集划分一致及相同实验环境下各算法模型的检测性能，主要评估对象有 IoU 阈值交互比 0.5 到 0.95 均值条件下各算法模型的单个类别识别精确率 AP@0.5:0.95 与召回率 AR@0.5:0.95，如表 4、5 所示。在不同 IoU 阈值交互比条件下各算法模型的精确率 mAP(0.5:0.95)、mAP(0.5)、mAP(0.75)、召回率 mAR(0.5:0.95)，以及针对小、中、大不同尺度目标在 IoU 阈值交互比 0.5 至 0.95 均值情况下的模型算法精确率 mAP(small)、mAP(medium)、mAP(large)和召回率 mAR(small)、mAR(medium)、mAR(large)，如表 6 所示。

(a) Deformable-DETR　　(b) CDAQ-DDETR

图 11　最终改进算法与基准模型检测结果对比

Fig. 11　Comparison of detection results between the final improved algorithm and the baseline model

表 4 改进算法与各主流算法在细分类目标上的精确率 AP@0.5:0.95

Tab. 4 Precision of the improved algorithm vs. mainstream algorithms on fine-grained object categories: AP@0.5:0.95

算法	AP@0.5:0.95/%									
	行人	人	自行车	汽车	面包车	卡车	三轮车	遮阳三轮车	公交车	摩托车
Faster-RCNN (NeurIPS 2015)	13.6	6.5	1.8	42.8	21.3	12.5	8.6	4.2	20.1	12.9
SSD (ECCV 2016)	3.2	2.9	0.7	28.9	11.7	10.4	4.0	1.6	16.3	4.1
RetinaNet (ICCV 2017)	9.6	3.1	0.9	43.7	21.3	15.6	7.7	4.1	19.9	10.0
Cascade RCNN (CVPR 2018)	15.1	6.3	1.3	48.6	23.1	12.1	5.7	3.0	17.0	11.4
FCOS (ICCV 2019)	15.2	9.1	5.2	49.1	23.7	19.5	7.3	5.9	30.1	11.0
GFL (NeurIPS 2020)	19.9	6.8	5.2	52.9	27.8	20.7	10.1	5.4	29.8	13.5
TOOD (ICCV 2021)	19.2	8.7	6.2	52.5	29.8	22.6	13.6	8.2	33.3	16.6
ViTDet (ECCV 2022)	25.5	14.4	10.5	59.0	36.7	36.9	24.4	15.7	47.5	26.5
RT-DETR (ICCV 2023)	26.6	21.3	14.0	60.2	38.4	32.5	23.3	14.0	46.6	30.3
YOLOv8 (ICCV 2023)	17.5	11.1	6.4	49.5	28.8	20.1	13.0	7.3	38.6	19.1
DINO (ICLR 2023)	32.1	23.6	18.3	60.9	39.2	32.2	23.9	13.3	49.4	31.3
CDAQ-DDETR (改进算法)	36.6	25.7	22.1	66.9	46.4	38.9	29.0	19.3	57.7	36.0

表 5 改进算法与各主流算法在细分类目标上的召回率 AR@0.5:0.95

Tab. 5 Recall of the improved algorithm vs. mainstream algorithms on fine-grained object categories: AR@0.5:0.95

算法	AR@0.5:0.95/%									
	行人	人	自行车	汽车	面包车	卡车	三轮车	遮阳三轮车	公交车	摩托车
Faster-RCNN (NeurIPS 2015)	19.7	12.0	3.4	49.5	34.7	23.9	18.1	15.7	25.9	22.3
SSD (ECCV 2016)	8.6	8.3	2.7	36.6	20.6	18.2	9.6	7.2	20.6	10.5
RetinaNet (ICCV 2017)	15.0	7.1	4.9	49.9	35.6	29.7	18.0	13.8	31.3	18.6
Cascade RCNN (CVPR 2018)	21.7	11.6	2.2	54.3	35.7	31.9	11.9	8.9	27.6	18.9
FCOS (ICCV 2019)	19.3	18.2	12.4	55.2	41.6	41.6	21.6	21.7	44.2	20.4
GFL (NeurIPS 2020)	26.7	11.2	8.7	59.0	43.3	36.7	21.6	18.1	38.2	21.8
TOOD (ICCV 2021)	27.2	14.1	10.3	58.0	43.6	34.5	24.9	21.5	38.5	23.6
ViTDet (ECCV 2022)	30.8	20.9	15.5	64.2	49.3	47.0	38.1	31.7	53.0	34.9
RT-DETR (ICCV 2023)	36.7	35.4	30.2	65.9	59.0	49.6	42.4	38.7	54.6	43.9
YOLOv8 (ICCV 2023)	26.4	21.3	20.4	56.2	48.6	40.3	31.3	28.8	48.2	32.3
DINO (ICLR 2023)	43.4	40.0	36.4	68.0	60.3	51.1	46.4	41.7	59.2	47.1
CDAQ-DDETR (改进算法)	48.7	41.7	44.0	73.5	71.9	63.5	55.3	52.9	70.5	52.2

表 6　改进算法与各主流算法在不同阈值和多尺度目标上的预测性能

Tab. 6　Predictive performance of the improved algorithm vs. mainstream algorithms across different thresholds and multi-scale objects　%

算法	mAP (0.5:0.95)	mAP (0.5)	mAP (0.75)	mAP (small)	mAP (medium)	mAP (large)	mAR (0.5:0.95)	mAR (small)	mAR (medium)	mAR (large)
Faster-RCNN (NeurIPS 2015)	14.4	27.5	13.3	8.5	21.6	24.4	22.5	15.0	31.4	34.7
SSD (ECCV 2016)	8.4	16.6	7.6	1.7	13.8	33.1	14.3	5.2	24.3	45.8
RetinaNet (ICCV 2017)	13.6	23.9	13.7	6.1	22.6	27.5	22.4	11.6	37.0	40.9
Cascade RCNN (CVPR 2018)	14.4	24.9	14.3	8.7	21.6	25.7	22.5	15.0	32.2	34.4
FCOS (ICCV 2019)	17.6	29.1	18.4	9.4	28.7	39.5	29.6	18.1	46.9	60.4
GFL (NeurIPS 2020)	19.2	30.6	20.0	10.6	29.9	46.4	28.5	17.9	43.1	63.1
TOOD (ICCV 2021)	21.1	35.0	22.1	12.6	31.7	31.4	29.6	19.6	43.5	41.1
ViTDet (ECCV 2022)	29.7	47.5	31.0	20.1	42.1	53.1	38.5	28.6	52.6	66.1
RT-DETR (ICCV 2023)	30.7	51.2	30.7	20.3	43.6	69.1	45.7	35.6	59.9	79.2
YOLOv8 (ICCV 2023)	21.1	35.4	21.1	11.7	32.7	44.7	35.4	25.3	50.0	57.9
DINO (ICLR 2023)	32.4	54.6	32.6	23.3	43.8	56.8	49.4	40.7	61.6	76.5
CDAQ-DDETR (改进算法)	37.9	59.3	40.0	30.2	48.5	44.2	57.4	50.7	67.9	67.8

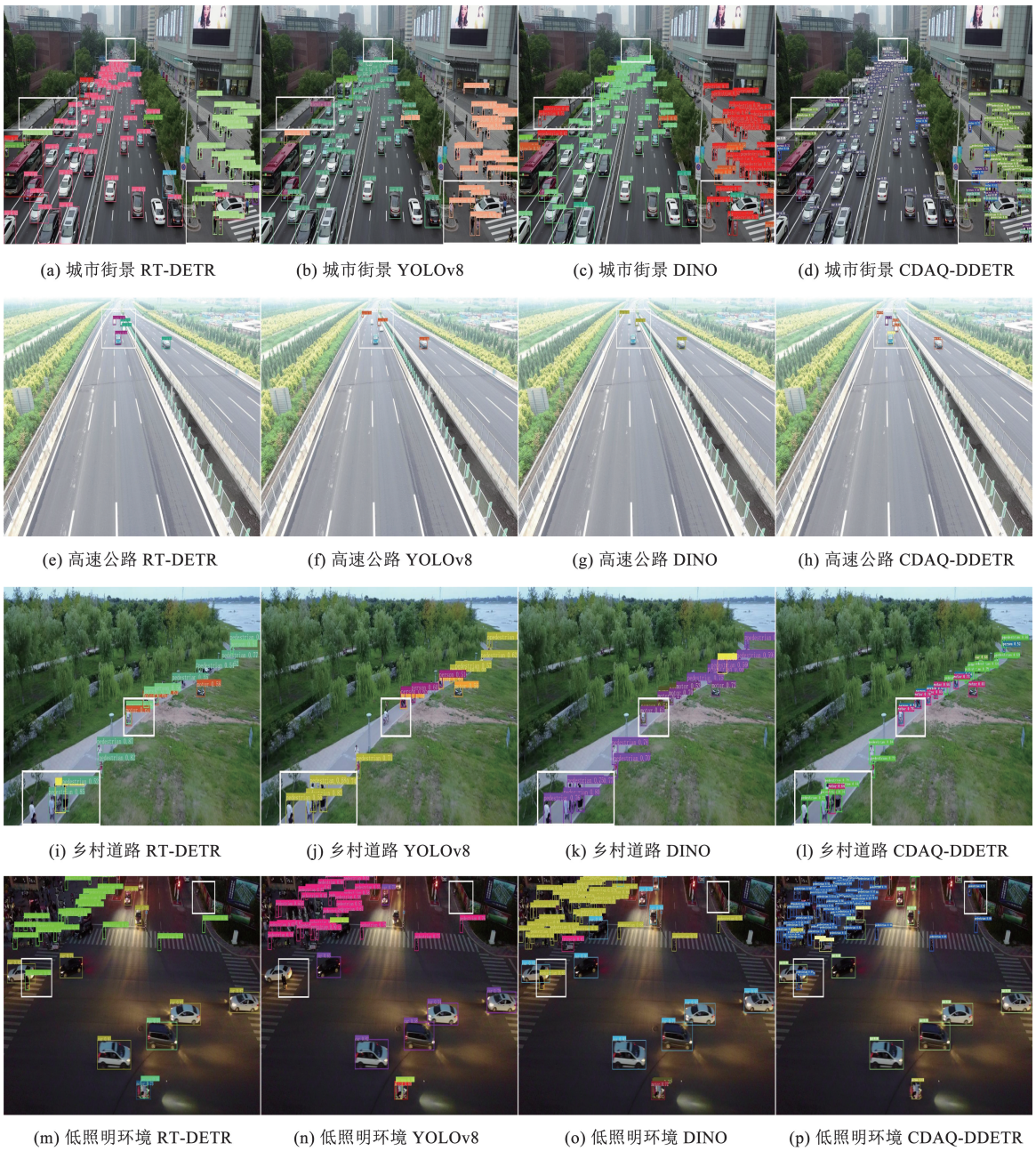
表 4、5 对比分析无人机航拍交通目标检测算法显示，Faster-RCNN、Cascade RCNN 作为 Two-stage 二阶段目标检测算法的典型代表，其核心基于区域候选框的检测方式相较于经典的一阶段 One-stage 算法（如 SSD、RetinaNet），在识别如行人（pedestrain）、汽车（car）这些常见且数据较充足的物体时，确实能学习训练到更好的预测效果，但对于诸如三轮车（tricycle）、摩托车（motor）这些尺寸较小且经常处于多密集场景下的交通目标，Two-stage 算法在密集小目标检测中受限。One-stage 算法如 FCOS、GFL、YOLOv8 提出解决方案。FCOS 去锚点框，用中心度优化密集场景检测；GFL 引入动态标签分配和质量感知损失；YOLOv8 以注意力机制提升小目标检测性能。可见表 4、5 上述 3 类算法在小目标检测性能上确实优于之前的经典算法，自 2021 年开始，新型 Transformer 架构（如 TOOD、ViTDet 等）在低数据量目标上识别性能提升，但对极小目标仍不足。本文改进算法 CDAQ-DDETR 通过 CBAM 双塔注意力及 DCNv2 可变形卷积提高算法对密集场景目标的识别精度，同时针对无人机航拍多尺度目标特点引入注意力感知融合金字塔，使不同类别不同尺度的目标效果均获得提升，同时对于多样化实际交通场景内遮挡干扰严重的目标，所提改进算法通过动态查询机制模块主动分析每张图片的具体内容，进而做到模型网络对特定目标的针对性提升，表 6 中的相关数据显示，本文改进算法 CDAQ-DDETR 在各尺度不同阈值下的精确率与召回率均高于现阶段主流检测算法。

3.4.3　可视化实验

为了更加直观地验证本文改进算法 CDAQ-DDETR 对于交通目标的检测性能，选取城市街景、高速公路、乡村道路、低照明环境，4 种代表性验证集图片进行可视化实验，选择近年份性能优越且富有代表性的 RT-DETR、YOLOv8、DINO 目标检测算法与本文改进算法 CDAQ-DDETR 进行直观对照，结果如图 12 所示。

在城市街景密集条件下，CDAQ-DDETR 针对相互遮挡冗余的交通目标检测性能更加优越，尤其面对远端及其微小尺寸的目标，本文改进算法相较其他 SOTA 算法检测效果有明显的改善，在高速公路

场景中,CDAQ-DDETR 算法对于高光照条件下与背景融合度高的小尺寸交通目标,同样具备高精准的识别定位,在乡村道路场景中,对于边缘化目标及车辆与人体遮挡度高的交通类别(如骑在摩托车上的人),CDAQ-DDETR 算法也能通过优越的算法网络对两种类别实现精准分割识别及定位,在低照明环境中,其他主流算法会因光感和亮度不强导致识别目标不精,而 CDAQ-DDETR 算法克服此类光照条件引发的复杂背景干扰,对于视觉远方复杂场景内的微小行人目标也能准确识别,这些能直接观察到对比结果充分验证了本文改进算法 CDAQ-DDETR 的突出性能,更加适用于实际密集场景中多尺度交通小目标的检测任务。



(a) 城市街景 RT-DETR (b) 城市街景 YOLOv8 (c) 城市街景 DINO (d) 城市街景 CDAQ-DDETR

(e) 高速公路 RT-DETR (f) 高速公路 YOLOv8 (g) 高速公路 DINO (h) 高速公路 CDAQ-DDETR

(i) 乡村道路 RT-DETR (j) 乡村道路 YOLOv8 (k) 乡村道路 DINO (l) 乡村道路 CDAQ-DDETR

(m) 低照明环境 RT-DETR (n) 低照明环境 YOLOv8 (o) 低照明环境 DINO (p) 低照明环境 CDAQ-DDETR

图 12 可视化实验对比

Fig. 12 Visualization of experimental comparison

4 结 论

本文提出了一种基于 Deformable DETR 改进的结合注意力变形和动态查询机制的小目标检测算法 CDAQ-DDETR,将 CBAM 注意力双塔机制和 DCNv2 可变形卷积引入原始残差网络 ResNet-50 中,增强了算法对密集区域交通小目标的语义获取能力;同时借助 AFN 网络思想添加低层特征的同时构建注意力感知融合金字塔模块,提高了原始算法对多尺度中小交通目标的检测性能;最后依靠在原解码器前向集成动态查询机制模块,通过动态结合输入图像匹配目标特性以构建最佳查询向量,提升了算法

对多样化背景干扰的适应泛化能力。实验数据评估及可视化检测结果表明，相较于现阶段主流目标检测算法，本文改进算法针对高密度多样化场景下中小尺度交通目标检测具有更好的准确性及优越性，能够适应实际交通运输场景中各类交通目标的高效精准识别及定位任务。接下来将主要针对模型算法的计算效率性能进行更深层次的研究，进一步优化算法架构。

参考文献

[1] 黎茂盛，李杭聪. 基于交叉口车牌识别数据的网络交通状态分类方法[J]. 哈尔滨工业大学学报，2023，55(11)：82
LI Maosheng，LI Hangcong. Method for network traffic state classification based on intersection license plate recognition data[J]. Journal of Harbin Institute of Technology，2023，55(11)：82. DOI：10.11918/202202003

[2] 夏春艳，黄松，郑长友，等. 自动驾驶交叉路口测试场景建模及验证方法[J]. 软件学报，2023，34(7)：3002
XIA Chunyan，HUANG Song，ZHENG Changyou，et al. Modeling and verification method for autonomous driving intersection test scenarios[J]. Journal of Software，2023，34(7)：3002. DOI：10.13328/j.cnki.jos.006855

[3] 杨天麟，王卫杰，康楠，等. 采用改进暗通道先验算法的高速公路能见度检测[J]. 哈尔滨工业大学学报，2023，55(3)：100
YANG Tianlin，WANG Weijie，KANG Nan，et al. Highway visibility detection using an improved dark channel prior algorithm[J]. Journal of Harbin Institute of Technology，2023，55(3)：100. DOI：10.11918/202111066

[4] REN Shaoqing，HE Kaiming，GIRSHICK R，et al. Faster R-CNN：towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence，2017，39(6)：1137. DOI：10.1109/TPAMI.2016.2577031

[5] CAI Zhaowei，VASCONCELOS N. Cascade R-CNN：delving into high quality object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake：IEEE，2018：6154. DOI：10.1109/CVPR.2018.00644

[6] LI Xiang，WANG Wenhai，WU Lijun，et al. Generalized focal loss：learning qualified and distributed bounding boxes for dense object detection[J]. Advances in Neural Information Processing Systems，2020，33：21002. DOI：10.48550/arXiv.2006.04388

[7] RAHMAN S，RONY J H，UDDIN J，et al. Real-time obstacle detection with YOLOv8 in a WSN using UAV aerial photography[J]. Journal of Imaging，2023，9(10)：216. DOI：10.3390/jimaging9100216

[8] 黄继鹏，史颖欢，高阳. 面向小目标的多尺度 Faster-RCNN 检测算法[J]. 计算机研究与发展，2019，56(2)：319
HUANG Jipeng，SHI Yinghuan，GAO Yang. Multi-scale faster-RCNN detection algorithm for small objects[J]. Journal of Computer Research and Development，2019，56(2)：319. DOI：10.7544/issn1000-1239.2019.20170749

[9] SIMONYAN K，ZISSERMAN A. Very deep convolutional networks for large-scale imagerecognition[EB/OL]. (2014-09-04). https://doi.org/10.48550/arXiv.1409.1556

[10] 杨彪，闫国成，刘占文，等. 基于异构图学习的交通场景运动目标感知[J]. 交通运输工程学报，2022，22(3)：238
YANG Biao，YAN Guocheng，LIU Zhanwen，et al. Motion target perception in traffic scenes based on heterogeneous graph learning[J]. Journal of Transportation Engineering，2022，22(3)：238. DOI：10.19818/j.cnki.1671-1637.2022.03.019

[11] 曾筠程，邵敏华，孙立军，等. 基于有向图卷积神经网络的交通预测与拥堵管控[J]. 中国公路学报，2021，34(12)：239
ZENG Yuncheng，SHAO Minhua，SUN Lijun，et al. Traffic prediction and congestion management based on directed graph convolutional neural networks[J]. China Journal of Highway and Transport，2021，34(12)：239. DOI：10.19721/j.cnki.1001-7372.2021.12.018

[12] VASWANI A，SHAZEER N，PARMAR N，et al. Attention is all you need[EB/OL]. (2017-06-12). https://doi.org/10.48550/arXiv.1706.03762

[13] 田永林，王雨桐，王建功，等. 视觉 Transformer 研究的关键问题：现状及展望[J]. 自动化学报，2022，48(4)：957
TIAN Yonglin，WANG Yutong，WANG Jiangong，et al. Key issues in visual Transformer research：current status and prospects[J]. Acta Automatica Sinica，2022，48(4)：957. DOI：10.16383/j.aas.c220027

[14] DOSOVITSKIY A，BEYER L，KOLESNIKOV A，et al. An image is worth 16×16 words：transformers for image recognition at scale[EB/OL]. (2020-10-22). https://doi.org/DOI：10.48550/arXiv.2010.11929

[15] 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

[16] ZHU Xizhou，SU Weijie，LU Lewei，et al. Deformable DETR：deformable transformers for end-to-end object detection[EB/OL]. (2020-10-08). https://doi.org/10.48550/arXiv.2010.04159

[17] LIN T Y，DOLLÁR P，GIRSHICK R，et al. Feature pyramid networks for object detection[EB/OL]. (2016-12-09). https://doi.org/10.48550/arXiv.1612.03144

[18] HE Kaiming，ZHANG Xiangyu，REN Shaoqing，et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas：IEEE，2016：770

[19] DU Dawei，ZHU Pengfei，WEN Longyin，et al. VisDrone-DET2019：the vision meets drone object detection in image challenge results[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. Seoul：IEEE，2019：213