

# 基于全局局部协同的非均匀图像去雾方法

罗小同<sup>1</sup> 杨汶锦<sup>1</sup> 曲延云<sup>1</sup> 谢源<sup>2</sup>

**摘要** 近年来, 基于卷积神经网络 (Convolutional neural network, CNN) 的图像去雾方法在合成数据集上取得了显著的进展, 但由于真实场景中存在雾分布不均的问题, 卷积运算的局部感受野难以有效捕获到上下文指导信息, 从而导致全局结构信息丢失. 因此, 真实场景下的图像去雾任务面临着巨大的挑战. 考虑到 Transformer 具有捕获长距离语义信息依赖关系的优势, 有利于引导全局结构信息重建. 然而, 标准 Transformer 结构的高计算复杂度阻碍了其在图像恢复中的应用. 针对上述提到的问题, 提出一个由 Transformer 和卷积神经网络组成的双分支协同非均匀图像去雾网络 Dehazeformer. Transformer 分支用于提取全局结构信息, 同时设计稀疏自注意力模块 (Sparse self-attention modules, SSM) 以降低计算复杂度. 卷积神经网络分支用于获取局部信息, 从而恢复纹理细节. 在真实非均匀有雾场景下的实验结果表明, 该方法不管是在客观评价还是在主观视觉效果方面均达到优异的性能.

**关键词** 图像去雾, 卷积神经网络, Transformer, 特征融合, 稀疏自注意力

**引用格式** 罗小同, 杨汶锦, 曲延云, 谢源. 基于全局局部协同的非均匀图像去雾方法. 自动化学报, 2024, 50(7): 1333–1344

**DOI** 10.16383/j.aas.c230567

## Dehazeformer: Nonhomogeneous Image Dehazing With Collaborative Global-local Network

LUO Xiao-Tong<sup>1</sup> YANG Wen-Jin<sup>1</sup> QU Yan-Yun<sup>1</sup> XIE Yuan<sup>2</sup>

**Abstract** In recent years, image dehazing methods based on convolutional neural network (CNN) have made remarkable progress in synthetic datasets, but the local receptive field of convolution operation is difficult to effectively capture contextual guidance information due to the uneven distribution of haze in the real scene, resulting in the loss of global structure information. Therefore, the image dehazing task in the real scene still faces great challenges. Considering that Transformer has the advantage of capturing long-range semantic information dependency relationships, it can facilitate global structure information reconstruction. However, the high computational complexity of the standard Transformer structure hinders its application in image restoration. To solve the problems mentioned above, this paper proposes a double-branch collaborative nonhomogeneous image dehazing network, which is called Dehazeformer and composed of Transformer and convolutional neural network. The Transformer branch is used to extract global structure information, and sparse self-attention modules (SSM) are designed to reduce computational complexity. Besides, the convolutional neural network branch is used to obtain local information to recover texture details. Extensive experiments in the real nonhomogeneous haze scene show that the proposed method achieves excellent performance in both objective evaluation and subjective visual effects.

**Key words** Image dehazing, convolutional neural network (CNN), Transformer, feature fusion, sparse self-attention

**Citation** Luo Xiao-Tong, Yang Wen-Jin, Qu Yan-Yun, Xie Yuan. Dehazeformer: Nonhomogeneous image dehazing with collaborative global-local network. *Acta Automatica Sinica*, 2024, 50(7): 1333–1344

雾霾是由空气中的烟雾、灰尘等漂浮颗粒造成

的一种恶劣大气现象, 这些漂浮在空中的颗粒会对光线进行吸收和散射, 导致图像成像质量下降, 如模糊、对比度低、颜色失真等. 低质量图像无疑会对下游的视觉任务 (如自动驾驶和监控) 造成严重的影响. 对于自动驾驶系统, 雾的出现对其检测其他车辆、行人等造成了巨大的困难; 对于监控等设备, 雾会导致成像退化, 从而使监控内容丢失.

图像去雾旨在提高图像的对比度、恢复图像的细节以及对图像颜色进行校正, 重建出干净清晰图像. 随着深度学习的发展, 图像去雾领域取得了巨大的突破. 通常使用大气散射模型<sup>[1]</sup>来对雾的产生

收稿日期 2023-09-12 录用日期 2024-03-17

Manuscript received September 12, 2023; accepted March 17, 2024

国家自然科学基金 (62176224, 62222602) 资助

Supported by National Natural Science Foundation of China (62176224, 62222602)

本文责任编辑 赫然

Recommended by Associate Editor HE Ran

1. 厦门大学信息学院计算机科学与技术系 厦门 361005 2. 华东师范大学计算机科学与技术学院 上海 200063

1. Department of Computer Science and Technology, School of Informatics, Xiamen University, Xiamen 361005 2. School of Computer Science and Technology, East China Normal University, Shanghai 200063

进行物理建模. 如图 1 所示, 空气中的烟雾等漂浮颗粒对光源发出的光线进行散射, 从而导致成像目标返回到成像设备的光线被衰减, 因此在有雾天气下, 成像设备得到的是退化图像. 大气散射模型的计算公式可以表示为

$$I(x) = J(x)t(x) + A(1 - t(x)) \quad (1)$$

其中,  $x$  为像素点的坐标位置;  $I$  是有雾图像, 由成像设备所得;  $J$  为对应的干净无雾图像;  $A$  表示大气光;  $t$  表示透射率, 其计算公式可以表示为

$$t(x) = e^{-\beta d(x)} \quad (2)$$

其中,  $d$  表示成像设备到成像目标之间的距离, 即景深;  $\beta$  为散射系数. 由式 (2) 可知, 透射率  $t$  与景深  $d$  呈负相关. 结合式 (1) 和式 (2) 可知, 图像去雾的关键在于求解大气光  $A$  和透射率  $t$ , 并且该过程是一个不适定问题.

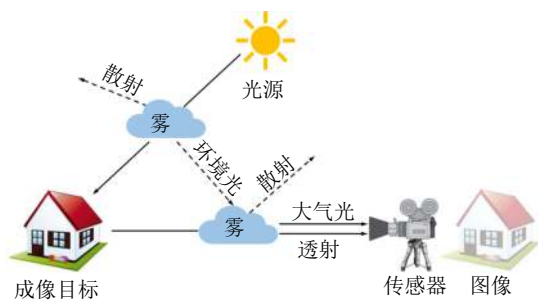


图 1 大气散射模型示意图

Fig. 1 Schematic diagram of atmospheric scattering model

针对不同类型的雾气, 需要采用不同的去雾方法. 一般来说, 可以将雾化程度分为均匀雾和非均匀雾两种类型. 均匀雾一般是由于大范围干湿度差异而形成的, 所以其浓度基本上是恒定的; 而非均匀雾则更加复杂, 它往往是由于环境中某些物质 (如汽车尾气、火灾烟雾等) 导致局部空气中含量过高而产生的. 针对这两种类型的雾气, 在处理图像时需要使用不同的去雾原理. 对于均匀雾, 一般采用透射率模型来处理; 而对于非均匀雾, 则需要采用更复杂的网络模型来学习复杂场景下的光传输、散射情况. 因此, 处理非均匀雾的方法比均匀雾更加复杂, 需要更多的技术手段和经验.

当前一些主流的图像去雾方法<sup>[2-5]</sup>无需估计雾成像参数, 而是直接通过卷积神经网络 (Convolutional neural network, CNN) 从合成的均匀有雾-清晰成对数据中学习有雾图像到清晰图像之间的映射. 但由于卷积核感受野受限, 这一类方法在非均匀有雾场景下表现不佳. 其主要原因为: 由于合成

数据集是由大气散射模型合成的均匀有雾图像, 而真实世界的有雾场景大多是比较复杂的, 导致这些方法难以泛化到真实有雾场景.

近年来, 真实场景去雾任务受到越来越多的关注, 图像恢复和增强新趋势以及图像和视频处理挑战 (New trends in image restoration and enhancement workshop and challenges on image and video processing, NTIRE) 在 2020 年和 2021 年均组织了非均匀图像去雾挑战赛, 并提供了真实世界非均匀有雾数据集. 其中优胜参赛方案<sup>[6-9]</sup>利用卷积神经网络强大的纹理细节恢复能力, 在低浓度有雾区域获得了良好的重建效果, 但是在高浓度有雾区域的表现却不尽如人意. 主要原因如下: 1) 卷积运算是一个局部特征表示, 缺乏全局特征表示; 2) 卷积神经网络滤波器是空间不变的, 即相同的卷积核处理所有位置的特征, 这会导致在高浓度的有雾区域出现伪影.

Transformer 在自然语言处理 (Natural language processing, NLP) 领域取得巨大成功. 受此启发, 许多高层计算机视觉任务都尝试使用 Transformer 替代卷积神经网络, 其可以很好地获取全局特征和远程语义信息依赖. 由于高层计算机视觉任务仅需提取可判别性信息用于决策, 不需要对图像进行纹理细节重建, 因此基于 Transformer 的方法在高层计算机视觉任务上取得了优异的性能. 由于缺乏局部信息提取能力, Transformer 在纹理细节上恢复不足, 而纹理细节的恢复在底层视觉任务中是至关重要的, 直接影响重建图像的质量.

因此, Transformer 和 CNN 结合有利于非均匀有雾图像的结构和纹理重建. 已有 Transformer 和 CNN 结合的非均匀去雾方法, 如 Dehazer<sup>[10]</sup>、ITB-dehaze<sup>[11]</sup>, 主要以图像块作为 Transformer 的输入, 具有较高的计算复杂度. 另外, 其中自注意力模块专注于建立 token 之间的远程依赖关系, 并利用 MLP 来建模, 忽略了空间维度上的局部性, 无法充分扩展感受野, 从而影响非均匀去雾模型的优化能力.

本文主要针对非均匀图像去雾任务, 其图像不同区域包含浓雾和薄雾, 符合真实场景下的雾成像情况. 为解决上述问题, 本文设计一个由 Transformer 分支和卷积神经网络分支组成的双分支结构 Dehazeformer, 其可以充分发挥卷积神经网络的局部纹理细节重建能力和 Transformer 的全局结构理解与远程语义信息依赖能力, 并通过特征融合操作, 实现高精度的雾去除. 具体地, 本文首先利用卷积操作提取浅层特征, 将学习到的特征映射作为 Transformer 的输入序列, 同时减小输入尺寸以降

低训练难度, 并使用稀疏自注意模块 (Sparse self-attention modules, SSM) 缓解计算复杂度高的问题. 自注意力机制可以远程捕捉不同输入序列之间的信息依赖, 从而利用清晰无雾的特征序列引导有雾特征序列. 此外, 与标准 Transformer 层相比, 本文方法将 MLP 替换为卷积块, 即在 Transformer 层中融合卷积块用于弥补自注意中局部信息提取能力的不足. 对于卷积神经网络分支, 使用在 ImageNet 上预训练的 ResNet 作为先验信息, 然后在非均匀的有雾数据集上对其进行微调, 从而适应去雾任务. 实现结果表明, 本文方法在主客观评价上均实现了优异的性能.

## 1 相关工作

随着深度学习的兴起, 图像去雾任务取得了重大突破, 现有非均匀图像去雾方法可大致分为三类.

### 1.1 基于 CNN 的方法

DMPHN<sup>[12]</sup> 提出一种快速的深度多图像块层次化去雾网络, 从不同的图像块中聚集有效特征. KTDN<sup>[7]</sup> 提出一种知识蒸馏去雾方法, 利用清晰图像来训练教师网络, 以提供强大的鲁棒图像先验, 通过教师模型和学生模型中间特征之间的监督, 鼓励去雾网络模仿教师网络. AtJwD<sup>[13]</sup> 设计一个共享的 DenseNet 编码器和四个不同的解码器来联合估计场景信息, 即大气光和透射率, 还使用通道注意力来生成不同的特征, 并在解码器上使用空洞 Inception 模块来生成非局部密集雾区域丢失的特征. ERRANet<sup>[14]</sup> 提出一种基于多分支注意力模块的端到端高效重参数化残差去雾网络, 包括空间注意力和通道注意力机制, 并采用颜色衰减损失和拉普拉斯金字塔损失对模型进行优化. 为快速有效地去除真实图像中的非均匀雾, MPSHAN<sup>[15]</sup> 提出一种多图像块和多尺度层级聚合网络, 通过层级融合模块、层级结构扩张卷积以及频域损失对算法进行优化, 在模型保持快速运行的情况下, 图像去雾效果得到了明显提升.

### 1.2 基于生成对抗网络的方法

DW-GAN<sup>[9]</sup> 提出一种离散小波变换去雾网络, 通过 DWT 保留更多的高频知识; 为防止过拟合, 在知识自适应分支中采用 ImageNet 预训练的 Res2Net, 来提高网络的泛化能力; 最后, 采用基于图像块的鉴别器对恢复后的图像进行去伪影处理. DB-CGAN<sup>[16]</sup> 提出一种双分支生成器的条件生成对抗网络, 包括 U-net 结构分支和多层残差结构分支,

并设计一个简单有效的子网络实现特征融合.

### 1.3 基于 CNN 和 Transformer 相结合的方法

近年来, 基于 CNN 的端到端深度学习方法在图像去雾方面取得了显著进展, 但在非均匀去雾方面往往效果提升有限. Transformer 和 CNN 结合有利于非均匀有雾图像的结构和纹理重建. Dehamer<sup>[10]</sup> 提出通过学习以 Transformer 特征为条件的调制矩阵来调制 CNN 特征, 还设计了传输感知三维位置嵌入模块, 将雾密度相关先验信息引入 Transformer 中, 该模块不仅提供相对位置, 还可以显示不同空间区域的雾密度. ITBdehaze<sup>[11]</sup> 设计一个专门的数据预处理策略来进行数据增强以及一个基于 Swin Transformer<sup>[17]</sup> 和 RCAB<sup>[18]</sup> 的双分支模型结构. 虽然目前已有 Transformer 与 CNN 结合的工作, 但在非均匀图像去雾上仍缺乏足够的探索.

## 2 本文方法

针对非均匀图像去雾任务, 本文提出基于全局局部协同的非均匀图像去雾网络 Dehazeformer, 旨在联合卷积神经网络和 Transformer, 同时学习局部纹理细节信息和全局图像结构表示. 如图 2 所示, 整个网络是一个编码器-解码器结构, 其中编码器包括 Transformer 分支和卷积神经网络分支. 对于非均匀有雾图像输入, 一方面通过 Transformer 编码器分支提取图像的全局结构表示, 充分利用图像中的清晰图像块引导网络重建有雾图像块; 另一方面, 通过卷积神经网络分支提取图像的局部纹理信息用于重建图像纹理细节. 最后, 融合两个分支的编码特征后输入到解码器从而得到去雾图像.

### 2.1 结构重建分支

#### 2.1.1 混合 Transformer 层

自然图像拥有丰富的纹理细节和颜色信息, 仅凭图像块级别的重建不能完全恢复图像的纹理细节. 针对这个问题, Han 等<sup>[19]</sup> 提出内部 Transformer 模块, 从像素级别的输入中提取局部纹理信息, 以弥补 Transformer 在纹理细节恢复方面的不足. 由于 Transformer 本身巨大的计算复杂度, 再增加一个内部 Transformer 必然会带来更大的计算压力. 为解决这个问题, 本文提出在 Transformer 层中嵌入卷积层, 用于充分发挥卷积神经网络在局部纹理信息提取中的优势.

标准 Transformer<sup>[20]</sup> 将图像均分成图像块作为输入, 导致图像中构成一些基本结构的底层特征 (如角和边) 很难提取; 另外, 其中自注意力模块专



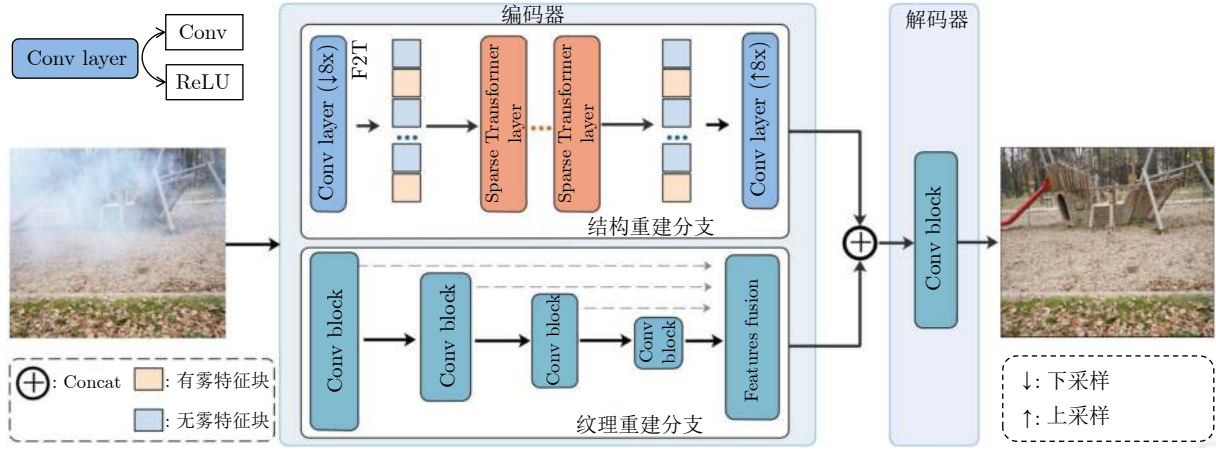


图2 基于全局局部协同的非均匀图像去雾网络 Dehazeformer 结构示意图

Fig.2 Schematic diagram of architecture of collaborative global-local network Dehazeformer for nonhomogeneous image dehazing

注于建立 token 之间的远程依赖关系, 忽略了空间维度上的局部性. 这样的做法使得有效的感受野不能被充分扩展, 从而影响 Transformer 的优化效率.

为解决上述问题, 本文将图像特征标记化作为 Transformer 层的输入. 与图像块作为 Transformer 输入不同, 本文方法使用多个卷积层提取的小尺度图像特征块作为 Transformer 的固定位置序列, 将其记为 F2T, 可以表示为

$$X' = \text{F2T}(X) = \text{ReLU}_n(\text{Conv}_n(X)) \quad (3)$$

其中,  $X$  为输入的非均匀有雾图像,  $X' \in \mathbf{R}^{\frac{H}{n} \times \frac{W}{n} \times D}$  为特征块,  $W$  和  $H$  分别表示输入的宽和高,  $D$  是通道的数量,  $\text{ReLU}(\cdot)$  表示线性整流激活函数,  $\text{Conv}(\cdot)$  表示卷积操作,  $n$  是卷积块的数量. 将学习到的特征映射作为 Transformer 的输入序列, 充分发挥卷积神经网络提取低层特征的优势, 同时减小输入尺寸以降低训练难度.

受 Transformer 可以捕获全局特征和远程语义信息依赖的启发, 本文方法应用 Transformer 的自注意力机制解决非均匀图像去雾问题, 即利用清晰的图像块引导网络重建有雾图像块的结构信息. 在 F2T 后, 使用多个 Transformer 层作为全局分支的编码器. 对于第  $l$  个 Transformer 层, 可以表示为

$$F^{l-1} = \text{LN}(\text{MSA}(E((X')^{l-1}))) + E((X')^{l-1}) \quad (4)$$

$$(X')^l = \text{LN}(\text{ConvBlock}(F^{l-1})) + F^{l-1} \quad (5)$$

其中,  $F$  表示中间层特征,  $E(\cdot)$  为输入的固定位置序列,  $\text{MSA}(\cdot)$  为多头自注意力,  $\text{LN}(\cdot)$  为层归一化,  $\text{ConvBlock}(\cdot)$  为卷积模块. 自注意力机制可以远程捕捉不同输入序列之间的信息依赖, 从而利用清晰无雾的特征序列引导有雾特征序列. 如图 3 所示,

与标准 Transformer 层相比, 本文方法将 MLP 替换为卷积块, 即在 Transformer 层中融合卷积块用于弥补自注意中局部信息提取能力的不足.

### 2.1.2 稀疏自注意力机制

自注意力机制是 Transformer 的核心模块, 主要用于学习图像全局结构表示, 并建立远程语义信息依赖, 但它也是 Transformer 计算复杂度高的主要原因. 尽管本文使用多层卷积块对输入的有雾图像进行下采样, 从而降低 Transformer 层输入的尺寸, 但该部分仍然具有较高的计算复杂度.

本文针对的是非均匀图像去雾任务, 图像不同区域会存在浓雾和薄雾的差异. 在全局范围内进行特征建模可能会引入额外的噪声, 因为每个位置都会占据一部分权重为当前 query 进行加权. 为此, 本文仅考虑当前查询点较为接近位置的自相似性计算, 然后进行特征聚合, 增大感受野. 受 Guo 等<sup>[21]</sup> 和 Zhao 等<sup>[22]</sup> 的启发, 本文使用稀疏自注意力模块减少 Transformer 层的计算复杂度. 如图 3(b) 所示, 与全局每个 query 进行注意力计算不同, 本文中稀疏自注意力模块是从水平和竖直两个方向上进行自注意力计算, 将注意矩阵简化为稀疏表示, 从而降低 Transformer 的计算复杂度. 对于水平自注意力, 只计算行序列之间的远程语义信息依赖关系, 并在水平维度上聚合上下文全局信息. 对于垂直自注意力, 只计算列序列之间的远程语义信息依赖关系, 并在垂直维度上聚合上下文全局信息. 多头自注意力 MSA 可以表示为

$$\text{head}_j = \text{Softmax}\left(\frac{Q^j(K^j)^T}{\sqrt{d}}\right)V^j \quad (6)$$

$$\text{MSA} = [\text{head}_1; \cdots; \text{head}_h] \quad (7)$$

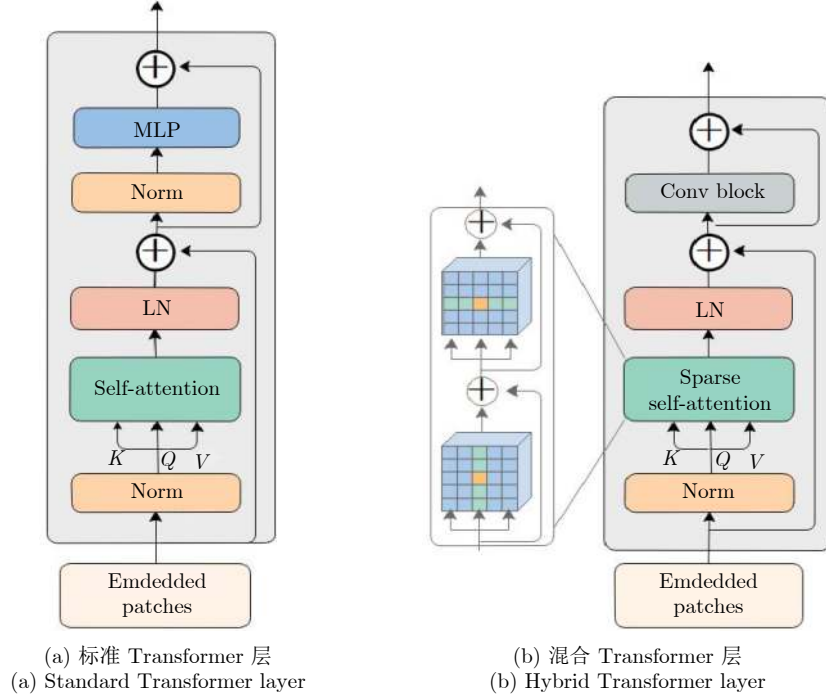


图3 标准 Transformer 层和本文提出的混合 Transformer 层结构对比

Fig.3 Architecture comparison of the standard Transformer layer and the proposed hybrid Transformer layer

其中,  $head$  表示注意力头的输出,  $K$ ,  $Q$  和  $V$  表示线性投影层,  $j$  是一个取值范围为  $1 \leq j \leq h$  的索引,  $d$  是嵌入的尺寸,  $h$  是多头自注意力中头的数量. 此外, SOTR<sup>[21]</sup> 主要用于解决实例分割问题, 其引入可学习的位置嵌入编码并添加到每个 token 以保持位置信息, 从而学习每个特征块的最佳位置编码. 其中, 对于实例分割任务, 输入序列的顺序是非常重要的线索, 对于同一物体通常是分布在图像中的相邻位置, 而模型只能用位置编码来获取这一顺序, 进而保证序列的置换不变性和平移不变性. 不同于 SOTR, 考虑到图像去雾作为回归任务, 需要恢复结构和纹理细节, 对图像特征的位置编码并不敏感, 因此本文在稀疏自注意力模块中去除了位置嵌入部分, 仅建模水平和垂直尺度远距离依赖关系, 用于图像结构重建.

给定输入图像  $I \in \mathbf{R}^{H \times W \times 3}$ , 首先通过三层卷积块提取特征  $F \in \mathbf{R}^{(H/8) \times (W/8) \times C}$  ( $C$  表示通道数), 然后 Transformer 层将特征输入分成  $N \times N$  块序列  $P_i \in \mathbf{R}^{N \times N \times C}$ . 在自注意力模块中, 水平自注意力模块和垂直自注意力模块输入的嵌入空间分别为  $1 \times N \times C$  和  $N \times 1 \times C$ . 为完成残差连接和后处理, 这两个维度的注意力产生  $N \times N \times C$  的输出.

综上所述, 本文 Transformer 分支中自注意力机制的计算复杂度为  $O((H/8) \times (W/8)^2 + (W/8) \times (H/8)^2)$ , 远低于标准 Transformer 的计算复杂度

$$O((H/8)^2 \times (W/8)^2).$$

## 2.2 纹理重建分支

Johnson 等<sup>[23]</sup> 利用预训练好的 VGG 模型提取特征作为损失约束, 从而提升纹理细节的重建性能, 表明分类模型作为纹理特征先验是可行的. 因此, 本文考虑使用在 ImageNet 数据集上预训练的 ResNet101 模型作为卷积神经网络分支的先验信息, 并对其在有雾数据集上进行微调.

如图 4 所示, 通过 ResNet101 模型可以提取到  $1/4$ 、 $1/8$ 、 $1/16$ 、 $1/32$  四个维度的特征. 为了充分利用不同维度上的特征信息, 受特征金字塔网络<sup>[24]</sup> 的启发, 将四个维度的特征进行上采样和特征增强后在通道维度上进行特征融合, 并将融合后的特征通过两个  $3 \times 3$  大小的卷积以对齐特征的维度. 如图 5(a) 所示, 特征增强模块主要由卷积层、激活函数 ReLU、通道自注意力模块和像素自注意力模块构成, 主要用于增强浓雾区域的重建能力. 如图 5(b) 所示, 通道自注意力模块主要由平均池化层、卷积、激活函数 ReLU、激活函数 Sigmoid 构成, 该模块负责学习每个通道的权重, 这些权重的大小在一定程度上反映了对应通道特征的重要程度. 如图 5(c) 所示, 像素自注意力模块主要由卷积、激活函数 ReLU、激活函数 Sigmoid 构成, 该模块负责学习每个像素点的权重, 这些权重的大小在一定程度上反映了对

应像素点特征的重要程度。

### 2.3 损失函数

本文采用的损失函数主要包括两个部分: 重建损失  $L_r$  和感知损失  $L_p$ , 以下对各项损失函数进行详细介绍。

#### 2.3.1 重建损失

本文采用平均绝对误差 (Mean absolute error, MAE) 测量去雾结果与对应清晰图像之间的距离

$$L_r = \|\hat{I} - J\|_1 \quad (8)$$

其中,  $\|\cdot\|_1$  表示  $L_1$  范数,  $\hat{I}$  是去雾结果,  $J$  是对应的清晰图像。

#### 2.3.2 感知损失

为进一步提高感知质量和语义保真度, 本文还

使用感知损失测量去雾结果与对应清晰图像之间的感知特征距离

$$L_p(\hat{I}, J) = \frac{1}{C_i H_i W_i} \|\phi_i(\hat{I}) - \phi_i(J)\|_1 \quad (9)$$

其中,  $\phi_i$  为从预训练 VGG16 网络中提取第  $i$  层感知特征的操作,  $C_i$ ,  $H_i$  和  $W_i$  分别为第  $i$  层提取到的特征通道数、高度和宽度。本文提取 VGG16 的第 2、7、12 层的感知特征用于计算感知损失。

综上所述, 网络的损失函数可表示为

$$L = L_r + L_p \quad (10)$$

本文提出 Transformer 和卷积神经网络双分支融合的非均匀图像去雾网络, 训练流程见算法 1。

#### 算法 1. Dehazeformer 训练流程

输入. 有雾数据集和对应的清晰数据集  $D_{\text{hazy}}$ ,  $D_{\text{clear}}$ ;

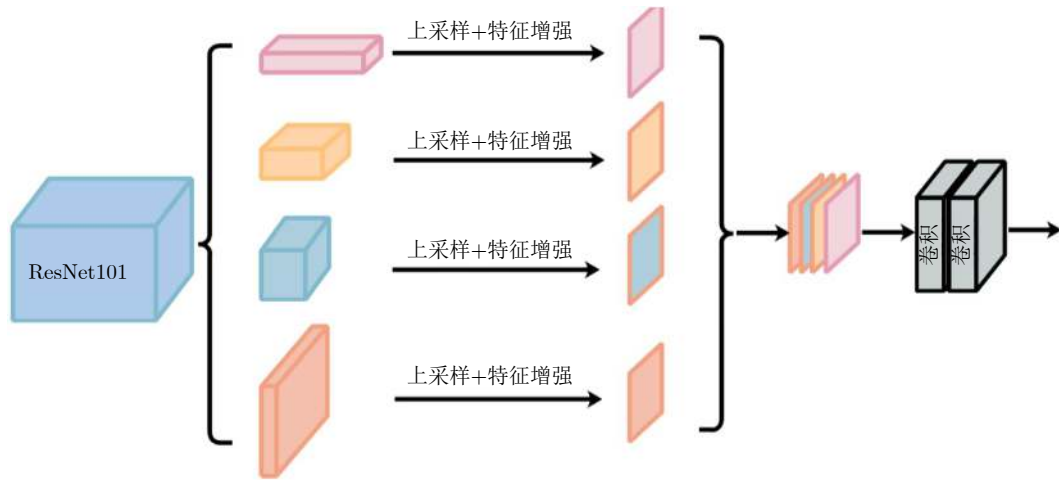


图 4 纹理重建分支的特征融合模块

Fig. 4 Feature fusion module of texture reconstruction branch

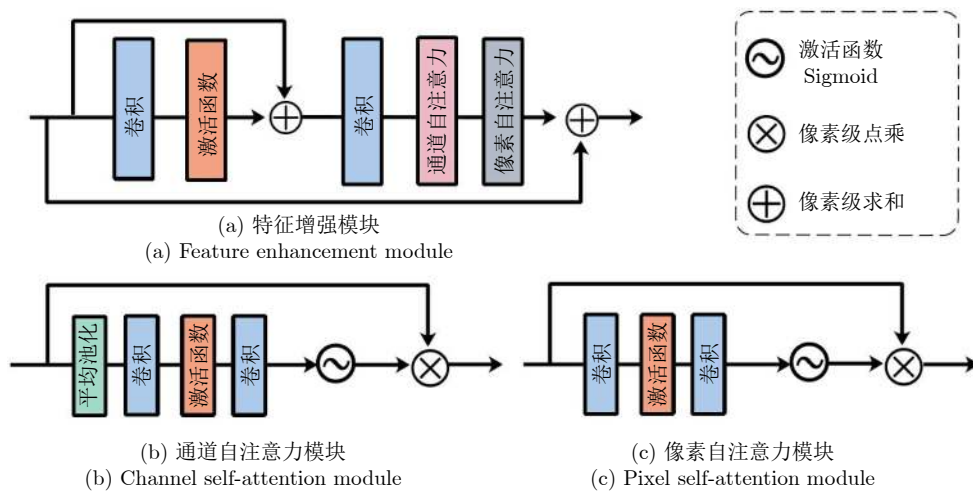


图 5 特征增强模块及注意力模块

Fig. 5 Feature enhancement module and attention modules



有雾图像  $X$ , 清晰图像  $Y$ ; 批处理大小  $m$ , 超参数  $\lambda$ ; 结构重建分支  $En_{trans}(\cdot)$ , 纹理重建分支  $En_{CNN}(\cdot)$ , 解码器  $De(\cdot)$ .

**输出.** 去雾图像  $y'$ .

- 1) **while** 未收敛, **do**
- 2) 从成对有雾数据集  $D_{hazy}$ ,  $D_{clear}$  中随机采样样本对  $(x_i)_{i=1}^m, (y_i)_{i=1}^m$ ;
- 3) 从 Transformer 主导的图像结构重建编码器提取全局特征:  $F_{global_x} \leftarrow En_{trans}(x)$ ;
- 4) 从 CNN 主导的纹理细节重建编码器提取局部特征:  $F_{local_x} \leftarrow En_{CNN}(x)$ ;
- 5) 对两个特征进行融合  

$$F_{fusion_x} = \text{Concat}(F_{global_x}, F_{local_x})$$
- 6) 融合后的特征输入解码器重建出清晰图像:  $y' \leftarrow De(F_{fusion_x})$ ;
- 7) 通过损失函数 (10) 更新 Dehazeformer 模型的梯度.
- 8) **end while**

### 3 实验结果与分析

#### 3.1 数据集及实验设置

##### 3.1.1 数据集

本文采用 NTIRE 挑战赛提出的非均匀图像去雾数据集 NH-HAZE20、NH-HAZE21 进行训练和测试, 其包含使用专业雾机生成的非均匀雾气的真实室外场景, 模拟真实的雾气条件. NH-HAZE20 数据集包含 55 对有雾-清晰图像, 其中有 45 对训练集、5 张验证集和 5 张测试集, 图像分辨率为  $1600 \times 1200$ , 包含各种户外场景下雾度分布不均匀的图片. NH-HAZE21 包含 25 对有雾-清晰图像、5 张验证集和 5 张测试集, 因其验证集和测试集目前还没有公开, 使用图像 1~20 作为训练集, 21~25 作为验证集. 由于数据量较少, 采用随机翻转、裁剪等方式进行数据增强.

##### 3.1.2 实验细节

在训练过程中, 将样本尺寸随机裁剪到  $256 \times 256$  大小, 并选用 Adam ( $\beta_1 = 0.900$ ,  $\beta_2 = 0.999$ ) 作为优化器, 批处理大小设为 6. 初始学习率设置为  $1 \times 10^{-4}$ , 对于 NH-HAZE20 和 NH-HAZE21 两个数据集一共训练 60 轮, 同时学习率在每 20 轮进行一次 0.5 倍的衰减. 由于训练集包含的图像较少, 通过随机旋转 ( $90^\circ$ 、 $180^\circ$  和  $270^\circ$ ) 和随机水平翻转等方式进行数据增强. 根据 NTIRE 非均匀图像去雾挑战赛所使用的评价指标, 本文采用 PSNR、SSIM 和 LPIPS 作为图像去雾质量评价指标, 其中 PSNR 和 SSIM 指标越高越好, LPIPS 指标越低越好. 本

文选用深度学习框架 PyTorch 来训练模型, 整个实验过程在 NVIDIA RTX 2080 GPU 上实现.

#### 3.2 消融实验

为验证本文方法 Dehazeformer 中每个模块的有效性, 本节在 NH-HAZE21 测试集上进行消融实验, 主要针对混合 Transformer 层 (Hybrid Transformer layer, HTL)、稀疏自注意力模块、结构重建分支 (Structure reconstruction branch, SRB) 和纹理重建分支 (Texture reconstruction branch, TRB) 构造了 6 种网络变体, 并且在同等实验设置下进行训练和测试. 具体实验设置如下:

- 1) Baseline: 基础网络, 即标准 Transformer 分支网络;
- 2) Baseline+HTL: 用混合 Transformer 层替换标准 Transformer 层;
- 3) Baseline+SSM: 用稀疏自注意力模块替换标准自注意力模块;
- 4) SRB: 结构重建分支, 即在基础网络中加入混合 Transformer 层和稀疏自注意力模块;
- 5) TRB: 纹理重建分支, 即在 ImageNet 上进行预训练的 ResNet 网络;
- 6) Ours: 本文方法, 即融合结构重建分支和纹理重建分支.

##### 3.2.1 混合 Transformer 层有效性分析

混合 Transformer 层通过将标准 Transformer 层中的 MLP 替换成卷积块, 用于增强 Transformer 层的局部信息提取能力. 如表 1 所示, 加入混合 Transformer 层相对于基础网络获得了 0.64 dB 的 PSNR 增益, 同时 SSIM 和 LPIPS 也分别获得 0.0089 和 0.0083 的增益. 此外, 在稀疏自注意力模块的基础上加入混合 Transformer 层获得了 0.54 dB 的 PSNR 增益, 同时 SSIM 和 LPIPS 也分别获得 0.0097 和 0.0085 的增益. 如图 6(b) 和图 6(c)、图 6(d) 和图 6(e) 所示, 加入混合 Transformer 层后, 对去雾结果的局部纹理细节重建有一定的提升. 此外, 表 2 中还给出了本文采用的混合 Transformer 层与 SOTR<sup>[21]</sup> 中的 Twin Transformer 层的对比结果, 分别为本文方法+位置嵌入编码、本文方法中混合 Transformer 层替换成 Twin Transformer 层、本文的混合 Transformer 层, 可以看到本文方法取得了较好的结果. 综上所述, 混合 Transformer 层能有效提升 Transformer 层的局部信息提取能力.

##### 3.2.2 稀疏自注意力模块有效性分析

稀疏自注意力模块是对标准自注意力模块的简

表 1 在 NH-HAZE21 测试集上的消融实验定量比较

Table 1 Quantitative comparison of ablation experiments on the NH-HAZE21 test dataset

| 方法           | HTL | SSM | SRB | TRB | PSNR (dB) ↑  | SSIM ↑        | LPIPS ↓       |
|--------------|-----|-----|-----|-----|--------------|---------------|---------------|
| Baseline     | ×   | ×   | ✓   | ×   | 21.24        | 0.8339        | 0.1849        |
| Baseline+HTL | ✓   | ×   | ✓   | ×   | 21.88        | 0.8428        | 0.1766        |
| Baseline+SSM | ×   | ✓   | ✓   | ×   | 21.54        | 0.8361        | 0.1837        |
| SRB          | ✓   | ✓   | ✓   | ×   | 22.08        | 0.8458        | 0.1752        |
| TRB          | ×   | ×   | ×   | ✓   | 21.62        | 0.8566        | 0.1740        |
| Ours         | ✓   | ✓   | ✓   | ✓   | <b>22.44</b> | <b>0.8631</b> | <b>0.1597</b> |

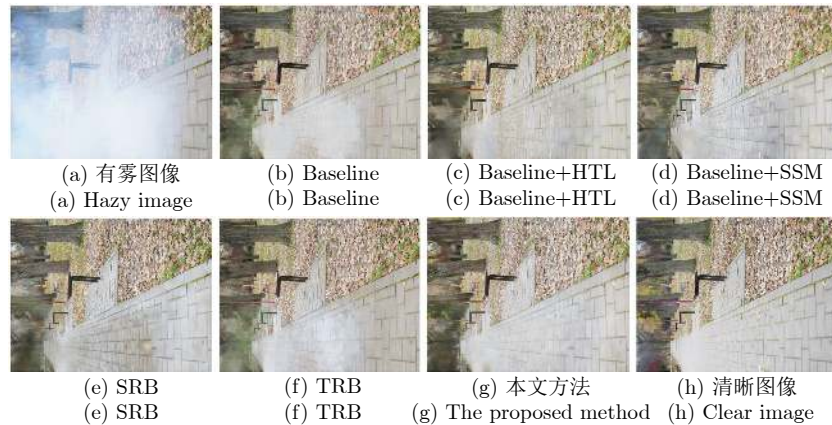


图 6 在 NH-HAZE21 测试集上消融实验的视觉比较

Fig. 6 Visual comparison of ablation experiments on the NH-HAZE21 test dataset

表 2 混合 Transformer 层与 Twin Transformer 层在 NH-HAZE21 测试集上的消融实验定量比较

Table 2 Quantitative comparison of ablation experiments between hybrid Transformer layer and Twin Transformer layer on the NH-HAZE21 test dataset

| 方法                 | PSNR (dB) ↑ | SSIM ↑ | LPIPS ↓ |
|--------------------|-------------|--------|---------|
| 本文方法+位置嵌入编码        | 22.33       | 0.8611 | 0.1613  |
| Twin Transformer 层 | 22.29       | 0.8611 | 0.1613  |
| 混合 Transformer 层   | 22.44       | 0.8631 | 0.1597  |

化,从而减少 Transformer 层的计算复杂度.如表 1 所示,加入稀疏自注意力模块相对于基础网络获得了 0.30 dB 的 PSNR 增益.同时,SSIM 和 LPIPS 也分别获得 0.002 2 和 0.001 2 的增益.相比于方法 Baseline+HTL,方法 SRB 中加入稀疏自注意力模块获得了 0.20 dB 的 PSNR 增益,同时 SSIM 和 LPIPS 也分别获得 0.003 0 和 0.001 4 的增益.如图 6(b) 和图 6(d)、图 6(c) 和图 6(e) 所示,加入稀疏自注意力模块后能有效提升去雾性能.此外,还对模型复杂度进行测试,本文模型使用标准 Transformer 的每秒浮点运算次数为 146 G,而采用稀疏自注意力则为 93 G,能有效减少计算复杂度.

### 3.2.3 结构重建分支有效性分析

结构重建分支能有效提升去雾结果的结构信息,如表 1 所示,与仅在纹理重建分支 TRB 上进行训练相比,加入结构重建分支的本文方法获得了 0.82 dB 的 PSNR 增益,同时 SSIM 和 LPIPS 也分别获得 0.0065 和 0.0143 的增益.如图 6(e) 和图 6(g) 所示,加入图像结构重建分支后,提升了去雾结果的全局重建能力.综上所述,图像结构重建分支能有效提升去雾效果.

### 3.2.4 纹理重建分支有效性分析

纹理重建分支有效提升去雾结果的局部细节恢复能力,如表 1 所示,与仅在图像结构重建分支 SRB 上进行训练相比,加入纹理重建分支的本文方法获得了 0.36 dB 的 PSNR 增益,同时 SSIM 和 LPIPS 也分别获得 0.017 3 和 0.015 5 的增益.如图 6(f) 和图 6(g) 所示,加入纹理重建分支后,提升了去雾结果的局部纹理细节重建性能.综上所述,纹理重建分支能有效提升去雾效果.

## 3.3 与主流方法的对比实验结果分析

为进一步说明本文方法 Dehazformer 的有效



性, 本节讨论本文方法与当前主流去雾方法的结果对比和分析, 这些主流的去雾方法包括 DCP<sup>[25]</sup>、CAP<sup>[26]</sup>、AOD-Net<sup>[27]</sup>、GridDehaze<sup>[28]</sup>、FFA-Net<sup>[4]</sup>、MSBDN<sup>[2]</sup>、KDDN<sup>[3]</sup>、AECR<sup>[5]</sup>、MPSHAN<sup>[15]</sup>、TransWeather<sup>[29]</sup>、Res2Net+RCAN<sup>[8]</sup>、DB-CGAN<sup>[16]</sup>、FADDehaze<sup>[30]</sup>、BiN-Flow<sup>[31]</sup>、PFONet<sup>[32]</sup>、SDD<sup>[33]</sup>、TUSR-Net<sup>[34]</sup> 和 ITBDehaze<sup>[11]</sup>.

### 3.3.1 与主流方法在 NH-HAZE20 数据集的对比

如表 3 所示, Dehazeformer 在 PSNR、SSIM 和 LPIPS 上的值为 22.16 dB、0.7345 和 0.2501, 与主流去雾方法相比均取得了最佳的成绩. 与 ITBDehaze 相比, Dehazeformer 在 PSNR 和 SSIM 上分别获得了 0.72 dB 和 0.0245 的增益. 因此, 本文方法在性能上达到当前非均匀去雾领先水平, 展示了方法的先进性. 在图 7 中, 还比较了 Dehazeformer 和其他主流去雾方法的主观视觉去雾效果. 传统方法 DCP 存在严重的颜色失真和纹理细节丢失问题, 其主要原因是该有雾场景不满足暗通道先验. 基于学习的主流图像去雾方法 GridDehaze、FFA-

Net、KDDN、AECR 等都存在颜色失真问题, GridDehaze、FFA-Net 和 KDDN 还存在一些伪影. 与 MSBDN 和 TransWeather 方法相比, 本文方法可以实现更好的细节恢复.

### 3.3.2 与主流方法在 NH-HAZE21 数据集的对比

为进一步验证 Dehazeformer 的有效性, 本文还与其他的主流去雾方法在 NH-HAZE21 数据集上进行对比. 如表 3 所示, 本文提出的 Dehazeformer 在 PSNR、SSIM 和 LPIPS 上的值为 22.44 dB、0.8631 和 0.1597, 与主流去雾方法相比均取得了最佳的表现. 与 ITBDehaze 相比, Dehazeformer 在 PSNR 和 SSIM 上分别获得了 0.77 dB 和 0.0251 的增益. 因此, 本文方法在性能上达到当前非均匀去雾领先水平, 展示了方法的先进性. 在图 8 中, 还比较了 Dehazeformer 和其他主流去雾方法的主观视觉去雾效果. 同样, 传统方法 DCP 存在严重的颜色失真和纹理细节丢失问题. 本文方法的去雾结果比基于学习的其他主流去雾方法 GridDehaze、FFA-Net、MSBDN、KDDN、AECR 和 TransWe-

表 3 在 NH-HAZE20 和 NH-HAZE21 测试集上与主流去雾方法的定量比较 (注: — 表示该方法未提供源码)

Table 3 Quantitative comparison with mainstream dehazing methods on the NH-HAZE20 and NH-HAZE21 test datasets (Note: — indicates that the method does not provide source code)

| 方法                           | NH-HAZE20    |               |               | NH-HAZE21    |               |               | NH-HAZE 平均值  |               |               |
|------------------------------|--------------|---------------|---------------|--------------|---------------|---------------|--------------|---------------|---------------|
|                              | PSNR (dB) ↑  | SSIM ↑        | LPIPS ↓       | PSNR (dB) ↑  | SSIM ↑        | LPIPS ↓       | PSNR (dB) ↑  | SSIM ↑        | LPIPS ↓       |
| DCP <sup>[25]</sup>          | 11.64        | 0.4533        | 0.5365        | 11.57        | 0.6278        | 0.4486        | 11.61        | 0.5674        | 0.4926        |
| CAP <sup>[26]</sup>          | 11.54        | 0.4188        | 0.5724        | 11.56        | 0.5848        | 0.4865        | 11.55        | 0.5018        | 0.5295        |
| AOD-Net <sup>[27]</sup>      | 13.44        | 0.4130        | —             | 15.20        | 0.6413        | 0.3103        | 14.32        | 0.5272        | —             |
| GridDehaze <sup>[28]</sup>   | 17.63        | 0.6668        | 0.3046        | 20.08        | 0.8134        | 0.2332        | 18.86        | 0.7401        | 0.2689        |
| FFA-Net <sup>[4]</sup>       | 17.44        | 0.6543        | 0.3340        | 20.51        | 0.8139        | 0.2315        | 18.98        | 0.7341        | 0.2828        |
| MSBDN <sup>[2]</sup>         | 19.01        | 0.7033        | 0.2858        | 20.89        | 0.8207        | 0.2393        | 19.95        | 0.7620        | 0.2626        |
| KDDN <sup>[3]</sup>          | 17.25        | 0.6602        | 0.3121        | 20.64        | 0.8156        | 0.2170        | 18.95        | 0.7379        | 0.2646        |
| AECR <sup>[5]</sup>          | 18.58        | 0.6575        | 0.2809        | 20.81        | 0.8269        | 0.1865        | 19.70        | 0.7422        | 0.2337        |
| MPSHAN <sup>[15]</sup>       | 18.13        | 0.6410        | —             | 18.97        | 0.7810        | —             | 18.55        | 0.7110        | —             |
| TransWeather <sup>[29]</sup> | 19.60        | 0.6990        | 0.2699        | 21.72        | 0.8368        | 0.1972        | 20.66        | 0.7679        | 0.2336        |
| Res2Net+RCAN <sup>[8]</sup>  | 21.44        | 0.7040        | —             | 21.66        | 0.8430        | —             | 21.55        | 0.7735        | —             |
| DB-CGAN <sup>[16]</sup>      | 18.29        | 0.6330        | —             | 19.33        | 0.7910        | —             | 18.81        | 0.7120        | —             |
| FADDehaze <sup>[30]</sup>    | 17.44        | 0.6300        | —             | 20.50        | 0.8400        | —             | 18.97        | 0.7350        | —             |
| BiN-Flow <sup>[31]</sup>     | 18.63        | 0.6340        | —             | —            | —             | —             | —            | —             | —             |
| PFONet <sup>[32]</sup>       | 20.09        | 0.6583        | —             | —            | —             | —             | —            | —             | —             |
| SDD <sup>[33]</sup>          | —            | —             | —             | 22.15        | 0.8350        | —             | —            | —             | —             |
| TUSR-Net <sup>[34]</sup>     | 21.96        | 0.7254        | —             | —            | —             | —             | —            | —             | —             |
| ITBDehaze <sup>[11]</sup>    | 21.44        | 0.7100        | —             | 21.67        | 0.8380        | —             | 21.56        | 0.7740        | —             |
| 本文方法                         | <b>22.16</b> | <b>0.7345</b> | <b>0.2501</b> | <b>22.44</b> | <b>0.8631</b> | <b>0.1597</b> | <b>22.30</b> | <b>0.7988</b> | <b>0.2049</b> |

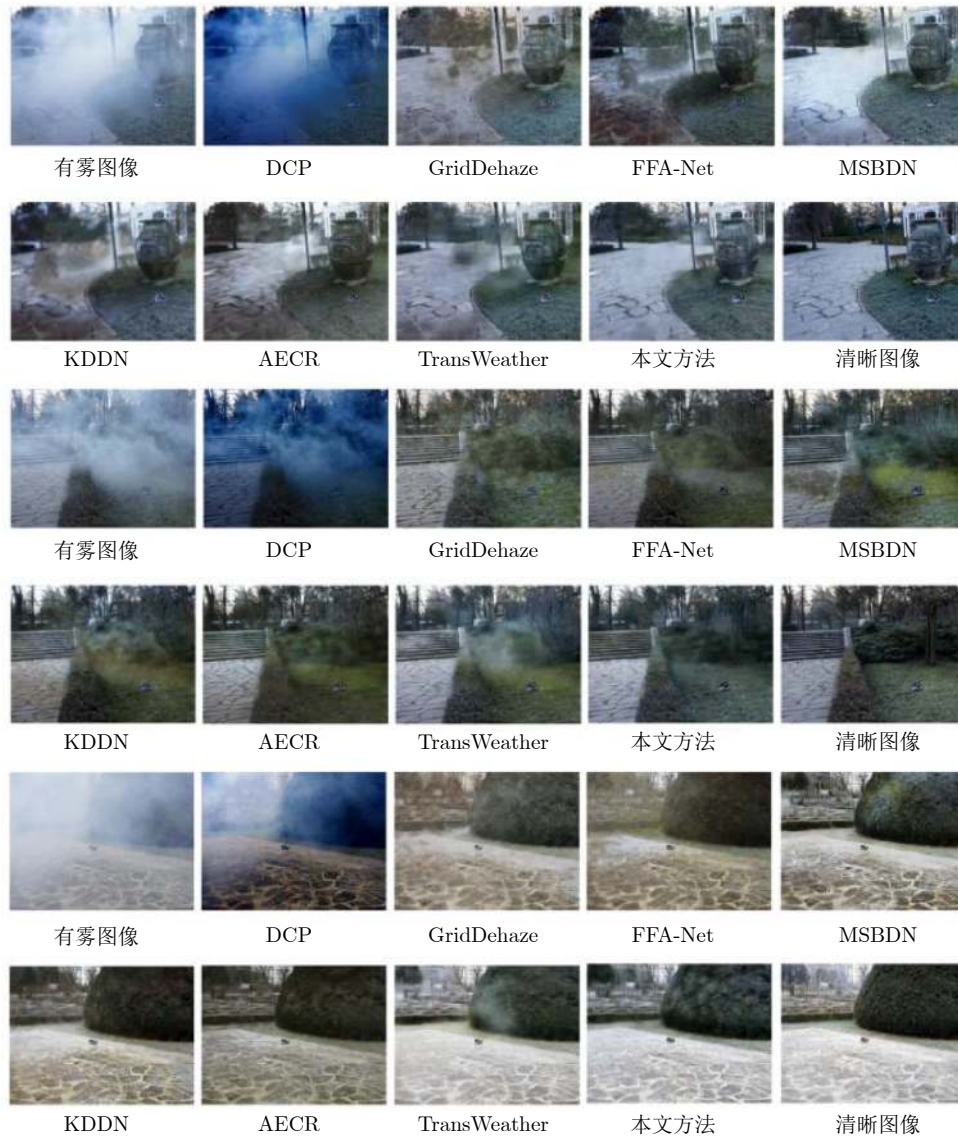


图 7 在 NH-HAZE20 测试集上与主流去雾方法的视觉比较

Fig. 7 Visual comparison with mainstream dehazing methods on the NH-HAZE20 test dataset

ather 的去雾结果拥有更多的纹理细节。

### 3.3.3 与 NTIRE 2021 去雾挑战赛优胜方案对比

如表 4 所示, 本文还与 NTIRE 2021 非均匀去雾挑战赛的优胜方案进行比较, 包括 DWT dehaze<sup>[9]</sup>、Mac dehaze<sup>[8]</sup>、Bilibili AI & FDU<sup>[6]</sup>、VIP UNIST<sup>[35]</sup>、Buaa colab<sup>[36]</sup>。可以看到, 本文提出的方法取得了最佳性能, 与 NTIRE 2021 非均匀图像去雾挑战赛的冠军方案相比, 本文方法在 PSNR 和 SSIM 上分别获得了 0.45 dB 和 0.0071 的增益。

## 4 结束语

本文提出一个由 Transformer 和卷积神经网络组成的双分支全局局部协同非均匀图像去雾网络。对于 Transformer 分支, 充分发挥 Transformer 的

远程语义信息依赖能力和全局信息表示能力。与标准 Transformer 层相比, 主要做了以下改进: 1) 使用卷积层提取较小尺度的特征作为输入, 充分利用卷积在提取特征上的优势, 缓解网络训练难度; 2) 使用混合 Transformer 层提高局部特征表示能力; 3) 使用稀疏自注意力模块缓解网络训练难度。对于卷积神经网络分支, 充分发挥卷积的纹理细节恢复能力和局部信息表示能力。本文提出的方法通过融合这两个分支, 充分发挥各自的优势。实验结果表明, 所提方法在非均匀有雾场景下的去雾效果优于当前主流去雾方法和 NTIRE 2021 的优胜方案。针对非均匀图像去雾任务, 如何设计高效的自注意力来建模远距离依赖关系是极为重要的, 在今后的工作中, 我们也将进一步对其进行研究。



图 8 在 NH-HAZE21 测试集上与主流去雾方法的视觉比较

Fig.8 Visual comparison with mainstream dehazing methods on the NH-HAZE21 test dataset

表 4 与 NTIRE 2021 非均匀图像去雾挑战赛优胜方案的定量比较

Table 4 Quantitative comparison with winning schemes of the nonhomogeneous image dehazing challenge in NTIRE 2021

| 方法                | PSNR (dB) $\uparrow$ | SSIM $\uparrow$ |
|-------------------|----------------------|-----------------|
| DWT dehaze        | 21.99                | 0.8560          |
| Mac dehaze        | 21.66                | 0.8430          |
| Bilibili AI & FDU | 21.24                | 0.7882          |
| VIP UNIST         | 21.17                | 0.8360          |
| Buaa colab        | 20.13                | 0.8034          |
| 本文方法              | <b>22.44</b>         | <b>0.8631</b>   |

## References

- Nayar S K, Narasimhan S G. Vision in bad weather. In: Proceedings of the Seventh IEEE/CVF International Conference on Computer Vision (ICCV). Kerkyra, Greece: IEEE, 1999. 820-827
- Dong H, Pan J S, Xiang L, Hu Z, Zhang X Y, Wang F, et al. Multi-scale boosted dehazing network with dense feature fusion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE, 2020. 2154-2164
- Hong M, Xie Y, Li C H, Qu Y Y. Distilling image dehazing with heterogeneous task imitation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE, 2020. 3459-3468
- Qin X, Wang Z L, Bai Y C, Xie X D, Jia H Z. FFA-Net: Feature fusion attention network for single image dehazing. In: Proceedings of the Thirty-fourth AAAI Conference on Artificial Intelligence. New York, USA: AAAI Press, 2020. 11908-11915
- Wu H Y, Qu Y Y, Lin S H, Zhou J, Qiao R Z, Zhang Z Z, et al. Contrastive learning for compact single image dehazing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, USA: IEEE, 2021. 10546-10555
- Liu J, Wu H Y, Xie Y, Qu Y Y, Ma L Z. Trident dehazing network. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Seattle, USA: IEEE, 2020. 1732-1741
- Wu H Y, Liu J, Xie Y, Qu Y Y, Ma L Z. Knowledge transfer dehazing network for nonhomogeneous dehazing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Seattle, USA: IEEE, 2020. 1975-1983
- Yu Y K, Liu H, Fu M H, Chen J, Wang X Y, Wang K Y. A two-branch neural network for non-homogeneous dehazing via ensemble learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Nashville, USA: IEEE, 2021. 193-202
- Fu M H, Liu H, Yu Y K, Chen J, Wang K Y. DW-GAN: A discrete wavelet transform GAN for nonhomogeneous dehazing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Nashville, USA: IEEE, 2021. 203-212
- Guo C L, Yan Q X, Anwar S, Cong R M, Ren W Q, Li C Y. Image dehazing Transformer with transmission-aware 3D position embedding. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). New Orleans, USA: IEEE, 2022. 5802-5810
- Liu Y Y, Liu H, Li L Y, Wu Z J, Chen J. A data-centric solution to nonhomogeneous dehazing via vision Transformer. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Vancouver, Canada: IEEE, 2023. 1406-1415
- Das S D, Dutta S. Fast deep multi-patch hierarchical network for nonhomogeneous image dehazing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Seattle, USA: IEEE, 2020. 1994-2001



- 13 Metwaly K, Li X L, Guo T T, Monga V. Nonlocal channel attention for nonhomogeneous image dehazing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Seattle, USA: IEEE, 2020. 1842–1851
- 14 Ye T, Chen E K, Huang X R, Chen P. Efficient re-parameterization residual attention network for nonhomogeneous image dehazing. arXiv preprint arXiv: 2109.05479, 2021.
- 15 Yang Kun, Zhang Juan, Fang Zhi-Jun. Multi-patch and multi-scale hierarchical aggregation network for fast nonhomogeneous image dehazing. *Computer Science*, 2021, **48**(11): 250–257 (杨坤, 张娟, 方志军. 基于多补丁和多尺度层级聚合网络的快速非均匀图像去雾. 计算机科学, 2021, **48**(11): 250–257)
- 16 Zhu Li-An, Zhang Hong. Nonhomogeneous image dehazing based on dual-branch conditional generative adversarial network. *Journal of Computer Applications*, 2023, **43**(2): 567–574 (朱利安, 张鸿. 基于双分支条件生成对抗网络的非均匀图像去雾. 计算机应用, 2023, **43**(2): 567–574)
- 17 Liu Z, Lin Y T, Cao Y, Hu H, Wei Y X, Zhang Z, et al. Swin Transformer: Hierarchical vision Transformer using shifted windows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, Canada: IEEE, 2021. 9992–10002
- 18 Zhang Y L, Li K P, Li K, Wang L C, Zhong B N, Fu Y. Image super-resolution using very deep residual channel attention networks. In: Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer, 2018. 294–310
- 19 Han K, Xiao A, Wu E H, Guo J Y, Xu C J, Wang Y H. Transformer in Transformer. arXiv preprint arXiv: 2103.00112, 2021.
- 20 Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X H, Unterthiner T, et al. An image is worth 16×16 words: Transformers for image recognition at scale. arXiv preprint arXiv: 2010.11929, 2021.
- 21 Guo R H, Niu D T, Qu L, Li Z B. SOTR: Segmenting objects with Transformers. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, Canada: IEEE, 2021. 7137–7146
- 22 Zhao G X, Lin J Y, Zhang Z Y, Ren X C, Su Q, Sun X. Explicit sparse Transformer: Concentrated attention through explicit selection. arXiv preprint arXiv: 1912.11637, 2019.
- 23 Johnson J, Alahi A, Li F F. Perceptual losses for real-time style transfer and super-resolution. In: Proceedings of the 14th European Conference on Computer Vision. Amsterdam, The Netherlands: Springer, 2016. 694–711
- 24 Lin T Y, Dollár P, Girshick R, He K M, Hariharan B, Belongie S. Feature pyramid networks for object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE, 2017. 936–944
- 25 He K M, Sun J, Tang X O. Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, **33**(12): 2341–2353
- 26 Zhu Q S, Mai J M, Shao L. A fast single image haze removal algorithm using color attenuation prior. *IEEE Transactions on Image Processing*, 2015, **24**(11): 3522–3533
- 27 Li B Y, Peng X L, Wang Z Y, Xu J Z, Feng D. AOD-Net: All-in-one dehazing network. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Venice, Italy: IEEE, 2017. 4780–4788
- 28 Liu X H, Ma Y R, Shi Z H, Chen J. GridDehazeNet: Attention-based multi-scale network for image dehazing. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, South Korea: IEEE, 2019. 7313–7322
- 29 Valanarasu J M J, Yasarla R, Patel V M. TransWeather: Transformer-based restoration of images degraded by adverse weather conditions. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, USA: IEEE, 2022. 2343–2353
- 30 Wu Zheng-Ping, Cheng Jie-Ying, Lei Bang-Jun, Zhao Jun-Chen. Fast nonhomogeneous image dehazing algorithm based on feature attention. *Foreign Electronic Measurement Technology*, 2023, **42**(9): 9–18 (吴正平, 程洁莹, 雷帮军, 赵俊臣. 基于特征注意力的快速非均匀雾图像去雾算法. 国外电子测量技术, 2023, **42**(9): 9–18)
- 31 Wu Y Q, Tao D P, Zhan Y B, Zhang C Y. BiN-Flow: Bidirectional normalizing flow for robust image dehazing. *IEEE Transactions on Image Processing*, 2022, **31**: 6635–6648
- 32 Li S S, Zhou Y, Ren W Q, Xiang W. PFONet: A progressive feedback optimization network for lightweight single image dehazing. *IEEE Transactions on Image Processing*, 2023, **32**: 6558–6569
- 33 Kim G, Kwon J. Self-parameter distillation dehazing. *IEEE Transactions on Image Processing*, 2022, **32**: 631–642
- 34 Song X B, Zhou D F, Li W, Dai Y C, Shen Z L, Zhang L J, et al. TUSR-Net: Triple unfolding single image dehazing with self-regularization and dual feature to pixel attention. *IEEE Transactions on Image Processing*, 2023, **32**: 1231–1244
- 35 Chen T Y, Fu J H, Jiang W T, Gao C, Liu S. SRKTDN: Applying super resolution method to dehazing task. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Nashville, USA: IEEE, 2021. 487–496
- 36 Jo E, Sim J Y. Multi-scale selective residual learning for non-homogeneous dehazing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Nashville, USA: IEEE, 2021. 507–515



罗小同 厦门大学信息学院博士研究生. 主要研究方向为计算机视觉与图像处理.

E-mail: [xiaotluo@stu.xmu.edu.cn](mailto:xiaotluo@stu.xmu.edu.cn)

(LUO Xiao-Tong Ph.D. candidate at the School of Informatics, Xiamen University. Her research interest covers computer vision and image processing.)



杨汶锦 厦门大学信息学院硕士研究生. 主要研究方向为计算机视觉与图像处理.

E-mail: [wjyang6@stu.xmu.edu.cn](mailto:wjyang6@stu.xmu.edu.cn)

(YANG Wen-Jin Master student at the School of Informatics, Xiamen University. His research interest covers computer vision and image processing.)



曲延云 厦门大学信息学院教授. 主要研究方向为模式识别, 计算机视觉和机器学习. 本文通信作者.

E-mail: [yyqu@xmu.edu.cn](mailto:yyqu@xmu.edu.cn)

(QU Yan-Yun Professor at the School of Informatics, Xiamen University. Her research interest covers pattern recognition, computer vision, and machine learning. Corresponding author of this paper.)



谢源 华东师范大学计算机科学与技术学院教授. 主要研究方向为模式识别, 计算机视觉和机器学习.

E-mail: [yxie@cs.ecnu.edu.cn](mailto:yxie@cs.ecnu.edu.cn)

(XIE Yuan Professor at the School of Computer Science and Technology, East China Normal University. His research interest covers pattern recognition, computer vision, and machine learning.)