文章编号:1673-0291(2026)02-0174-16 **DOI**:10.11860/j.issn.1673-0291.20250066

基于改进ViT网络的交通图像去雾研究

秦雅琴,童唯乔,赵仕林
(昆明理工大学 交通工程学院,昆明 650500)

摘 要:在雾霾天气条件下,交通监控图像常因对比度低、边缘模糊和远景信息退化而严重影响道路目标检测与交通事件识别的准确性.现有图像去雾方法多基于物理模型或端到端网络结构,在结构细节保持、边缘连续性增强以及复杂交通场景的适应性方面存在不足.针对上述问题,提出一种融合渐进最优滤波与引导滤波特性的改进型视觉转换器(Vision Transformer, ViT)交通图像去雾方法.首先,在输入端构建一种AO-Guided融合滤波算法,对图像进行增强处理,通过渐进最优极值滤波与引导滤波的协同作用机制增强图像边缘与结构敏感性;其次,在编码阶段嵌入AO-Guided特征增强模块,以提升网络对中远景雾化区域细节的建模能力;再次,设计联合损失函数与中间监督机制,引导网络实现全局结构保持与局部细节还原的协同优化;最后,在RESIDE真实有雾交通图像数据集上开展去雾实验.研究结果表明:与基于ViT架构的典型去雾方法DehazeFormer相比,改进ViT网络的交通图像去雾方法在图像可见边集合数目比、平均梯度比和黑色像素占比三项结构敏感性指标上分别提升13.1%、提升40.5%、降低25%,可明显提升雾霾下交通图像的去雾效果.

关键词:信号与信息处理;图像去雾;特征增强;交通图像;Vision Transformer算法
中图分类号:TP391.41 **文献标志码**:A

Research of traffic image dehazing based on an improved vision transformer network

QIN Yaqin, TONG Weiqiao, ZHAO Shilin
(School of Traffic Engineering, Kunming University of Science and Technology, Kunming 650500, China)

Abstract: Under hazy weather conditions, traffic surveillance images often suffer from low contrast, blurred edges, and degraded distant visual information, which severely compromise the accuracy of road target detection and traffic event recognition. Existing image dehazing methods, which are primarily based on physical models or end-to-end network architectures, continue to exhibit limitations in structural detail preservation, edge continuity enhancement, and adaptability to complex traffic scenarios. To address these issues, this paper proposes an improved Vision Transformer (ViT)-based traffic image dehazing method that integrates the characteristics of progressive optimal filtering and guided filtering. First, an AO-Guided method is introduced at the input stage for image enhancement

收稿日期:2025-05-20;**修回日期**:2025-10-31
基金项目:国家自然科学基金(72261021);云南省科学技术厅基础研究项目(202501AS070152)
Foundation items: National Natural Science Foundation of China (72261021); Basic Research Project of the Department of Science and Technology of Yunnan Province (202501AS070152)
第一作者:秦雅琴(1972—),女,湖南平江人,教授,博士,博士生导师.研究方向为交通安全与图像处理. **email**:qinyaqin@kust.edu.cn.
通信作者:赵仕林(1995—),男,重庆奉节人,博士生. **email**:zhaoshilin@stu.kust.edu.cn.

引用格式:秦雅琴,童唯乔,赵仕林.基于改进ViT网络的交通图像去雾研究[J].北京交通大学学报,2026,50(2):174-189.
QIN Yaqin, TONG Weiqiao, ZHAO Shilin. Research of traffic image dehazing based on an improved vision transformer network[J]. Journal of Beijing Jiaotong University, 2026, 50(2): 174-189. (in Chinese)

by leveraging the synergistic mechanism of progressive optimal extreme filtering and guided filtering, this approach enhances the edge and structural sensitivity of the images. Second, an AO-Guided feature enhancement module is embedded during the encoding stage to improve the network's ability to model detailed information in mid- and long-range fog-degraded regions. Third, a joint loss function and an intermediate supervision mechanism are designed to guide the network in achieving the collaborative optimization of global structure preservation and local detail restoration. Finally, dehazing experiments are conducted on the RESIDE real-world hazy traffic image dataset. The results demonstrate that, compared to DehazeFormer, a representative ViT-based dehazing method, the proposed approach improves three key structural sensitivity metrics: the ratio of visible edges and the average gradient ratio increase by 13.1% and 40.5%, respectively, while the proportion of black pixels decreases by 25%. Consequently, the proposed method significantly enhances the dehazing performance of traffic images under hazy conditions.

Keywords: signal and information processing; image dehazing; feature enhancement; traffic image; vision Transformer algorithm

随着智慧交通与车路协同系统的持续发展,图像感知技术正逐步成为城市道路管理、交通流监测及异常预警的核心手段之一.然而,受限于雾霾等复杂气象条件的影响,前端视觉采集易产生图像模糊、对比度下降、边缘信息丢失等现象,严重干扰交通目标识别与行为理解的准确性.因此,交通场景下有雾图像退化问题的处理方法研究,已成为当前智能交通感知领域亟需突破的关键技术方向之一[1-2].

早期图像去雾方法多基于物理建模,通过构建图像退化过程中的透射率与大气光模型进行图像重建.典型方法如文献[3]提出的暗通道先验(Dark Channel Prior,DCP)及文献[4]提出的颜色衰减先验(Color Attenuation Prior,CAP),在合成图像上取得了一定效果,但由于过于依赖固定先验,在复杂交通场景中易出现图像颜色失真、细节还原不足等问题,难以满足工程应用中对实时性与通用性的双重要求.

近年来,随着深度学习技术的发展,端到端图像去雾方法逐步替代传统模型成为主流.文献[5]提出去雾网络(Dehazing Network, DehazeNet);文献[6]设计了全一去雾网络(All-in-One Dehazing Network,AOD-Net),通过引入卷积神经网络完成透射率估计,有效提升了去雾精度;文献[7]提出的特征融合注意力网络(Feature Fusion Attention Network, FFA-Net)结构进一步引入注意力机制与多尺度特征融合策略,在保持图像细节与结构信息方面取得了显著改进.然而,基于卷积模型天然存在感受野受限与局部建模能力不足的问题,难以有效捕捉全局结构与远距离目标之间的关联,尤其在长距离交通场景中远距离目标信息建模方面存在明显性能瓶颈.

随着 Transformer 架构在图像识别领域的成功应用,其在低层视觉任务中的研究也逐渐受到关注.文献[8]提出的 Restormer 以及文献[9]提出的 DehazeFormer 引入了窗口注意力、归一化重构与跨层融合等机制,显著提升了图像复原质量;文献[10]通过建立基于细节增强注意力机制的去雾网络,在复杂场景下显著提升了图像结构恢复能力;文献[11]通过多层特征融合与自适应策略提升了模型在不同雾强条件下的泛化性能;文献[12]结合多尺度 Transformer 与对比学习策略,在复杂雾霾条件下实现了更优的特征表达能力;文献[13]基于频域特征建模构建 Transformer 去雾网络;文献[14]系统地对扩散模型在图像复原任务中的应用进展进行总结,为复杂退化图像处理提供了新的研究范式.

与自然场景相比,交通监控图像在目标分布、结构形态和应用需求等方面具有明显差异.首先,交通场景中大量存在以车道线、道路边界和标志线为代表的细长型高频边缘结构,这些结构往往决定车辆的行驶路径与交通秩序,但在有雾条件下极易被模糊,若缺乏针对性增强,去雾后仍可能出现边缘断裂或模糊不清;其次,交通图像中包含数量有限的远景小目标,如远处的车辆、信号灯和交通标志,其像素尺度较小,受雾霾散射影响更严重,恢复难度显著高于近景目标.

鉴于现有 Transformer 类图像去雾模型在交通场景下仍存在边缘模糊、细节还原不足及训练不稳定等问题,本文基于交通图像中道路边缘清晰、结构规则、细节纹理密集的特点,建立一种融合 AO-Guided 特征增强机制的改进型 ViT 去雾网络结构.

首先,通过在输入端引入图像级结构增强模块,突出有雾图像中的高频响应区域;其次,在编码结构中构建AO-Guided滤波通道,引导主干网络自适应关注交通目标边缘与局部纹理;再次,在训练阶段引入中间监督机制,并融合结构相似性、感知一致性、边缘保持等多元损失约束,提升网络收敛稳定性与结构保持能力;最后,通过可见边集合数目比、平均梯度比与图像黑像素比例等结构敏感指标对不同去雾模型的性能进行定量对比分析,验证所提方法在交通图像清晰化处理任务中的有效性与实用性.

1 交通图像特征增强方法

1.1 融合型滤波特征增强算法

1.1.1 渐进最优极值滤波机理

渐进最优极值滤波(Asymptotically Optimal Max-Min Filter,AO-MMF)是一种通过前缀与后缀递推实现快速极值的计算方法,相比传统最大/最小滤波方法能显著降低运算复杂度[15].其基本形式为:对于输入图像$I(x,y)$,定义其在二维空间中的局部邻域窗口为$\Omega(x,y)$,该窗口内像素的最大值与最小值分别表示为

$$I_{\max}(x,y)=\max\{S_{\max}(x,y),P_{\max}(x,y)\} \quad (1)$$

$$I_{\min}(x,y)=\min\{S_{\min}(x,y),P_{\min}(x,y)\} \quad (2)$$

式中:(x,y)为图像中像素点的空间坐标;$S_{\max}(x,y)$与$S_{\min}(x,y)$为沿扫描方向通过前缀递推得到的局部最大值与最小值;$P_{\max}(x,y)$与$P_{\min}(x,y)$为通过后缀递推得到的局部最大值与最小值.

在随机输入情况下,AO-MMF算法的期望比较次数可达渐进最优值,每个输入像素仅需1.5次左右的比较[15],在高分辨率交通图像的实时去雾预处理中具备较强实用性.

1.1.2 引导滤波机理

引导滤波(Guided Filtering,GF)能够在保持边缘结构的同时实现平滑抑噪,可有效去除图像中的噪声,较好地保持图像原始的结构特征与边缘信息,适合道路图像的清晰度恢复与结构细节的增强任务[16].

引导滤波的核心思想是:在以像素点k为中心的局部邻域窗口ω_k内,利用局部线性模型对输出图像q与引导图像G之间的关系进行建模[17],其表达式为

$$q_i=a_kG_i+b_k,\ \forall i\in\omega_k \quad (3)$$

式中:q_i为输出图像在像素点i处的值;G_i为引导图像在像素点i处的值;a_k和b_k为局部窗口ω_k内待估计的线性系数,通过最小化残差平方和并引入正则项,可得到较为稳定的边缘保持结果.

1.1.3 算法设计

基于AO-MMF和GF滤波策略,本文提出一种融合型滤波特征增强算法,即构建具有结构自适应与尺度感知特性的AO-Guided融合滤波器,通过权重函数在GF与AO-MMF之间自适应切换,在保持图像关键边缘细节的同时提高整体计算效率.

设输入图像为$I_f\in\mathbb{R}^{H\times W\times C}$,融合滤波输出特征图为$F_{AO\text{-}GF}\in\mathbb{R}^{H\times W\times C}$,$H$,$W$,$C$分别为图像的高度、宽度和通道数.该算法基于如下耦合优化模型设计,表达式为

$$F_{AO-GF}(x,y)=\alpha(x,y)\cdot GF(I_f,I_g)+(1-\alpha(x,y))\cdot AO(I_f) \quad (4)$$

式中:(x,y)为图像空间中的像素坐标位置,$x\in[1,H]$,$y\in[1,W]$;$GF(I_f,I_g)$为以I_g为引导图像对输入图像I_f进行引导滤波,输出的边缘保持特征;$AO(I_f)$为在输入图像I_f上应用AO-MMF后的特征图;$\alpha(x,y)\in[0,1]$为自适应融合权重,控制两种滤波特征对最终融合特征的影响程度.其中$GF(I_f,I_g)$近似于对局部透射率场的平滑估计,强调在强结构边界处抑制噪声、保留几何轮廓.通过最大/最小滤波近似估计局部极值响应,及对雾浓度随空间变化引起的灰度极值偏移敏感性.

自适应融合权重建模表达式为

$$\alpha(x,y)=\frac{\exp(-\lambda\cdot\nabla(GF(I_f,I_g))(x,y))}{\exp(-\lambda\cdot\nabla(GF(I_f,I_g))(x,y))+\exp(-\lambda\cdot|R_{GF}(x,y)-R_{AO}(x,y)|} \quad (5)$$

式中:$\nabla(GF(I_f,I_g))(x,y)$为引导滤波输出特征在像素点(x,y)处的梯度幅值,用以刻画图像边缘结构信息;$R_{GF}(x,y)$,$R_{AO}(x,y)$分别为引导滤波与渐进最优极值滤波结果与输入图像I_f在像素点(x,y)处的残差;λ为控制融合偏向程度的尺度因子,设置为正值;指数函数用于增强边缘响应对融合权重的调控能力,使得边缘区域偏向GF,纹理区域偏向AO-MMF处理[18].

$\alpha(x,y)$中引入的残差项是引导滤波结果$R_{GF}(x,y)$与渐进最优极值滤波结果$R_{AO}(x,y)$相对于原始观

测在道路边缘、车道线、护栏等高对比度区域的物理偏移量.

引导滤波能够较好地贴合真实场景结构,其相对于原始观测的偏移量较小,即 $R_{GF}(x,y)$ 较小;AO-MMF 由于将局部灰度推向局部极值,其输出结果相对于原始观测的偏移量较大,即对应残差 $R_{AO}(x,y)$ 较大.当 $R_{GF}(x,y)<R_{AO}(x,y)$ 时,$\alpha(x,y)\to 1$,使融合结果偏向于结构保持能力更强的引导滤波结果.

在远景均匀雾场或纹理较弱区域,局部极大值或极小值更能反映雾霾厚度及大气光叠加效应,此时 $R_{AO}(x,y)$ 相对较小,当 $R_{AO}(x,y)<R_{GF}(x,y)$ 时,$\alpha(x,y)\to 0$,融合结果由 AO-MMF 主导,从而增强弱对比度区域的层次变化.

1.2　特征增强性能评估

实验选取 RESDIE 数据集中 250 幅不同场景内的交通图像[19]开展特征增强性能评估,每幅图像统一调整为 512×512 像素.分别采用导向滤波[15],最大/最小滤波[16],及 AO-Guided 滤波融合算法进行特征增强效果验证.交通近景和交通远景下不同算法的特征增强效果见图 1.



图 1　不同算法滤波后交通近景、远景特征增强效果对比图

Fig.1　Comparison of feature enhancement effects in near-field and far-field traffic scenes using different filtering algorithms

由图 1 可知:不同滤波方法在交通近景与远景场景中的特征增强效果存在差异;相比引导滤波与最大/最小滤波方法,AO-Guided 方法在边缘保持、细节恢复及远景清晰度提升方面表现更优.

为量化 AO-Guided 滤波模型在特征增强方面的有效性,本文选取图像平均梯度(Average Gradient, AG)、信息熵(Information Entropy, IE)、边缘像素数(Total Edge Number, TEN)及图像平均亮度(Mean Energy, ME)4 项具有代表性的图像质量评价指标进行对比评价.其中,AG 用于衡量图像整体清晰度与边缘锐利程度,是判断边缘细节强化效果的重要依据[20];IE 反映图像灰度分布的复杂程度,可评估图像纹理信息的丰富性[21];TEN 通过 Canny 边缘检测算子计算边缘像素总数,以量化图像结构复杂度[22];ME 则用于评估图像亮度提升效果,间接体现图像视觉质量的改善程度[23].同时引入算法运行时间(Processing Time, PT)作为计算效率评价指标.

不同算法评价指标的平均值见表 1.由表 1 可知:相较于原始图像、传统最大/最小滤波及引导滤波方法,AO-Guided 滤波融合算法在 AG、IE、TEN、ME 等多个评价指标上均取得最优或次优表现,体现出其在边缘细节保留、信息丰富性与图像亮度平衡方面的综合优势.特别是在保证结构清晰与纹理完整的前提下,AO-Guided 滤波融合算法有效提升了图像的判别性特征,为后续网络模型提供了更具学习价值的输入先验.

表 1　图像特征增强的质量评价指标

Tab.1　Quality evaluation metrics for image feature enhancement

滤波方法	AG	IE	TEN	ME	PT/ms
原始图像	5.722 6	7.248 6	12 400	131.61	18.01
最大滤波	6.891 2	6.932 1	13 584	137.84	20.15
最小滤波	4.124 8	6.314 5	9 823	125.77	19.52
引导滤波	6.431 7	7.501 2	14 203	132.90	26.34
AO-Guided	6.982 1	7.625 4	14 512	134.27	23.56

2　网络框架

2.1　图像输入增强模块设计

在实际的图像去雾任务中,输入图像中的雾气

会严重遮蔽道路边缘、交通标志、车道线等关键结构特征，从而影响深度网络对图像语义的理解与学习效果．为提升 Transformer 网络对雾化模糊图像中边缘与结构信息的感知能力，本文在网络输入阶段引入一种融合增强策略，将 AO-Guided 融合输出的特征图像 $F_{\text{AO-GF}}$ 与原始输入图像 I_{f} 进行双通道拼接输入，构成增强特征输入张量 X_{input}，其表达式为

$$X_{\text{input}}=\text{Concat}\left(I_{\text{f}},F_{\text{AO-GF}}\right)\in\mathbb{R}^{H\times W\times 2C} \tag{6}$$

式中：Concat(·)为特征拼接算子．

$F_{\text{AO-GF}}$ 包含更多边缘响应与纹理细节，具备更强的可学习性．在预处理阶段对 I_{f} 和 $F_{\text{AO-GF}}$ 进行归一化处理后拼接输入，可有效提升 Transformer 网络对浅层空间特征的感知能力，缓解雾霾影响下的信息退化问题．

为了在网络初始阶段实现低级特征融合与压缩，本文设计了一个卷积特征压缩层，压缩融合后的增强特征图 F_0 为

$$F_0=\text{ReLU}\left(\text{BN}\left(K_{3\times 3}\bullet X_{\text{input}}+b\right)\right) \tag{7}$$

式中：$K_{3\times 3}\in\mathbb{R}^{3\times 3\times 2C\times C}$ 为卷积核权重参数；ReLU(·)为线性整流激活函数；BN(·)为批量归一化操作；b 为偏置项，用于对卷积输出进行线性平移．

此外，为了增强模型在初始阶段对雾化区域的敏感性，引入残差连接以保持原始图像的细节，经过残差增强后的输入特征图 F_{enhanced} 为

$$F_{\text{enhanced}}=F_0+P_{\text{s}}\bullet I_{\text{f}} \tag{8}$$

式中：$P_{\text{s}}\in\mathbb{R}^{1\times 1\times C\times C}$ 为残差通道映射权重，实现特征维度对齐与通道重构．

图像输入增强模块结构见图 2．由图 2 可知，该模块通过通道拼接、卷积压缩与残差融合，实现多源特征的有效融合与增强，可提升输入特征的表达能力．

2.2 特征增强模块设计

由于 Transformer 对局部纹理细节建模能力较弱，难以准确恢复雾天图像中模糊边缘和退化区域．为进一步提升 Transformer 网络中间特征的表达能力，在编码器路径中引入 AO-Guided 特征增强模块，作为可插拔式增强结构，用于对中间特征进行局部与全局感知的融合优化．该模块以 AO-Guided 滤波模型为基础，将其融合机制集成于主干网络，实现对雾天图像中边缘信息、暗通道响应和纹理结构的精准提取．

将 AO-Guided 特征增强模块嵌入于 ViT 网络

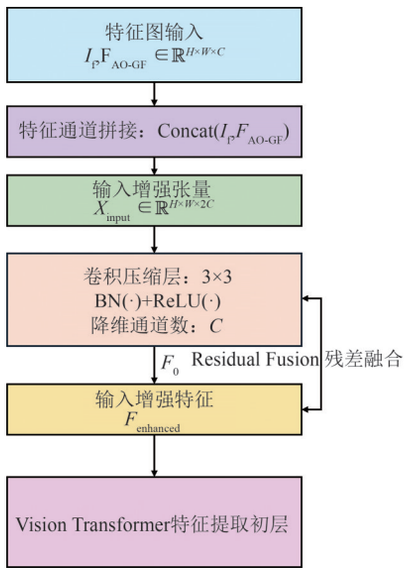


图 2 图像输入模块结构

Fig.2 Structure of the image input module

主干中各 DehazeFormer Block 之前，用于实现特征通道级别的自适应增强．该模块以 Transformer 编码器阶段的中间特征图 $F_{\text{in}}\in\mathbb{R}^{H\times W\times C}$ 作为输入，输出增强后的中间特征图 F_{out}，并作为后续 Transformer 模块进行深层特征建模的输入．

首先，对输入的中间特征 F_{in} 同时采用两种滤波策略进行处理：一是自引导滤波策略，即以输入中间特征自身作为引导图像进行滤波处理；二是渐进最优极值滤波策略．得到中间特征图 F_{GF}，F_{AO} 分别为

$$F_{\text{GF}}=\text{GF}\left(F_{\text{in}},F_{\text{in}}\right) \tag{9}$$

$$F_{\text{AO}}=\text{AO}\left(F_{\text{in}}\right) \tag{10}$$

式中：GF(·)表示引导滤波算子；AO(·)表示渐进最优极值滤波算子．

其次，考虑到不同区域在图像中对两类 GF 与 AO-MMF 两类增强方式的敏感度存在差异，本文引入注意力加权融合机制构造最终的增强特征图 F_{out}，其表达式为

$$F_{\text{out}}\left(x,y\right)=\gamma\left(x,y\right)\bullet F_{\text{GF}}\left(x,y\right)+\beta\left(x,y\right)\bullet F_{\text{AO}}\left(x,y\right) \tag{11}$$

式中：$F_{\text{out}}\left(x,y\right)$ 表示在像素位置 $\left(x,y\right)$ 处的融合增强特征；$F_{\text{GF}}\left(x,y\right)$ 与 $F_{\text{AO}}\left(x,y\right)$ 分别表示像素位置 $\left(x,y\right)$ 处的引导滤波特征与渐进最优极值滤波特征；对应像素位置的融合系数 $\gamma\left(x,y\right)$ 与 $\beta\left(x,y\right)$ 由显著性映射函数计算得到，其中 $\gamma\left(x,y\right)$ 为引导滤波特征的融合系数，$\beta\left(x,y\right)$ 为渐进最优极值滤波特征的融合系数．

局部方差是对小窗口内灰度波动强度的度量,会受到场景结构变化和雾浓度非均匀性的共同影响.局部方差较大,说明这一位置既承载了关键几何信息,也反映出雾场梯度的急剧变化;在远景大面积天空、均匀雾帘或路面平坦区,局部方差较小,说明场景内容和雾密度都较为平滑.

为实现两支路输出的有效融合,引入局部方差显著性映射作为加权机制.在中间特征图 F_{in} 的每一个像素位置 (x,y) 上,设定一个以 (x,y) 为中心、大小为 $k\times k$ 的局部邻域窗口 $N(x,y)$,统计每个位置在该窗口内的局部方差,引导滤波特征的局部方差 $\delta^2_{\text{GF}}(x,y)$ 以及渐进最优机制滤波特征的局部方差 $\delta^2_{\text{AO}}(x,y)$ 分别为

$$\delta^2_{\text{GF}}(x,y)=\frac{1}{\left|N(x,y)\right|}\sum_{(i,j)\in N(x,y)}\left(F_{\text{GF}}(i,j)-\mu_{\text{GF}}(x,y)\right)^2 \tag{12}$$

$$\delta^2_{\text{AO}}(x,y)=\frac{1}{\left|N(x,y)\right|}\sum_{(i,j)\in N(x,y)}\left(F_{\text{AO}}(i,j)-\mu_{\text{AO}}(x,y)\right)^2 \tag{13}$$

式中:$\left|N(x,y)\right|$ 表示以 (x,y) 为中心的局部窗口 $N(x,y)$ 内的像素总数;$\mu_{\text{AO}}(x,y)$,$\mu_{\text{GF}}(x,y)$ 分别表示 $F_{\text{AO}}(x,y)$,$F_{\text{GF}}(x,y)$ 在窗口 $N(x,y)$ 内的局部均值.

则对应的引导滤波特征的融合系数为

$$\gamma(x,y)=\frac{\exp\left(\delta^2{}_{\text{GF}}(x,y)\right)}{\exp\left(\delta^2{}_{\text{GF}}(x,y)\right)+\exp\left(\delta^2{}_{\text{AO}}(x,y)\right)} \tag{14}$$

$$\beta(x,y)=1-\gamma(x,y) \tag{15}$$

$\gamma(x,y)$ 将“局部结构显著性”映射为两分支的选择偏好:在高方差区域,$\gamma(x,y)$ 趋向于增强 GF 分支的权重,使输出特征在物理上更侧重轮廓和边缘的保持;在低方差区域,$\gamma(x,y)$ 趋向 AO-MMF 分支,通过对局部极值的放大来拉伸灰度动态范围,提升远景纹理和弱对比结构的可见性;$\beta(x,y)$ 由 $\gamma(x,y)$ 补偿得到,用于控制 AO 分支的权重分配.

AO-Guided 特征增强模块的结构见图 3.图 3 中,该模块以中间特征图作为输入,首先通过引导滤波与渐进最优极值滤波两条分支提取互补特征,然后引入基于局部方差的显著性加权机制对取得的两类特征实现自适应融合,最终完成特征的有效增强与重构.

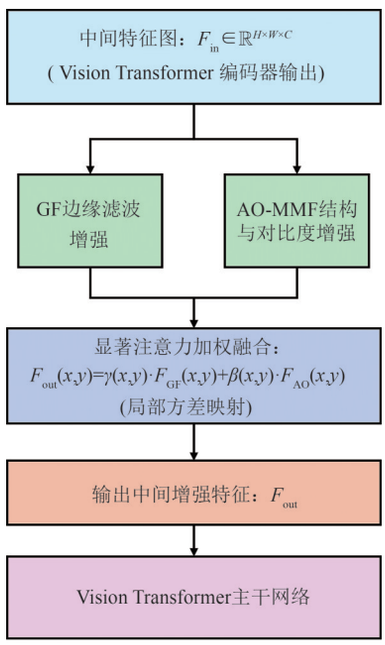


图 3　特征增强模块结构

Fig.3　Feature enhancement module structure

2.3　改进 ViT 图像去雾网络设计

为充分挖掘图像层面和特征层面的多维信息,提升 Transformer 模型对图像雾化区域的理解能力,在 ViT 去雾网络框架基础上,引入图像输入增强模块与 AO-Guided 特征增强模块,构建具备双重增强机制的图像去雾网络结构,见表 2.

由表 2 可知,模型以图像级增强输入模块为起点,通过卷积压缩完成特征初步融合,然后采用多层 Transformer 编码与解码结构进行层次化特征提取与重建,并结合下采样与上采样实现多尺度信息建模.在解码阶段引入 SK 注意力融合模块与 AO-Guided 增强模块,对特征进行自适应融合与强化,最终通过重建卷积层与软重建模块(Soft Reconstruction)输出最终结果.

2.3.1　网络架构

改进的 ViT 网络架构见图 4.图 4 中:L1~L5 表示网络不同深度下的多尺度特征层级;层级越深,特征分辨率越低、感受野越大,语义表达能力越强;浅层特征主要保留边缘和纹理信息,深层特征主要刻画全局结构与雾化退化特征.由图 4 可知,整个网络采用类 U-Net 的编码-解码结构,其中主干骨架使用 ViT 结构[9],并在输入阶段引入图像增强模块,在 Transformer 主干模块之间嵌入 AO-Guided 特征增强模块,从而实现端到端的多尺度图像去雾处理.

网络主要由 5 个模块组成.

模块 1:图像级增强模块.与传统单一通道输入

表 2 网络结构

Tab.2 Network structure

模块名称	类型	输入尺寸	输出尺寸
图像级增强输入模块	拼接＋融合	$H\times W\times 2C$	$H\times W\times 2C$
3×3卷积层	卷积	$H\times W\times 2C$	$H\times W\times C$
浅层特征提取阶段	DehazeFormer Block	$H\times W\times C$	$H\times W\times C$
特征下采样	下采样	$H\times W\times C$	$H/2\times W/2\times 2C$
Transformer编码模块	DehazeFormer Block	$H/2\times W/2\times 2C$	$H/2\times W/2\times 2C$
特征下采样	下采样	$H/2\times W/2\times 2C$	$H/4\times W/4\times 4C$
Transformer编码模块	DehazeFormer Block	$H/4\times W/4\times 4C$	$H/4\times W/4\times 4C$
Transformer解码模块	DehazeFormer Block	$H/2\times W/2\times 2C$	$H/2\times W/2\times 2C$
特征上采样	上采样	$H/4\times W/4\times 4C$	$H\times W\times C$
SK注意力融合模块	SK注意力融合	$H/2\times W/2\times 2C$	$H\times W\times C$
特征上采样	上采样	$H/2\times W/2\times 2C$	$H\times W\times C$
SK注意力融合模块	通道融合	$H\times W\times C$	$H\times W\times C$
AO-Guided增强模块	模块嵌入	$H\times W\times C$	$H\times W\times C$
Transformer解码模块	DehazeFormer Block	$H\times W\times C$	$H\times W\times C$
重建卷积层	卷积	$H\times W\times C$	$H\times W\times C$
Soft Reconstruction模块	非线性重建	$H\times W\times C$	$H\times W\times C$

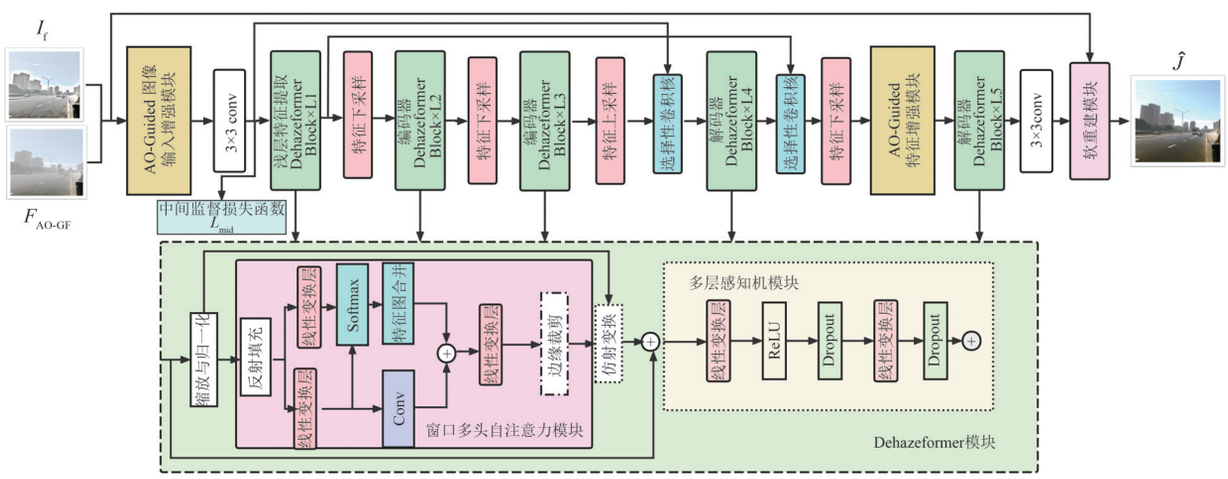


图 4 改进的 ViT 网络架构

Fig.4 Improved Vision Transformer network architecture

方式不同，本文首先将原始雾图 $I_f\in\mathbb{R}^{H\times W\times C}$ 与由 AO-Guided 图像增强算法得到的融合特征图 $F_{AO-GF}\in\mathbb{R}^{H\times W\times C}$，进行通道级拼接，得到双通道输入特征 $X_{input}\in\mathbb{R}^{H\times W\times 2C}$. 将该融合特征作为输入经过一个 3×3 卷积操作映射至标准通道维度 C，以匹配主干网络的输入要求，经残差连接计算，得到输入增强特征 $F_{enhanced}$，然后，网络进入浅层特征提取阶段与多尺度编码器处理流程.

模块 2：编码器模块. 编码器部分由三组 DehazeFormer Block 和两个下采样模块构成，通过逐步将特征图分辨率由原始尺寸下采样至 $\frac{W}{4}\times\frac{C}{4}$，实现图像特征压缩. 每个 DehazeFormer Block 包含窗口多头自注意力和多层感知机（Multi-Layer Perceptron，MLP）子结构. 窗口多头自注意力机制（Window Multi-Head Self-Attention，W-MHSA）的表达式为

$$W-\mathrm{MHSA}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=\mathrm{Softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d}}\right)\boldsymbol{V} \quad (16)$$

式中：$W-\mathrm{MHSA}(\cdot)$ 为窗口多头自注意力机制；$\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}$ 分别为查询（Query）、键（Key）和值（Value）矩阵，由输入特征通过线性映射得到；$\mathrm{Softmax}(\cdot)$ 为归一化指数函数，用于计算注意力权重分布；d 为特征向量的维度.

为进一步增强特征的非线性表达能力，将 W-MHSA 输出的中间特征图 F_{in} 输入 MLP 模块，通过线性映射、非线性激活和再次线性映射完成特征变

换,其表达式为

$$\mathrm{MLP}(F_{\mathrm{in}})=\mathrm{Linear}(\mathrm{ReLU}(\mathrm{Linear}(F_{\mathrm{in}}))) \tag{17}$$

式中:MLP(·)为多层感知机;Linear(·)为线性映射操作.

模块3:AO-Guided特征增强模块.AO-Guided特征增强模块在解码阶段的中间特征流中进行嵌入,其输入为当前层的中间特征图F_{in},经引导滤波与渐进最优极值滤波双分支处理后,分别得到结构增强特征与对比度增强特征,并通过基于局部方差的注意力加权机制进行自适应融合,生成特征增强图F_{out}.将该增强特征输入至后续ViT主干网络层,以实现对解码过程中细节信息与边缘结构的持续强化.

模块4:解码器模块与SK融合机制.解码器结构与编码器完全对称,包含三组上采样模块与Transformer解码模块.解码阶段通过选择性卷积核(Selective Kernel, SK)融合模块对不同尺度特征进行自适应选择,得到融合多尺度特征F_{SK}为

$$F_{\mathrm{SK}}=\sum_{i=1}^{n}a_i\cdot F_i \tag{18}$$

$$\sum_{i=1}^{n}a_i=1 \tag{19}$$

式中:a_i为通过注意力机制自适应学习得到的第i个尺度特征的融合权重;F_i为第i个尺度的输入特征图;n为不同尺度的分支数量.

模块5:输出重建模块.特征解码后通过1×1卷积恢复图像空间通道,后续采用软重建模块(Soft Reconstruction)进行非线性映射,提升去雾图像细节表现,最终输出的无雾图像$\hat{J}$为

$$\hat{J}=\mathrm{Sigmoid}(\mathrm{ReLU}(\mathrm{Conv}_{1\times1}(F_{\mathrm{out}}))) \tag{20}$$

式中:Sigmoid(·)为非线性激活函数;$\mathrm{Conv}_{1\times1}(\cdot)$为卷积操作.

2.3.2　损失函数

在图像去雾任务中,损失函数被用来衡量去雾后的图像与真实无雾图像间的差异,从而优化模型训练过程.

为优化所提出的基于AO-Guided模块集成的图像去雾网络,构建联合损失函数进行网络训练,以提升去雾图像在细节保留、结构保持与感知质量等方面的综合表现,联合损失函数L_{total}为

$$L_{\mathrm{total}}=\lambda_1L_1+\lambda_2L_{\mathrm{SSIM}}+\lambda_3L_{\mathrm{Percep}}+\lambda_4L_{\mathrm{Edge}}+\lambda_5L_{\mathrm{mid}} \tag{21}$$

式中:λ_m为各损失项的损失系数,$m=1,2,\cdots,5$;L_1,L_{SSIM},L_{Percep},L_{Edge},L_{mid}为各损失项.

L_1为重建损失,用于最小化网络最终输出的无雾图像$\hat{J}$与真实无雾图像J之间的像素级绝对误差,表达式为

$$L_1=\frac{1}{HW}\sum_{x=1}^{H}\sum_{y=1}^{W}\left|\hat{J}(x,y)-J(x,y)\right| \tag{22}$$

式中:H,W分别为输入图像的高度和宽度;$\hat{J}(x,y)$,$J(x,y)$分别为由网络输出的无雾图像与真实无雾图像在像素位置(x,y)处的像素值.

结构相似性指数(Structural Similarity Index, SSIM)用于衡量去雾图像与真实无雾图像在亮度、对比度及结构信息方面的相似性,取值范围为$[0,1]$,其值越大表示去雾图像与真实无雾图像在结构上越接近.由于SSIM本质上为相似性度量,而模型训练过程以最小化损失函数为目标,因此构造基于SSIM的损失函数L_{SSIM},其表达式为

$$L_{\mathrm{SSIM}}=1-\mathrm{SSIM}(\hat{J},J) \tag{23}$$

式中:$\mathrm{SSIM}(\hat{J},J)$表示去雾图像$\hat{J}$与真实无雾图像J的结构相似性指标值.

L_{Percep}为感知损失,其基于预训练的VGG网络特征层提取结果,衡量预测图像与真实图像在高层语义特征空间的差异,表达式为

$$L_{\mathrm{Percep}}=\sum_{l\in\Lambda}\left\|\phi_l(\hat{J})-\phi_l(J)\right\|_2^2 \tag{24}$$

式中:$\phi_l(\cdot)$表示第l层的特征提取器;Λ表示选定的特征层集合.

L_{Edge}为边缘保持损失,通过拉普拉斯算子提取图像边缘,确保去雾过程中图像细节得以保持,其损失函数表达式为

$$L_{\mathrm{Edge}}=\left\|\nabla^2\hat{J}-\nabla^2J\right\|_1 \tag{25}$$

式中:∇^2为拉普拉斯算子.

此外,考虑到输入阶段引入了融合图像$F_{\mathrm{GO\text{-}GF}}$,故在训练中加入中间监督机制,在输入增强模块与浅层特征提取之间的增强输出F_{enhanced}处,引入一个轻量级的辅助预测头,该分支采用1×1卷积层和像素重排上采样结构(PixelShuffle),将浅层特征上采样至图像分辨率并计算与真实无雾图像J的差异,对中间特征图进行引导训练,定义其损失函数为

$$L_{\mathrm{mid}}=\left\|J_{\mathrm{mid}}-J\right\|_1 \tag{26}$$

式中:J_{mid}为由中间特征通过辅助重建分支得到的预测图像.训练时,L_{mid}与主输出损失一并构成联合优化目标,推理时该分支关闭,不参与前向计算.

3 实验设置与结果对比

3.1 实验设置

3.1.1 数据集选取与构建

本文研究场景为道路交通图像，故基于公开的 RESIDE 图像去雾数据集[19]，将是否包含道路、车道线、车辆或交通标志等交通语义特征作为筛选标准，筛选出 1 880 张道路交通场景的清晰图像，基于大气散射模型，使用不同的散射系数和全局大气光值对每张清晰图像生成 15 张不同的雾霾图像，合成流程见图 5. 具体合成分为 6 个步骤.

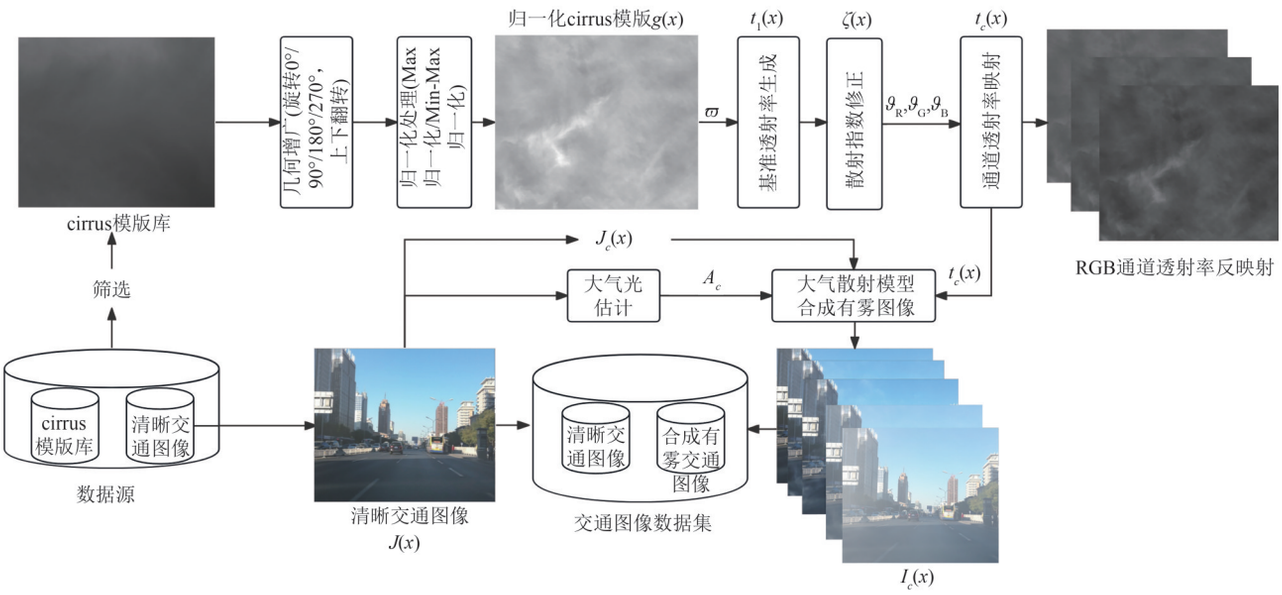


图 5 交通图像雾化合成方法流程

Fig.5 Flowchart of the haze synthesis method for traffic images

步骤 1：构建大气散射成像模型. 模型合成过程遵循经典大气散射模型[24]，其表达式为

$$I_c(x)=J(x)\cdot t(x)+A\cdot(1-t(x)) \tag{27}$$

式中：$J(x)$为清晰图像；$I_c(x)$为合成有雾图像；$A=(A_R,A_G,A_B)$为各通道大气光分量，其中 R、G、B 分别对应红、绿、蓝 3 个颜色通道；$t(x)\in(0,1]$为透射率.

步骤 2：构建雾场模版并进行数据增广. 选取卷云灰度影像(cirrus)作为雾场纹理. 经 4 种旋转角度(0°/90°/180°/270°)与上下翻转组合形成 8 种几何增广，并缩放至与清晰图相同尺寸. 模板经最大值归一化与 Min-Max 归一化两种方式标准化，以增加多样性.

步骤 3：生成基准透射率并进行雾强采样. 以归一化模版$g(x)$构造基准透射率$t_1(x)$，其表达式为

$$t_1(x)=1-\varpi\cdot g(x) \tag{28}$$

式中：$\varpi\in[0.05,1]$为全局雾强控制系数，用于调节基准透射率的衰减程度，ϖ取值越大，$t_1(x)$越小，图像雾化程度越强. 为更贴近实际中雾霾常见分布，采用对数均匀采样生成ϖ为

$$\begin{cases}\varpi=\exp(u)\\ u\sim U(\ln 0.05,\ln 1)\end{cases} \tag{29}$$

步骤 4：引入散射指数并进行非均匀性修正. 考虑不同能见度下的散射差异，对每个像素引入散射指数$\zeta(x)$为

$$\zeta(x)=\mathrm{clip}\left(p_3\tilde{t}^3+p_2\tilde{t}^2+p_1\tilde{t}+p_0,0,4\right) \tag{30}$$

$$\tilde{t}=1-t_1(x) \tag{31}$$

式中：clip(·,0,4)为将ζ限制在[0,4]；$(p_3,p_2,p_1,p_0)=(-21.547,41.224,-27.465,6.537)$.

步骤 5：构建多通道透射率映射关系. 以蓝通道中心波长为基准，$(\vartheta_R,\vartheta_G,\vartheta_B)=(0.655\,0,0.562\,5,0.482\,5)$，分别表示 R、G、B 三通道的中心波长，μm. 将t_1映射为各通道透射率$t_c(x)$，其表达式为

$$t_c(x)=\exp\left(\left(\frac{\vartheta_B}{\vartheta_c}\right)^{\zeta(x)}\cdot\ln t_1(x)\right)\quad c\in\{R,G,B\} \tag{32}$$

步骤 6：估计大气光参数. 大气光通过清晰图像的高亮像素统计得到. 对每个通道，从像素强度分布中选取最亮的 0.1% 像素并取均值作为A_c，其表达式为

$$A_c=E\left[J_c(x)\,\middle|\,J_c(x)\in\mathrm{Top}\,0.1\%\right] \tag{33}$$

式中：$J_c(x)$表示清晰图像在通道$c\in\{R,G,B\}$下像素位置x的像素值；$E[\cdot]$为期望算子，对选取的高

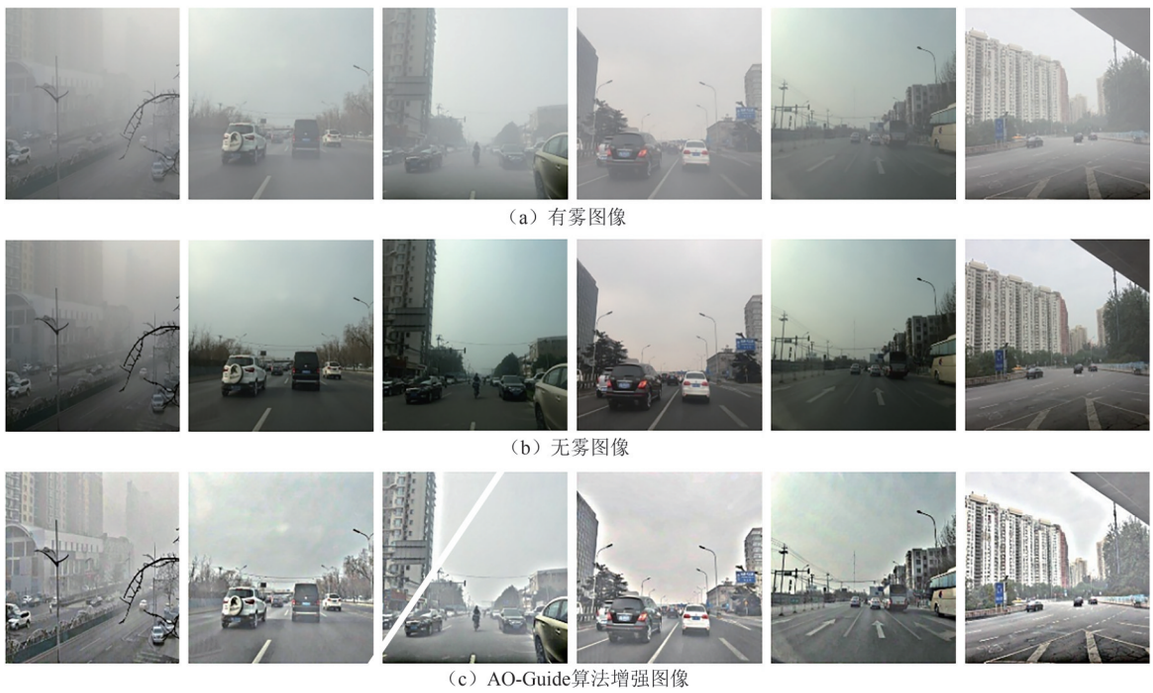
亮像素集合进行均值计算.

步骤 7:生成合成有雾图像.将 $t_c(x)$ 与 A_c 代入大气散射模型生成合成雾图,其表达式为

$$I_c(x)=J_c(x)\cdot t_c(x)+A_c\cdot(1-t_c(x)) \quad (34)$$

每幅清晰交通图像均对应生成 15 幅具有不同退化水平与空间分布特征的有雾图像,共 28 200 张组成数据集,本文以清晰图像的序列号为单位,确保同一图像的多退化版本仅出现在训练集或测试集一侧,按照 8∶2 划分为模型的训练集与测试集.

对数据集进行预处理:将图像统一调整为 512×512 像素,并将像素值归一化至[0,1]区间,同时训练阶段的输入为合成雾图与 AO-Guided 预处理增强图的通道拼接图像,监督标签为对应的清晰图像,部分训练输入图像见图 6.由图 6 可知,合成雾图在不同退化程度与空间分布下呈现出明显的能见度降低与对比度衰减,而经 AO-Guided 增强后,图像在边缘细节、结构信息及整体清晰度方面均得到有效改善.



(a) 有雾图像

(b) 无雾图像

(c) AO-Guide算法增强图像

图 6　无雾图像、合成雾图与算法增强图像

Fig.6　Haze-free images, synthetic hazy images, and algorithm-enhanced images

3.1.2　实验环境与训练参数设置

在网络训练阶段,使用的实验环境配置见表 3.训练过程中统一采用 Adam 优化器,以初始学习率 1×10^{-4} 进行训练,为防止训练过程中模型出现过拟合现象,设置批处理大小为 16,训练轮次为 8 个 Epoch,采用 Early Stopping 机制,当验证集连续损失 5 个 Epoch 且未改善时,提前终止训练过程.

在损失函数整体构成中,各项权重参数分别对应不同的优化目标:λ_1 主要用于约束像素级重建精度,λ_2 用于保持结构相似性,λ_3 用于增强感知质量,λ_4 用于强化边缘细节特征,λ_5 则作为中间监督项以引导浅层特征的学习.为获得合理的权重配置,本文通过多组候选权重组合进行对比实验,最终选取综合性能最优的一组参数作为模型的权重设置,其中,λ_1、λ_2、λ_3、λ_4 分别取 0.3、0.4、0.01、0.05.λ_5 在训练过程中采用由 0.5 衰减至 0.1 的余弦衰减策略[25],在训练前期增强模块输出对边缘与结构信息的刻画能力,加速模型收敛;在训练后期逐渐减弱约束,以保证整体去雾效果的优化主要依赖最终输出分支,从而在细节保持与全局一致性之间取得平衡.

表 3　实验环境设置

Tab.3　Experimental environment setting

环境	名称	参数
硬件环境	操作系统	Ubuntu 18
	CPU	13th Gen Intel(R)@2.20 GHz
	GPU	NVIDIA GeForce RTX 4060
	显存	16 GB
软件环境	CUDA	Cuda 12.7
	学习框架	PyTorch 2.3.1
	开发环境	Python 3.11

为验证模型在交通实际场景下的可行性,本文对比评估了 Baseline、Baseline+AO-Guided 以及本文最终模型在 RESIDE 交通场景子集上的参数量 Params、乘加运算次数(Multiply-Accumulate Opera-

tions, MACs)、浮点运算次数(Floating Point Operations, FLOPs)、推理时延和每秒处理帧数(Frames Per Second, FPS).其中,Params表示模型中可学习参数的规模;MACs表示模型完成一次前向推理所需的乘加运算次数;FLOPs表示模型推理过程中的浮点运算总量;推理时延表示单张图像完成去雾处理所需的平均时间;FPS表示模型单位时间内可处理的图像帧数.实验环境见表3,结果见表4.

表4 各模型在RESIDE交通场景子集上的推理效率对比

Tab.4 Comparison of inference efficiency across models on the RESIDE traffic scenario subset

模型结构	分辨率/像素	Params/M	MACs/次	FLOPs	时延/ms	FPS
Baseline	512×512	0.686	26.632	53.264	15.8	63.3
Baseline+AO-Guided	512×512	0.704	28.740	57.480	16.9	59.2
改进ViT网络的交通图像去雾方法	512×512	0.705	28.910	57.820	17.0	58.8

表4中:Baseline为DehazeFormer-T[9]骨干网络;“Baseline+AO-Guided模块”为在Baseline的编码层中嵌入AO-Guided特征增强模块;改进ViT网络去雾模型则是在嵌入AO-Guided特征增强模块的基础上,进一步增加输入增强通道拼接与训练期中间监督(推理期关闭).

由表4可知:Baseline的参数量为0.686 M,FLOPs为53.264 G,对应单帧时延15.8 ms,帧率约63 FPS;在引入AO-Guided模块后,参数量增至0.704 M,FLOPs增加至57.480 G,时延上升至16.9 ms,但帧率仍保持在59 FPS以上,计算负担可控,验证了该模块在性能提升与开销增加之间的平衡性;Proposed模型在此基础上引入图像级增强通道拼接与训练期中间监督,参数量为0.705 M,FLOPs仅增加了0.6%,时延变化极小,帧率约59 FPS,其推理性能与含AO-Guided的配置几乎一致.由此可知,本文方法能维持约60FPS的处理能力,满足交通场景的实时处理需求.

本文记录了训练过程中的峰值信噪比(Peak Signal-to-Noise Ratio, PSNR)与结构相似性指数(Structural Similarity Index, SSIM)两个指标的迭代变化情况见图7.

图7中,横轴为评估次数,模型在训练过程中以5 000次迭代为间隔进行性能评估.由图7可知:训练开始阶段,模型性能快速提升,体现为PSNR与SSIM指标迅速上升,表明网络能够快速学习到初始的图像去雾特征;随着训练的进行,两项指标进入相对平稳且持续缓慢上升的阶段,反映出网络逐步收敛并逐渐逼近性能最优点;在约第94次评估后,两项指标趋于稳定,PSNR稳定在36.55 dB左右,SSIM稳定在0.990 5左右.结果表明,本文提出的融合AO-Guided滤波模块的改进ViT网络具有良好的收敛性能,能够有效且稳定地提升雾化图像的去雾质量.

基于改进的ViT网络模型训练完成后,部分测试集的去雾效果图和深度图见图8.由图8可知:所提方法能够有效提升图像对比度与细节清晰度,并在复杂场景下保持良好的结构一致性;同时,深度分布结果进一步验证了模型对空间层次信息的恢复能力.

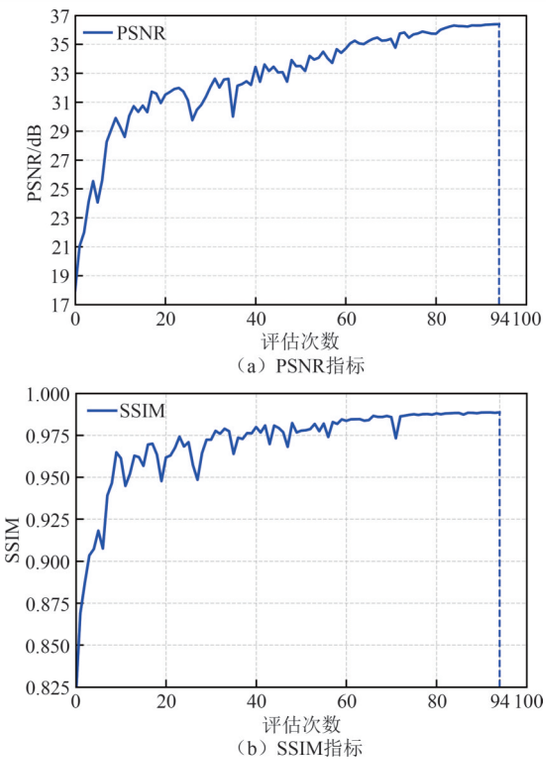


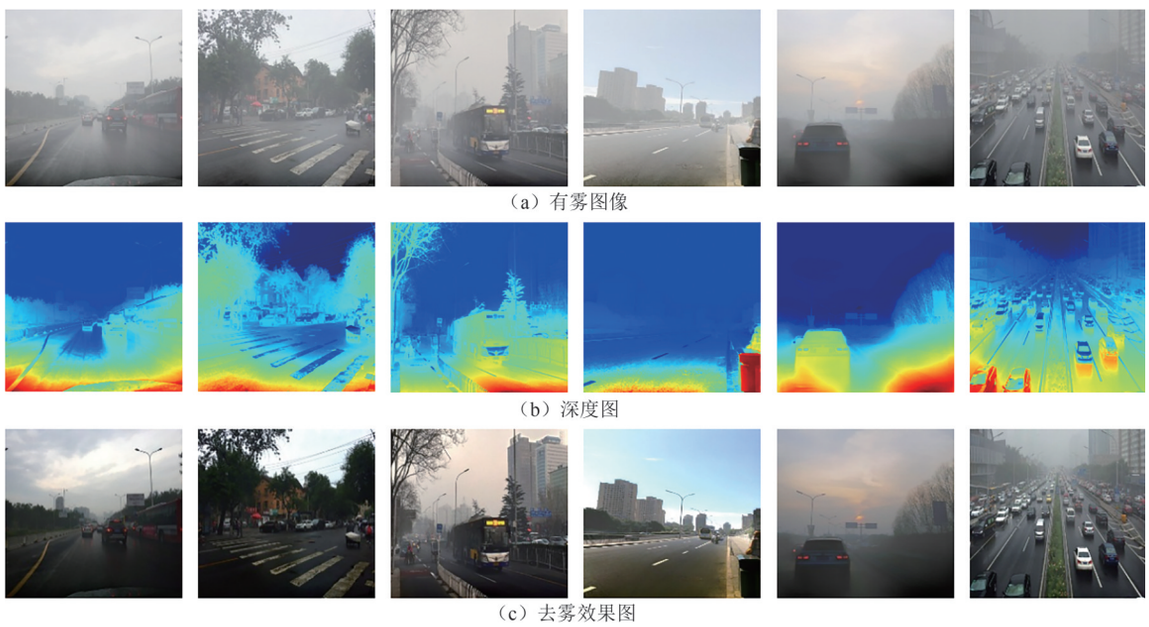
图7 PSNR和SSIM指标迭代曲线

Fig.7 Iterative curves of PSNR and SSIM indicators

3.2 对比实验结果

为验证基于改进ViT网络的图像去雾方法在交通场景中的效果,本文设计了与经典及当前先进去雾方法的对比实验.包括传统经典方法DCP[3]、经典深度学习方法DehazeNet[5],以及近年来先进的DehazeFormer[9]、GCANet[26]等主流深度去雾网络.所有参与对比实验的方法均统一使用本文所构建的实验数据集进行测试,并采用文献[3,5,9,26]中的最优参数配置,使用HSTS数据集中的300张真实交通有雾图像进行去雾效果对比.

对比实验中,针对交通图像去雾任务的特殊需



（a）有雾图像

（b）深度图

（c）去雾效果图

图 8　改进 ViT 网络的去雾效果

Fig.8　Improved ViT network dehazing effect diagram

求,引入文献[27]提出的可见边梯度法,采用原有雾图像与去雾图像的新可见边集合数目比(e)、平均梯度比($\bar{r}$)、黑色像素点所占图像大小百分比(δ)3个指标,以及文献[28]提出的雾感知密度评价指标(Fog Aware Density Evaluator, FADE),从不同角度客观评价图像的去雾效果.

e,$\bar{r}$,δ的表达式分别为

$$e=\frac{n_r-n_0}{n_0} \tag{35}$$

$$\bar{r}=\frac{\bar{g}_r}{\bar{g}_0} \tag{36}$$

$$\delta=\frac{n_s}{H\times W} \tag{37}$$

式中:n_0和n_r分别为原有雾图像和去雾复原图像中可见边的数目;$\bar{g}_r$为去雾图像的平均梯度;$\bar{g}_0$为原有雾图像的平均梯度;n_s为koheler图像分割方法后变为黑色像素点的像素数目.

FADE指标表达式为

$$D=\frac{D_f}{D_{ff}+1} \tag{38}$$

式中:D_f为去雾图像与无雾参考模型之间的马氏类距离;D_{ff}为去雾图像与有雾参考模型之间的马氏类距离.FADE值越小,表示图像中残余雾气越少,远景目标的可见性越好.

不同算法在真实有雾图像下的去雾效果及其对应的深度估计结果见图9.

由图9可知:①各算法在细节还原、亮度保持及边缘一致性等方面仍存在较大差异.DCP算法虽然能够恢复部分图像细节,但在天空区域易产生明显的色彩失真与边缘光晕现象,导致整体观感较差;GCANet在处理浓雾区域时出现局部团雾残留,影响图像的清晰度和一致性;DehazeNet虽然能有效去除浅层雾气,但存在图像亮度偏暗、对远景恢复能力不足的问题;DehazeFormer作为具有代表性的Vision Transformer去雾网络,在全局对比度恢复上较强,但在交通场景中对细长结构(如道路线、护栏)存在局部过度增强与边缘不连续的现象;②本文所提出的AO-Guided滤波与改进ViT网络的去雾方法在图像亮度方面,能够有效恢复图像的光照层次,避免整体亮度过暗或残余雾气未完全消除的情况;在纹理与边缘结构还原方面,道路线条、车道边缘得视觉可读性较好;在远近景的分层结构上表现出较好的连续性与稳定性,但在远距离小目标轮廓恢复上仍存在不足.

在HSTS数据集300张真实交通有雾图像上进行去雾实验后,得到不同算法对应的e、$\bar{r}$、δ以及FADE 4个指标的平均值,结果见表5.

由表5可知:①与GCANet、DehazeNet等传统网络相比,改进ViT网络的交通图像去雾方法不仅在细节还原上具备更强的结构保持能力,还有效避免了亮度不足、团雾残留等现象;相较DCP等基于先验的方法,可显著提升图像的整体自然性与视觉一致性;与DehazeFormer网络相比,改进ViT网络的去雾模型在交通场景的细长边缘与车辆可辨性方面表现更为稳定,在图9的定性结果中展现了更清

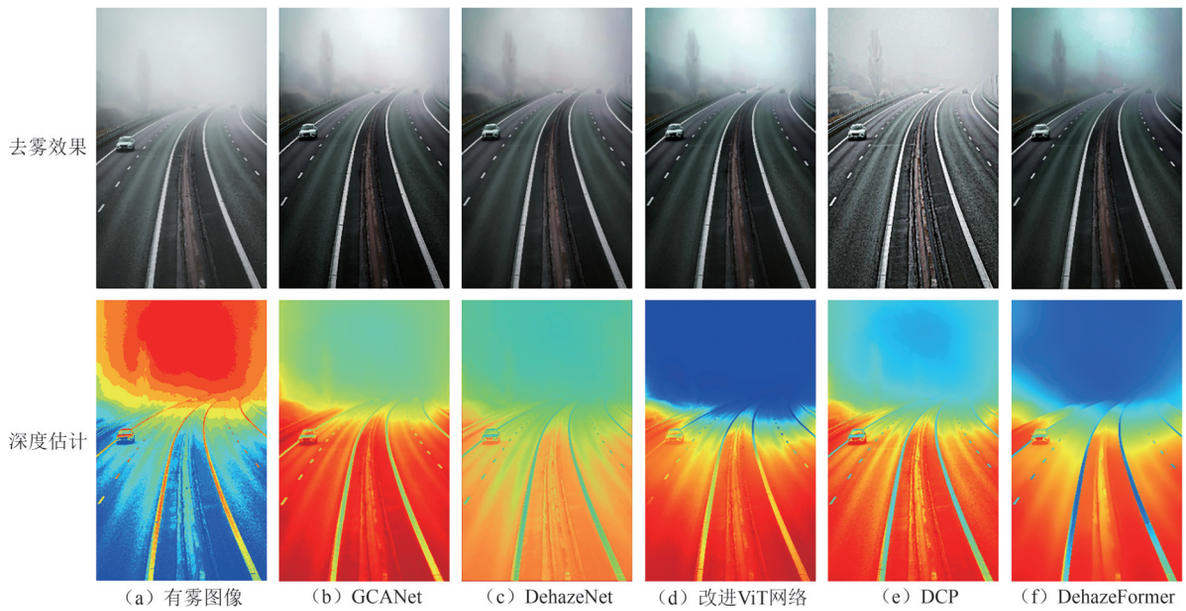


图 9 不同算法去雾结果对比

Fig.9 Comparison of dehazing results for different algorithms

表 5 主要算法在真实雾图上的量化指标值

Tab.5 Quantitative metrics of the main algorithms on real-world hazy images

去雾方法	e	$\bar{r}$	δ/%	FADE
GCANet	0.98	1.95	0.78	0.489 4
DehazeNet	1.03	1.81	1.02	0.543 8
DCP	0.34	1.26	3.38	0.958 9
DehazeFormer	1.28	1.31	0.48	0.374 6
改进 ViT 网络的交通图像去雾方法	1.45	1.84	0.36	0.420 8

晰的车辆轮廓与交通标线.②新可见边集合数目比 e 从 1.28 提升至 1.44，提升幅度为 12.5%，说明本文模型在捕捉雾天图像中的边缘结构特征方面表现更优；平均梯度比从 1.31 提升至 1.84，提升约 40.5%，显著增强了图像的清晰度与纹理还原能力；黑色像素占比从 0.48% 降低至 0.36%，减少了图像空洞区域和信息丢失现象.改进 ViT 网络的交通图像去雾方法在平均梯度比上虽优于 DehazeFormer 与 DCP，但略低于 GCANet 的 1.95.这是由于 GCANet 依赖局部卷积堆叠来强化梯度响应，对纹理锐化更具针对性；而本文模型注重边缘结构性与整体一致性，因此，在单一梯度指标上低于 GCANet，但在亮度、结构与视觉自然性等综合指标上表现更为均衡.③在 FADE 指标上，改进 ViT 网络的交通图像去雾方法的得分为 0.420 8，高于 DehazeFormer 的 0.374 6，表明其在全局雾密度抑制方面表现相对不足.由于 FADE 对远景残雾和小尺度目标具有较高敏感性，该差异主要归因于极端浓雾条件下远处车辆、交通标志等目标的有效像素尺度有限，其结构信息在强散射作用下被严重压缩.同时，当前损失函数未对远景区域施加显式权重，使得模型在高雾浓度背景下难以充分恢复远景小目标的精细轮廓，表现为边界锐度降低与对比度不足.尽管如此，改进 ViT 网络的交通图像去雾方法在边缘连续性和交通要素几何可辨性方面依然保持更强稳定性，因而在整体交通场景可读性上具备更强的适用性.

综上，实验结果验证了所提出改进策略在提升图像质量和增强交通场景可辨性方面的有效性与合理性.

3.3 消融实验分析

3.3.1 增强模块及分支作用分析

为评估图像输入与特征增强模块对整体去雾性能的贡献，在所构建的 RESIDE 交通场景子集上分别针对图像输入增强模块、AO-Guided 特征增强模块以及特征增强模块的两条分支（GF/AO-MMF）进行消融实验.

输入增强与特征增强模块的 4 种网络变体的量化指标统计见表 6.表 6 中："＋输入增强模块"表示在 Baseline 基础上引入输入增强模块；"＋特征增强模块"表示在 Baseline 基础上引入特征增强模块.

由表 6 可知：与 Baseline 模型相比，加入图像输入增强模块后，e、$\bar{r}$、δ、SSIM 和 PSNR 分别得到提升；边缘结构指标 e 从 0.85 提高到 1.15，提升幅度达到 35.3%，表明该模块有效增强了网络在初始阶段对图像结构特征与边缘细节的感知与捕获能力；当单独引入 AO-Guided 特征增强模块时，PSNR 提升至 34.88 dB，结构相 SSIM 达到 0.978，

表 6　不同模块组合配置下的消融实验结果

Tab.6　Ablation study results under different module combination configurations

模型变体	＋输入增强模块	＋特征增强模块	e	$\bar{r}$	δ/%	SSIM	PSNR/dB
Baseline	—	—	0.85	1.43	1.02	0.956	31.26
Baseline	√	—	1.15	1.60	0.68	0.965	33.72
Baseline	—	√	1.02	1.55	0.57	0.978	34.88
Baseline	√	√	1.45	1.84	0.36	0.989	36.20

表明了 AO-Guided 模块在模型中间特征优化与细节补偿上的重要作用;本文模型在融合两个模块后,e、$\bar{r}$、δ、SSIM 和 PSNR 分别达到 1.45、1.84、0.36%、0.989 和 36.20 dB,性能表现最优.此结果证明了图像级和特征级的增强策略之间存在明显的协同效应.

在固定“Baseline＋输入增强模块”条件下,仅改变 AO-Guided 模块内部两条分支(GF/AO-MMF)设置进行消融实验,结果见表 7.

表 7　AO-Guided 模块不同分支配置下的消融实验结果

Tab.7　Ablation study results under different branch configurations of the AO-Guided module

GF 分支	AO－MMF 分支	e	$\bar{r}$	δ/%	FADE
—	—	1.15	1.60	0.68	0.51
√	—	1.24	1.66	0.51	0.44
—	√	1.35	1.77	0.76	0.55
√	√	1.46	1.84	0.36	0.41

表 7 中:“GF 分支”表示是否启用引导滤波分支;“AO-MMF 分支”表示是否启用渐进最优极值滤波分支;“√”表示对应分支被启用.

由表 7 可知:仅启用 GF 分支时,e 和 $\bar{r}$ 分别提升至 1.24 和 1.66,δ 降至 0.51%,FADE 为 0.44,体现了其在边缘结构保持方面的优势;仅启用 AO-MMF 分支时,e 和 $\bar{r}$ 进一步提升至 1.35 和 1.77,但 δ 和 FADE 分别上升至 0.76% 和 0.55,说明该分支在结构锐化方面表现突出,而在远景去雾和残余雾气抑制方面的效果相对有限;同时启用两条分支时,模型在各项指标上均达到最优,在边缘连续性、纹理恢复与全局一致性之间实现了更均衡的性能.

综上,输入增强与特征增强模块在整体上有效提升了去雾性能,而 GF 与 AO-MMF 分支在特征细粒度建模层面形成互补,从而在边缘保持与全局一致性之间实现了模型性能均衡.

3.3.2　损失函数权重敏感性分析

为评估联合损失函数中各组成项对模型性能的影响,在保持训练数据集与实验设置参数不变的条件下,采用单因素敏感分析法,固定其他损失项权重为最优组合,调整某一损失函数的权重 λ_m,对 L_1,L_{SSIM},L_{Percep},L_{Edge},L_{mid} 进行权重敏感性分析,结果见图 10.

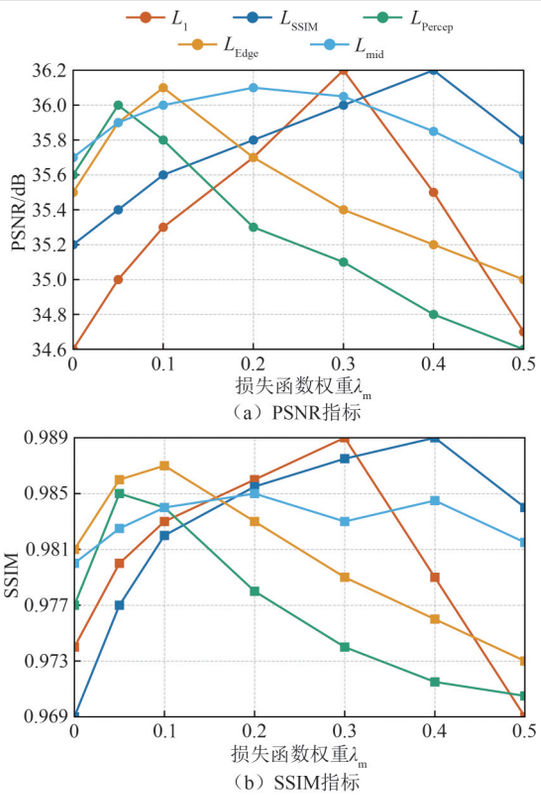


图 10　不同损失函数权重对 PSNR 与 SSIM 的敏感性分析

Fig.10　Sensitivity analysis of PSNR and SSIM to different loss function weights

由图 10 可知:①像素重建损失 L_1 在低权重区间,PSNR 提升至约 35.9 dB,在 $\lambda_m > 0.2$ 后逐步下降至 34.7 dB 左右,SSIM 保持在约 0.982 的水平,表明 L_1 对像素级精度的提升具有正向作用,但过高的权重会削弱结构一致性.②结构相似性损失 L_{SSIM} 随着权重增加,性能稳步提升,在 λ_m 时达到峰值(PSNR=36.2 dB,SSIM=0.989),表明 L_{SSIM} 在维持全局结构与视觉一致性方面具有关键作用.③感知损失 L_{Percep} 在低权重下 SSIM 提升至 0.985,PSNR 达到 35.8 dB;当 $\lambda_m > 0.3$ 后,性能下降趋势明显.该结果表明,适度的感知约束可提升语义一致性,但过高权重会抑制低层纹理与边缘信息的恢复.④边缘保持损失 L_{Edge} 在 λ_m 附近性能最优(PSNR=36.0 dB,SSIM=0.985),体现出其在保持轮廓与局部细节方面的重要性;当权重超过 0.3 时,整体性能出现下降,表明过强的边缘约束

可能影响图像的整体自然性.⑤中间监督损失L_{mid}在各权重区间均保持稳定增益,PSNR最高达到36.1 dB,SSIM达到0.986,曲线变化平缓,表明L_{mid}在训练过程中能提供稳定的正向增益,加速收敛并提升整体一致性.

综上所述,联合损失函数应以结构相似性与边缘保持为主要约束,并通过合理配置像素重建、感知及中间监督损失的权重,实现精度、结构与感知之间的均衡,从而有效提升去雾性能.

3.3.3 跨分辨率下模块鲁棒性分析

为评估模型不同变体在不同输入尺度下的适应性,本文对512×512像素分辨率下的交通场景子集样本分别下采样至128×128像素、256×256像素两种分辨率进行消融对比实验,结果见表8.

表8 不同分辨率下模型性能的消融实验结果

Tab.8 Ablation study results on model performance at different resolutions

分辨率	Baseline (PSNR/SSIM)	+输入增强 (PSNR/SSIM)	+GF分支 (PSNR/SSIM)	+AO−MMF分支 (PSNR/SSIM)	本文模型 (PSNR/SSIM)
128×128	28.93/0.952	29.15/0.956	29.27/0.961	29.19/0.957	29.44/0.960
256×256	30.71/0.966	31.28/0.969	31.30/0.970	31.36/0.967	31.53/0.972
512×512	31.26/0.956	33.72/0.965	34.32/0.973	34.54/0.975	36.20/0.989

表8中:“+输入增强”表示在Baseline基础上引入图像级输入增强模块;“+GF分支”表示在Baseline基础上引入引导滤波特征增强分支;“+AO-MMF分支”表示在Baseline基础上引入渐进最优极值滤波的特征增强分支.

由表8可知:在低分辨率128×128像素条件下,输入增强模块对性能提升作用最为显著,PSNR/SSIM由28.93 dB/0.952提升至29.15 dB/0.956,有效补偿了弱纹理条件下的结构信息;在中等分辨率256×256像素条件下,特征增强模块成为主要影响因素,GF与AO-MMF分支分别达到31.30 dB/0.970和31.36 dB/0.967,表明其能够有效增强结构恢复与纹理细化;在高分辨率512×512像素条件下,AO-MMF分支性能最佳(34.54 dB/0.975),在全局一致性与远景还原方面收益明显.

综上,输入增强模块可在低分辨率下提升结构表达,GF分支可强化边缘特征,AO-MMF分支在高分辨率下可改善全局一致性,二者结合使模型在多尺度条件下保持稳定性能增益与较高鲁棒性.

4 结论

1)采用AO-Guided融合滤波增强策略,实现对交通图像边缘结构与弱对比纹理信息的自适应增强.实验结果表明,该方法在图像平均梯度、信息熵和边缘像素数等指标上均优于传统最大滤波、最小滤波及单一引导滤波方法,有效提升了雾天交通图像的结构表达能力.

2)引入AO-Guided特征增强模块嵌入Vision Transformer编码结构,通过局部方差显著性映射实现Guided滤波与极值滤波特征的自适应融合,有效提升了网络对远景雾化区域细节与道路结构信息的恢复能力,改善了交通图像中车道线与道路边缘等结构的可辨识性.

3)构建基于AO-Guided模块集成的改进Vision Transformer图像去雾网络,在RESIDE真实交通雾图数据集上开展对比实验.实验结果表明,与典型ViT去雾模型DehazeFormer相比,该方法在可见边集合数目比和平均梯度比指标上分别提升13.1%和40.5%,黑色像素比例降低约25%,有效提升了雾霾条件下交通图像的清晰度.

当前方法在极端浓雾及远景信息严重退化场景下仍存在一定的恢复精度下降.未来研究可进一步引入区域自适应损失与远景显著性引导以增强小目标与远处轮廓的还原效果,结合多尺度长程依赖建模提升远景细节的表达能力,同时通过模型结构轻量化设计降低计算负担,以进一步提升复杂交通场景下的实时处理性能.

参考文献(References):

[1] 尹宏鹏, 陈波, 柴毅, 等. 基于视觉的目标检测与跟踪综述[J]. 自动化学报, 2016, 42(10): 1466−1489.
YIN Hongpeng, CHEN Bo, CHAI Yi, et al. Vision-based object detection and tracking: A review[J]. Acta Automatica Sinica, 2016, 42(10): 1466−1489. (in Chinese)

[2] 马永杰, 程时升, 马芸婷, 等. 卷积神经网络及其在智能交通系统中的应用综述[J]. 交通运输工程学报, 2021, 21(4): 48−71.
MA Yongjie, CHENG Shisheng, MA Yunting, et al. Review of convolutional neural network and its application in intelligent transportation system [J]. Journal of Traffic and Transportation Engineering, 2021, 21(4): 48−71. (in Chinese)

[3] HE K M, SUN J, TANG X O. Single image haze removal using dark channel prior[J]. IEEE Transactions on Pattern

Analysis and Machine Intelligence, 2011, 33(12): 2341－2353.

[4] ZHU Q S, MAI J M, SHAO L. A fast single image haze removal algorithm using color attenuation prior[J]. IEEE Transactions on Image Processing, 2015, 24(11): 3522－3533.

[5] 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.

[6] LI B Y, PENG X L, WANG Z Y, et al. AOD-net: all-in-one dehazing network[C]//2017 IEEE International Conference on Computer Vision (ICCV). Venice, 2017: 4780－4788.

[7] QIN X, WANG Z L, BAI Y C, et al. FFA-net: feature fusion attention network for single image dehazing[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 11908－11915.

[8] ZAMIR S W, ARORA A, KHAN S, et al. Restormer: efficient transformer for high-resolution image restoration[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, 2022: 5718－5729.

[9] SONG Y D, HE Z Q, QIAN H, et al. Vision transformers for single image dehazing[J]. IEEE Transactions on Image Processing, 2023, 32: 1927－1941.

[10] CHEN Z X, HE Z W, LU Z M. DEA-net: single image dehazing based on detail-enhanced convolution and content-guided attention[J]. IEEE Transactions on Image Processing, 2024, 33: 1002－1015.

[11] CHEN J W, ZHAO G H. Contrastive multiscale transformer for image dehazing[J]. Sensors, 2024, 24(7): 2041.

[12] YI W C, DONG L Q, LIU M, et al. MFAF-Net: image dehazing with multi-level features and adaptive fusion[J]. The Visual Computer, 2024, 40(4): 2293－2307.

[13] ZHANG Y Q, HE X, ZHAN C X, et al. Frequency-oriented transformer for remote sensing image dehazing[J]. Sensors, 2024, 24(12): 3972.

[14] LI X, REN Y L, JIN X, et al. Diffusion models for image restoration and enhancement: a comprehensive survey[J]. International Journal of Computer Vision, 2025, 133(11): 8078－8108.

[15] LI M M, LIANG H Y, LIU S X, et al. Asymptotically optimal algorithms for running max and Min filters on random inputs[J]. IEEE Transactions on Signal Processing, 2018, 66(13): 3421－3435.

[16] HE K M, SUN J, TANG X O. Guided image filtering[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(6): 1397－1409.

[17] ZHANG C X, GE L Y, CHEN Z, et al. Guided filtering: toward edge-preserving for optical flow[J]. IEEE Access, 2018, 6: 26958－26970.

[18] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4): 834－848.

[19] LI B Y, REN W Q, FU D P, et al. Benchmarking single-image dehazing and beyond[J]. IEEE Transactions on Image Processing, 2019, 28(1): 492－505.

[20] XUE W F, ZHANG L, MOU X Q, et al. Gradient magnitude similarity deviation: a highly efficient perceptual image quality index[J]. IEEE Transactions on Image Processing, 2014, 23(2): 684－695.

[21] CHEN X Q, ZHANG Q Y, LIN M H, et al. No-reference color image quality assessment: from entropy to perceptual quality[J]. EURASIP Journal on Image and Video Processing, 2019(1): 77.

[22] NI Z K, MA L, ZENG H Q, et al. ESIM: edge similarity for screen content image quality assessment[J]. IEEE Transactions on Image Processing, 2017, 26(10): 4818－4831.

[23] HUANG S C, CHENG F C, CHIU Y S. Efficient contrast enhancement using adaptive gamma correction with weighting distribution[J]. IEEE Transactions on Image Processing, 2013, 22(3): 1032－1041.

[24] NARASIMHAN S G, NAYAR S K. Vision and the atmosphere[J]. International Journal of Computer Vision, 2002, 48(3): 233－254.

[25] ZHANG C, SHAO Y C, SUN H J, et al. The WuC-Adam algorithm based on joint improvement of Warmup and cosine annealing algorithms[J]. Mathematical Biosciences and Engineering, 2023, 21(1): 1270－1285.

[26] CHEN D D, HE M M, FAN Q N, et al. Gated context aggregation network for image dehazing and deraining[C]//2019 IEEE Winter Conference on Applications of Computer Vision (WACV). Waikoloa Village, 2019: 1375－1383.

[27] HAUTIÈRE N, TAREL J P, AUBERT D, et al. Blind contrast enhancement assessment by gradient ratioing at visible edges[J]. Image Analysis & Stereology, 2008, 27(2): 87.

[28] CHOI L K, YOU J, BOVIK A C. Referenceless prediction of perceptual fog density and perceptual image defogging[J]. IEEE Transactions on Image Processing, 2015, 24(11): 3888－3901.