

# 基于 Transformer 和门控融合机制的图像去雾算法<sup>①</sup>



王 燕, 陈燕燕, 刘晶晶, 胡津源

(兰州理工大学 计算机与通信学院, 兰州 730050)

通信作者: 王 燕, E-mail: [wangyan@lut.edu.cn](mailto:wangyan@lut.edu.cn)

**摘要:** 针对现有的图像去雾算法仍然存在去雾不彻底、去雾后的图像边缘模糊、细节信息丢失等问题, 本文提出了一种基于 Transformer 和门控融合机制的图像去雾算法。通过改进的通道自注意力机制提取图像的全局特征, 提高模型处理图像的效率, 设计多尺度门控融合块捕获不同尺度的特征, 门控融合机制通过动态调整权重, 提高模型对不同雾化程度的适应能力, 同时更好地保留图像边缘及细节信息, 并使用残差连接增强特征的重用性, 提高模型泛化能力。经实验验证, 所提出的去雾算法可以有效恢复真实有雾图像中的内容信息, 在合成的有雾图像数据集 SOTS 上的峰值信噪比达到了 34.841 dB, 结构相似性达到了 0.984, 去雾后的图像内容信息完整且没有出现细节信息模糊和去雾不彻底等现象。

**关键词:** 图像去雾; Transformer; 自注意力机制; 门控融合机制; 多尺度特征融合

引用格式: 王燕, 陈燕燕, 刘晶晶, 胡津源. 基于 Transformer 和门控融合机制的图像去雾算法. 计算机系统应用, 2025, 34(2):1-10. <http://www.c-s-a.org.cn/1003-3254/9780.html>

## Image Dehazing Algorithm Based on Transformer and Gated Fusion Mechanism

WANG Yan, CHEN Yan-Yan, LIU Jing-Jing, HU Jin-Yuan

(School of Computer and Communication, Lanzhou University of Technology, Lanzhou 730050, China)

**Abstract:** Aiming at the existing image dehazing algorithms which still have problems such as incomplete dehazing, blurred edges of dehazed images, and detail information loss, this study presents an image dehazing algorithm based on Transformer and gated fusion mechanism. Global features of the image are extracted by the improved channel self-attention mechanism to improve the efficiency of the model in processing images. A multi-scale gated fusion block is designed to capture features of different scales. The gated fusion mechanism improves the adaptability of the model to different degrees of dehazing by dynamically adjusting weights while better preserving the image edges and detail information. Residual connections are used to enhance the reusability of features and improve the generalization ability of the model. Experimental verification shows that the proposed dehazing algorithm can effectively restore the content information in real hazy images. On the synthesized hazy image dataset SOTS, the peak signal-to-noise ratio reaches 34.841 dB, and the structural similarity reaches 0.984. The dehazed image has complete content information without blurred detail information and incomplete dehazing.

**Key words:** image dehazing; Transformer; self-attention mechanism; gated fusion mechanism; multi-scale feature fusion

① 基金项目: 国家自然科学基金 (62266030)

收稿时间: 2024-07-22; 修改时间: 2024-09-03; 采用时间: 2024-09-19; csa 在线出版时间: 2024-12-09

CNKI 网络首发时间: 2024-12-09

雾霾主要由悬浮在空气中的颗粒物和化学物质组成,它们通过吸收或散射大气光,最终导致采集到的图像质量下降。退化的图像对后续视觉任务带来了巨大挑战,因此,对图像去雾技术的研究具有重要的现实意义。图像去雾算法的目的是从降质图像中去除雾霾噪声,提高图像对比度、改善图像细节和提升图像质量,以满足特定应用场景的需求<sup>[1]</sup>。传统的去雾算法通常需要依赖物理模型,但这类方法很容易引入误差。主流的基于卷积神经网络(convolutional neural network, CNN)的算法虽然取得了比较好的效果,但其对图像的全局关系建模存在局限性,使得去雾图像的局部特征间缺乏连续性,从而出现边缘模糊、细节信息丢失、去雾不彻底等现象。近年来,随着Transformer模型被引入图像分类任务中,其优秀的远程建模能力被运用到了各种视觉任务中<sup>[2-4]</sup>。虽然Transformer的自注意力机制可以有效捕获图像的全局特征,但巨大的计算复杂度限制了其在视觉任务中的发展。本文通过分析这些问题产生的原因,提出了一种基于Transformer和门控融合机制的图像去雾算法。

本文主要贡献如下。

(1) 为了降低标准Transformer在空间自注意力上的高计算复杂度,本文改进了其结构,引入了跨通道的自注意力机制,通过计算通道间的互协方差,以较低的计算复杂度捕获全局特征,提高了模型对图像的处理效率。

(2) 针对去雾后的图像存在边缘模糊和细节信息丢失问题,本文设计了门控融合机制,通过动态地学习和调整输入特征的权重,使网络更精确地关注和保留重要的边缘和细节信息,从而提高去雾模型对细节的捕获和保留能力。

(3) 针对图像去雾不彻底问题,本文通过对输入图像的不同尺度特征进行选择与融合,增强特征之间的互补性,突出差异性特征,提高了模型对图像中各种尺度信息的捕获和融合能力,提升了去雾的彻底性和准确性。

## 1 相关工作

### 1.1 传统的图像去雾算法

传统去雾算法主要通过消除图像噪声和增强图像对比度来改善雾图,如直方图均衡化<sup>[5]</sup>、同态滤波<sup>[6]</sup>和基于Retinex理论<sup>[7]</sup>的算法等,但此类方法没有考虑图

像退化的原因,可能导致图像的视觉效果失真,应用范围受限。随着大气散射模型<sup>[8]</sup>的提出,图像去雾算法有了较大的突破性进展,通过应用物理模型从退化图像中估计透射图和全球大气光值来恢复无雾的清晰图像。例如,He等人<sup>[9]</sup>提出了基于暗通道先验的去雾算法,Zhu等人<sup>[10]</sup>提出了基于颜色衰减先验的去雾算法等,但这类算法通常依赖于特定的先验假设,可能导致在多变的实际环境中,图像恢复效果不理想。

### 1.2 基于CNN的图像去雾算法

目前,基于CNN的图像去雾算法主导了图像去雾研究领域,去雾效果迎来了质的提升,经过处理的图像在对比度、颜色饱和度和细节处理方面更完善,雾霾去除更彻底<sup>[11,12]</sup>。其按照是否估计中间参数可分为两类:第1类基于物理模型,首先利用深度卷积模型估计介质传输图和大气光图像,然后通过大气散射模型重建清晰图像。Cai等人<sup>[13]</sup>通过构建卷积网络模型Dehaze-Net来估计介质传输图,根据大气散射模型完成清晰图像的重构。Ren等人<sup>[14]</sup>设计了一种逐步细化的多尺度卷积神经网络MSCNN来估计介质传输图,然后再进行局部的细化。Li等人<sup>[15]</sup>提出了一种一体化去雾网络AOD-Net,将介质传输图和大气光图像统一成一个变量K,然后对其进行估计,省去中间步骤从而减小参数预估中的累计误差。但这类方法受限于参数估计不准确问题,这种不准确的结果会在训练过程中不断累积,最终导致模型的去雾效果变差。第2类不需要估计中间参数,利用端到端的卷积神经网络建立有雾图像到清晰图像的直接映射。Ren等人<sup>[16]</sup>提出了一种门控融合网络GFN,通过学习门控机制来融合不同尺度的图像特征,有效提升模型的细节恢复能力。Liu等人<sup>[17]</sup>提出了GridDehazeNet,将基于网格的多尺度提取与注意力机制相结合,在图像的细节和纹理恢复方面显示出了优异能力。Li等人<sup>[18]</sup>提出一种独立的细节恢复网络Dehazing-DRN,有效提升去雾图像的细节清晰度和视觉质量。Yi等人<sup>[19]</sup>提出了高效的多尺度拓扑网络MSTN,显著提高了去雾模型的处理效率和去雾结果的精确度。然而,由于局部感受野和卷积的相互作用,卷积神经网络建模远程依赖信息的能力有限<sup>[20,21]</sup>。

### 1.3 基于Transformer的图像去雾算法

不同于卷积操作缺乏获取图像全局信息的能力,Transformer能够通过全局计算建模特征之间的依赖关系<sup>[22]</sup>。Chen等人<sup>[23]</sup>将Transformer引入图像复原任务,

提出一种图像处理 Transformer, 采用多头和多尾的结构处理多种图像复原任务. 为解决 Transformer 自注意力操作导致的计算成本高的问题, Swin Transformer<sup>[24]</sup>在窗口内执行局部注意, 并为跨窗口连接引入位移窗口技术. SwinIR<sup>[25]</sup>做了进一步的改进, 通过将卷积与 Transformer 结合, 有效降低计算量. 此外, UFormer<sup>[26]</sup>和 Restormer<sup>[27]</sup>均基于 Transformer 构建了编解码结构实现图像复原. Zhao 等人<sup>[28]</sup>提出一种结合特征增强网络和 vision Transformer 的网络 CFEN-ViT, 有效增强图像细节和颜色真实性. Song 等人<sup>[29]</sup>基于 Swin Transformer, 针对图像去雾任务的特性改进归一化层、激活函数和空间信息聚合方案, 提出 Dehaze-Former 并取得了良好的去雾性能.

## 2 模型设计

本文设计了一种基于 Transformer 和门控融合机

制的图像去雾算法, 总体架构如图 1 所示. 该网络包含 3 个模块: 浅层特征提取、深层特征提取和特征重建模块. 其中, 浅层特征提取模块通过简单的卷积层将输入图像转换为特征图; 深层特征提取模块通过堆叠多个残差 Transformer 块 (residual Transformer block, RTB) 和一个卷积层提取深层特征; 特征重建模块采用卷积层和像素洗牌上采样重构清晰图像. 对于 RTB, 则是由多个通道注意 Transformer 块 (channel attention Transformer block, CATB) 和一个卷积层通过残差连接进行构建. 每个 CATB 由层归一化 (layer normalization, LN)、多头通道自注意力 (multi-head channel self-attention, MCSA) 和多尺度门控融合块 (multi-scale gating fusion block, MGFB) 组成. MCSA 采用通道自注意力机制, 通过跨通道的互协方差计算降低计算复杂度, 提高模型效率. MGFB 通过提取和融合不同尺度的特征, 进一步提高模型的表示能力.

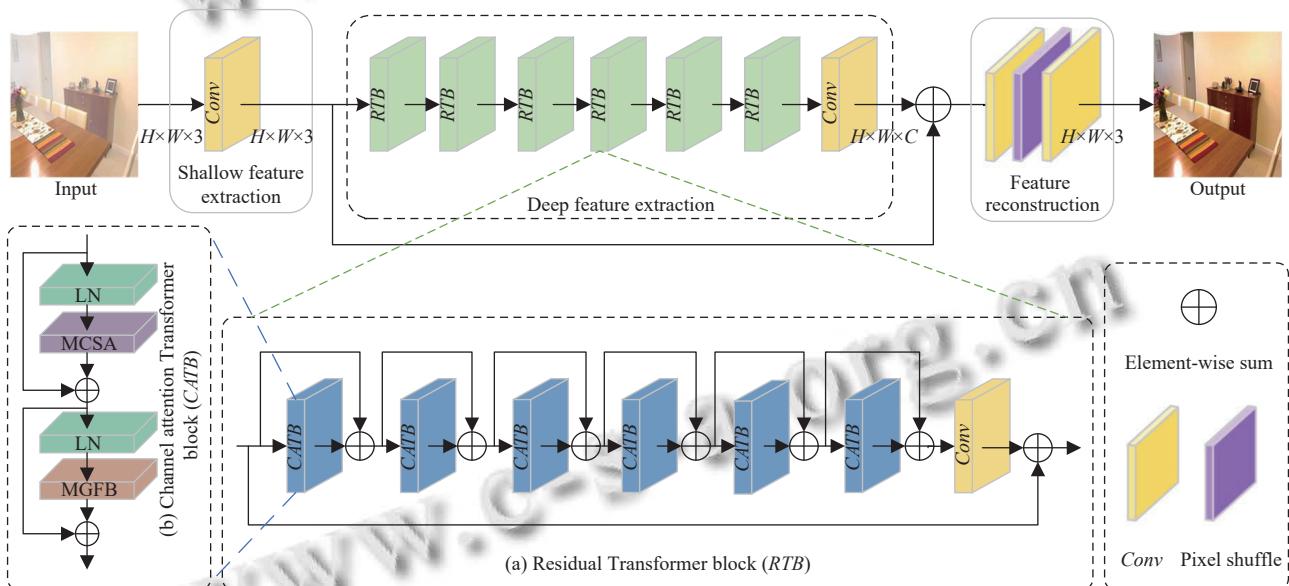


图 1 网络架构图

### 2.1 浅层特征提取

给定一个有雾图像  $I \in \mathbb{R}^{H \times W \times 3}$ , 其中,  $H \times W$  表示输入图像的空间分辨率. 首先使用 1 个  $3 \times 3$  的卷积层提取浅层特征  $F_0 \in \mathbb{R}^{H \times W \times C}$ , 如式 (1) 所示:

$$F_0 = \text{Conv}_{3 \times 3}(I) \quad (1)$$

其中,  $\text{Conv}_{3 \times 3}(\cdot)$  表示卷积操作. 通过该卷积层, 将输入图像映射到更高维的特征空间, 为后续的深层特征处理提供简单而有效的手段.

### 2.2 深层特征提取

浅层特征图  $F_0$  经过  $K$  个 RTB 和 1 个  $3 \times 3$  卷积层提取深层特征. 中间特征  $F_1, F_2, \dots, F_K$  和输出的深层特征  $F_d$  的提取过程如式 (2) 和式 (3) 所示:

$$F_i = \text{RTB}_i(F_{i-1}), i = 1, 2, \dots, K \quad (2)$$

$$F_d = \text{Conv}(F_K) \quad (3)$$

其中,  $\text{RTB}_i(\cdot)$  表示第  $i$  个 RTB. 对于第  $i$  个 RTB, 由  $L$  个 CATB 和 1 个  $3 \times 3$  卷积层组成. 中间特征  $F_i^1, F_i^2, \dots, F_i^L$

和第  $i$  个 RTB 的提取过程如式(4)和式(5)所示:

$$F_i^{(j)} = CATB_{i,j}(F_i^{(j-1)}) + F_i^{(j-1)}, j = 1, 2, \dots, L \quad (4)$$

$$RTB_i(F_i) = Conv_{3 \times 3}(F_i^L) \quad (5)$$

其中,  $CATB_{i,j}(\cdot)$  表示第  $i$  个 RTB 内的第  $j$  个 CATB.

### 2.2.1 多头通道自注意力

Transformer 中的计算开销主要来自于自注意力层, 在标准的多头自注意力<sup>[30]</sup>中,  $Q$ (查询) 和  $K$ (键) 之间点积操作的时间和内存复杂度随着输入空间分辨率的增加呈二次方增长. 即对于空间分辨率为  $H \times W$ , 通道数为  $d$  图像, 计算复杂度为  $O((WH)^2 \cdot d)$ , 这使得它对计算资源的要求非常高. 为了解决这一问题, 提出了 MCSA, 如图 2 所示. MCSA 的关键在于在通道之间运用自注意力, 通过计算跨通道的互协方差代替空间维度的注意力计算, 生成编码全局上下文的转置注意力图  $A$ . 这一改进使得计算复杂度降为  $O(WH \cdot d^2)$ , 减少了自注意力层的计算负担, 尤其在高分辨率图像处理任务中, 显著降低了时间和内存消耗, 可以有效提高模型处理图像的效率. 此外, MCSA 的另一个关键设计是引入深度可分离卷积, 能够突显更丰富的局部上下文信息. 这一设计使得模型能够在学习远程建模能力的同时增强局部表征能力, 进一步优化模型去雾性能, 使得模型在保持优异性能的同时, 具备更高的计算效率.

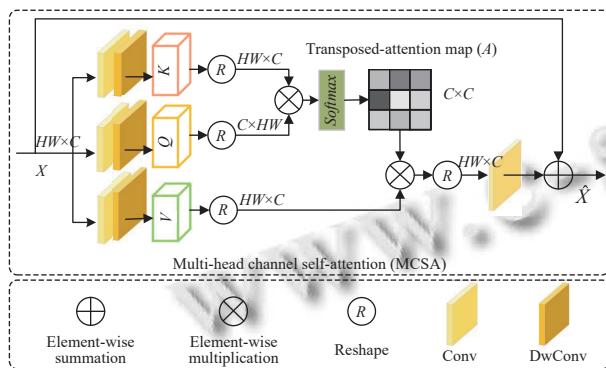


图 2 多头通道自注意力

输入特征  $X$  经过层归一化后, 通过 MCSA 将通道数划分为多个“头”, 并行学习每个“头”的注意力图. 具体来说, 首先通过  $1 \times 1$  逐点卷积降低输入特征维度, 然后使用  $3 \times 3$  深度可分离卷积捕获多尺度信息, 生成  $Q$ 、 $K$ 、 $V$  投影. 其中, 深度可分离卷积的操作被分解为深度卷积和逐点卷积两步, 深度卷积在通道之间进行特征提取, 逐点卷积引入跨通道的交互信息, 这使得网

络能够更好地聚合不同尺度的信息, 同时减少模型参数数量, 降低计算复杂度, 为去雾任务提供更好的性能.  $Q$ 、 $K$ 、 $V$  的生成过程如式(6)所示:

$$\begin{cases} Q = DwConv_{3 \times 3}^Q(PwConv_{1 \times 1}^Q(X)) \\ K = DwConv_{3 \times 3}^K(PwConv_{1 \times 1}^K(X)) \\ V = DwConv_{3 \times 3}^V(PwConv_{1 \times 1}^V(X)) \end{cases} \quad (6)$$

其中,  $X$  是输入特征,  $DwConv_{3 \times 3}^{(\cdot)}$  是  $3 \times 3$  深度卷积,  $PwConv_{1 \times 1}^{(\cdot)}$  是  $1 \times 1$  逐点卷积. 接着对  $Q$  和  $K$  重塑, 并计算它们的点积, 生成大小为  $\mathbb{R}^{C \times C}$  的转置注意力图  $A$ , 避免大小为  $\mathbb{R}^{HW \times HW}$  的注意力图带来的巨大计算开销. 最后将注意力图  $A$  的值与  $V$  进行点积操作, 得到的特征经过  $1 \times 1$  卷积进行线性映射并与输入的特征进行加和得到最终输出. 这一过程如式(7)所示:

$$\begin{cases} \hat{X} = Conv_{1 \times 1}(Attention(Q, K, V)) + X \\ Attention(Q, K, V) = V \cdot Softmax(K \cdot Q / \sqrt{d_k}) \end{cases} \quad (7)$$

其中,  $X$  和  $\hat{X}$  是输入和输出特征图,  $\sqrt{d_k}$  是一个可学习的缩放参数, 用于控制点积的程度.

### 2.2.2 多尺度门控融合块

门控单元的核心是通过一个或多个神经网络来学习输入特征的重要性权重, 并动态地对这些特征进行加权, 以适应不同的去雾场景. 首先, 门控单元使用 1 个 Sigmoid 函数来生成门控信号, 将每个特征通道的权重值限制在 0~1 之间, 表示特征被抑制或增强的程度. 通过门控单元生成的权重(或门控信号)应用于输入特征中. 门控机制通过逐元素相乘的方式, 将门控信号与输入特征相结合, 从而实现对不同特征通道的加权调控, 通过这种加权机制, 不同特征通道根据任务需求得到不同的权重, 这使得模型能够适应不同的去雾情况. 对于严重的雾霾区域, 模型可能会对全局信息给予更高的权重, 而在细节丰富的区域, 模型则倾向于加强局部细节特征.

本文为了更好地处理不同程度的雾霾分布, 有效解决去雾不彻底和细节信息丢失问题, 设计了 MGFB, 如图 3 所示. 通过在传输过程中插入两个不同尺度的深度可分离卷积路径提取和融合更加丰富的特征信息. 同时设计了一种混合门控机制, 通过引入具有可学习参数的门控单元, 动态地学习和调整两个分支的权重, 使网络更加关注有雾图像中需要更多处理的区域, 从而更有效地去除雾霾, 解决去雾不彻底问题; 引入门控

机制也能使网络更精确地关注和保留重要的边缘和细节信息,从而提高去雾质量和去雾准确性。

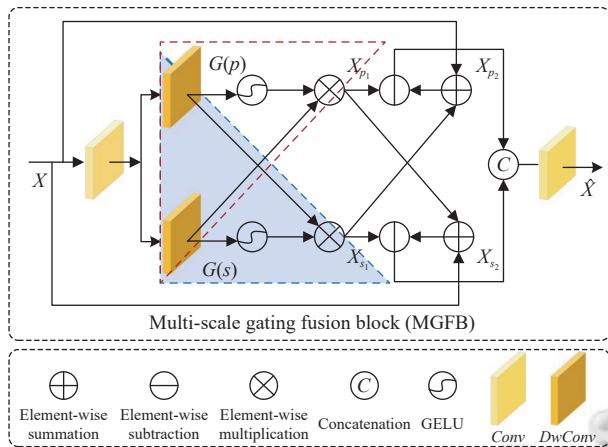


图3 多尺度门控融合块

具体地,给定1个输入特征 $X \in \mathbb{R}^{H \times W \times C}$ ,经过层归一化之后,首先应用 $1 \times 1$ 的逐点卷积实现跨通道的特征交互,然后将其输入两个并行分支,其中 $3 \times 3$ 和 $5 \times 5$ 的深度可分离卷积用来捕获不同尺度的局部特征。之后两个分支进行逐元素乘积,其中一个分支通过GELU非线性激活,形成门控 $G(p)$ 和 $G(s)$ ,构成一种混合门控单元。通过门控单元动态地调整权重从而生成特征 $X_{p_1}$ 和 $X_{s_1}$ 。然后将特征 $X_{s_1}$ 与 $X$ 逐元素相加来增强特征的互补性,再和特征 $X_{p_1}$ 逐元素相减来突出差异性特征,从而得到特征 $X_{p_2}$ 。同样将特征 $X_{p_1}$ 与 $X$ 逐元素相加后,和特征 $X_{s_1}$ 逐元素相减得到 $X_{s_2}$ ,通过这种特征之间的相加减操作来更好的恢复图像。最后 $X_{p_2}$ 和 $X_{s_2}$ 进行通道拼接,经过 $1 \times 1$ 逐点卷积得到输出特征 $\hat{X} \in \mathbb{R}^{H \times W \times C}$ 。MGFB的特征提取和融合过程如式(8)所示:

$$\begin{cases} G(X_{p_1}) = \sigma(DwConv_{3 \times 3}Conv(X)) \odot DwConv_{5 \times 5}Conv(X) \\ G(X_{s_1}) = \sigma(DwConv_{5 \times 5}Conv(X)) \odot DwConv_{3 \times 3}Conv(X) \\ X_{p_2} = G(X_{s_1}) + X - G(X_{p_1}) \\ X_{s_2} = G(X_{p_1}) + X - G(X_{s_1}) \\ \hat{X} = Conv[X_{p_2}, X_{s_2}] \end{cases} \quad (8)$$

其中,  $\sigma(\cdot)$ 是GELU激活,  $G(\cdot)$ 表示门控单元,  $\odot$ 表示哈达玛积,  $[ \cdot ]$ 表示通道级联。

### 2.3 特征重建模块

在特征重建模块,为了重建去雾后的图像 $J$ ,采用了卷积和像素洗牌上采样操作来聚合浅层特征和深层

特征。这一操作如式(9)所示:

$$J = Conv(PixelShuffler(Conv(F_0 + F_d))) \quad (9)$$

其中,  $PixelShuffler(\cdot)$ 是像素洗牌上采样操作。浅层特征 $F_0$ 主要包含图像低频信息,这些信息对于保持图像的基本结构和纹理细节至关重要。相比之下,深层特征 $F_d$ 侧重于补充在去雾过程中可能丢失的高频信息,这些信息对于恢复图像的清晰度和细节精度非常重要。为了进一步提升重建质量和模型的稳定性,网络采用了残差连接策略,将低频信息直接传输到重建模块,保证了图像基本内容的完整性和连续性,同时帮助深层特征提取模块专注于处理高频信息,确保训练的稳定性。

### 2.4 损失函数

在训练过程中,使用Charbonnier损失来计算去雾后的图像与标签图像之间的损失。如式(10)所示:

$$L = \sqrt{\|J - J'\|^2 + \varepsilon^2} \quad (10)$$

其中,  $J'$ 是标签图像,  $J$ 是去雾后的图像,  $\varepsilon$ 在实验中设置为 $10^{-3}$ 的常数,用来保持训练的稳定性。Charbonnier损失又被称为损失的平方根,相较于和损失函数,Charbonnier损失在梯度较小的区域采用平方根函数替代平方函数,能够在保持平滑性的同时减少对噪声的敏感性,有效避免过拟合问题。并且Charbonnier损失函数中的常数是可调节的,可以根据实际情况进行调整,从而达到更好的鲁棒性。

## 3 实验分析

### 3.1 实验设置

为了验证所提出的图像去雾算法的有效性,设计并进行了一系列实验。本文使用Python 3.8 和 PyTorch 1.11.0 搭建深度学习环境,所有实验(训练和测试)均在NVIDIA GeForce RTX 3090 (24 GB) GPU上开展。训练图像被随机裁剪为 $224 \times 224$ 大小作为网络的输入,网络模型训练使用Adam优化器,批量大小为16,初始学习率设为0.0001,整个网络的迭代次数为500。

### 3.2 数据集

本文使用合成数据集RESIDE的子集ITS(indoor training set)、OTS(outdoor training set)、SOTS(synthetic objective testing set)和真实世界数据集O-HAZE、NH-HAZE、Dense-HAZE评估所提出的模型。其中,ITS包

含由 1399 张清晰图像生成的 13990 张室内有雾图像; OTS 含由 2061 张清晰图像生成的 72135 张室外有雾图像。为了加快模型的收敛速度, 本文选取整个 ITS 数据集和 OTS 数据集中的 24045 张合成图像进行训练。SOTS 测试集包含 500 对室内图像和 500 对室外图像。真实世界数据集 O-HAZE 包含 45 张户外场景图像对、NH-HAZE 包含 45 张户外不均匀图像对、Dense-HAZE 包含 55 张高浓度雾霾图像对, 本文随机选取每个真实图像数据集的 5 张图像作为测试集, 其余图像作为训练集。

### 3.3 评价指标

为了客观评价本文算法的优劣, 采用峰值信噪比 (peak signal-to-noise ratio, *PSNR*) 和结构相似性 (structural similarity, *SSIM*) 作为去雾图像质量评价的客观指标。其中, *PSNR* 用于衡量图像重构或恢复的质量, *PSNR* 越大, 表示图像恢复的误差越小。其计算过程如式(11)所示:

$$PSNR = 10 \log \left( \frac{MAX^2}{MSE} \right) \quad (11)$$

其中, *MAX* 表示图像的最大像素值, *MSE* 表示图像处理前后的均方误差, 其计算过程如式(12)所示:

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i, j) - J(i, j)]^2 \quad (12)$$

其中, *I* 表示处理前的图像, *J* 表示处理后的图像, 图像大小为  $m \times n$ 。

*SSIM* 用于衡量两幅图像的视觉相似度, 它考量了图像的亮度、对比度和结构信息, *SSIM* 越大, 表示处理后的图像与原始图像更相似。其计算过程如式(13)所示:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (13)$$

其中,  $\mu_x$  和  $\mu_y$  分别表示图像 *x* 和 *y* 的像素均值,  $\sigma_x^2$  和  $\sigma_y^2$  分别表示图像 *x* 和 *y* 的方差,  $\sigma_{xy}$  表示图像 *x* 和 *y* 的协方差,  $c_1$ 、 $c_2$  为常数。

### 3.4 实验结果及对比分析

#### 3.4.1 合成有雾图像数据集上的结果

将所提出方法的实验结果与基于物理的方法 DCP<sup>[9]</sup>、基于 CNN 的方法 GFN<sup>[16]</sup>、GridDehazeNet<sup>[17]</sup>、Dehazing-DRN<sup>[18]</sup>、MSTN<sup>[19]</sup> 和基于 Transformer 的方

法 SwinIR<sup>[25]</sup>、CFEN-ViT<sup>[28]</sup>、DehazeFormer-T (DehazeFormer<sup>[29]</sup> 的变体) 等的实验结果在 *PSNR* 和 *SSIM* 上进行了比较, 表 1 显示了不同图像去雾方法在合成有雾图像数据集 SOTS-indoor、SOTS-outdoor 上的定量结果。表 1 中数据为所有测试结果均值, 最优结果使用粗体字表示, 次优结果用下划线表示。

表 1 不同方法在 SOTS 数据集上的定量结果

Methods	SOTS-indoor		SOTS-outdoor		#Params (M)
	PSNR (dB)	SSIM	PSNR (dB)	SSIM	
DCP <sup>[9]</sup>	16.369	0.803	19.016	0.819	—
GFN <sup>[16]</sup>	22.156	0.810	21.089	0.802	0.51
GridDehazeNet <sup>[17]</sup>	32.026	0.971	30.710	0.963	0.96
Dehazing-DRN <sup>[18]</sup>	34.235	0.971	33.036	0.980	12.60
MSTN <sup>[19]</sup>	34.634	<u>0.978</u>	32.523	<u>0.981</u>	18.91
SwinIR <sup>[25]</sup>	33.026	0.869	32.521	0.901	0.88
CFEN-ViT <sup>[28]</sup>	32.150	0.966	31.036	0.969	—
DehazeFormer-T <sup>[29]</sup>	<b>34.983</b>	0.976	<u>33.974</u>	0.963	0.68
Ours	34.841	<b>0.984</b>	<b>34.074</b>	<b>0.982</b>	1.31

通过表 1 的实验数据可见, 在所有的比较方法中, 本文方法的 *SSIM* 值在 SOTS-indoor 和 SOTS-outdoor 数据集上均取得最优值; *PSNR* 值在 SOTS-outdoor 数据集上取得最优值, 在 SOTS-indoor 数据集上仅比 DehazeFormer-T 低 0.142 dB。这是由于 DehazeFormer-T 方法是基于 Swin Transformer 构建的, 其滑动窗口机制在局部信息提取方面表现较好, 而本文所提出方法的目的在于融合全局和局部信息, 并以较低的计算资源消耗实现较好的图像处理效果; 同时通过门控单元对不同的图像区域进行更灵活的权重分配, 使模型对复杂场景有更好的适应能力。因此, 本文所提方法在局部信息提取方面相比 Swin Transformer 稍显逊色, 导致在 SOTS-indoor 数据集上的 *PSNR* 值比 DehazeFormer-T 略低。此外, 为了更全面地评价所提出的模型, 将所提出模型与其他去雾模型的参数量进行比较。由对比结果可知, 所提出的模型在参数量上与其他基于深度学习的方法相当或更好。

图 4 展示了不同方法在合成有雾数据集 SOTS 上的定性结果, 第(1)–(3)行为 SOTS-indoor 上的定性结果, 第(4)–(6)行为 SOTS-outdoor 上的定性结果。可以看出, DCP 的复原效果较为彻底, 这是由于暗通道先验存在局限性, 导致去雾后的图像整体发暗, 并且存在颜色失真现象, 如图 4(b); GridDehazeNet 存在一定的颜色偏移, 如图 4(c); SwinIR 和 CFEN-ViT 存在边缘模糊

和去雾不彻底现象,如图4(f)和(g);与真实图像GT相比,Dehazing-DRN、MSTN和DehazeFormer-T的去雾

效果与本文提出方法的结果较为接近,但Dehazing-DRN和MSTN方法在边缘细节上恢复效果略差。

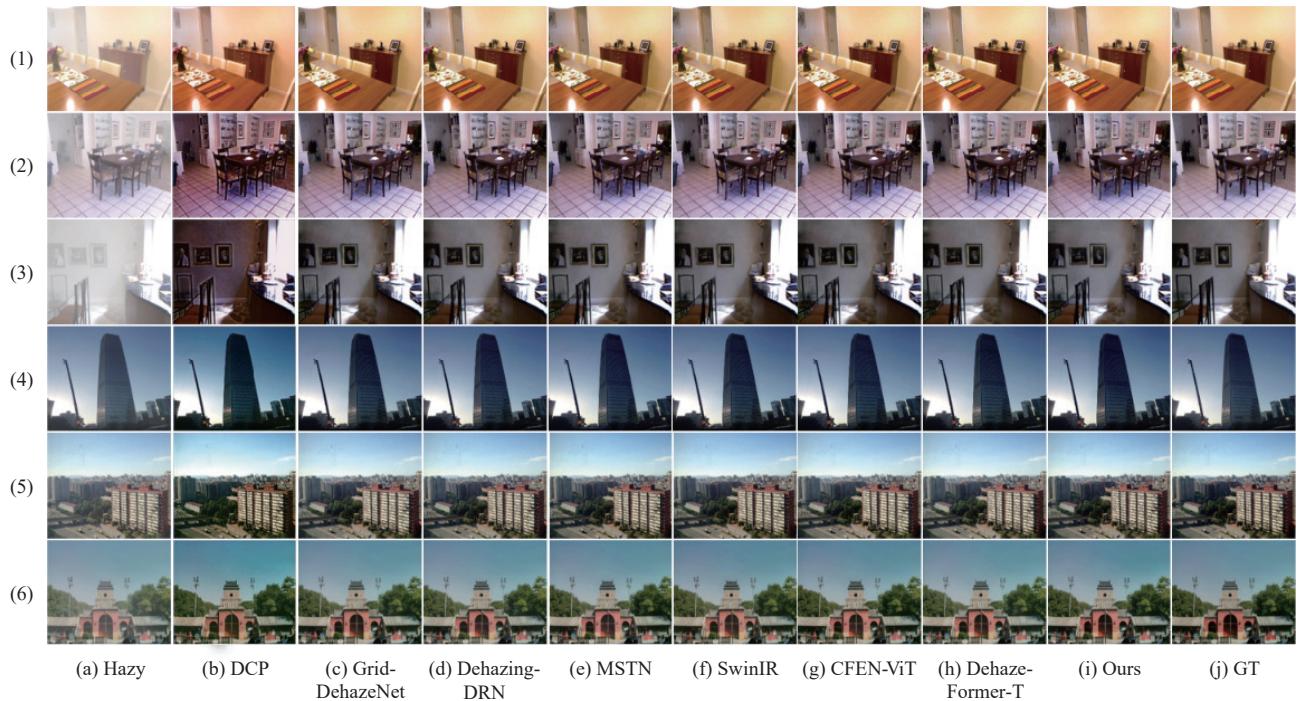


图4 不同算法在SOTS数据集上的定性结果

#### 3.4.2 真实有雾图像数据集上的结果

表2显示了不同方法在真实有雾图像数据集O-HAZE、NH-HAZE和Dense-HAZE上的定量结果。表2中数据为所有测试结果均值,最优结果使用粗体字表示,次优结果用下划线表示。

表2 不同方法在真实有雾图像数据集上的定量结果

Methods	O-HAZE		NH-HAZE		Dense-HAZE	
	PSNR (dB)	SSIM	PSNR (dB)	SSIM	PSNR (dB)	SSIM
DCP <sup>[9]</sup>	16.364	0.631	18.269	0.513	12.325	0.439
GFN <sup>[16]</sup>	18.136	0.616	11.231	0.512	10.465	0.248
GridDehazeNet <sup>[17]</sup>	25.377	0.844	17.364	0.723	13.169	0.441
Dehazing-DRN <sup>[18]</sup>	27.369	0.826	19.023	<u>0.730</u>	14.391	0.431
MSTN <sup>[19]</sup>	26.259	0.762	18.452	0.630	13.259	0.420
SwinIR <sup>[25]</sup>	26.532	0.765	18.369	0.691	13.897	0.406
CFEN-ViT <sup>[28]</sup>	27.871	0.758	17.562	0.721	13.268	0.410
DehazeFormer-T <sup>[29]</sup>	<u>28.032</u>	<u>0.861</u>	<u>20.026</u>	0.713	<u>15.126</u>	<u>0.438</u>
Ours	<b>28.193</b>	<b>0.864</b>	<b>20.207</b>	<b>0.733</b>	<b>15.227</b>	<b>0.452</b>

通过实验数据可以看出,本文方法的PSNR和SSIM值在所有的比较方法中都是最高的。证明了所提出的模型在复杂多样的真实雾霾场景中有更好的适应性和去雾性能。相比之下,在O-HAZE和Dense-HAZE数据

集上,DehazeFormer-T的PSNR和SSIM仅次于本文方法;在NH-HAZE数据集上,DehazeFormer-T的PSNR取得次优值,Dehazing-DRN的SSIM取得次优值。

图5展示了不同方法在真实的有雾图像数据集O-HAZE上的定性结果,可以看出,DCP存在图像偏暗等颜色失真问题,如图5(b);GridDehazeNet存在边缘模糊和去雾不彻底现象,如图5(c);MSTN和SwinIR存在一定的雾霾残留现象,如图5(e)和(f)的第2、第3幅图;Dehazing-DRN和CFEN-ViT存在一定的颜色偏移,如图5(d)和(g)的第2幅图;与真实图像GT相比,DehazeFormer-T去雾效果与本文提出方法的结果较为接近,但在边缘细节上恢复效果略差。

图6展示了不同方法在真实的有雾图像数据集NH-HAZE上的定性结果,可见DCP存在图像偏暗等颜色失真问题以及严重的雾霾残留现象,如图6(b);GridDehazeNet存在颜色失真和显著的雾霾残留现象,如图6(c);MSTN、SwinIR和CFEN-ViT存在一定的颜色偏移和去雾不彻底现象,如图6(e)、(f)和(g);Dehazing-DRN存在一定的颜色偏移,如图6(d)的第2幅图;与真实图像GT相比,DehazeFormer-T去雾效果与本文提出方法的结果较为接近,但仍然存在边缘

细节模糊的问题,如图6(h)的第2、3幅图。

综上比较,本文方法的处理效果对颜色失真、细

节模糊、去雾不彻底等现象有一定的改善,在色彩细节方面与真实图像最为接近,取得了最佳视觉体验。

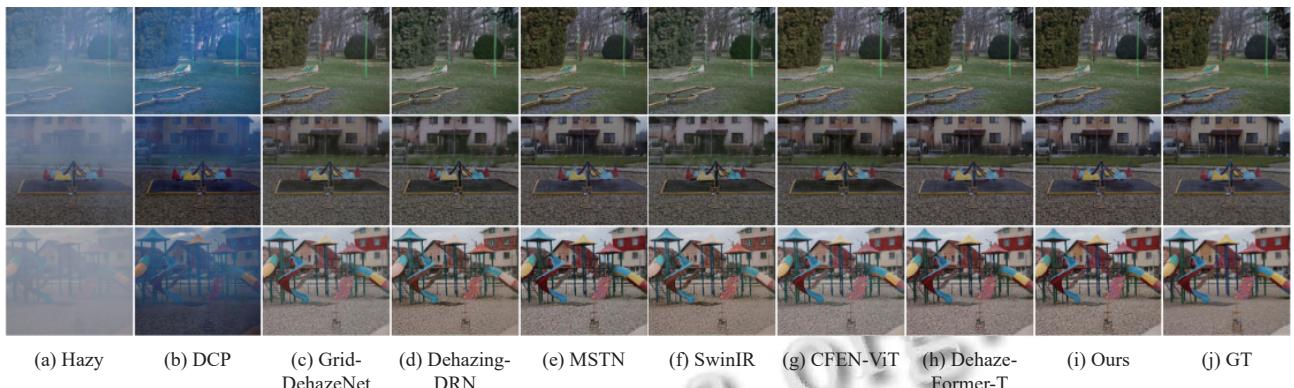


图5 不同方法在O-HAZE数据集上的定性结果

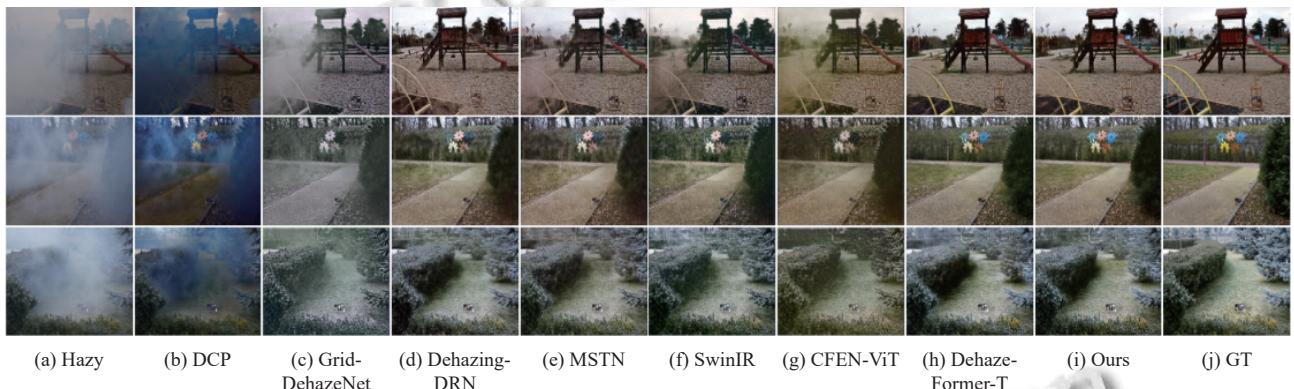


图6 不同方法在NH-HAZE数据集上的定性结果

### 3.5 消融实验

为了验证Transformer内部MCSA、MGFB的有效性,设计了如表3所示的具有不同模块组合的消融实验。在实验中需要保持其余结构和参数不变,以NH-HAZE数据集为例给出证明。其中M1表示模型中其他部分保留,删除MGFB; M2表示模型中其他部分保留,删除MCSA; M3为本文所提出模型。

表3 Transformer内部各模块的消融实验结果				
Module	MCSA	MGFB	PSNR (dB)	SSIM
M1	√	—	19.691	0.721
M2	—	√	19.453	0.696
M3	√	√	<b>20.207</b>	<b>0.733</b>

从表3可以看出,本文所提模型的PSNR和SSIM取得了最优结果。删除MGFB后,PSNR和SSIM值分别降低了0.561 dB和0.012;删除MCSA后,PSNR和SSIM值分别降低了0.754 dB和0.037。证明了MCSA

和MGFB的有效性。

为了验证RTB数量对网络性能的影响,设计了如表4所示的消融实验。以NH-HAZE数据集为例给出证明。从实验数据可以看出,对于RTB的数量,性能增益逐渐饱和,本文选择RTB数量为6以获得相对较小的模型。

表4 RTB数量对网络性能的影响

Number	PSNR (dB)	SSIM	#Params (M)
×2	18.903	0.685	0.96
×4	19.102	0.702	1.10
×6	20.207	<b>0.733</b>	1.31
×8	<b>20.210</b>	<b>0.733</b>	2.49

### 3.6 网络收敛性分析

所提出网络的收敛性分析结果如图7所示,图7(a)和(b)分别展示了在O-HAZE数据集上训练网络的PSNR和SSIM随迭代次数的收敛过程,图7(c)展示了

损失函数随迭代次数的收敛过程。网络训练的总迭代次数为 500，从图中可以看出，随着迭代次数增加，Loss

曲线随着迭代次数的增加逐渐下降，在 400 次之后，Loss 曲线趋于平缓，训练模型逐渐收敛。

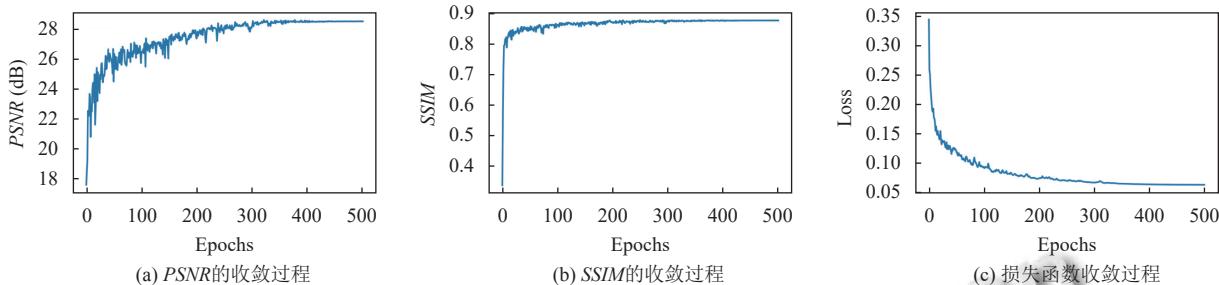


图 7 网络收敛性分析

## 4 结论

本文提出了一种基于 Transformer 和门控融合机制的图像去雾算法，通过改进标准 Transformer 的空间自注意力，降低计算复杂度，同时设计了一种多尺度门控融合块来有效聚合不同尺度的特征，使网络在学习远程建模能力的同时关注局部特征；门控融合机制通过动态调整权重，使网络更好地适应不同的雾霾分布情况，有效解决了图像边缘模糊、细节信息丢失和去雾不彻底等问题。实验结果表明，所提出的算法在合成有雾数据集和真实世界有雾数据集上均取得了较为理想的性能。然而，如何在多变的雾霾环境下有效处理不同浓度和分布的雾霾，实现更高质量的图像恢复性能仍有待研究。因此，在后续工作中，将继续研究有效应对不同雾霾环境的去雾算法，优化去雾性能，提高泛化能力。

## 参考文献

- Kaplan NH. Real-world image dehazing with improved joint enhancement and exposure fusion. *Journal of Visual Communication and Image Representation*, 2023, 90: 103720. [doi: [10.1016/j.jvcir.2022.103720](https://doi.org/10.1016/j.jvcir.2022.103720)]
- You JY, Korhonen J. Transformer for image quality assessment. *Proceedings of the 2021 IEEE International Conference on Image Processing (ICIP)*. Anchorage: IEEE, 2021. 1389–1393. [doi: [10.1109/ICIP42928.2021.9506075](https://doi.org/10.1109/ICIP42928.2021.9506075)]
- Cheng BW, Misra I, Schwing AG, et al. Masked-attention mask Transformer for universal image segmentation. *Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. New Orleans: IEEE, 2022. 1280–1289. [doi: [10.1109/CVPR52688.2022.00135](https://doi.org/10.1109/CVPR52688.2022.00135)]
- Cai YH, Bian H, Lin J, et al. Retinexformer: One-stage Retinex-based Transformer for low-light image enhancement. *Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV)*. Paris: IEEE, 2023. 12470–12479. [doi: [10.1109/ICCV51070.2023.01149](https://doi.org/10.1109/ICCV51070.2023.01149)]
- Aswini B, Josemartin MJ. Adaptive heuristic histogram equalisation approach combined with gamma transformation for image contrast enhancement. *Proceedings of the 4th International Conference on Microelectronics, Signals & Systems (ICMSS)*. Kollam: IEEE, 2021. 1–6. [doi: [10.1109/ICMSS53060.2021.9673652](https://doi.org/10.1109/ICMSS53060.2021.9673652)]
- Di SY, Sun WS. Research on low illumination image processing algorithm based on adaptive parameter homomorphic filtering. *Proceedings of the 2022 Asia Conference on Algorithms, Computing and Machine Learning (CACML)*. Hangzhou: IEEE, 2022. 681–685. [doi: [10.1109/CACML55074.2022.00118](https://doi.org/10.1109/CACML55074.2022.00118)]
- 游达章, 陶加涛, 张业鹏, 等. 基于灰度变换及改进 Retinex 的低照度图像增强. *红外技术*, 2023, 45(2): 161–170.
- Cantor A. Optics of the atmosphere: Scattering by molecules and particles. *IEEE Journal of Quantum Electronics*, 1978, 14(9): 698–699. [doi: [10.1109/JQE.1978.1069864](https://doi.org/10.1109/JQE.1978.1069864)]
- He KM, Sun J, Tang XO. Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, 33(12): 2341–2353. [doi: [10.1109/TPAMI.2010.168](https://doi.org/10.1109/TPAMI.2010.168)]
- Zhu QS, Mai JM, Shao L. A fast single image haze removal algorithm using color attenuation prior. *IEEE Transactions on Image Processing*, 2015, 24(11): 3522–3533. [doi: [10.1109/TIP.2015.2446191](https://doi.org/10.1109/TIP.2015.2446191)]
- 杨振舰, 尚佳美, 张众维, 等. 基于残差注意力机制的图像去雾算法. *西北工业大学学报*, 2021, 39(4): 901–908. [doi: [10.1051/jnwpu/20213940901](https://doi.org/10.1051/jnwpu/20213940901)]
- 寇大磊, 钱敏, 权冀川, 等. 基于多尺度卷积网络的快速图

- 像去雾算法. 计算机工程与应用, 2020, 56(20): 191–198. [doi: [10.3778/j.issn.1002-8331.1909-0165](https://doi.org/10.3778/j.issn.1002-8331.1909-0165)]
- 13 Cai BL, Xu XM, Jia K, et al. DehazeNet: An end-to-end system for single image haze removal. *IEEE Transactions on Image Processing*, 2016, 25(11): 5187–5198. [doi: [10.1109/TIP.2016.2598681](https://doi.org/10.1109/TIP.2016.2598681)]
- 14 Ren WQ, Liu S, Zhang H, et al. Single image dehazing via multi-scale convolutional neural networks. *Proceedings of the 14th European Conference on Computer Vision*. Amsterdam: Springer, 2016. 154–169. [doi: [10.1007/978-3-319-46475-6\\_10](https://doi.org/10.1007/978-3-319-46475-6_10)]
- 15 Li BY, Peng XL, Wang ZY, et al. AOD-Net: All-in-one dehazing network. *Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV)*. Venice: IEEE, 2017. 4780–4788. [doi: [10.1109/ICCV.2017.511](https://doi.org/10.1109/ICCV.2017.511)]
- 16 Ren WQ, Ma L, Zhang JW, et al. Gated fusion network for single image dehazing. *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018. 3253–3261. [doi: [10.1109/CVPR.2018.00343](https://doi.org/10.1109/CVPR.2018.00343)]
- 17 Liu XH, Ma YR, Shi ZH, et al. GridDehazeNet: Attention-based multi-scale network for image dehazing. *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. Seoul: IEEE, 2019. 7313–7322. [doi: [10.1109/ICCV.2019.00741](https://doi.org/10.1109/ICCV.2019.00741)]
- 18 Li Y, Cheng D, Zhang DW, et al. Single image dehazing with an independent detail-recovery network. *Knowledge-based Systems*, 2022, 254: 109579. [doi: [10.1016/j.knosys.2022.109579](https://doi.org/10.1016/j.knosys.2022.109579)]
- 19 Yi QS, Li JC, Fang FM, et al. Efficient and accurate multi-scale topological network for single image dehazing. *IEEE Transactions on Multimedia*, 2022, 24: 3114–3128. [doi: [10.1109/TMM.2021.3093724](https://doi.org/10.1109/TMM.2021.3093724)]
- 20 Xiao J, Fu XY, Liu AP, et al. Image de-raining Transformer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(11): 12978–12995. [doi: [10.1109/TPAMI.2022.3183612](https://doi.org/10.1109/TPAMI.2022.3183612)]
- 21 Zhao HY, Gou YB, Li BY, et al. Comprehensive and delicate: An efficient Transformer for image restoration. *Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Vancouver: IEEE, 2023. 14122–14132. [doi: [10.1109/CVPR52729.2023.01357](https://doi.org/10.1109/CVPR52729.2023.01357)]
- 22 张凯丽, 王安志, 熊娅维, 等. 基于Transformer的单幅图像去雾算法综述. *计算机科学与探索*, 2024, 18(5): 1182–1196. [doi: [10.3778/j.issn.1673-9418.2307103](https://doi.org/10.3778/j.issn.1673-9418.2307103)]
- 23 Chen HT, Wang YH, Guo TY, et al. Pre-trained image processing Transformer. *Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Nashville: IEEE, 2021. 12294–12305. [doi: [10.1109/CVPR46437.2021.01212](https://doi.org/10.1109/CVPR46437.2021.01212)]
- 24 Liu Z, Lin YT, Cao Y, et al. Swin Transformer: Hierarchical vision Transformer using shifted windows. *Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. Montreal: IEEE, 2021. 9992–10002. [doi: [10.1109/ICCV48922.2021.00986](https://doi.org/10.1109/ICCV48922.2021.00986)]
- 25 Liang JY, Cao JZ, Sun GL, et al. SwinIR: Image restoration using swin Transformer. *Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*. Montreal: IEEE, 2021. 1833–1844. [doi: [10.1109/ICCVW54120.2021.00210](https://doi.org/10.1109/ICCVW54120.2021.00210)]
- 26 Wang ZD, Cun XD, Bao JM, et al. UFormer: A general U-shaped Transformer for image restoration. *Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. New Orleans: IEEE, 2022. 17662–17672. [doi: [10.1109/CVPR52688.2022.01716](https://doi.org/10.1109/CVPR52688.2022.01716)]
- 27 Zamir SW, Arora A, Khan S, et al. Restormer: Efficient Transformer for high-resolution image restoration. *Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. New Orleans: IEEE, 2022. 5718–5729. [doi: [10.1109/CVPR52688.2022.00564](https://doi.org/10.1109/CVPR52688.2022.00564)]
- 28 Zhao D, Li J, Li HY, et al. Complementary feature enhanced network with vision Transformer for image dehazing. *arXiv:2109.07100*, 2021.
- 29 Song YD, He ZQ, Qian H, et al. Vision Transformers for single image dehazing. *IEEE Transactions on Image Processing*, 2023, 32: 1927–1941.
- 30 Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv:2010.11929v2*, 2021.

(校对责编: 王欣欣)