基于改进RT-DETR的雾天环境目标检测*

王赵嘉文　付锐　曲思柔　孙秦豫

（长安大学汽车学院，西安 710054）

【摘要】针对雾天图像对比度下降和目标纹理模糊导致的检测性能退化问题，以实时检测Transformer（RT-DETR）为基准，提出雾天目标检测算法Fog-RT-DETR。该算法引入渐进特征金字塔网络（AFPN）优化多尺度特征融合，利用方差增强卷积块注意力机制（VE-CBAM）感知局部低纹理区域，并通过特征退化增强模块（FDEM）生成空间退化权重并调节特征增强强度。在基于DAWN和ACDC构建的雾天数据集上，Fog-RT-DETR的mAP@0.5和mAP@0.5∶0.95分别达到81.45%和58.78%，较基准模型分别提高2.12百分点和3.89百分点，召回率提高0.49百分点。消融实验表明，3个模块的联合使用取得了试验中的最佳检测结果；通道压缩试验进一步展示了模型在精度与计算开销之间的调整空间。

主题词：目标检测　雾天天气　RT-DETR　注意力机制　可变形卷积

中图分类号：TP391.41　　**文献标志码：**A　　**DOI:** 10.19620/j.cnki.1000-3703.20250758

Object Detection in Foggy Environments Based on Improved RT-DETR

Wang Zhaojiawen, Fu Rui, Qu Sirou, Sun Qinyu

（School of Automobile, Chang'an University, Xi'an 710054）

【Abstract】To address the degradation of object detection caused by reduced contrast and blurred textures in foggy images, this paper proposes Fog-RT-DETR based on the Real-Time Detection Transformer (RT-DETR). The method integrates an Asymptotic Feature Pyramid Network (AFPN) for multi-scale feature fusion, a Variance-Enhanced Convolutional Block Attention Module (VE-CBAM) for local low-texture perception, and a Feature Degradation Enhancement Module (FDEM) for spatially weighted feature enhancement. Experiments on a foggy-weather dataset constructed from DAWN and ACDC show that Fog-RT-DETR achieves 81.45% mAP@0.5 and 58.78% mAP@0.5∶0.95, improving over the baseline by 2.12 and 3.89 percentage points, respectively. Recall increases by 0.49 percentage points. Ablation experiments show that combining all three modules yields the best detection results in the reported experiments. Channel compression experiments further demonstrate a trade-off between detection accuracy and computational cost.

Key words: Object detection, Dense fog weather, RT-DETR, Attention mechanism, Deformable convolution

【引用格式】王赵嘉文, 付锐, 曲思柔, 等. 基于改进RT-DETR的雾天环境目标检测[J]. 汽车技术, 2026(9): 28-36.

WANG Z J W, FU R, QU S R, et al. Object Detection in Foggy Environments Based on Improved RT-DETR[J]. Automobile Technology, 2026(9): 28-36.

1　前言

随着高级驾驶辅助系统（Advanced Driver Assistance Systems，ADAS）和自动驾驶技术的发展，交通场景中的视觉环境感知成为车辆安全运行的重要环节。摄像头能够提供丰富的道路图像信息，但雾天成像易受大气散射影响，出现对比度降低、边缘模糊和颜色失真等问题，使车辆、行人等目标的识别与定位更加困难。因此，研究雾天环境下的目标检测方法，有助于提高复杂天气条件下交通环境感知的可靠性。

基于卷积神经网络（Convolutional Neural Network，CNN）的目标检测方法已得到广泛应用，但正常天气下训练的模型在雾天场景中可能受到图像退化和数据分布变化的影响。针对这一问题，一类研究采用图像去雾作为检测前的预处理步骤：DehazeNet[1]利用卷积网络估计图像透射率；门控融合网络（Gated Fusion Network，

*基金项目：汽车智能驾驶功能标准及评价技术研究（SQ2024YFB2500059）。

通信作者：付锐（1965—），女，教授，研究方向为交通安全，Furui@chd.edu.cn。

GFN)[2]融合不同图像处理分支的信息；Yang等[3]结合物理模型解耦与对抗训练开展图像去雾；Li[4]则利用生成对抗网络(Generative Adversarial Network，GAN)进行多域天气图像转换。然而，“先去雾、后检测”的处理方式增加了系统环节与计算开销，去雾结果中的伪影也可能影响后续检测。因此，直接在检测网络内部改善退化特征表达，是值得研究的另一种技术路径。

基于CNN的目标检测器主要包括两阶段和一阶段方法。两阶段检测器以Faster R-CNN[5]为代表，其区域卷积神经网络(Region-Based Convolutional Neural Network，R-CNN)框架先生成候选区域，再完成分类与边界框回归。一阶段检测器包括YOLO(You Only Look Once)系列[6]和单次多框检测器(Single Shot MultiBox Detector，SSD)[7]，直接从图像特征预测目标类别与位置。针对恶劣天气，图像自适应YOLO(Image-Adaptive YOLO，IA-YOLO)[8]引入可微图像处理模块(Differentiable Image Processing，DIP)，联合优化图像增强与检测；多尺度域自适应YOLO(Multiscale Domain Adaptive YOLO，MS-DAYOLO)[9]则通过多尺度域适应缓解训练域与测试域之间的差异。

近年来，基于Transformer的目标检测方法为端到端检测提供了新的实现路径。Carion等[10]提出检测Transformer(Detection Transformer，DETR)，将目标检测建模为集合预测任务，利用二分图匹配实现预测与真实目标之间的一一对应，省去手工设计的锚框和非极大值抑制(Non-Maximum Suppression，NMS)。针对DETR收敛较慢和计算开销较高的问题，可变形DETR(Deformable DETR)[11]通过关注稀疏采样位置改进注意力计算。RT-DETR[12]进一步通过高效混合编码器和查询选择策略兼顾检测精度与推理效率，为实时端到端检测提供了基础框架。

本文旨在缓解雾天图像低对比度和局部非均匀退化对目标检测的影响，在RT-DETR基础上构建Fog-RT-DETR。具体而言，引入AFPN改善不同尺度特征之间的信息交互；设计VE-CBAM，利用局部方差信息调节注意力权重；构建FDEM，通过退化感知与门控机制调节特征增强强度。本文采用雾天目标检测对比试验验证整体效果，通过模块消融实验分析各模块的作用，并结合通道压缩和跨天气测试考察计算开销与泛化表现。

2 计算方法

2.1 RT-DETR算法

本文研究重点是在控制额外计算开销的同时，提高退化场景中的目标识别与定位能力。整体网络结构如图1所示。

Transformer利用自注意力机制建模特征之间的依赖关系。DETR[10]将该机制引入目标检测，并通过集合预测与二分图匹配实现端到端训练，减少对锚框设计和NMS后处理的依赖。RT-DETR[12]在此基础上改进多尺度特征处理和查询初始化，以降低计算开销。本文以RT-DETR为基准框架，在特征融合和特征增强环节进行改进。

RT-DETR主要包括骨干网络、高效混合编码器和

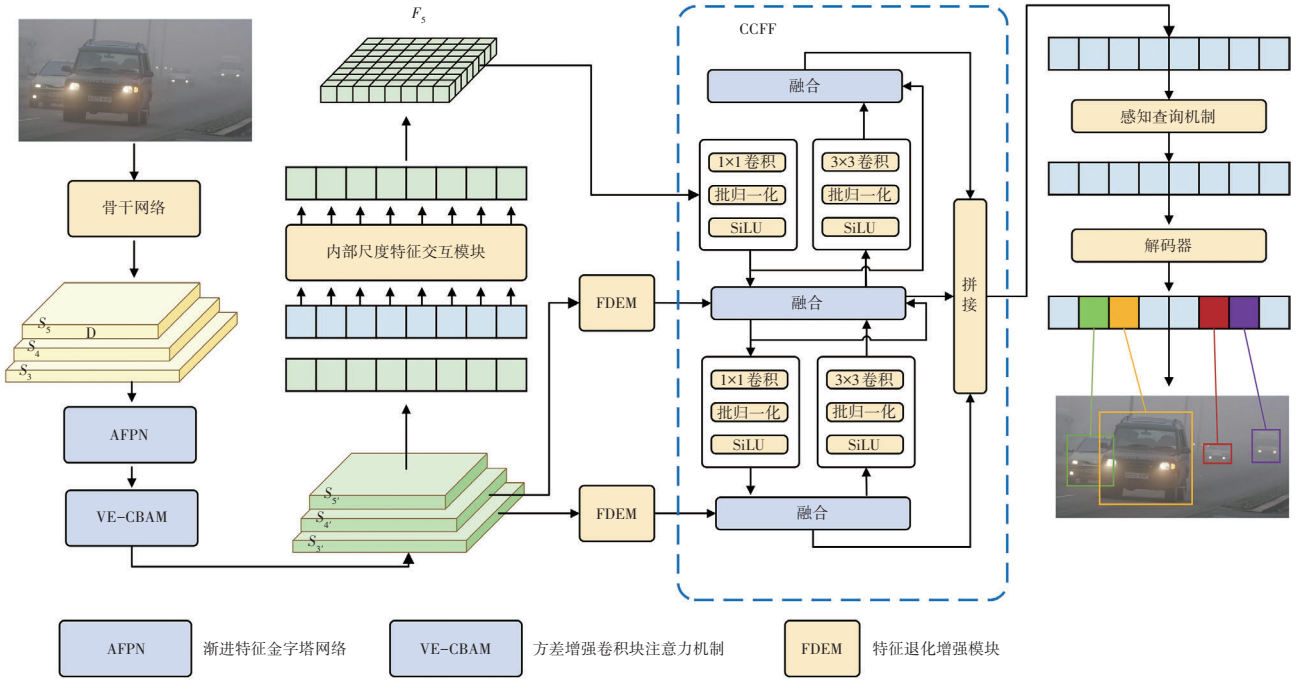


图1 Fog-DETR网络结构

带辅助预测头的Transformer解码器，并利用不确定性最小化查询选择机制（Uncertainty-Minimal Query Selection）为解码器提供初始查询。

骨干网络提取多尺度特征，高效混合编码器随后分别处理尺度内交互与跨尺度融合。尺度内交互采用基于注意力的尺度内特征交互模块（Attention-Based Intra-Scale Feature Interaction，AIFI），主要作用于高层语义特征；跨尺度融合采用基于CNN的跨尺度特征融合模块（CNN-Based Cross-Scale Feature Fusion，CCFF）[12]。为与式（3）的符号保持对应，本文以CCFM表示该跨尺度融合模块的函数标记。前者增强高层特征的上下文联系，后者整合不同分辨率的特征信息。

高效混合编码器的计算过程可以概括为：

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

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

$$Output = f_{\text{CCFM}}(\{S_3, S_4, F_5\}) \tag{3}$$

式中：S_3、S_4和S_5分别为骨干网络第3、4、5阶段输出的特征图，下标表示特征层级；f_{Flatten}表示将空间维度展平为序列的操作；Q、K和V分别表示自注意力中的查询、键和值；f_{Attn}表示AIFI中的注意力交互运算；f_{Reshape}将序列恢复为与S_5相同的空间排列，得到增强特征F_5；f_{CCFM}表示跨尺度特征融合运算，将S_3、S_4和F_5融合为编码器输出Output。

查询初始化需要同时考虑分类与定位质量。仅按分类置信度选择特征，可能将分类分数较高但定位质量较低的候选项送入解码器。RT-DETR通过在训练目标中考虑二者的不一致性，改善编码器特征的质量，从而为后续查询选择提供更可靠的候选特征[12]。

特征不确定性U用于描述定位预测分布P与分类预测分布C之间的差异，其表达式为：

$$U\left(\hat{\boldsymbol{X}}\right) = \left\| P\left(\hat{\boldsymbol{X}}\right) - C\left(\hat{\boldsymbol{X}}\right) \right\|, \quad \hat{\boldsymbol{X}} \in \mathbf{R}^D \tag{4}$$

RT-DETR将上述不确定性纳入训练目标，通过基于梯度的优化改善查询特征的分类与定位一致性：

$$L\left(\hat{\boldsymbol{X}}, \hat{Y}, Y\right) = L_{\text{box}}\left(\hat{b}, b\right) + L_{\text{cls}}\left(U\left(\hat{X}\right), \hat{c}, c\right) \tag{5}$$

式中：$\hat{\boldsymbol{X}}$为编码器特征向量，$\mathbf{R}$为实数域，D为其维度，P和C分别为定位预测分布与分类预测分布，$\|\cdot\|$为衡量两者差异的范数，$\hat{Y}$和Y分别为预测结果与真实标注，$\hat{c}$和c为预测类别信息与真实类别，$\hat{b}$和b为预测边界框与真实边界框，L为训练损失，L_{box}为边界框回归项，L_{cls}为考虑不确定性的分类项。

该目标促使分类置信度与定位质量更加一致；编码器中选出的固定数量特征随后用于初始化解码器查询，并经迭代更新生成最终检测结果。

2.2 渐进特征金字塔网络（AFPN）

为改善多尺度特征之间的信息交互，本文在检测框架中引入AFPN。该网络采用逐步融合的思路，使不同层级的特征在交互过程中逐渐整合，以减小跨层语义差异对特征融合的影响。

传统特征金字塔网络（Feature Pyramid Network，FPN）[13]通过自顶向下的路径和横向连接，将高层语义信息传递至较高分辨率的特征层。不同层级特征的空间分辨率与语义内容存在差异，因此融合时需要协调细节信息与语义信息。AFPN的渐进融合思路为逐步整合多层特征提供了另一种方式。

在本文所描述的融合路径中，首先对高层特征S_4和S_5进行融合，再将所得特征与S_3整合。这里S_3、S_4和S_5与式（1）~式（3）中的符号含义一致，均表示骨干网络不同阶段的输出。通过分阶段的信息交互，高层语义与较高分辨率的局部细节得以逐步结合，为后续检测提供多尺度特征。融合结构如图2所示。

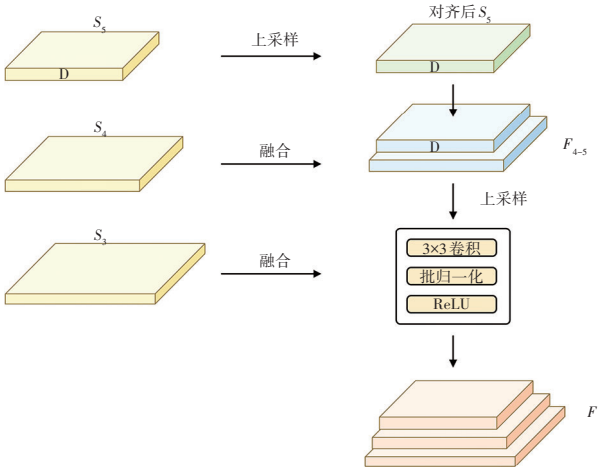


图2 渐进特征金字塔网络结构

2.3 方差增强卷积块注意力机制（VE-CBAM）

为增强雾天图像中的有效特征，本文设计VE-CBAM，在通道与空间注意力的基础上引入局部方差信息。该模块利用特征局部变化程度作为辅助线索，调节不同区域的增强权重，以缓解低对比度和纹理模糊对特征表达的影响。

卷积块注意力模块（Convolutional Block Attention Module，CBAM）[14]依次计算通道注意力和空间注意力。通道分支通过全局平均池化、全局最大池化及共享多层感知机（Multilayer Perceptron，MLP）生成权重；空间分支沿通道维度聚合信息并通过卷积生成空间权重。将其用于雾天检测时，仍有以下方面可进一步改进：

a. 通用注意力权重未显式引入雾天局部退化的辅

助线索；

b. 全局池化压缩了局部空间差异，可能难以充分描述不同区域的退化程度；

c. 空间注意力缺少对局部纹理变化的直接约束。

因此，本文引入局部方差权重，对原有注意力进行补充。局部方差描述特征在邻域内的波动程度，可作为局部纹理强弱的辅助指标。雾天退化可能降低边缘和纹理对比度，但低方差也可能来自天空、路面等本身较平滑的区域。因此，局部方差不能直接等同于雾浓度或退化程度。VE-CBAM将该信息与可学习注意力结合使用，而不是仅凭方差判定区域是否受雾影响。

通道分支先通过全局平均池化和全局最大池化获得描述向量，经共享MLP和Sigmoid函数生成通道权重，再与局部方差权重结合。空间分支对加权特征沿通道维度进行最大池化与平均池化，拼接后经7×7卷积生成空间注意力图。模块结构如图3所示。

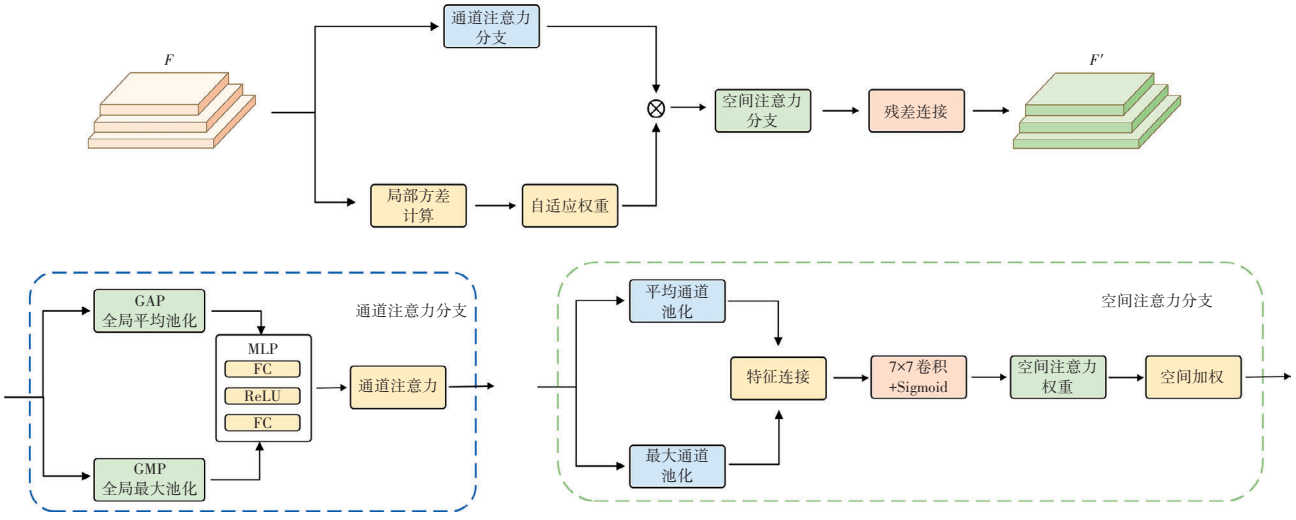


图3　方差增强卷积块注意力机制（VE-CBAM）结构

VE-CBAM对每个尺度的输入特征分别进行处理，其核心计算过程可表示为：

$$W_{\text{fog}} = \sigma\left(\frac{1}{Var_{\text{local}}(F)}\right) \tag{6}$$

$$M_c = \sigma(MLP(GAP(F)) + MLP(GMP(F))) \tag{7}$$

$$M_c^{\text{fog}} = M_c \odot W_{\text{fog}} \tag{8}$$

$$F_{\text{CAM}} = M_c^{\text{fog}} \odot F \tag{9}$$

$$F_s = [MaxPool_c(F_{\text{CAM}}); AvgPool_c(F_{\text{CAM}})] \tag{10}$$

$$M_s = \sigma(Conv_{7\times7}(F_s)) \tag{11}$$

$$F_i^{\text{enhanced}} = M_s \odot F_{\text{CAM}} + F_i \tag{12}$$

式中：F为当前尺度的输入特征，与式（12）中的F_i对应，i为尺度索引；Var_{local}为在特征图局部邻域内计算的方差；σ为Sigmoid函数，其值域为(0，1)；W_{fog}为由局部方差得到的辅助权重，式（6）按方差大于0的情形定义，方差为0时可按右极限将权重定义为1；GAP和GMP分别为全局平均池化与全局最大池化；MLP为共享多层感知机；M_c为通道注意力权重；M_c^{fog}为结合局部方差后的权重；F_{CAM}为加权特征；$MaxPool_c$和$AvgPool_c$分别为沿通道维度执行最大池化和平均池化；方括号内的分号表示沿通道维度拼接；F_s为拼接后的空间描述特征；$Conv_{7\times7}$为7×7卷积；M_s为空间注意力权重；F_i^{enhanced}为当前尺度的输出；符号⊙表示按维度广播后的逐元素乘法；M_c在空间维度广播，M_s在通道维度广播。

残差连接用于保留输入信息，避免输出仅依赖注意力加权分支。

2.4　特征退化增强模块（FDEM）

本文采用FDEM对空间分布不均匀的退化特征进行增强，如图4所示。该模块包含并行的退化感知分支与特征修正分支：前者生成空间权重，后者生成用于增强的特征，随后由门控机制融合两条分支的信息。该设计使不同空间位置获得不同的增强强度，以改善局部退化条件下的特征表达。

2.4.1　特征退化感知网络

退化感知分支利用可变形卷积网络[15]提取特征。可变形卷积通过学习采样偏移量调整采样位置，从而适应局部不规则结构。提取到的特征经卷积投影和激活函数处理后，映射为单通道空间权重。按照式（14）和式（15）的记号，两层卷积核尺寸分别为3×1和1×1，上标（1）、（2）表示层的顺序，其计算过程为：

$$F_{\text{dcn}} = DCN(F_{\text{original}}) \tag{13}$$

$$F_{\text{activated}} = ReLU(Conv_{3\times1}^{(1)}(F_{\text{dcn}})) \tag{14}$$

$$M_d = \sigma(Conv_{1\times1}^{(2)}(F_{\text{activated}})) \tag{15}$$

式中：F_{original}为输入特征图；DCN表示可变形卷积特征

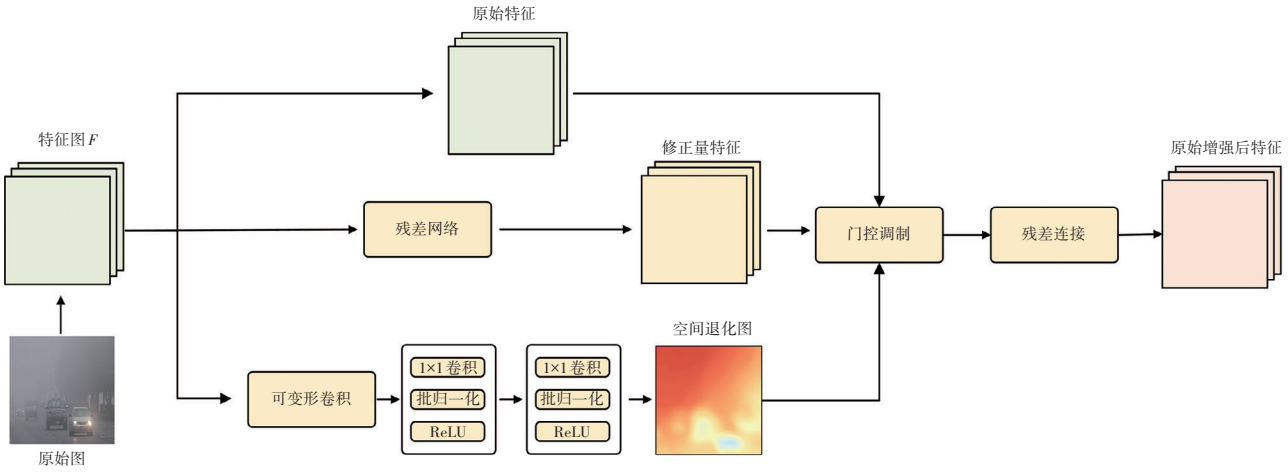


图4　特征退化增强模块（FDEM）结构

提取运算，F_{dcn}为其输出；$Conv_{3\times1}^{(1)}$和$Conv_{1\times1}^{(2)}$分别为第一层3×1卷积和第二层1×1卷积；*ReLU*为修正线性单元（Rectified Linear Unit），用于引入非线性；$F_{\mathrm{activated}}$为激活后的特征；σ为Sigmoid函数，M_d为单通道空间权重图，M_d中的数值用于控制增强幅度，不具有能见度或雾浓度等物理量的单位。

2.4.2　特征修正量估算模块

特征修正分支与退化感知分支并行，用于生成供门控增强使用的特征。该分支采用残差块堆叠，通过输入特征与各残差映射的累加得到分支输出。分支输出F_{residual}包含原始特征与各残差映射，可在保留输入信息的同时引入学习到的特征变化。其计算过程为：

$$F_{\mathrm{residual}} = F_{\mathrm{original}} + \sum_{l=1}^{L} F\left(F_{l-1}, W_l\right) \tag{16}$$

式中：F_{original}为输入特征，亦为首个残差块的输入；F_{residual}为残差分支输出；l为残差块索引；L为残差块总数；F_{l-1}为第l个残差块的输入；F在求和项中表示残差映射函数，与式（6）中的特征张量符号应按上下文区分；W_l为该残差块的可学习参数集合。

各残差映射的输出与输入特征保持相容的形状，以便执行逐元素相加。

2.4.3　自适应门控增强

自适应门控增强根据退化感知分支的空间权重，对残差分支输出逐位置加权，再与原始特征相加。该机制使增强分支在不同位置具有不同的贡献。该门控融合同时保留原始特征的直接路径与残差网络处理后的分支信息，利用空间权重控制后者对最终输出的贡献。

该自适应增强过程的核心数学表达式为：

$$F_{\mathrm{out}} = F_{\mathrm{original}} + M_d \odot F_{\mathrm{residual}} \tag{17}$$

式中：F_{original}表示输入特征图；F_{out}表示增强后的输出特征图；F_{residual}为式（16）定义的残差分支输出；M_d为式（15）生成的单通道空间权重，在通道维度广播后与F_{residual}相乘；⊙表示逐元素乘法。

最终加法要求两条特征路径具有相同的空间尺寸和通道数。

3　试验设置与结果分析

3.1　数据集构建

本文基于恶劣天气车辆检测数据集DAWN（Vehicle Detection in Adverse Weather Nature）[16]和带对应图像的恶劣条件数据集（Adverse Conditions Dataset with Correspondences，ACDC）[17]构建雾天目标检测实验数据。所用图像涉及不同雾气程度与道路场景，用于评估模型在退化条件下的检测表现。本文试验数据共计2 300张。

试验数据按8∶2划分为训练集与验证集，分别包含1 840张和460张图像，检测对象包括车辆、行人及两轮交通工具。数据集图像示例如图5所示。



图5　雾天数据集图像示例

3.2　试验设置

试验在配备NVIDIA RTX 4090图形处理器（Graphics Processing Unit，GPU）的服务器上进行，采用PyTorch 2.0和Python 3.8。批次大小设为8，使用AdamW优化器；学习率与权重衰减系数均设为1×10^{-4}，总训练轮数为100。骨干网络采用ResNet-50，输入图

像尺寸为640像素×640像素。

3.3 评价指标

本文采用召回率（Recall）和平均精度均值（mean Average Precision，mAP）评价检测性能。交并比（Intersection over Union，IoU）定义为预测框与真实框交集面积和并集面积之比；单类别平均精度（Average Precision，AP）用于概括精确率—召回率关系，mAP为各类别AP的平均值。mAP@0.5表示IoU阈值为0.5时的mAP；mAP@0.5∶0.95表示在0.5至0.95、间隔0.05的10个IoU阈值下取平均。召回率为正确检出的真实目标数占全部真实目标数的比例。前两项反映不同定位要求下的检测精度，召回率反映目标的检出程度。

3.4 损失函数

训练损失由分类损失、边界框坐标损失和广义交并比（Generalized Intersection over Union，GIoU）损失加权组成，对应配置项loss_vfl、loss_bbox、loss_giou的权重分别为1、5、2。分类项采用Varifocal Loss（VFL），以定位质量参与分类监督，改善分类分数与定位质量的一致性；其中α=0.75、γ=2.0，分别控制负样本权重与调制强度。坐标回归项采用L_1损失，约束预测框与真实框的坐标差异。GIoU损失同时考虑两框的重叠关系与最小包围区域，即使两框没有交集，也能提供与位置关系有关的优化信号。上述损失共同用于分类与定位训练，不将其作用限定为雾天特有的优化效果。

3.5 对比试验结果与分析

为评估Fog-RT-DETR的检测性能，本文选取Faster R-CNN[5]、YOLOv5、YOLOv8、IA-YOLO[8]、MS-DAYOLO[9]、DETR[10]、DINO及RT-DETR[12]作为对比模型，覆盖两阶段检测、一阶段检测、恶劣天气适应及Transformer检测等方法。各模型的检测结果如表1所示。

表1 不同目标检测算法性能对比

模型	骨干网络	召回率	mAP@0.5∶0.95	mAP@0.5
Faster-RCNN	ResNet-50	0.723 4	0.485 2	0.741 8
YOLOv5	CSPDarknet	0.736 7	0.504 5	0.762 3
YOLOv8	CSPDarknet	0.748 9	0.527 3	0.779 1
IA-YOLO	Darknet-53	0.754 5	0.536 5	0.784 2
MS-DAYOLO	CSPDarknet	0.751 2	0.532 8	0.781 5
DETR	ResNet-50	0.712 8	0.468 9	0.726 5
DINO	ResNet-50	0.741 2	0.518 3	0.769 7
RT-DETR	ResNet-50	0.760 6	0.548 9	0.793 3
Fog-RT-DETR	ResNet-50	0.765 5	0.587 8	0.814 5

表1表明，Fog-RT-DETR在本组试验的3项检测指标上均取得最高数值。其召回率为0.765 5，mAP@0.5为0.814 5，mAP@0.5∶0.95为0.587 8；RT-DETR对应数值为0.760 6、0.793 3和0.548 9。3项指标分别提高0.49%、2.12%和3.89%。其中，较严格的多IoU阈值平均精度增幅更大，说明改进对本试验中的定位质量具有积极作用。

IA-YOLO和MS-DAYOLO在本试验中的检测指标高于所列YOLOv5和YOLOv8结果，表明面向恶劣天气的针对性设计具有应用价值。DETR、DINO和RT-DETR之间也存在性能差异，但仅凭表1数据不能将这些差异全部归因于收敛速度或某一网络组件。Fog-RT-DETR的结果支持所提出改进在当前试验设置下的有效性；各模块的独立作用还需结合消融实验分析。表1未报告推理时延或帧率，因此不据此判定模型的实时性。

进一步比较，Fog-RT-DETR相较于IA-YOLO的召回率、mAP@0.5和mAP@0.5:0.95分别提高1.10百分点、3.03百分点和5.13百分点；相较于MS-DAYOLO，相应增幅为1.43百分点、3.30百分点和5.50百分点。在本试验所用数据与配置下，本文方法在目标检出和多IoU阈值综合评价上均取得较好的结果。此外，RT-DETR的两项mAP指标相差24.44百分点，本文模型的差值缩小至22.67百分点，但各类别、各IoU阈值下的变化仍需结合细分结果分析。

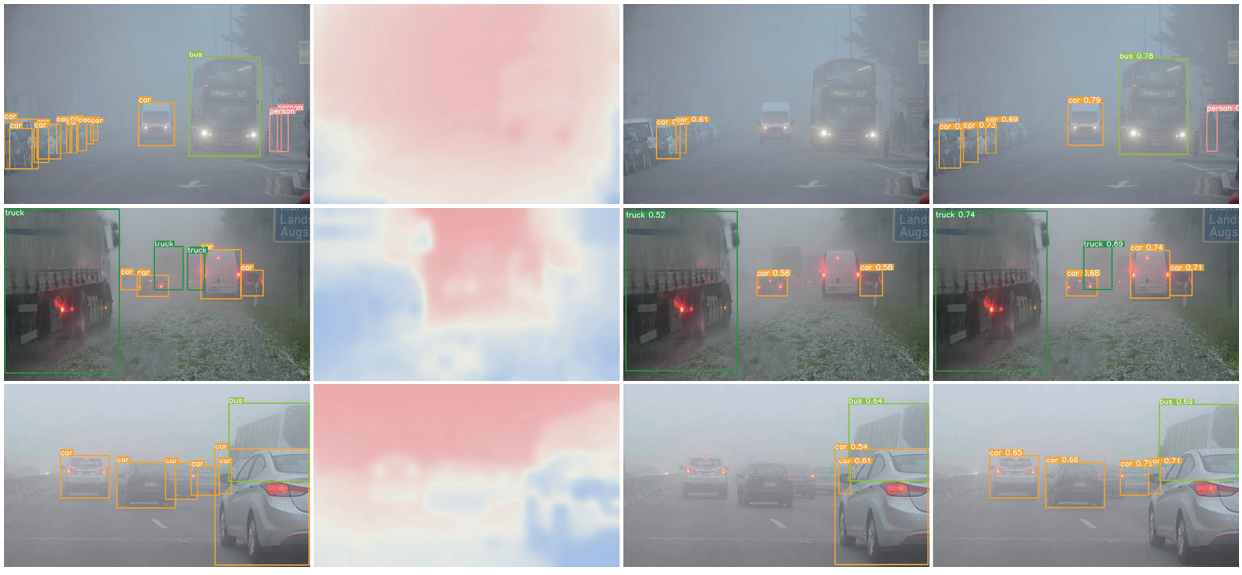
图6给出3组雾天场景的定性检测结果，以及FDEM生成的空间权重可视化。该权重图用于反映模型对不同位置的增强分配，不能直接视为经过标定的能见度分布或雾浓度测量。与基准RT-DETR相比，Fog-RT-DETR在所展示样例中检出了部分基准模型遗漏的目标。定性结果可辅助说明模型行为，整体性能仍以检测指标为依据。

3.6 消融实验

表2给出了模块消融实验结果。实验以RT-DETR为基准，分别启用AFPN、VE-CBAM和FDEM，并考察两两组合与全部组合，以比较不同模块配置对检测指标、参数量和计算量的影响。

单独引入AFPN、VE-CBAM和FDEM后，mAP@0.5较基准分别提高0.54百分点、0.31百分点和0.79百分点，mAP@0.5∶0.95分别提高1.23百分点、0.8百分点和1.56百分点。FDEM在单模块配置中取得最高增益；全部模块联合使用时，两项mAP分别达到0.814 5和0.587 8，高于任一单模块和双模块配置。上述结果支持各模块的互补作用，但现有结果未报告多次独立重复实验，尚不足以给出统计显著性结论。

模型复杂度方面，基准模型的参数量为42.75×10^6个、计算量为67.92×10^9次，完整模型分别为48.63×10^6个和



（a）原始标签　（b）空间退化图　（c）RT-DETR　（d）Fog-RT-DETR

图6　雾天场景下的检测结果对比

表2　不同模块组合对RT-DETR性能影响的消融实验

AFPN	VE-CBAM	FDEM	召回率	mAP@0.5∶0.95	mAP@0.5	参数量/×10⁶个	FLOPs/×10⁹次
–	–	–	0.760 6	0.548 9	0.793 3	42.75	67.92
√	–	–	0.762 8	0.561 2	0.798 7	44.28	72.15
–	√	–	0.761 9	0.557 3	0.796 4	43.43	67.93
–	–	√	0.763 4	0.564 5	0.801 2	44.97	72.16
√	√	–	0.764 1	0.572 8	0.805 8	47.08	82.89
√	–	√	0.764 7	0.579 1	0.808 4	48.61	87.12
–	√	√	0.763 8	0.576 4	0.807 1	47.10	82.89
√	√	√	0.765 5	0.587 8	0.814 5	48.63	87.13

87.13×10⁹次，增加5.88×10⁶个和19.21×10⁹次。FLOPs为浮点运算次数（Floating-Point Operations），用于表征计算工作量。单独加入VE-CBAM后，计算量由67.92×10⁹次变为67.93×10⁹次，在表2中的精度下增量较小。参数量与FLOPs可用于比较复杂度，但实际推理速度还取决于硬件、实现和数据精度，需通过时延或帧率测量确认。

表3比较了不同FDEM特征通道压缩比。压缩比由1.00降至0.25时，通道数由256降为64，参数量由48.63×10⁶个降为47.60×10⁶个，FLOPs由87.13×10⁹次降为84.15×10⁹次，计算量减少2.98×10⁹次，约为3.42%。在表2中保留2位小数的结果中，mAP@0.5均为0.81，mAP@0.5∶0.95由0.59降至0.58；受舍入精度限制，不能据此确定精确的指标降幅。该结果表明通道压缩能够减少部分计算开销，但是否适用于具体移动端设备，还需结合实测时延、内存占用与功耗评价。

3.7　跨场景鲁棒性分析

为考察模型在非雾天气中的表现，本文使用2 178张未参与模型训练的恶劣天气图像开展跨天气测试，数据来源为DAWN[16]和ACDC[17]。所考察场景包括雨、雪和沙尘等退化条件。这里的跨天气测试用于分析雾天训练模型对其他退化条件的适应情况，并不等同于针对未见目标类别的零样本检测。不同天气的物理成因和成像影响并不相同，因此测试结果只能支持所覆盖场景中的泛化表现，不能直接证明模型已经获得适用于所有退化类型的修复机制。

表3　特征通道压缩比敏感性分析

压缩比	通道数	召回率	mAP@0.5∶0.95	mAP@0.5	参数量/×10⁶个	FLOPs/×10⁹次
1.00	256	0.77	0.59	0.81	48.63	87.13
0.75	192	0.77	0.59	0.81	48.35	86.25
0.50	128	0.76	0.59	0.81	48.02	85.30
0.25	64	0.76	0.58	0.81	47.60	84.15

图7展示了Fog-RT-DETR与RT-DETR在不同恶劣天气样例中的检测结果。部分样例中，Fog-RT-DETR检出了基准模型遗漏的目标，说明本文方法对这些退化条件具有一定适应性。由于可视化样例数量有

限，该结果主要作为定性补充，不替代完整测试集上的定量评价。

图8给出了Fog-RT-DETR与基准模型在不同天气场景下的检测性能对比。图示结果中，本文方法的各场景指标均高于基准，可作为跨天气性能改善的补充证据。但该图不能单独解释性能提升的具体机制，模型对更多天气类型和不同数据来源的泛化能力仍需进一步评估。



(a)原始标签　　(b)RT-DETR　　(c)Fog-RT-DETR

图7　不同恶劣天气场景下的检测效果对比

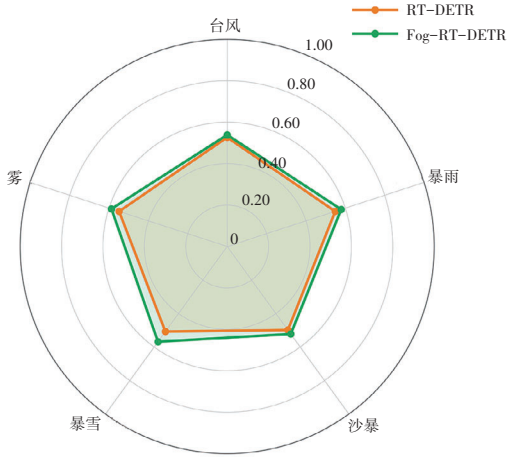


图8　不同恶劣天气场景下的检测性能对比

4　结束语

本文针对雾天图像退化引起的目标检测困难问题，在RT-DETR基础上构建了Fog-RT-DETR，通过多尺度特征融合、局部方差辅助注意力及空间门控增强改善特征表达。在本文试验设置下，模型mAP@0.5达到81.45%，mAP@0.5∶0.95达到58.78%，较基准分别提高2.12百分点和3.89百分点，召回率提高0.49百分点。消融实验支持3个模块联合使用的有效性，通道压缩实验展示了检测精度与计算开销之间的调整空间。当前结论主要基于有限规模数据，尚需通过更多独立数据、重复实验及统一测速验证模型的稳定性、泛化能力和部署

效率。后续将进一步研究不同天气条件下的适应能力，并结合边缘设备实测评估其应用价值。

参 考 文 献

[1] CAI B L, XU X M, JIA K, et al. DehazeNet: An End-to-End System for Single Image Haze Removal[J]. IEEE Transactions on Image Processing, 2016, 25(11): 5187-5198.

[2] REN W Q, MA L, ZHANG J W, et al. Gated Fusion Network for Single Image Dehazing[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 3253-3261.

[3] YANG X T, XU Z, LUO J B. Towards Perceptual Image Dehazing by Physics-Based Disentanglement and Adversarial Training[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2018, 32(1): 7485-7492.

[4] LI X L, KOU K, ZHAO B. Weather GAN: Multi-Domain Weather Translation Using Generative Adversarial Networks [EB/OL]. (2021-03-09)[2026-09-07]. https://arxiv.org/abs/2103.05422.

[5] REN S 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, 2017, 39(6): 1137-1149.

[6] 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. Piscataway: IEEE, 2016: 779-788.

[7] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single Shot MultiBox Detector[C]// Computer Vision-ECCV 2016. Cham: Springer, 2016: 21-37.

[8] LIU W Y, REN G F, YU R S, et al. Image-Adaptive YOLO for Object Detection in Adverse Weather Conditions[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2022, 36(2): 1792-1800.

[9] HNEWA M, RADHA H. Multiscale Domain Adaptive YOLO for Cross-Domain Object Detection[C]// Proceedings of the IEEE International Conference on Image Processing. Piscataway: IEEE, 2021: 3323-3327.

[10] CARION N, MASSA F, SYNNAEVE G, et al. End-to-End Object Detection with Transformers[C]// Computer Vision-ECCV 2020. Cham: Springer, 2020: 213-229.

[11] ZHU X Z, SU W J, LU L W, et al. Deformable DETR: Deformable Transformers for End-to-End Object Detection[C/OL]// Proceedings of the International Conference on Learning Representations. Vienna, Austria: IMLS, 2021[2026-09-07]. https://openreview.net/forum?id=gZ9hCDWe6ke.

[12] ZHAO Y A, LV W Y, XU S L, et al. DETRs Beat YOLOs on Real-Time Object Detection[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2024: 16965-16974.

[13] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature Pyramid Networks for Object Detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 2117-2125.

[14] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional Block Attention Module[C]// Computer Vision-ECCV 2018. Cham: Springer, 2018: 3-19.

[15] DAI J F, QI H Z, XIONG Y W, et al. Deformable Convolutional Networks[C]// Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 764-773.

[16] KENK M A, HASSABALLAH M. DAWN: Vehicle Detection in Adverse Weather Nature Dataset[EB/OL]. (2020-08-12)[2026-09-07]. https://arxiv.org/abs/2008.05402.

[17] SAKARIDIS C, DAI D X, VAN GOOL L. ACDC: The Adverse Conditions Dataset with Correspondences for Semantic Driving Scene Understanding[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2021: 10765-10775.

（责任编辑　天下溪）

修改稿收到日期为2026年9月8日。